Special Issue Reprint

Featured Papers in Mathematics and Finance

Edited by
W. Brent Lindquist and Svetlozar (Zari) Rachev

MDPI

Featured Papers in Mathematics and Finance

Featured Papers in Mathematics and Finance

Guest Editors

W. Brent Lindquist
Svetlozar (Zari) Rachev

Guest Editors

W. Brent Lindquist
Department of Mathematics
and Statistics
Texas Tech University
Lubbock, TX
USA

Svetlozar (Zari) Rachev
Department of Mathematics
and Statistics
Texas Tech University
Lubbock, TX
USA

This is a reprint of the Special Issue, published open access by the journal *Journal of Risk and Financial Management* (ISSN 1911-8074), freely accessible at: https://www.mdpi.com/journal/jrfm/special_issues/O6S87482Y7.

For citation purposes, cite each article independently as indicated on the article page online and as indicated below:

Lastname, Firstname, Firstname Lastname, and Firstname Lastname. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

Contents

About the Editors

W. Brent Lindquist

Dr. W. Brent Lindquist is a professor in the Department of Mathematics and Statistics at Texas Tech University, specializing in computational mathematical finance. He received his PhD in theoretical physics from Cornell and transitioned to a career in computational mathematics at the Courant Institute for Mathematical Sciences at New York University. Prior to joining Texas Tech, Dr. Lindquist served as professor and chair in the Department of Applied Mathematics and Statistics at Stony Brook University, where he helped lead the transition of that department into one of the top 10 applied math programs in the country. As a computational mathematician, Brent has developed theory and numerical methods for portfolio optimization; option pricing; PDEs; flow in porous media; automated 3D image analysis for porous media and neuron and fiber morphology; Riemann problems in 2D; hierarchy formation in social animal groups; and the numerical solution of Feynman diagrams. He is a co-recipient of the Lee Segal prize from the Society of Mathematical Biology. He was one of the founding developers of the Frontier package used to study reservoir flow at field scales and is the principal architect of the 3DMA-Rock code for studying flow at the pore scale. Dr. Lindquist has over 100 publications, has presented his research in 25 countries on five continents, and participated as a PI or co-PI in projects receiving USD 20M of grant funding. He has supervised 40 PhD students.

Svetlozar (Zari) Rachev

Svetlozar (Zari) Rachev is a professor at Texas Tech University who specializes in mathematical finance, probability theory, and statistics. He is recognized for his significant contributions to probability metrics, derivative pricing, financial risk modeling, and econometrics. In risk management, he is credited as the creator of the methodology behind FinAnalytica's flagship product which received several awards, including the "Best Market Risk Solution Provider" at the Waters Rankings in 2010, 2012, and 2015 and the "Most Innovative Specialist Vendor" at the Risk Awards in 2014. Rachev earned an MSc from the Faculty of Mathematics at Sofia University in 1974, a PhD from Lomonosov Moscow State University in 1979, and a Dr Sci degree from Steklov Mathematical Institute in 1986. Rachev's contributions to mathematical finance include his work on the application of non-Gaussian models for risk assessment, option pricing, and their applications in portfolio theory. He is also known for introducing a new risk/return ratio, the "Rachev Ratio", which measures the reward potential relative to tail risk in a non-Gaussian setting. In probability theory, Rachev's books on probability metrics and mass transportation problems are widely cited. In the 2023 edition of the Research.com Ranking of Top Scientists in the field of Economics and Finance, Rachev was ranked 540 in the world and 364 in the United States. In the same edition, he was also ranked number 386 among the Best Mathematics Scientists in the world.

Editorial

Editorial for the Special Issue of Journal of Risk and Financial Management: Featured Papers in Mathematics and Finance

Svetlozar (Zari) T. Rachev * and W. Brent Lindquist

Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409, USA;
brent.lindquist@ttu.edu
* Correspondence: zari.rachev@ttu.edu

We are privileged to present this Special Issue of the Journal of Risk and Financial Management (JRFM), focused on the intersection of mathematics and finance. The 27 contributions to this issue explore theoretical advancements and practical applications in financial risk management, asset pricing, and portfolio optimization. These papers collectively emphasize the ongoing synergy between mathematical rigor and financial innovation.

Key Highlights of the Issue

This Special Issue encompasses a variety of topics, reflecting the extensive research at the intersection of mathematics and finance:

1. Asset Pricing and Portfolio Optimization: Several papers tackle challenges in pricing and managing portfolios, including enhancements to classical models, such as CAPM and the Fama–French three-factor model, and new approaches utilizing entropy and Bayesian analysis. One notable work introduces a novel metric for analyzing downside risks in portfolio selection.

2. Green Finance and ESG: The growing importance of sustainability is evident, with studies investigating the resilience of green investments amid policy uncertainties and ESG reporting's influence on financial performance.

3. Advanced Statistical and Econometric Techniques: Innovations in econometrics are represented by studies on multivariate GARCH models, penalized regression, and improved methods for estimating implied volatility. These papers highlight how refined mathematical tools can better capture market dynamics and dependencies.

4. Cryptocurrency and Financial Technology: Several contributions delve into cryptocurrency markets, examining interconnected risks, uncertainty indices, and volatility spillovers. These studies showcase the evolving challenges digital assets pose and their integration into traditional finance.

5. Risk Measures and Financial Stability: Papers on value at risk (VaR) and conditional VaR (CVaR) offer insights into managing extreme financial risks. Additionally, the stability of banking systems and non-performing loans under fiscal constraints is rigorously analyzed.

6. Financial Engineering: Advanced derivative pricing models, including those for path-dependent and step barrier options, demonstrate the application of complex mathematical frameworks to real-world financial instruments.

7. Macro-Financial Linkages: Studies exploring the interaction of monetary policy, climate change, and energy uncertainty on financial markets underline the importance of integrating macroeconomic considerations into financial modeling.

Implications for Research and Practice

This collection underscores the indispensable role of mathematical modeling in addressing contemporary financial challenges. From enhancing portfolio optimization to managing risks in volatile markets, these studies provide actionable insights for academics and practitioners alike. They also highlight emerging areas of inquiry, such as integrating sustainability metrics and quantifying systemic risks in interconnected markets.

Conflicts of Interest: The authors declare no conflict of interest.

Article

Large Drawdowns and Long-Term Asset Management

Eric Jondeau [1,*,†] **and Alexandre Pauli** [2,†]

1   Faculty of Business and Economics (HEC Lausanne), Swiss Finance Institute and CEPR,
    University of Lausanne, CH 1015 Lausanne, Switzerland
2   Ecole Polytechnique Fédérale de Lausanne, Route Cantonale, CH 1015 Lausanne, Switzerland;
    alexandre.pauli@epfl.ch
*   Correspondence: eric.jondeau@unil.ch
†   The authors contributed equally to this work.

Abstract: Long-term investors are often hesitant to invest in assets or strategies prone to significant drawdowns, primarily due to the challenge of predicting these drawdowns. This study presents a multivariate Markov-switching model for small- and large-cap returns in the U.S. equity markets, demonstrating that three distinct regimes are necessary to capture the negative trends in expected returns during financial crises. Our findings indicate that this framework enhances the prediction of conditional drawdowns compared to standard alternative models of financial returns. Furthermore, out-of-sample analysis shows that investment strategies based on these predictions outperform those relying on models with one or two regimes.

Keywords: large drawdowns; stock-market returns; Markov-switching model; portfolio allocation model

1. Introduction

In recent decades, the recurrence of disasters has raised the issue of protecting investors' portfolios from large market drawdowns.[1] In the first quarter of 2020, the U.S. equity market experienced a 36% decline—one of the five largest one-quarter drawdowns in the last century. Such substantial drawdowns are particularly significant for long-term asset managers, including insurers, pension funds, and sovereign wealth funds, as these losses can not only impact annual portfolio performance but may also threaten the survival of the fund. To mitigate these losses, some managers incorporate drawdown objectives into their portfolio optimization processes. However, modeling and predicting the temporal evolution of large drawdowns remains a considerable challenge, particularly as these events often unfold over extended periods (e.g., a quarter or a year), a time scale inconsistent with most conventional financial econometric models.

In this paper, we address both modeling and prediction challenges related to market drawdowns. We introduce a model for daily financial returns based on regime switches, which facilitates the prediction of significant market drawdowns. Our findings demonstrate that three distinct regimes are necessary to capture the long-term dynamics of U.S. stock market returns. Leveraging this model, we design an investment strategy aimed at minimizing the expected value of a large drawdown measure, possibly subject to an expected return target. We then evaluate this strategy out of sample over the last 30 years, providing evidence that investing according to this approach effectively allows investors to reduce their losses due to large market downturns, outperforming standard alternative models such as GARCH-type models.

Most of the literature exploring the prediction of large drawdowns relies on some form of historical simulation. For instance, Chekhlov et al. (2005) simulate scenarios based on historical data to predict next-period drawdowns. This approach is likely to work well insample but may not perform well when market conditions vary over time, possibly due to climate change. A key challenge in using parametric models to predict large drawdowns is

that most models fail to capture large losses accumulating over relatively extended periods. In particular, standard GARCH models, even with nonnormal distributions, are not able to generate the negative trend that we observe during financial market crises. Consequently, reproducing large drawdowns similar to those observed in the 1999 dotcom crisis, the 2008 subprime crisis, or the 2020 COVID-19 pandemic becomes highly challenging with these models. In contrast, Markov-Switching (MS) models can generate large losses associated with a crisis because they allow for different drifts in market returns across regimes. Therefore, when a given return process enters a bear regime, it can accumulate negative values for relatively long periods of time. Papers describing these models for financial returns often overlook this long-term property because their main focus is on other features of the returns' data generating process. In particular, Ang and Bekaert (2002), Ang and Chen (2002), and Guidolin and Timmermann (2004, 2007, 2008) used MS models to capture the nonnormality or the asymmetric correlation of returns. More recently, similar approaches were adopted to analyze the consequences in financial markets of disasters arising from climate change (Karydas and Xepapadeas 2019, and Barnett et al. 2020) or from the COVID-19 pandemic (Pagano et al. 2023). For shorter prediction horizons (e.g., 10 days), Peng et al. (2022) find that an MS model with two regimes is sufficient for measuring large drawdowns. For longer horizons, we provide evidence that three regimes are necessary to capture the occurrence of financial crises and subsequent drawdowns effectively.

The literature defines several concepts of large drawdowns. One widely accepted measure is the period maximum drawdown (MDD), which corresponds to the largest loss from peak to trough over a given period. Investment strategies based on MDD are analyzed by Grossman and Zhou (1993) in a continuous-time framework or by Reveiz and Leon (2008) in discrete time. Chekhlov et al. (2005) define the concept of conditional drawdown (CDD), which corresponds to an average of the largest drawdowns in a given period, with the average drawdown (ADD) and the MDD as limiting cases. More recently, Goldberg and Mahmoud (2016) define the conditional expected drawdown (CED), which corresponds to an average of the largest period MDD values over a long sample. These measures have interesting properties, as CDD and CED are convex measures of risk and as such can be reduced by portfolio diversification.

In this paper, we first evaluate the number of regimes necessary to predict large drawdowns. We perform this analysis using a long sample of daily returns for both small- and large-cap stocks, covering nearly 100 years, with the last 30 years used for out-of-sample evaluation.[2] Given that small caps are more vulnerable to drawdowns than large caps, we assess whether a strategy focused on minimizing an expected large drawdown measure produces optimal weights that differ from those minimizing the portfolio variance. We estimate multivariate MS-GARCH models with one to four regimes and different distributions for the innovation process. We empirically demonstrate that three regimes are statistically necessary to fit the data while four regimes would not provide any additional information. In the three-regime model, one of the regimes has large negative expected returns, which allows us to generate large windfalls consistent with actual crises. In contrast, standard GARCH models fail to generate such extensive drawdowns.

We proceed by running an out-of-sample investment exercise. Using a rolling window, we estimate the various models and allocate the portfolio for the subsequent period. Although this experiment is time-consuming, it effectively mimics an investor's real-time allocation process, penalizes overparameterized models, and helps mitigate the winner's curse problem (Hansen 2009). We find that models with three regimes provide the best predictions of large drawdowns in this experiment. When allocating investor wealth by minimizing the predicted large drawdown measure for the upcoming period, the ex post drawdown of the portfolio is consistently lower with the three-regime model compared to the GARCH model or the two-regime model. Investors using the three-regime model tend to allocate long positions in large caps and short positions in small caps—an allocation that proves effective in mitigating large drawdowns.

Finally, we consider an investor aiming to maximize an expected return—large drawdown criterion, which also takes into account the model's ability to predict expected returns. The out-of-sample results again demonstrate that one- and two-regime models are nearly always dominated by the three-regime models, particularly for extreme drawdowns.

2. Materials and Methods

2.1. Definitions and Measurement

The maximum drawdown is the most widely used concept of large drawdowns. It represents the maximum loss from peak to trough over a given time period. As MDD increases for longer time series, it is customary to measure MDD over a given fixed time period of one quarter or one year, for example. This section defines the notations and the various concepts of large drawdowns that we will analyze in the remainder of the paper.

Let $p_t = (p_{t,1}, \cdots, p_{t,H})$ be a sample path associated with the stochastic process P with a continuous and strictly increasing distribution, where $p_{t,h}$ denotes the log-price on day h in period t (e.g., a given quarter or year).[3] The drawdown within this period on day h corresponds to the return loss between the last peak and the current price:

$$DD_{t,h} = \max_{1 \le j \le h} p_{t,j} - p_{t,h}.$$

We denote the drawdowns within the period as $DD_t = (DD_{t,1}, \cdots, DD_{t,H})'$. The maximum drawdown is defined as the largest drawdown of the period:

$$MDD_t = \max_{1 \le h \le H} DD_{t,h}.$$

Chekhlov et al. (2003, 2005) define the conditional drawdown (CDD) as the average of the largest drawdowns in a given period exceeding a quantile of the drawdown distribution, which mitigates the impact of outliers. For a probability θ, CDD is given by the average of the worst $(1 - \theta) \times 100\%$ drawdowns. Formally, we define the drawdown threshold $Th_\theta(DD_t)$ as the θ-quantile of the drawdown distribution: $Th_\theta(DD_t) = \inf\{s \mid \Pr(DD_t > s) \le 1 - \theta\}$. CDD corresponds to the tail conditional expectation of the drawdown distribution, i.e., the average of the drawdowns above the threshold:

$$CDD_{\theta,t} = E_t[DD_t \mid DD_t > Th_\theta(DD_t)], \tag{1}$$

where the expectation is applicable over the sample path. When $\theta = 0$, $CDD_{\theta,t}$ is equal to the average drawdown, $ADD_t = E_t[DD_t]$. When $\theta \to 1$, $CDD_{\theta,t}$ coincides with MDD_t. For a given sample path p_t, we have $ADD_t \le CDD_{\theta,t} \le MDD_t$.

Finally, Goldberg and Mahmoud (2015, 2016) introduce the concept of conditional expected drawdown (CED) as the average of the MDD values exceeding a quantile of the MDD distribution. For a probability $\tilde{\theta}$, $CED_{\tilde{\theta}}$ corresponds to the average of the worst $(1 - \tilde{\theta}) \times 100\%$ maximum drawdowns. Consequently, the threshold $Th_{\tilde{\theta}}(MDD)$ is determined by the $\tilde{\theta}$-quantile of the MDD distribution: $Th_{\tilde{\theta}}(MDD) = \inf\{s \mid \Pr(MDD > s) \le 1 - \tilde{\theta}\}$ and the $CED_{\tilde{\theta}}$ is therefore given by:

$$CED_{\tilde{\theta}} = E[MDD \mid MDD > Th_{\tilde{\theta}}(MDD)], \tag{2}$$

where the expectation is taken over the full sample. When $\tilde{\theta} = 0$, $CED_{\tilde{\theta}}$ is equal to the sample average period MDD, i.e., $E[MDD]$.

Similar to the well-known expected shortfall, which measures the tail conditional expectation of the return distribution, CDD and CED correspond to the tail conditional expectation of the drawdown distribution and the tail conditional expectation of the maximum drawdown distribution, respectively. Importantly, as shown by Chekhlov et al. (2005) and Goldberg and Mahmoud (2015), MDD, CDD, and CED satisfy the properties of deviation measures, i.e., nonnegativity, shift invariance, positive homogeneity, and convexity. Convexity of a measure of risk implies that this measure can be reduced by

diversification and used in quantitative optimization. As a consequence, if an investor minimizes this measure, the minimum, if it exists, is a global minimum.

Our subsequent analysis will focus on the three large drawdown measures: CDD, MDD, and CED. An essential feature of these measures is that drawdowns can develop over different time frames.[4] To cope with this feature, we consider three different horizons H, corresponding to one quarter, two quarters, and four quarters.

2.2. Empirical Measures of Large Drawdowns

We now briefly describe how to measure the ex post drawdown of an asset or a portfolio of assets. We match the horizon of the drawdown measures to a long-term investor's horizon H. For instance, an asset manager rebalancing the portfolio every quarter wants to control for large drawdowns occurring throughout the quarter. Therefore, the full sample is divided into T nonoverlapping subsamples of length H, with the sequence of log-prices in subsample t given by $p_t = (p_{t,1}, \cdots, p_{t,H})$ for $t = 1, \cdots, T$. For each subsample, the vector of drawdowns is denoted by $DD_t = (DD_{t,1}, \cdots, DD_{t,H})$, with $DD_{t,h} = \max_{1 \leq j \leq h} p_{t,j} - p_{t,h}$, as before.

We obtain the drawdown-based measures for each subsample as follows. The period ADD is simply given by the sample mean of the drawdowns, $ADD_t = \frac{1}{H} \sum_{h=1}^{H} DD_{t,h}$; the period MDD is given by the maximum drawdown of the subsample, $MDD_t = \max_{1 \leq h \leq H} DD_{t,h}$; and the period CDD is calculated as the average of the drawdowns over the θ-quantile:

$$CDD_{\theta,t} = \frac{1}{(1-\theta)H} \sum_{h=1}^{H} DD_{t,h} \, I_{(DD_{t,h} > Th_{\theta,t})}, \tag{3}$$

where $I_x = 1$ if x is true and 0 otherwise, and $Th_{\theta,t} = \inf\{s \mid \frac{1}{H} \sum_{h=1}^{H} I_{(DD_{t,h} > s)} \leq 1 - \theta\}$.

Finally, CED is based on the distribution of the period MDD measures over the full sample: we collect the MDD values over all the subsamples $MDD = (MDD_1, \cdots, MDD_T)$ and the sample CED corresponds to the average of the worst $(1 - \tilde{\theta}) \times 100\%$ MDD values over the full sample:

$$CED_{\tilde{\theta}} = \frac{1}{(1-\tilde{\theta})T} \sum_{t=1}^{T} MDD_t \, I_{(MDD_t > Th_{\tilde{\theta}})}, \tag{4}$$

where $Th_{\tilde{\theta}} = \inf\{s \mid \frac{1}{T} \sum_{t=1}^{T} I_{(MDD_t > s)} \leq 1 - \tilde{\theta}\}$.

2.3. Investor's Problem

We now consider the investment strategy of a long-term investor with investment horizon H. At the end of period t, n risky assets are available. We denote by $r_{i,t+1,h}$ the log-return of asset i on day h of the period $t+1$. The vector of cumulated log-returns over h days is denoted by $R_{t+1,h} = \{R_{i,t+1,h}\}_{i=1}^{n}$, where $R_{i,t+1,h} = \sum_{j=1}^{h} r_{i,t+1,j}$. Portfolio weights, determined at the end of period t, are denoted by $\alpha_t = (\alpha_{1,t}, \cdots, \alpha_{n,t})$ with $\sum_{i=1}^{n} \alpha_{i,t} = 1$. The (unknown) value of the portfolio at the end of period $t+1$ is $P_{t+1}(\alpha_t) = \alpha_t' \exp(R_{t+1,H})$ in the absence of rebalancing during period $t+1$. The sequence of daily log-values of the portfolio in period $t+1$ is given by: $p_{t+1}(\alpha_t) = (p_{t+1,1}(\alpha_t), \cdots, p_{t+1,H}(\alpha_t))$, where $p_{t+1,h}(\alpha_t) = \log P_{t+1,h}(\alpha_t)$ and $P_{t+1,h}(\alpha_t) = \alpha_t' \exp(R_{t+1,h})$.

For a given weight vector α_t, we compute daily log-values in period $t+1$ and obtain large drawdown measures as described in Section 2.2.[5] The investment criterion consists of minimizing the expected value of one of the large drawdown measures for the next investment period (period CDD, MDD, or CED), which we denote generically as $XDD_{t+1}(\alpha_t)$. The optimal weight is:

$$\alpha_t^* \in \arg\min_{\{\alpha_t\}} E_t[XDD_{t+1}(\alpha_t)]. \tag{5}$$

We also consider a criterion based on the trade-off between the expected return and the risk of a large drawdown:

$$\alpha_t^* \in \arg\max_{\{\alpha_t\}} \; E_t[R_{p,t+1}(\alpha_t)] - \frac{\lambda}{2} E_t[XDD_{t+1}(\alpha_t)], \tag{6}$$

where λ denotes the aversion for large drawdowns.

The minimization problem (5) corresponds to the case where the aversion for large drawdowns λ goes to infinity. Although the optimization problem (6) may be more attractive for an investor willing to combine risk and return, problem (5) allows us to investigate more specifically the ability of the various models to predict large drawdowns. In Section 3.4, we will focus our comments on the results based on problem (5) and briefly discuss problem (6).

To obtain predictions of portfolio large drawdown measures, i.e., $E_t[XDD_{t+1}(\alpha_t^*)]$, we proceed as follows: First, we assume a multivariate MS-GARCH model to describe the data generating process (DGP) for daily log-returns and endogenize the path dependence of the return process. Two-regime and three-regime models would capture the drawdown trend if switching probabilities are sufficiently low. Using this DGP, we simulate assets' daily returns for the next investment period. Second, for a given portfolio weight vector, we obtain simulated paths of the portfolio return from which we deduce the large drawdown measures. This approach provides us with predictions of the large drawdown measures for the next investment period as a function of the weight vector, thereby enabling us to pinpoint the optimal weight that minimizes the objective function (Equation (5)). The details of our approach are the subject of the next section.

2.4. Methodology
2.4.1. Multivariate MS-GARCH Model

To capture the possible impact of a large drawdown on the performance of the long-term portfolio, we assume a multivariate MS-GARCH model for the return process. The vector of daily log-returns for the n assets is denoted by $\tilde{r}_{d+1} = (\tilde{r}_{1,d+1}, \cdots, \tilde{r}_{n,d+1})$. The temporal index $d = 1, \cdots, D$, represents days and runs over the full sample.[6]

The model is written as follows:

$$\tilde{r}_{d+1} = \mu_{d+1}(S_{d+1}) + \varepsilon_{d+1},$$

where $\mu_{d+1}(S_{d+1})$ is the vector of expected returns, conditional on state S_{d+1}, and ε_{d+1} is the vector of unexpected returns. It is defined as:

$$\varepsilon_{d+1} = \Omega_{d+1}(S_{d+1})^{1/2} z_{d+1},$$

where $\Omega_{d+1}(S_{d+1})$ denotes the $(n \times n)$ covariance matrix of unexpected returns and z_{d+1} is a sequence of iid innovations with distribution $D(0, I_n)$ with zero mean and identity covariance matrix.

States are defined by the Markov chain $\{S_{d+1}\}$ with K regimes and transition matrix $P = (p_{kk'})_{k,k'=1,\cdots,K}$, where transition probabilities are $p_{kk'} = \Pr(S_{d+1} = k'|S_d = k)$, $k, k' \in \{1, \cdots, K\}$.

Expected returns are constant within each state: $\mu_{d+1}(S_{d+1}) = \mu^{(k)}$ when $S_{d+1} = k$.[7] The covariance matrix $\Omega_{d+1}(S_{d+1})$ is time- and state-dependent. In a given state k, it is driven by a multivariate GARCH process with state-dependent conditional correlation matrix, as in Pelletier (2006) or Haas and Liu (2018). The conditional variance of asset i in state k is defined as a standard univariate GARCH(1,1) process:[8]

$$\sigma_{i,d+1}^{(k)2} = \omega_i^{(k)} + \alpha_i^{(k)}(\tilde{r}_{i,d} - \mu_i^{(k)})^2 + \beta_i^{(k)}\sigma_{i,d}^{(k)2},$$

with different parameters for each state, as in Haas et al. (2004).

The $(n \times n)$ correlation matrix is constant in a given state: $\Gamma^{(k)} = \left(\rho_{ij}^{(k)} \right)_{i,j=1,\cdots,n}$, so that the covariance matrix $\Omega_{d+1}(S_{d+1}) = \Omega_{d+1}^{(k)}$ in state k is:

$$\Omega_{d+1}^{(k)} = (D_{d+1}^{(k)})^{1/2} \, \Gamma^{(k)} \, (D_{d+1}^{(k)})^{1/2},$$

where $D_{d+1}^{(k)}$ is the diagonal matrix with $\left(\sigma_{i,d+1}^{(k)2} \right)_{i=1,\cdots,n}$ on the diagonal.

We consider two types of multivariate distributions for innovations z_{d+1}: A Gaussian distribution $N(0, I_n)$ and a standardized Student's t distribution $t(0, I_n, \nu)$, where ν denotes the degree of freedom. The choice of the innovation distribution may matter for two reasons. First, for investors who care about higher moments, the investment criterion might involve metrics, such as large drawdowns, that depend on the properties of the innovation distribution. Second, in our model, large drawdowns can be captured in principle by a higher probability of being in a bear market or by negative expected returns in the bear regime. However, a lower degree of freedom of the Student's t distribution can also affect the dynamics of regime shifts and possibly result in lower returns.[9]

To make inferences about the regimes, we calculate the probability of being in each regime. We denote by $f(\tilde{r}_{d+1} \mid \underline{\tilde{r}}_d, \theta)$ the distribution of the daily log-return process conditional upon past log-returns, with $\underline{\tilde{r}}_d = \{\tilde{r}_d, \tilde{r}_{d-1}, \cdots\}$ and θ denoting the vector of unknown parameters. Parameters include expected returns ($\mu^{(k)}$), volatility parameters ($\omega^{(k)}, \alpha^{(k)}, \beta^{(k)}$), correlations ($\Gamma^{(k)}$), probabilities ($p_{kk'}$), and the degree of freedom (ν). Using Hamilton (1989)'s filter, we obtain the predicted probabilities $\pi_{k,d+1} = \Pr[S_{d+1} = k \mid \underline{\tilde{r}}_d]$ and the filtered probabilities $\phi_{k,d+1} = \Pr[S_{d+1} = k \mid \underline{\tilde{r}}_{d+1}]$ as:

$$\pi_{d+1} = P \, \phi_{d+1} \qquad \text{and} \qquad \phi_{d+1} = \frac{\pi_{d+1} \odot l_{d+1}}{e'(\pi_{d+1} \odot l_{d+1})},$$

$$\text{with } l_{d+1} = \begin{bmatrix} f(\tilde{r}_{d+1} \mid \underline{\tilde{r}}_d, S_{d+1} = 1; \theta) \\ \vdots \\ f(\tilde{r}_{d+1} \mid \underline{\tilde{r}}_d, S_{d+1} = K; \theta) \end{bmatrix} \quad \text{and} \quad e = (1, \cdots, 1)'.$$

The estimation of the model is based on standard likelihood maximization, where the log-likelihood is defined as: $\log L_D(\theta) = \sum_{d=1}^{D-1} f(\tilde{r}_{d+1} \mid \underline{\tilde{r}}_d, \theta) = \sum_{d=1}^{D-1} \log \left(\sum_{k=1}^{K} \pi_{d+1} \odot l_{d+1} \right)$. We impose stationarity conditions as described by Haas et al. (2004) and Abramson and Cohen (2007) in the univariate case and Haas and Liu (2018) in the multivariate case.[10]

While MS-GARCH models are valuable tools, they have certain limitations. To avoid overfitting, it is essential to carefully determine the number of regimes and ensure the model accurately captures the key patterns in the data. Estimating these models can be computationally demanding, especially because large datasets are needed for reliable results. This makes them less practical for systems with many variables or a high number of regimes. Additionally, the regimes must correspond to clear economic or financial conditions to make their interpretation meaningful. In our analysis, we address these challenges by using a simple model with only two processes, a long dataset, and by rigorously checking how well the model fits the data.

2.4.2. Minimizing the Expected Large Drawdown of a Portfolio

In some specifications of the multivariate MS model, analytical formulas for portfolio characteristics are available. For instance, Guidolin and Timmermann (2004, 2008) provide formulas for the high-order moments in a model with regime-dependent (but time-independent) means and variances. Other nonlinear characteristics, such as the VaR or the expected shortfall of the portfolio return distribution, cannot be computed analytically, even in this simple model (Guidolin and Timmermann 2004). Additionally, in models such as MS-GARCH, analytical expressions are usually not available because variances are path

dependent. For this reason, we compute expected large drawdowns with Monte Carlo simulations.

For ease of exposition, we again assume a quarterly investment horizon, with H representing the number of days in a quarter. We solve the allocation problem (Equation (5)) at the end of quarter t through the following steps:

1. We estimate the parameters of the MS-GARCH model using daily log-returns available in quarters $1, \cdots, t$. The last day of the estimation period is denoted by $d = t \times H$. The next quarter, $t + 1$, contains days $d + 1, \cdots, d + H$.

2. For a given estimated model, we simulate Q samples of length H of daily log-returns for the n assets: $\{r_{t+1,h}^{(q)}\}_{h=1}^{H}$, $q = 1, \cdots, Q$. As the probability of being in state k at the end of period t is given by predicted probabilities $\{\pi_{d+1}^{(k)}\}_{k=1}^{K}$, we simulate a fraction $\pi_{d+1}^{(k)}$ of the draws using $\Omega_{d+1}^{(k)}$ as an initial condition for the covariance matrix in period $t + 1$. From the simulated daily log-returns, we compute cumulative log-returns in quarter $t + 1$ as: $R_{t+1,h}^{(q)} = \sum_{j=1}^{h} r_{t+1,j}^{(q)}$ for $h = 1, \cdots, H$.

3. For a portfolio weight vector α_t, we obtain daily log-values of the portfolio: $p_{t+1}^{(q)}(\alpha_t) = (p_{t+1,1}^{(q)}(\alpha_t), \cdots, p_{t+1,H}^{(q)}(\alpha_t))$, where $p_{t+1,h}^{(q)}(\alpha_t) = \log(\alpha_t' \exp(R_{t+1,h}^{(q)}))$.

4. We predict the risk measures with simulated daily log-prices of the portfolio. For each simulation q, we compute the drawdown measures using the definitions given in Section 2.2, yielding $XDD_{t+1}^{(q)}(\alpha_t)$. The predictions of the drawdown measures are then given by the average over the Q simulations: $X\hat{D}D_{t+1}(\alpha_t) = \frac{1}{Q} \sum_{q=1}^{Q} XDD_{t+1}^{(q)}(\alpha_t)$, except for CED. To generate CED predictions, we rely on the MDD values obtained over all simulations $MDD_{t+1}(\alpha_t) = (MDD_{t+1}^{(1)}(\alpha_t), \cdots, MDD_{t+1}^{(Q)}(\alpha_t))$ and take the average of the worst $(1 - \tilde{\theta}) \times 100\%$ MDD values as in Equation (4).

5. We iterate points 3 and 4 over α_t until the optimal portfolio weight vector α_t^* is found for the investment problem (5). To solve investment problem (6), we also predict the expected return of the portfolio for the next quarter $t + 1$ as $E_t[R_{p,t+1}(\alpha_t)] = \frac{1}{Q} \sum_{q=1}^{Q} \left[\exp\left(p_{t+1,H}^{(q)}(\alpha_t) - p_{t,H}^{(q)}(\alpha_t) \right) - 1 \right]$.

To obtain accurate estimates of the optimal weights, we simulate a large number of draws ($Q = 50,000$).

3. Results

3.1. Data

Our empirical application is based on two size portfolios constructed using the Fama and French (1993) methodology. Small caps include the firms with the lowest market capitalization (bottom 30%), while large caps consist of the firms with the largest market capitalization (top 30%).[11] The portfolios comprise all NYSE, AMEX, and NASDAQ stocks for which market equity data are available. The sample spans from July 1926 to December 2020, encompassing a total of 24,896 daily returns.[12]

Size portfolios offer several advantages for analyzing the construction of a portfolio in the context of large drawdowns. First, data on these portfolios is available over an extensive period (nearly 100 years). Second, size portfolios have been the focus of large attention for decades, making their properties relatively well understood, particularly in terms of their risk and return characteristics. Third, numerous studies have examined small-cap and large-cap stocks due to their distinct behaviors during crises and varying market conditions, such as bull and bear markets. Perez-Quiros and Timmermann (2000) provide evidence that small caps exhibit a high degree of asymmetry between recessionary and expansionary states. During recessions, small caps are more strongly impacted than large caps by deteriorating credit market conditions. Ang and Chen (2002) and Patton (2004) investigate the dependence between small and large caps, focusing on their asymmetric behavior during bull and bear markets. Huang et al. (2012) report that small firms are more exposed

to extreme downside risks, and that the higher average returns of small caps actually compensate investors for the occurrence of larger drawdowns. The COVID-19 market crash illustrates this phenomenon: in the first quarter of 2020, small caps experienced a 45% decline, whereas large caps decreased by only 33%. From this perspective, we compare the optimal weights allocated to size portfolios based on the minimization of large drawdowns and the minimization of standard portfolio variance.

Table 1 reports statistics on small and large caps. Panel A corresponds to the full sample (1926–2020, 378 quarters), while Panel B focuses on the out-of-sample period that we used for the investment analysis (1990–2020, 124 quarters). The first part of the table displays descriptive statistics and standard risk measures. Small caps exhibit higher annual return and higher volatility on average. The Value-at-Risk (VaR) and expected shortfall (ES) measures demonstrate that small caps are more prone to large adverse shocks. The overall MDD is equal to 92% for small caps and 86.5% for large caps, corresponding in both cases to the stock market crash of 1929–1932.

The second part of the table reports the four sample measures of large drawdowns described in Section 2.2, for horizons of one quarter, two quarters, and four quarters. We compute CDD for a probability $\theta = 0.8$, i.e., we consider the average of the worst 20% drawdowns in a given subsample (e.g., the worst 12 drawdowns in a given quarter) (see Chekhlov et al. 2005). We compute CED with a probability $\tilde{\theta} = 0.9$, which corresponds to the worst 10% of MDD values in the sample (the worst 12 MDD values in the out-of-sample period) (see Goldberg and Mahmoud 2016). By examining the four period drawdown measures, we find that they also are all greater for small caps than large caps. On average, the one-quarter MDD on small caps is larger by approximately 2.3% and the two-quarter MDD is larger by 4%. ADD and CDD exhibit similar patterns. CED is also substantially higher for small caps than for large caps (by roughly 8% over one quarter and 11% over one year). This evidence suggests that, despite higher expected returns for small caps, investors may be reluctant to invest in small caps because they are more exposed to extreme downside risk, particularly over the long term, as suggested by Ang and Chen (2002) and Huang et al. (2012).

The table also reports relatively high first-order autocorrelation coefficients in ADD, CDD, and MDD measures in the full sample. The four-quarter MDD has AR(1) parameters equal to 48% for small caps and 45% for large caps. In the out-of-sample period (Panel B: 1990–2020), large drawdowns are much less persistent. The AR(1) parameter is usually low and close to 0 for small caps at all horizons, although it remains slightly higher for large caps. For the four-quarter MDD, the AR(1) parameters are equal to 11% for small caps and 38% for large caps.

Figure 1 presents the temporal evolution of the large drawdown measures. As expected, it reveals that, over the last century, four periods have been accompanied by large drawdowns: the Great Depression (1929–1933), the oil crisis (1973–1979), the subprime crisis (2008–2012), and the COVID-19 downturn (2020). The figure also displays the 10% CED for each horizon over the full sample (horizontal lines on the right-hand side plots). With the four-quarter CED as a threshold, we identify only two exceptional drawdowns (the Great Depression and the subprime crisis episodes), whereas with the two-quarter CED we would also include the COVID-19 downturn as an exceptional event.

3.2. Full-Sample Model Estimation

In this section, we evaluate the ability of MS-GARCH models to predict large drawdowns over the full sample (1926–2020). This unique estimation helps interpret the parameter estimates and formally test the number of regimes and the choice of the innovation distribution. Tables 2 and 3 report parameter estimates for the models with one, two, and three regimes, with normal and Student's t innovations. Table 4 reports likelihood ratio (LR) test statistics, which we use to identify the model that best reproduces the data properties. We first consider the one-regime model, i.e., the standard multivariate GARCH model with constant conditional correlation. Parameter estimates of expected returns, volatility dynam-

ics, and their correlation are fairly standard and similar for both innovation distributions. The degree of freedom of the Student's t distribution is equal to 5.68, which suggests that innovations have relatively heavy tails. The LR test rejects the null hypothesis that the distribution is normal.

The properties of the models with two regimes are very different depending on the distribution assumed. With normal innovations, the second regime has large negative expected returns for both assets. Two distinct regimes are clearly identified: the first regime pertains to normal conditions and the second regime corresponds to the bear state, possibly associated with market downturns. Our estimates are consistent with the high degree of asymmetry of small caps highlighted by Perez-Quiros and Timmermann (2000): In the bear market, expected returns are much lower for small caps than for large caps, probably reflecting tighter credit market conditions. The probability of remaining in the bear state is relatively low ($p_{22} = 68.3\%$), with a stationary probability equal to 22%.[13]

Table 1. Summary statistics on daily returns and period drawdowns for small caps and large caps.

	Panel A: 1926–2020				Panel B: 1990–2020			
	Small Caps		Large Caps		Small Caps		Large Caps	
Daily returns	Stat.		Stat.		Stat.		Stat.	
Annualized Mean	10.63		8.99		10.38		9.88	
Annualized Std dev.	19.50		17.15		20.36		18.06	
Skewness	−0.39		−0.48		−0.81		−0.40	
Kurtosis	23.61		21.83		13.40		14.03	
Maximum	20.42		14.15		8.02		11.16	
Minimum	−16.75		−20.94		−14.26		−12.57	
VaR (0.1%)	8.53		6.86		9.01		7.56	
VaR (1%)	3.70		3.11		3.72		3.26	
VaR (5%)	1.82		1.56		1.97		1.74	
ES (0.1%)	10.76		9.16		11.24		9.32	
ES (1%)	5.55		4.64		5.50		4.76	
ES (5%)	3.09		2.62		3.17		2.77	
Overall MDD	92.02		86.54		67.13		58.29	
Period drawdowns	Stat.	AR(1)	Stat.	AR(1)	Stat.	AR(1)	Stat.	AR(1)
ADD - 1Q	3.84	0.15	2.78	0.20	3.65	0.12	2.47	0.25
ADD - 2Q	5.62	0.29	3.85	0.42	5.19	0.00	3.30	0.35
ADD - 4Q	7.52	0.46	4.82	0.42	6.16	−0.10	3.83	0.35
20% CDD - 1Q	7.75	0.25	5.76	0.28	7.28	0.12	5.27	0.23
20% CDD - 2Q	11.80	0.41	8.20	0.46	10.81	0.05	7.14	0.30
20% CDD - 4Q	16.65	0.45	10.96	0.43	13.99	0.01	8.87	0.33
MDD - 1Q	9.87	0.33	7.55	0.34	9.34	0.17	7.10	0.26
MDD - 2Q	15.05	0.44	10.93	0.47	13.99	0.06	9.93	0.28
MDD - 4Q	21.52	0.48	15.23	0.45	18.67	0.11	13.08	0.38
10% CED - 1Q	31.22		22.84		25.97		20.60	
10% CED - 2Q	41.97		31.35		35.40		25.41	
10% CED - 4Q	51.37		40.25		38.36		31.19	

Note: This table reports statistics (in percentage) on daily returns for small caps and large caps. Panel A covers the period from 1926 to 2020. Panel B covers the period from 1990 to 2020. VaR(α) and ES(α) denote the Value-at-Risk and Expected Shortfall computed for probability α. "XDD - nQ" means that the XDD measure is computed over n quarters. "20% CDD" means that the CDD measure is computed for probability $\theta = 20\%$. "10% CED" means that the CED measure is computed for probability $\tilde{\theta} = 10\%$.
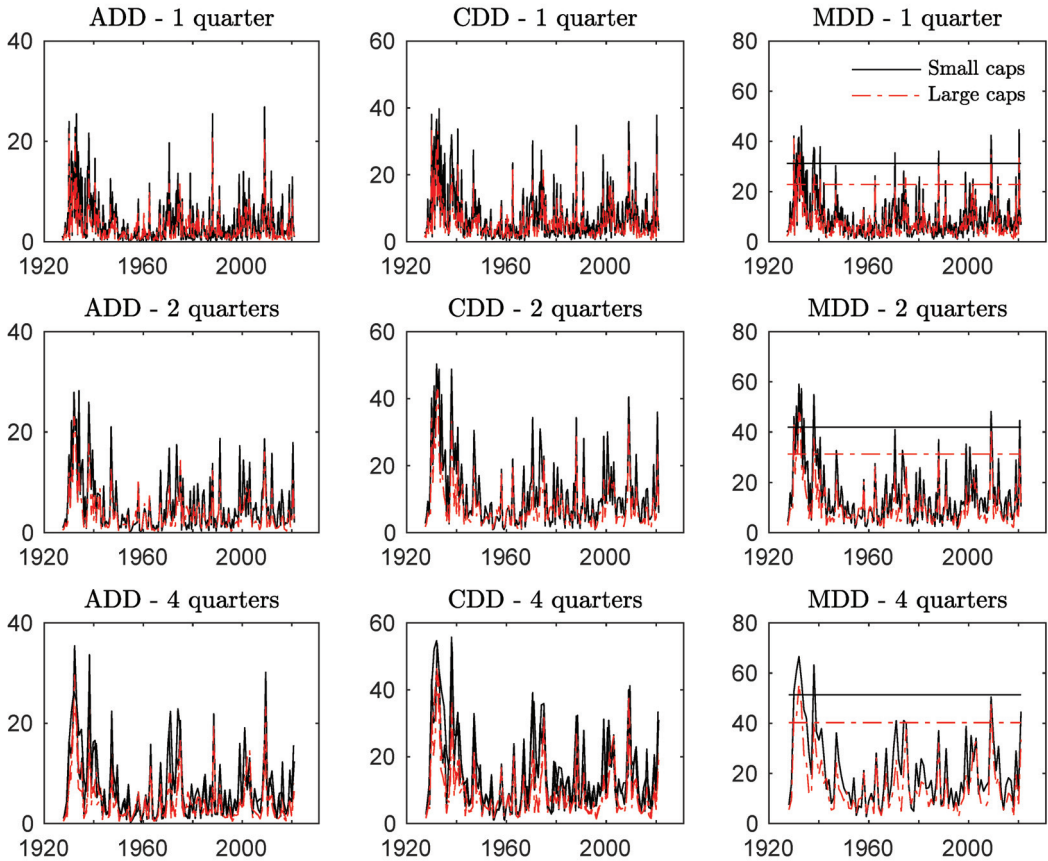
Figure 1. Evolution of ADD, CDD, and MDD over non-overlapping subsamples. The figure displays the evolution in percentage of ADD, 20% CDD, and MDD over various non-overlapping subsamples (from one to four quarters) between 1926 and 2020. The straight line on right plots corresponds to 10% CED. The black lines correspond to the small caps, the red dashed lines to the large caps. CED is computed with 376, 188, and 94 observations for the one-quarter, two-quarter, and four-quarter horizons, respectively.

In contrast, with Student's t innovations, the expected returns in Regime 2 are close to 0. The probability of remaining in the low regime is the same as the probability of remaining in the high regime ($p_{11} = 96.3\%$ and $p_{22} = 96.7\%$), so that the stationary probability of being in Regime 2 is as high as $\pi_{\infty,2} = 53\%$. As a consequence, Regime 2 cannot be interpreted as a pure bear state. These results suggest that, in this model, the occurrence of large drawdowns is not captured by large negative expected returns but instead by the heavy-tailed nature of the Student's t innovations.[14]

For the three-regime models, expected returns, volatility dynamics, and the correlation are similar for both innovation distributions. The results align with our expectations: Regime 1 captures the long periods during which the stocks are in a bull market (with high expected returns). Regime 2 corresponds to the slow growth or recovery regime, with intermediate expected returns. Regime 3 accounts for bear market conditions. As in the two-regime case, small caps have much higher expected returns than large caps in Regime 1 and much lower expected returns than large caps in Regime 3. Regime 3 also exhibits a

higher correlation than Regime 1, reflecting the asymmetry in dependence found by Ang and Chen (2002).

Table 2. Parameter estimates for one-regime and two-regime models.

	One regime—Normal distribution				One regime—Student's t distribution			
	Small caps		Large caps		Small caps		Large caps	
	param.	std err.	param.	std err.	param.	std err.	param.	std err.
Expected returns								
μ ($\times 100$)	0.0679	(0.006)	0.0588	(0.005)	0.0865	(0.004)	0.0687	(0.004)
Volatility dynamics								
ω ($\times 100$)	1.4840	(0.201)	0.9667	(0.122)	0.9902	(0.093)	0.7532	(0.071)
α	0.1298	(0.008)	0.0984	(0.006)	0.1220	(0.007)	0.0970	(0.004)
β	0.8654	(0.008)	0.8962	(0.006)	0.8764	(0.006)	0.9010	(0.004)
Correlation								
ρ	0.8291	(0.003)			0.8212	(0.002)		
Degree of freedom								
ν	–				5.6810	(0.154)		
Log-lik.	−48,291.3				−46,183.1			
BIC	96.6736				92.4674			
	Two regimes—Normal distribution				**Two regimes—Student's t distribution**			
	Small caps		Large caps		Small caps		Large caps	
	param.	std err.	param.	std err.	param.	std err.	param.	std err.
Expected returns								
$\mu^{(1)}$ ($\times 100$)	0.1265	(0.007)	0.0965	(0.006)	0.2093	(0.035)	0.1155	(0.012)
$\mu^{(2)}$ ($\times 100$)	−0.2706	(0.043)	−0.1771	(0.030)	−0.0276	(0.016)	0.0207	(0.010)
Volatility dynamics								
$\omega^{(1)}$ ($\times 100$)	0.2974	(0.052)	0.3132	(0.049)	1.2510	(0.293)	0.8278	(0.140)
$\alpha^{(1)}$	0.0418	(0.005)	0.0417	(0.004)	0.2055	(0.021)	0.1309	(0.017)
$\beta^{(1)}$	0.9209	(0.008)	0.9289	(0.006)	0.7968	(0.020)	0.8663	(0.019)
$\omega^{(2)}$ ($\times 100$)	0.1550	(0.380)	0.5501	(0.279)	0.1576	(0.169)	0.2482	(0.195)
$\alpha^{(2)}$	0.1868	(0.024)	0.1463	(0.019)	0.0411	(0.023)	0.0422	(0.021)
$\beta^{(2)}$	0.9173	(0.011)	0.9291	(0.008)	0.9563	(0.023)	0.9565	(0.022)
Correlation								
$\rho^{(1)}$	0.8179	(0.004)			0.7450	(0.014)		
$\rho^{(2)}$	0.8334	(0.006)			0.8884	(0.010)		
Transition probabilities								
p_{11}	0.9081	(0.009)			0.9626	(0.024)		
p_{22}	0.6828	(0.027)			0.9669	(0.019)		
Degree of freedom								
ν	–				6.2260	(0.252)		
Log-lik.	−45,974.7				−45,420.1			
BIC	92.1478				91.0507			

Note: See Table 3 for details.

The degree of freedom of the Student's t distribution is equal to 7.8, suggesting relatively thin tails compared to the two-regime model. Similar to the two-regime models, the Student's t innovation partly captures the occurrence of large negative returns, as expected returns in Regime 3 are higher with Student's t innovations. As a result, the stationary probability of being in Regime 3 is higher than in the model with normal innovations (30.4% versus 19.5%).

Table 3. Parameter estimates for three-regime model.

| | Three Regimes—Normal Distribution | | | | Three Regimes—Student's t Distribution | | | |
| | Small Caps | | Large Caps | | Small Caps | | Large Caps | |
	param.	std err.	param.	std err.	param.	std err.	param.	std err.
Expected returns								
$\mu^{(1)}$ ($\times100$)	0.3304	(0.022)	0.1174	(0.011)	0.3194	(0.017)	0.1251	(0.011)
$\mu^{(2)}$ ($\times100$)	0.0754	(0.008)	0.0966	(0.007)	0.1193	(0.008)	0.1133	(0.008)
$\mu^{(3)}$ ($\times100$)	−0.3451	(0.024)	−0.1897	(0.024)	−0.2316	(0.017)	−0.1049	(0.020)
Volatility dynamics								
$\omega^{(1)}$ ($\times100$)	0.0001	(0.191)	0.1797	(0.108)	0.5529	(0.205)	0.4921	(0.141)
$\alpha^{(1)}$	0.0403	(0.019)	0.0325	(0.009)	0.1303	(0.022)	0.0927	(0.013)
$\beta^{(1)}$	0.9480	(0.022)	0.9534	(0.012)	0.8585	(0.022)	0.8945	(0.013)
$\omega^{(2)}$ ($\times100$)	0.2705	(0.046)	0.3461	(0.060)	0.0201	(0.200)	0.0858	(0.040)
$\alpha^{(2)}$	0.0401	(0.005)	0.0439	(0.004)	0.0122	(0.003)	0.0177	(0.003)
$\beta^{(2)}$	0.9180	(0.008)	0.9233	(0.006)	0.9739	(0.006)	0.9691	(0.006)
$\omega^{(3)}$ ($\times100$)	0.0001	(0.640)	0.8492	(0.405)	0.1357	(0.310)	0.6340	(0.358)
$\alpha^{(3)}$	0.2262	(0.033)	0.1655	(0.022)	0.1328	(0.031)	0.1050	(0.022)
$\beta^{(3)}$	0.9005	(0.018)	0.9246	(0.011)	0.9145	(0.022)	0.9306	(0.017)
Correlation								
$\rho^{(1)}$	0.7814	(0.025)			0.7300	(0.016)		
$\rho^{(2)}$	0.8487	(0.006)			0.8465	(0.009)		
$\rho^{(3)}$	0.8462	(0.008)			0.8764	(0.007)		
Transition matrix								
$P_{1,:}$	0.9206	0.0169	0.0453		0.9402	0.0170	0.0266	
	(0.015)	(0.003)	(0.018)		(0.009)	(0.004)	(0.007)	
$P_{2,:}$	0.0350	0.8954	0.2663		0.0165	0.9116	0.1139	
	(0.016)	(0.008)	(0.026)		(0.007)	(0.011)	(0.013)	
$P_{3,:}$	0.0445	0.0877	0.6884		0.0433	0.0715	0.8595	
	(0.008)	(0.006)	(0.021)		(0.008)	(0.009)	(0.014)	
Degree of freedom								
ν	–				7.7750	(0.399)		
Log-lik.	−45,282.0				−44,852.6			
BIC	90.9244				90.0753			

Note: Tables 2 and 3 report parameter estimates for the models with daily returns on small caps and large caps. Table 2 reports estimates of the one-regime and two-regime models. Table 3 reports estimates of the three-regime models. The estimation is based on the period spanning from 1926 to 2020. BIC denotes the Bayesian Information Criterion.

A formal test of the number of regimes can be performed using the Likelihood-Ratio (LR) test but the usual asymptotic distribution of the test statistic does not hold. The reason is that, in the test of the null hypothesis that $n-1$ regimes are sufficient against the alternative of n regimes, parameters associated with the n-th regime are not identified under the null hypothesis and the regularity conditions justifying the χ^2 approximation to the LR test do not hold. Hansen (1992, 1996) proposed an LR test procedure that addresses this problem (see also Garcia 1998). Specifically, we adopt the strategy proposed by Ang and Bekaert (2002), which is based on Monte-Carlo simulations to obtain the finite-sample distribution of the LR test statistic.[15] We implement this approach for all tests of the number of regimes from 1 to 3 (see Table 4). For 1 and 2 regimes, we reject the null hypothesis, with p-values all below 0.5%, whatever the distribution of the innovation process. These tests indicate that the one-regime model should be rejected against the two-regime model and the two-regime model should be rejected against the three-regime model.

We also estimate four-regime models (with normal and Student's t innovations) with our data and test whether 3 regimes are sufficient to match the data. Compared to three regimes, the gain in likelihood with four regimes is relatively large, but using the simulation-

based LR distribution, we do not reject the null hypothesis of three regimes against four regimes, with a *p*-value equal to 20.2% with the normal distribution and to 19% with the Student's t distribution.[16]

Table 4. Likelihood ratio tests.

Null Hypothesis	Alternative Hypothesis	dof	LR Stat.	*p*-Value
H0(N1): 1 regime—normal	Ha(N1): 1 regime—Student's t	1	4216.3	<0.5%
H0(NR1): 1 regime—normal	Ha(NR1): 2 regimes—normal	11	4637.0	<0.5%
H0(TR1): 1 regime—Student's t	Ha(TR1): 2 regimes—Student's t	11	1528.0	<0.5%
H0(N2): 2 regimes—normal	Ha(N2): 2 regimes—Student's t	1	1107.3	<0.5%
H0(NR2): 2 regimes—normal	Ha(NR2): 3 regimes—normal	13	1385.3	<0.5%
H0(NT2): 2 regimes—Student's t	Ha(NT2): 3 regimes—Student's t	13	1137.2	<0.5%
H0(N3): 3 regimes—normal	Ha(N3): 3 regimes—Student's t	1	859.2	<0.5%
H0(NR3): 3 regimes—normal	Ha(NR3): 4 regimes—normal	15	924.0	20.2%
H0(NT3): 3 regimes—Student's t	Ha(NT3): 4 regimes—Student's t	15	772.0	19.0%

Note: The table reports the likelihood ratio test statistics for various tests of interest. The first two columns indicate the null and alternative hypotheses. The third column reports the degree of freedom (dof) of the test (number of restrictions under the null hypothesis). The fourth column reports the LR test statistics. As explained in the main text, the *p*-values in the fifth column are based on the asymptotic χ^2 distribution for the test of the null hypothesis of the normal distribution (N1, N2, and N3) and on simulations of the finite-sample distribution for the test of the null of $n - 1$ regimes against n regimes.

In Figure 2, we represent the filtered probability of being in the low regime for the two-regime and three-regime models. First, we note that the two-regime/Student's t model produces a high filtered probability (on average above 50%), suggesting that this regime actually does not capture bear markets. Second, the filtered probabilities in the two-regime and three-regime models with normal innovations exhibit similar temporal evolution. Peaks occur at the same times with similar probabilities. However, these peaks do not always coincide with actual market downturns. The first peak occurs in June 1932, after the Wall Street crash of October 1929. The second peak corresponds to the oil crash in mid-1973. The third peak in May 1984 could not be associated with any market downturn. The fourth peak, which occurred in October 1999, corresponds to the dotcom crash. The last peak in September 2014 again does not correspond to any large market decline. Consequently, models with normally distributed innovations do not accurately capture observed market downturns.

In the three-regime model with Student's t innovations, most peaks in the filtered probability actually correspond to market events associated with a long-lasting bear market. We can identify three main episodes: The first one corresponds to the bear market at the beginning of the period (with peaks in the second half of 1929 and at the end of 1937, associated with the Wall Street crash and the economic recession, respectively). The second episode corresponds to the inflationary bear market of the seventies (with peaks at the end of 1969 and mid-1973).[17] The third episode is associated with the market crashes at the turn of the new century (Russian crisis in 1998, dotcom crash in 2001 and financial crisis in 2008). In the more recent period, the filtered probability also increased in mid-2015 (associated with the stock market sell-off following the ending of quantitative easing by the Federal Reserve) and at the beginning of 2020 (associated with the COVID-19 market crash).[18]

The analysis of filtered probabilities clearly suggests that the three-regime/Student's t model provides a better description of the market downturns observed in the sample period than the other competing models.
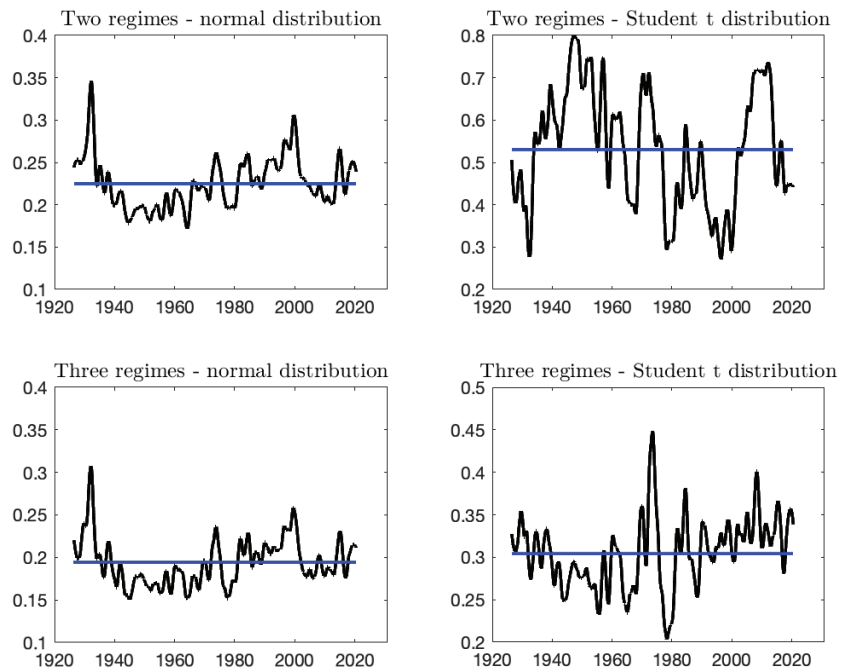
Figure 2. Filtered Probability of Being in the Bear State. The figure displays the filtered probability ϕ_{d+1} of being in the low expected return regime (bear state), for the two-regime and three-regime models. The horizontal blue line corresponds to the stationary probability of being in the bear state $\pi_{b,\infty} = \Pr[S_{d+1} = k_b]$, where k_b denotes the bear state.

Another related interesting feature of MS models is that they can generate asymmetry in the distribution of returns even if the volatility dynamics and the innovation distribution are symmetric. The reason for this property is the possible shift from one regime to another, which generates relatively large (positive or negative) events. To assess whether the various models are able to generate the asymmetry that we observe in the data (see Table 1), we simulate long trajectories of the small caps and large caps daily returns and compute the skewness of the simulated series. As reported in Table 5, the one-regime models do not generate any asymmetry because this property is absent from the model. Two-regime models are able to produce some asymmetry but it is clearly insufficient to match the data. For instance, in the normal distribution case, the skewness measures of the small caps and large caps returns are equal to -0.19 and -0.16, respectively. This is well above the sample skewness that we obtained over the full sample (-0.39 for small caps and -0.48 for large caps, respectively). In contrast, three-regime models produce skewness measures that are in the ball park of the sample measures. In the normal model, the skewness estimates are equal to -0.8 and -0.5 for small caps and large caps, respectively. In the Student's t model, the values are equal to -0.5 and -0.3. Again, this analysis suggests that three regimes are necessary to match the extreme behavior of actual returns.

The table also reports predictions of the large drawdown measures for the three horizons. Again, one-regime models and the two-regime/Student's t model fail at capturing the magnitude of large drawdown measures obtained with the data, by substantially underestimating drawdowns. The two-regime/normal model is able to capture the magnitude of drawdowns on average but fail at generating the asymmetry observed between small caps and large caps. The drawdowns are usually similar for both size portfolios. In contrast, three-regime models are able to generate this asymmetry, although it is often not as large as with the observed data.

Table 5. Predictions of Moments and Drawdown Measures.

	Sample Data (1926–2020)		1-Regime Normal		1-Regime Student t		2-Regime Normal		2-Regime Student t		3-Regime Normal		3-Regime Student t	
	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large	Small	Large
Skewness	−0.39	−0.48	0.01	0.00	0.00	0.00	−0.19	−0.17	0.17	0.01	−0.84	−0.47	−0.50	−0.26
Kurtosis	23.92	21.98	16.65	10.73	42.01	36.66	13.43	13.32	72.30	28.86	14.70	14.07	14.04	13.88
ADD														
1 quarter	3.84	2.78	2.55	2.32	2.16	2.13	3.76	3.43	2.40	2.21	3.66	3.29	3.42	2.91
2 quarters	5.62	3.85	3.42	3.11	2.68	2.75	5.62	5.36	3.14	2.91	5.34	4.92	5.33	4.74
4 quarters	7.52	4.82	4.53	4.08	3.22	3.46	7.88	7.68	3.86	3.66	7.31	6.85	7.69	7.18
20% CDD														
1 quarter	7.75	5.76	5.41	4.95	4.72	4.62	7.93	7.40	5.20	4.79	7.74	7.11	7.58	6.47
2 quarters	11.80	8.20	7.35	6.70	6.02	6.06	11.65	11.25	6.95	6.39	11.25	10.41	11.78	10.57
4 quarters	16.65	10.96	10.05	9.01	7.71	7.94	16.14	15.80	8.84	8.27	15.41	14.40	16.83	15.68
MDD														
1 quarter	9.87	7.55	7.05	6.44	6.22	6.05	10.20	9.62	6.80	6.25	9.94	9.25	10.04	8.67
2 quarters	15.05	10.93	9.84	8.92	8.27	8.18	14.86	14.41	9.25	8.51	14.38	13.46	15.49	14.13
4 quarters	21.52	15.23	13.96	12.41	11.26	11.23	20.54	20.15	12.17	11.43	19.78	18.63	21.98	20.68
10% CED														
1 quarter	31.22	22.84	12.02	10.63	11.25	10.37	20.59	19.29	12.14	10.64	20.74	19.10	20.85	17.48
2 quarters	41.97	31.35	17.22	15.06	15.87	14.74	27.90	26.86	16.81	14.83	27.82	25.75	29.74	26.97
4 quarters	51.37	40.25	20.59	17.89	17.73	17.16	31.19	30.59	18.40	16.89	30.64	28.55	33.57	31.51

Note: This table reports predictions of some statistics on interest based on simulations of the various models. The statistics the skewness, the kurtosis, the ADD, the 20% CDD, the MDD, and the 10% CED. For large drawdown measures, we consider three investment horizons: one quarter, two quarters, and four quarters. The simulations are based on the parameter estimates obtained over the full sample (1926–2020). Statistics are computed using 1000 draws.

3.3. Rolling-Window Model Estimation and Adequacy Tests

To implement the out-of-sample allocation strategy, we use a rolling window to estimate the parameters of the MS-GARCH models over subsample periods. For each model, the first set of parameters is estimated over the sample of daily returns from January 1927 to December 1989, while the last window covers the sample from January 1958 to December 2020. These parameter estimates will be used to simulate paths of daily log-returns of length H and predict next-period drawdown measures.

The temporal evolution of parameter estimates for the competing models is displayed in Appendix A. Comparison with the full sample estimates in Tables 2 and 3 reveals a remarkable match between the two sets of estimates on average. The figures demonstrate that parameter estimates are usually rather stable over time, although a few parameters exhibit trends. In particular, the correlation between small and large caps tends to decrease in Regimes 1 and 3 in the three-regime model. The degree of freedom of the Student's t distribution tends to increase in the two-regime and three-regime models.

We assess the adequacy of out-of-sample estimation of our models with respect to returns for small and large caps by backtesting predicted $C\hat{D}D_{\theta,t}$ and $M\hat{D}D_t$ obtained through simulations.[19] To perform these tests, we adopt the approach proposed by Acerbi and Szekely (2014) for testing expected shortfall estimates. We adjust their testing framework for the unconditional coverage of the CDD measure.

The methodology and the main results are reported in Appendix B. In a nutshell, our adequacy tests reveal that one- and two-regime models inaccurately predict drawdown measures for both small and large caps. Specifically, the one-regime and two-regime/Student's t models underestimate drawdown measures, while the two-regime/normal and three-regime/normal models overestimate realized CDD and MDD for large caps. The one-regime and two-regime/Student's t models systematically underestimate drawdown measures for both small and large caps. The only model that perform relatively well for both small and large caps is the three-regime/Student's t model, with the expected number

of exceedances and average drawdown above the threshold being close to the numbers observed in the data.

3.4. Out-of-Sample Analysis

We now consider investors who allocate their wealth in real time period by period, with an investment horizon from one quarter to one year between 1990 and 2020. The out-of-sample strategy is implemented as follows. We use the rolling-window estimation of the models to predict next-period drawdown measures. With these predictions, we determine the optimal portfolio weight that minimizes the expected drawdown measures. Next, we roll the window by one period (H days) and proceed in the same way until we reach the last subsample (ending in September 2020). This out-of-sample analysis corresponds to 124 nonoverlapping quarterly allocations and 31 nonoverlapping annual allocations.

Figure 3 displays the evolution of the optimal small-cap weights obtained by mini-mizing large drawdown measures based on the various models for a two-quarter horizon. As illustrated, the evolution of optimal weights exhibits notable differences across the approaches used for prediction. For the one-regime models and the two-regime/Student's t model, the optimal small-cap weights are all positive, regardless of the targeted large drawdown measure. This finding implies that a diversified portfolio, with weights close to 50%, could provide effective diversification against large drawdowns. This result indi-cates that these models do not effectively generate the large drawdowns observed in the small-cap portfolio. Conversely, the two-regime/normal and three-regime/normal models display similar patterns, with optimal small-cap weights close to 0. This suggests that the innovation process plays a critical role in generating sufficiently large drawdowns for small caps, potentially resulting in negative weights.

The three-regime model/Student's t model displays optimal weights that are usually negative. This finding can be explained by the ability of the three-regime/Student's t model to generate large and negative expected returns in a relatively long-lasting Regime 3, which allows large drawdowns in small caps to develop and therefore to identify that holding small caps implies higher tail risk. As a consequence, this model produces large negative weights for most allocation criteria and investment horizons.[20]

In general, the optimal weights are ordered in the same way for the various investment criteria: the weight of small caps is higher for CDD, then for MDD, and finally for CED. We compare these weights with those resulting from the minimization of the portfolio variance. Minimum variance (MV) portfolio weights are obtained using the same sim-ulation approach. Results indicate that, for all investment horizons, the optimal weight of the MV portfolio is always negative, in the range $[-40\%; -20\%]$ for the two-regime and three-regime models. Figure 3 also demonstrates that the weight of small caps is systematically lower for the MV criterion than for criteria based on large drawdowns.[21]

These patterns suggest that investors targeting large drawdowns tend to be even more cautious in their allocation than MV investors, as they are reluctant to take substantial short positions in small caps.

Table 6 reports results for the out-of-sample allocation when the investor minimizes the expected value of large drawdowns. We compare the performance of the allocation based on the various parametric models. The standard (one-regime) multivariate GARCH models (with normal and Student's t distributions) and the two-regime/Student's t models fail at capturing that the risk of large drawdowns is higher for small caps. Therefore, these models generate a large weight on average for small caps, for all investment horizons. As small caps experienced larger drawdowns than large caps in our sample, strategies based on these models tend to underperform and suffer from much higher ex post drawdowns on average. For instance, for the two-quarter CDD, strategies based on these models have average small caps weights equal to 0.46, 0.63, and 0.39, respectively. Their ex post two-quarter CDD values are equal to 8.47%, 9.16%, and 8.30% on average. In contrast, the two-regime/normal model and the three-regime models (with normal and Student's t

distributions) tend to have negative small cap weights on average, with substantially lower ex post two-quarter CDD values, equal to 7.23%, 7.26%, and 7.15%, respectively.
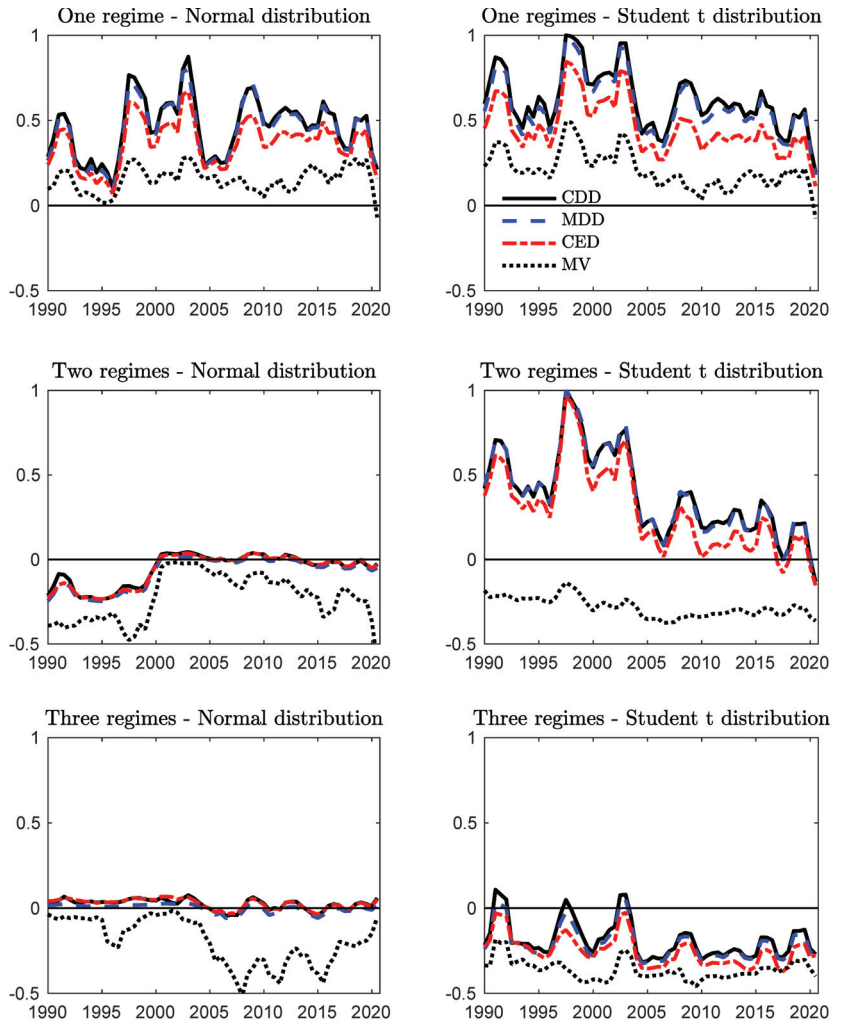
Figure 3. Out-of-Sample Optimal Weights—Two-quarter Horizon (1990–2020). The figure displays the temporal evolution of the optimal weight of small caps for the 20% CDD, MDD, 10% CED, and MV portfolios over the two-quarter horizon, when predictions are based on the one-regime, two-regime, and three-regime models.

It is worth emphasizing that, in all cases, the strategies based on the standard multivariate GARCH models result in much larger drawdown measures than strategies based on three-regime models. The gap is economically substantial for the three measures CDD, MDD, and CED. The ex post four-quarter CDD of the one-regime/normal model is 2.1 points higher that the same measure of the three-regime/Student's t model. The difference is equal to 2.3 points for the four-quarter MDD and 1.7 points for the four-quarter CED. The gaps are less severe for the two-regime/normal model but they remain large for long horizons. They are equal to 2.4 points for the four-quarter CDD, 0.6 point for the four-quarter MDD, 0.4 point for the four-quarter CED.

Overall, the three-regime/normal model performs well for the one-quarter horizon, while the three-regime/Student's t model performs the best for the two- and four-quarter horizons. In all cases, the best models generate an optimal weight that is negative or close to 0.

Table 6. Out-of-Sample Allocation: Minimization of Large Drawdowns.

Horizon	Statistics	1-Regime Normal	1-Regime Student t	2-Regime Normal	2-Regime Student t	3-Regime Normal	3-Regime Student t
Panel A: Minimization of CDD							
1 quarter	Weight	0.44	0.54	−0.11	0.41	−0.17	0.07
	20% CDD	5.75	5.90	5.24	5.66	**5.23**	5.27
2 quarters	Weight	0.46	0.63	−0.06	0.39	0.02	−0.18
	20% CDD	8.47	9.16	7.23	8.30	7.26	**7.15**
4 quarters	Weight	0.51	0.67	0.42	0.47	0.87	−0.07
	20% CDD	11.36	12.66	11.62	11.34	14.28	**9.22**
Panel B: Minimization of MDD							
1 quarter	Weight	0.43	0.53	−0.12	0.39	−0.17	0.05
	MDD	7.56	7.71	7.04	7.40	7.10	**7.02**
2 quarters	Weight	0.44	0.59	−0.08	0.39	0.00	−0.21
	MDD	11.31	11.95	10.10	11.14	10.06	**9.92**
4 quarters	Weight	0.48	0.64	0.09	0.49	0.18	−0.10
	MDD	16.08	17.15	14.42	16.16	14.80	**13.82**
Panel C: Minimization of CED							
1 quarter	Weight	0.37	0.44	−0.16	0.26	−0.21	−0.14
	10% CED	22.25	22.53	20.69	21.43	**20.30**	20.53
2 quarters	Weight	0.37	0.47	−0.06	0.30	0.03	−0.25
	10% CED	29.24	30.03	25.26	28.80	25.56	**24.84**
4 quarters	Weight	0.41	0.52	0.02	0.43	0.00	−0.06
	10% CED	35.22	36.40	33.96	36.51	33.96	**33.56**

Note: This table reports the optimal weight of small caps (α^*) and the value of the objective function at the optimum, when the investor minimizes 20% CDD, the MDD, and 10% CED (Panels A to C, respectively). We consider three investment horizons: one quarter, two quarters, and four quarters. These results are based on the simulation of the model estimated over subsamples between 1990 and 2020. Numbers in bold correspond to the smallest values of the large drawdown measures.

We also consider an investor maximizing the expected return—large drawdown criterion (Equation (6)), which also accounts for the ability of the models to predict expected returns. Results are reported in Table 7 with a degree of aversion for large drawdowns equal to $\lambda = 5$. Optimal weights for small caps are usually higher, reflecting the higher expected return of small caps. For the one-regime models and the two-regime/Student's t model, small caps weights are even often larger than one, indicating the large caps are shorted in these allocations. Turning to the two-regime/normal model and three-regime models, we find that weights are also higher than in the case of the minimization of large drawdowns but to a lesser extent. For the MDD and CED targets, small caps weights are negative for a short horizon but close to 50% for two-quarter and four-quarter horizons. These results indicate that investors also targeting the expected return would invest more in small caps because of their higher return on average.

Table 7. Out-of-Sample Allocation: Maximization of Expected Return—Large Drawdown Criterion.

Horizon	Statistics	1-Regime Normal	1-Regime Student t	2-Regime Normal	2-Regime Student t	3-Regime Normal	3-Regime Student t
Panel A: Expected Return—CDD criterion							
1 quarter	Weight	0.47	0.59	0.12	1.10	0.17	0.62
	Criterion	−0.020	−0.059	0.015	−0.037	0.029	−0.004
	Opp. cost (%)	4.87	8.76	1.43	6.62	0.00	3.32
2 quarters	Weight	0.48	0.64	0.56	1.40	0.62	0.29
	Criterion	−0.070	−0.161	−0.026	−0.119	−0.031	−0.004
	Opp. cost (%)	6.63	15.70	2.23	11.53	2.73	0.00
4 quarters	Weight	0.54	0.73	0.48	1.83	0.50	0.65
	Criterion	−0.175	−0.323	−0.055	−0.293	−0.058	−0.075
	Opp. cost (%)	11.66	26.43	−0.30	23.50	0.00	1.68
Panel B: Expected Return—MDD criterion							
1 quarter	Weight	0.62	0.85	−0.02	0.85	−0.03	0.31
	Criterion	−0.091	−0.111	−0.076	−0.113	−0.067	−0.074
	Opp. cost (%)	2.36	4.40	0.94	4.57	0.00	0.67
2 quarters	Weight	0.68	1.01	0.43	1.01	0.60	−0.01
	Criterion	−0.203	−0.258	−0.181	−0.254	−0.203	−0.147
	Opp. cost (%)	5.57	11.07	3.42	10.71	5.64	0.00
4 quarters	Weight	0.78	1.28	0.49	1.42	0.51	0.50
	Criterion	−0.319	−0.442	−0.281	−0.478	−0.284	−0.277
	Opp. cost (%)	4.23	16.49	0.45	20.14	0.70	0.00
Panel C: Expected Return—CED criterion							
1 quarter	Weight	0.89	1.33	−0.11	0.50	−0.15	−0.03
	Criterion	−0.463	−0.481	−0.424	−0.460	−0.406	−0.409
	Opp. cost (%)	5.66	7.48	1.78	5.41	0.00	0.28
2 quarters	Weight	1.05	1.73	0.45	0.57	0.69	−0.16
	Criterion	−0.648	−0.684	−0.616	−0.665	−0.681	−0.526
	Opp. cost (%)	12.21	15.87	9.03	13.95	15.49	0.00
4 quarters	Weight	1.28	1.97	0.50	0.78	0.51	0.48
	Criterion	−0.788	−0.850	−0.762	−0.889	−0.762	−0.789
	Opp. cost (%)	2.59	8.86	0.03	12.73	0.00	2.78

Note: This table reports the optimal weight of small caps (α^*), the value of the objective function at the optimum, and the opportunity cost ($\xi^{(sub)}$) relative to the optimal model, when the investor maximizes the expected return—large drawdown criterion, with 20% CDD, the MDD, and 10% CED (Panels A to C, respectively). The degree of aversion for large drawdowns is equal to $\lambda = 5$. We consider three investment horizons: one quarter, two quarters, and four quarters. These results are based on the simulation of the model estimated over subsamples between 1990 and 2020.

Regarding the performance of the strategies, we obtain essentially the same conclusions as before. One-regime models are always dominated by models with at least two regimes. In addition, three-regime models usually outperform other models, in particular

for extreme drawdowns (MDD and CED). To illustrate the gain of using a three-regime model to predict large drawdowns, we compute the opportunity cost of using a one-regime or two-regime model, using the relation:

$$E_t[R_{p,t+1}(\alpha_t^{(sub)}) - \xi^{(sub)}] - \frac{\lambda}{2}E_t[XDD_{t+1}(\alpha_t^{(sub)})] = E_t[R_{p,t+1}(\alpha_t^{(opt)})] - \frac{\lambda}{2}E_t[XDD_{t+1}(\alpha_t^{(opt)})],$$

where $\alpha_t^{(sub)}$ and $\alpha_t^{(opt)}$ denote the vector of weights obtained using the suboptimal model and the optimal model, respectively. The opportunity cost $\xi^{(sub)}$ is defined as the fraction of the expected return that an investor using the suboptimal strategy is ready to pay to get access to the optimal strategy.

The estimates of the opportunity cost are also reported in the table. For the one-quarter investment horizon, the best strategy is the one based on the three-regime/normal model. In this case, an investor using the two-regime/normal model would be ready to pay 1% if the objective function is targeting the MDD and 1.8% if the objective function is targeting the CED. For the two-quarter investment horizon, the best strategy is the one based on the three-regime/Student's t model. In this case, an investor using the two-regime/normal model would be ready to pay 3.4% if the objective function is targeting the MDD and up to 9% if the objective function is targeting the CED. As these estimates illustrate, the gain of using a three-regime model is economically substantial for investors targeting large drawdown measures.

This out-of-sample analysis demonstrates that three regimes are necessary in the multivariate MS-GARCH model to produce large and negative weights for small caps. Student's t innovations also help generate a longer-lasting bear state and therefore to produce more drawdowns for small caps. Importantly, as the allocation exercise is performed in a fully out-of-sample fashion, we account for the risk of over-parameterization of the three-regime models.

4. Conclusions

This paper addresses the modeling and prediction of large drawdowns in financial markets. Given that the standard GARCH model struggles to capture prolonged market declines, we explore multivariate MS-GARCH models, which can generate regimes characterized by both positive and negative market trends. Provided that the probability of remaining in a bear regime is sufficiently high, these models are capable of reproducing significant drawdown properties. Specifically, our findings show that in three-regime models, one regime is marked by substantial negative expected returns (representing a bear market) and the capacity to generate large windfalls.

In the out-of-sample investment analysis, as changes in the distribution of drawdowns are updated, three-regime models prove to be superior tools for predicting expected drawdowns by imposing restrictions that align well with the observed data. These models consistently outperform other parametric models featuring one regime (such as standard multivariate GARCH models) or two regimes.

To assess the predictive power of MS-GARCH models for large drawdowns, we intentionally keep our model specification straightforward within a well-established framework, ensuring robust results. Notably, we do not attempt to identify factors influencing the dynamics of state probabilities, which may be affected by variables related to government or monetary policies, geopolitical issues, or climate change, among others.[22] Exploring these factors represents an important avenue for future research.

Author Contributions: Conceptualization, E.J. and A.P.; methodology, E.J. and A.P.; validation, E.J. and A.P.; writing—original draft preparation, E.J. and A.P.; writing—review and editing, E.J. and A.P.; supervision, E.J. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: All data used in this paper are from Kenneth French's website: https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, accessed on 2 December 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Evolution of Model Parameters

Panel A: Normal distribution

Panel B: Student's t distribution
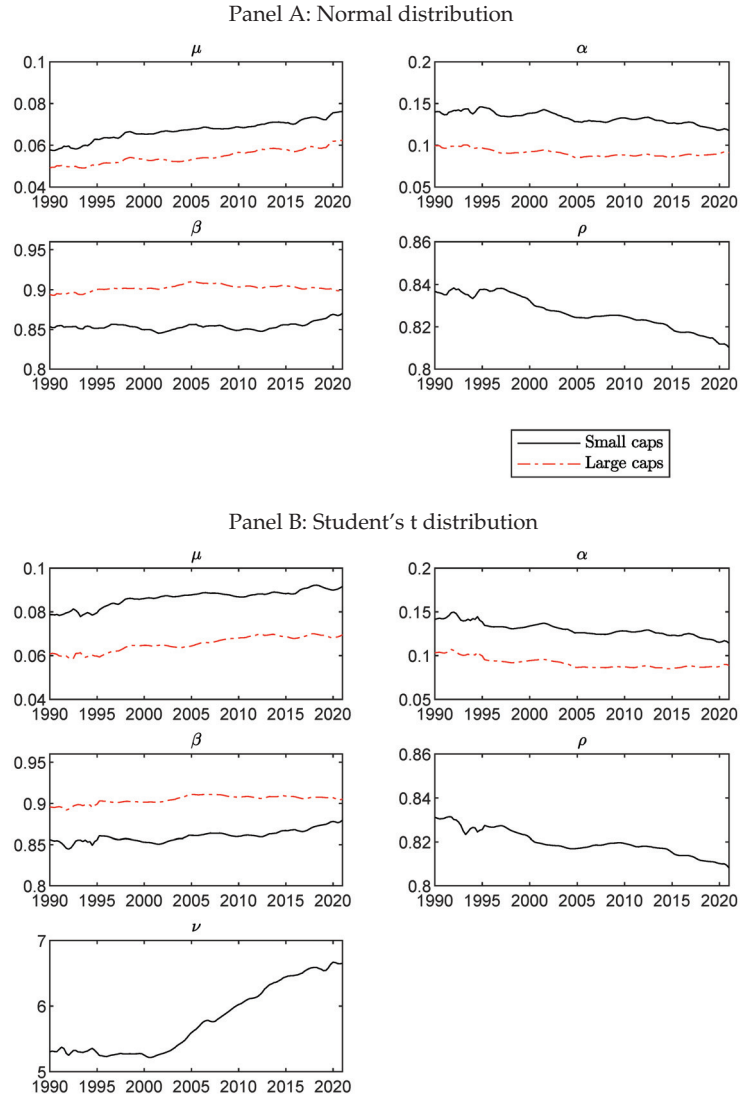
Figure A1. Model Parameters: One-regime Models.

Panel A: Normal distribution

Panel B: Student's t distribution
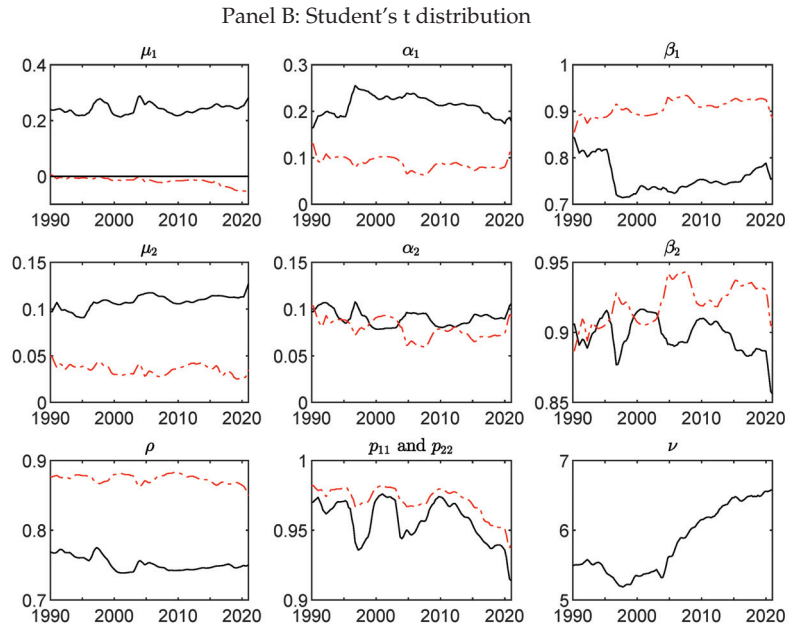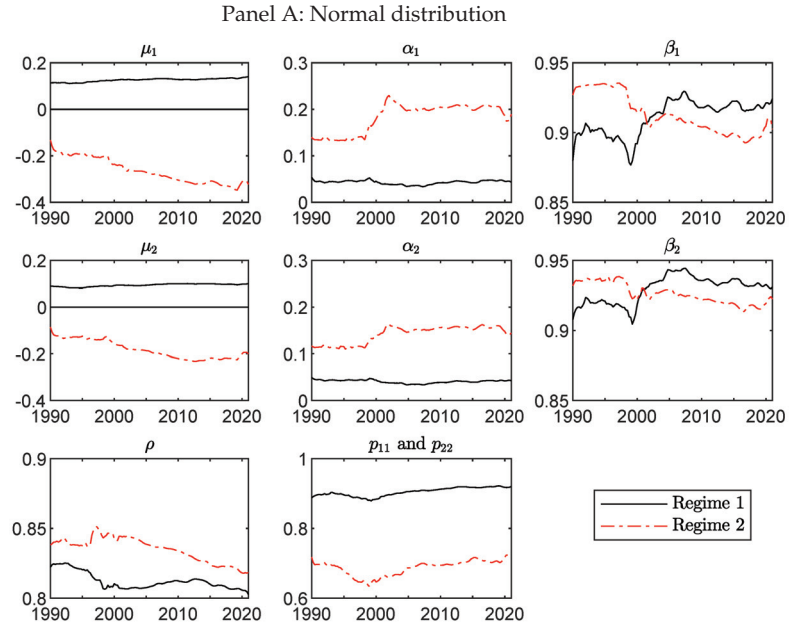
Figure A2. Model Parameters: Two-regime Models.

Panel A: Normal distribution

Panel B: Student's t distribution
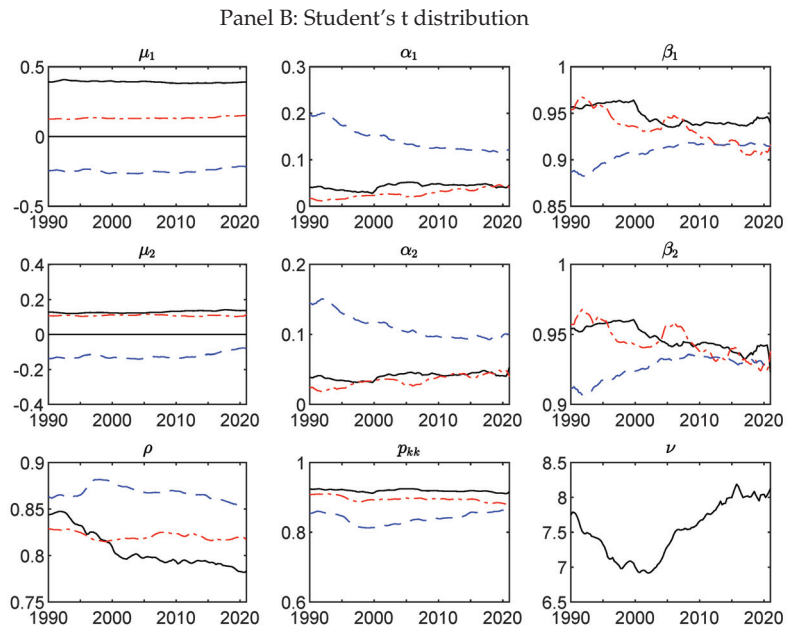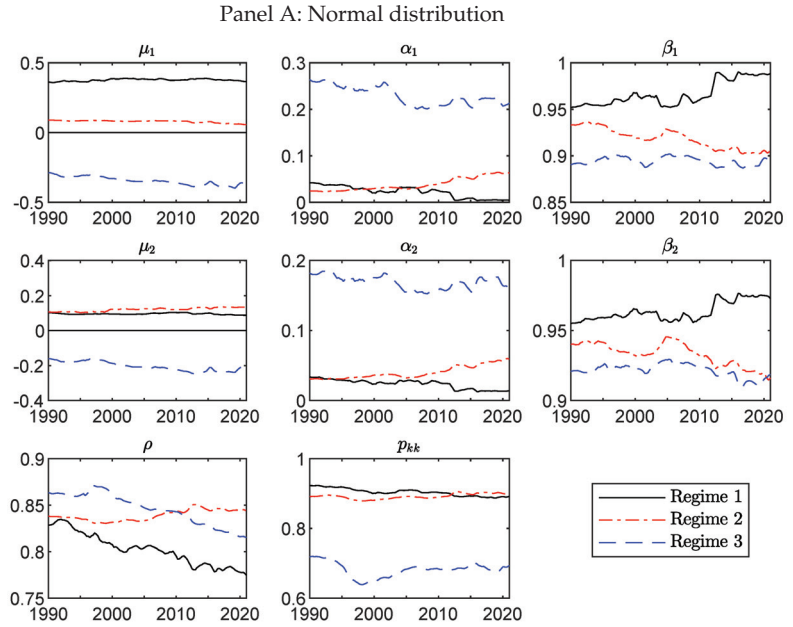
Figure A3. Model Parameters: Three-regime Models.

Appendix B. Adequacy Tests

Our predictions of the large drawdown risk measures, generically defined as $X\hat{D}D_t$, are obtained by simulations from the models. Therefore, by backtesting $X\hat{D}D_t$, we assess whether the model underlying a particular prediction is adequate. To backtest our predicted risk measures, we rely on the second test by Acerbi and Szekely (2014). Note that Acerbi and Szekely (2017) propose a more robust approach for backtesting a risk measure that depends on another statistic, such as ES or CDD, which depend on a threshold. The proposed ridge backtest accounts for the sensitivity of the statistic to the threshold estimation by penalizing errors in estimation. However, this test is only relevant for validating a model from a prudential perspective, that is, testing that the model does not underestimate the risk measure. Therefore, this procedure is not suitable for testing the CDD. As we want to assess if a model tends to overestimate or underestimate the CDD, we do not pursue with this test. The backtest for $M\hat{D}D_t$ does not suffer from the sensitivity to the threshold.

Appendix B.1. Logic of the Adequacy Test

The logic of the test is the following: For a given risk measure, we define a Z statistic that is equal to zero under the null hypothesis that the model correctly predicts this risk measure:

$$E_{H_0}[Z_{XDD}(X\hat{D}D, DD)] = 0,$$

where DD is a random variable representing drawdowns. Then, from the sequence of risk measure predictions and drawdowns realizations, we compute the realized Z statistic over the out-of-sample period:

$$\mathbf{z}_{XDD} = \frac{1}{T} \sum_{t=1}^{T} Z_{XDD}(X\hat{D}D_t, DD_t),$$

where DD_t is the vector of observed drawdowns defined in Section 2.2.

Finally, we test the adequacy of a given model by comparing the realized Z statistic obtained from this model to the distribution of the Z statistics under the null hypothesis. This distribution is obtained by computing Z statistics from simulations of the model over the out-of-sample period.[23]

Appendix B.2. Adequacy Test Statistics

For ease of exposition, we assume that the distribution is continuous and strictly increasing, as in Section 2.1.

For CDD, we build the Z statistic from the fact that $CDD_{\theta,t} = E_t[DD_t \, I_{(DD_t > Th_{\theta,t})}]$, where $Th_{\theta,t} = \inf\{s \mid \frac{1}{H} \sum_{h=1}^{H} I_{(DD_{t,h} > s)} \leq 1 - \theta\}$, with $E_t[Th_{\theta,t}] = (1 - \theta)H$. The test statistic is defined as (see Equation (3)):

$$\mathbf{z}_{CDD} = \frac{1}{T} \sum_{t=1}^{T} \sum_{h=1}^{H} \frac{DD_{t,h} \, I_{(DD_{t,h} > \hat{T}h_{\theta,t})}}{(1 - \theta) \, H \, C\hat{D}D_{\theta,t}} - 1,$$

where $C\hat{D}D_{\theta,t}$ and $\hat{T}h_{\theta,t}$ are predictions of $CDD_{\theta,t}$ and $Th_{\theta,t}$ based on simulations of the model estimated using data until quarter $t - 1$, and T corresponds to the number of quarters in the out-of-sample period.

The hypotheses for this test are as follows:

$$H_0 : CDD_{\theta,t} - C\hat{D}D_{\theta,t} = 0 \text{ for all } t$$
$$H_1 : |CDD_{\theta,t} - C\hat{D}D_{\theta,t}| > 0 \text{ for some } t.$$

Under the null hypothesis, we have $E_{H_0}[Z_{CDD}(C\hat{D}D, DD)] = 0$. Under the alternative hypothesis, if the model underestimates the CDD, we have $E_{H_1}[Z_{CDD}(C\hat{D}D, DD)] > 0$, and if the model overestimates the CDD, we have $E_{H_1}[Z_{CDD}(C\hat{D}D, DD)] < 0$. The test

actually jointly evaluates the frequency and the magnitude of the tail events because the test statistics do not impose that $\sum_{h=1}^{H} I_{(DD_t > \hat{r}h_{\theta,t})} = (1 - \theta)H$, i.e., the frequency is correct. Therefore, the test statistic will be close to 0 if both the predicted threshold is close to the realized threshold and the predicted CDD is close to the realized CDD.

For MDD, we simply define the Z statistic as:

$$\mathbf{z}_{MDD} = \frac{1}{T} \sum_{t=1}^{T} \frac{MDD_t}{M\hat{D}D_t} - 1,$$

where $M\hat{D}D_t = \frac{1}{Q} \sum_{q=1}^{Q} MDD_t^{(q)}$ is the prediction of MDD_t based on model's simulations.

The hypotheses of this test are:

$$H_0 : MDD_t - M\hat{D}D_t = 0 \text{ for all } t$$
$$H_1 : |MDD_t - M\hat{D}D_t| > 0 \text{ for some } t.$$

It follows that $E_{H_0}[Z_{MDD}(M\hat{D}D, DD)] = 0$ under the null hypothesis that the model correctly describes the realized MDD. When the model underestimates (overestimates) the risk measure, the Z statistic takes a positive (negative) value, i.e., $E_{H_1}[Z_{MDD}(M\hat{D}D, DD)] \gtrless 0$.

Appendix B.3. Adequacy Test Significance

To find the test significance, we generate Z statistics under the null hypothesis by simulating the model for which we test the adequacy. Then, by comparing the realized Z statistic to the simulated Z statistic under H_0, we evaluate the *p*-value, i.e., the probability to obtain the realized Z statistic from the model underlying the null hypothesis. We proceed as follows:

1. Using the model under H_0, we simulate Q samples of log-returns for all quarters of the backtesting period: $\{r_t^{(q)}\}_{h=1}^{H}$, $t = 1, \cdots, T$ and $q = 1, \cdots, Q$. The simulations are generated as in Section 2.4.2.
2. We calculate the vectors of drawdowns for each simulation, $DD_t^{(q)}$, $t = 1, \cdots, T$. Then, we compute Z statistic for each simulation, $\mathbf{z}_{XDD}^{(q)} = \mathbf{z}_{XDD}(X\hat{D}D, DD^{(q)})$, $q = 1, \cdots, Q$. They represent the distribution of the Z statistic under the null hypothesis.
3. We compute the realized Z statistic on the drawdowns $\mathbf{z}_{XDD}^{(q)} = \mathbf{z}_{XDD}(X\hat{D}D, DD)$.
4. We estimate the significance by comparing the realized Z statistic to the distribution of the simulated Z statistics and compute the *p*-value of the bilateral test as: p-val $= 1 - \frac{1}{Q} \sum_{q=1}^{Q} (|\mathbf{z}_{XDD}^{(q)}| < |\mathbf{z}_{XDD}|)$.

Appendix B.4. Adequacy Test Results

Results of the adequacy tests are reported in Table A1. The statistics are based on 50,000 simulations of the estimated models. The table indicates that most models fail to capture the time-series properties of large drawdown statistics. The one-regime models and the two-regime/Student's t model systematically underestimate the drawdown measures for both small and large caps. For CDD, the expected number of exceedances is equal to 1562 (i.e., 20% of the total number of daily observations in the out-of-sample period). The one-quarter number of exceedances for the two-regimes/Student's t model is equal to 2666 and 1961 for small caps and large caps, respectively, meaning that the estimated threshold is too low. As a consequence, the estimated CDD is too small, resulting in high values of the test statistic $Z_{CDD}(DD)$. For all three models and all three investment horizons, *p*-values for the CDD and MDD tests are below 1.5%. The reason for this failure is that these models do not allow for a negative trend in expected returns, and therefore, they cannot reproduce long-lasting market declines.

The two-regime/normal model and the three-regime/normal model provide a good description of the CDD and MDD of small caps, with *p*-values above 5% for all three

horizons. However, both models fail at correctly predicting the drawdown measures of large caps. They systematically overestimate the realized CDD and MDD for large caps. For large caps, the one-quarter number of exceedances is close to 1000 for both models, while the expected number is equal to 1562.

The only model that performs relatively well for small caps and large caps is the three-regime/Student's t model. For CDD, it produces a number of exceedances and an average drawdown above the threshold, which is close to the numbers observed in the data. For the one-quarter horizon, the numbers of exceedances equal 1800 and 1412 for the small caps and large caps, respectively, while the expected number is 1562. The *p*-value of the test statistic is equal to 2.8% for small caps and to 66.7% for large caps, suggesting that the 20% threshold on drawdowns is slightly too small for small firms. Similarly, this model performs relatively well for the two-quarter horizon, with a number of exceedances and an average drawdown above the threshold, which is in line with the data.

Table A1. Adequacy Tests for Out-of-sample Predictions (1990–2020).

	1-Regime Normal		1-Regime Student t		2-Regime Normal		2-Regime Student t		3-Regime Normal		3-Regime Student t	
	Small	**Large**	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**	**Small**	**Large**
Panel A: 20% CDD												
1 quarter												
Nb exc.	2597	1781	2961	1964	1490	1015	2666	1961	1512	1082	1800	1412
Stat.	1.693	0.653	2.310	0.876	0.052	−0.307	1.693	0.829	0.039	−0.282	0.308	0.055
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.666)	(0.007)	(0.000)	(0.000)	(0.762)	(0.018)	(0.028)	(0.667)
2 quarters												
Nb exc.	2930	1938	3443	2221	1168	697	2967	2173	1278	784	1442	1124
Stat.	2.165	0.687	3.134	0.987	−0.134	−0.545	2.167	0.891	−0.063	−0.484	0.037	-0.266
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.210)	(0.000)	(0.000)	(0.000)	(0.367)	(0.001)	(0.396)	(0.033)
4 quarters												
Nb exc.	3163	1728	3863	2117	1037	510	3243	2037	1244	613	1193	683
Stat.	2.061	0.449	3.285	0.794	−0.328	−0.649	2.237	0.662	−0.186	−0.567	−0.238	−0.545
p-val.	(0.000)	(0.015)	(0.000)	(0.001)	(0.066)	(0.000)	(0.000)	(0.003)	(0.218)	(0.002)	(0.138)	(0.002)
Panel B: MDD												
1 quarter												
Stat.	0.622	0.266	0.885	0.377	−0.047	−0.202	0.739	0.363	−0.033	−0.192	0.078	−0.009
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.327)	(0.000)	(0.000)	(0.000)	(0.538)	(0.000)	(0.176)	(0.852)
2 quarters												
Stat.	0.802	0.284	1.168	0.423	−0.057	−0.285	0.919	0.383	−0.010	−0.237	−0.033	−0.188
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.305)	(0.000)	(0.000)	(0.000)	(0.862)	(0.000)	(0.595)	(0.002)
4 quarters												
Stat.	0.873	0.290	1.286	0.435	−0.045	−0.299	1.017	0.363	0.021	−0.232	-0.059	−0.269
p-val.	(0.000)	(0.000)	(0.000)	(0.000)	(0.503)	(0.000)	(0.000)	(0.000)	(0.773)	(0.001)	(0.404)	(0.000)

Note: This table reports the test statistics and the *p*-value for 20% CDD and MDD. For CDD, the table also reports the number of exceedances, with an expected number equal to 1562. We consider three investment horizons: one quarter, two quarters, and four quarters. These results are based on the predictions of the models between 1990 and 2020.

Notes

1   Disasters may include severe macroeconomic and financial crises, pandemics, wars, or extreme weather and climate conditions.

2   We use a large sample to estimate model's parameters accurately and perform the out-of-sample analysis over a long period of time. Such a long sample would not be necessary in practice to estimate MS models and to predict subsequent large drawdowns.

3   For simplicity, we assume here a fixed number of days H per period. In the empirical analysis, we will use the actual number of days per period.

4   A drawdown may span over a short period (as for the COVID-19 crisis, with a 36% drawdown in 24 days) or over a window of more than a year (as for the subprime crisis, with a 60% drawdown in 355 days).

5   We note that the knowledge of the cumulated log-return at the end of the period is not sufficient to infer the large drawdown measures, as peaks and troughs are likely to occur on random days within the period.

6   The model is estimated over a long sample of daily returns, $(\tilde{r}_1, \cdots, \tilde{r}_D)$, where D is the number of days in the full sample. In contrast, large drawdown measures are computed over relatively short subsamples (e.g., one quarter or one year) with H days, which we have denoted by $r_t = (r_{t,1}, \cdots, r_{t,H})$, $t = 1, \cdots, T$ in Section 2. Since we use nonoverlapping subsamples, both notations define the same sample: $(\tilde{r}_1, \cdots, \tilde{r}_D) = (r_1, \cdots, r_T)$, where $D = H \times T$.

7   Assuming an autoregressive process would have a very limited effect for a long-term investment objective because the autocorrelation of daily returns is low. The first-order autocorrelation of the market return is equal to 0.05 over the 1926–2020 period and equal to -0.06 over the 1990–2020 period.

8   In this expression, we follow the suggestion of Klaassen (2002) and Haas et al. (2004) and define shocks with respect to a given state using the conditional mean $\mu_i^{(k)}$ instead of the unconditional mean μ_i adopted by Gray (1996).

9   In Section 3.2, we use simulations to demonstrate that a three-regime model with a (symmetric) Student's t innovations can generate some asymmetry in large drawdowns, as observed in the data.

10   We note that stationarity conditions apply to the complete distribution and not regime by regime. As a consequence, usual stationarity conditions in a GARCH model might not be satisfied for some regimes. In particular, global stationarity can be obtained even when $\alpha_i^{(k)} + \beta_i^{(k)} \geq 1$ for some asset i and regime k.

11   We have also analyzed the cases of firms in the bottom and top 20% and 10% market capitalization, with limited impact on the main results. The data are available on the website of Kenneth French at (https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html, accessed on 2 December 2024). Other long-sample portfolios, such as value versus growth or losers versus winners portfolios, are also available.

12   Our sample ends in December 2020, slightly after the end of the COVID-19 market crash. Therefore, the impact of this episode on the financial performance of our investment strategies is reflected in the out-of-sample analysis discussed in Section 3.4.

13   Stationary (or steady-state) probabilities are defined as: $\pi_\infty = P \, \pi_\infty$. In the two-regime case, this relation boils down to $\pi_{\infty,2} = \Pr[S_t = 2] = (1 - p_{11})/(1 - p_{11} - p_{22})$.

14   This conclusion appears very robust and not driven by the choice of starting values. We experimented with several sets of starting parameter values and obtained the same parameter estimates for both models. A similar phenomenon—where the probability of remaining in the bear state is lower with normal innovations than with Student's t innovations—was reported by Haas and Paolella (2012) and Haas and Liu (2018).

15   The finite-sample distribution of the LR statistic is obtained by simulating many samples of returns, using the estimated parameters of the $(n-1)$-regime model and estimating, for each simulated samples, the $(n-1)$-regime model and the n-regime model, from which we compute the LR test statistics. The finite-sample distribution of the LR test statistic is computed from the empirical distribution of the LR statistics based on the simulated samples.

16   Guidolin and Timmermann (2007) estimate MS model for U.S. small caps, large caps, and long-term bonds. In a specification with within-regime constant expected returns, volatilities, and correlations, they find that four regimes are necessary to match the data. In their model, the intermediate regime is further decomposed into a slow growth regime and a recovery regime.

17   The largest spike in the filtered probability of being in the bear state is associated with the 1973 oil crisis. The drawdown actually started in January 1973, accelerated in October with the surge in oil price and lasted until December 1974, with a drawdown of 48% over this period. A major feature of this 1973–1974 drawdown is its duration. From peak to through, the downturn lasted for almost 2 years, while the subprime crisis was associated with a 60% drawdown in slightly more than 1 year.

18   As in the models with normal innovations, one of the peaks in the filtered probability, in mid-1984, could not be associated with any particular stock market event.

19   Although we developed a similar test for CED, the number of observations was insufficient for robust conclusions.

20   We note that, even with our preferred model, i.e., the three-regime model with Student's t distribution, the relationship between the probability of being in the bear regime (Figure 2) and the optimal small-cap weight (Figure 3) is far from perfect. The reason is that the probability of being in the bear regime predicts the state for the next day, whereas the portfolio is allocated for a long horizon (from 1 quarter to 1 year) and the regime is likely to change over this period.

21   This result of a lower small-cap weights for the MV criterion does not seem to be driven by a theoretical relation between the variance and large drawdown measures but more likely by the properties of our data.

22   For instance, Gray (1996) allows state probabilities to depend on lagged interest rates in an MS-GARCH model for short-term interest rates.

23   Note that Acerbi and Szekely (2014) assume a one-sided test in line with Basel VaR tests, which are designed to detect excesses of VaR exceptions. In our case, we assume a two-sided test, as we test whether a given model correctly predicts large drawdown measures.

References

Abramson, Ari, and Israel Cohen. 2007. On the stationarity of Markov-switching GARCH processes. *Econometric Theory* 23: 485–500. [CrossRef]

Acerbi, Carlo, and Balazs Szekely. 2014. Backtesting Expected Shortfall. *Risk* 27: 76–81.

Acerbi, Carlo, and Balazs Szekely. 2017. *General Properties of Backtestable Statistics*. Working Paper. New York: MSCI Inc. Available online: https://ssrn.com/abstract=2905109 (accessed on 25 February 2023).

Ang, Andrew, and Geert Bekaert. 2002. International asset allocation with regime shifts. *Review of Financial Studies* 15: 1137–87. [CrossRef]

Ang, Andrew, and Joseph Chen. 2002. Asymmetric correlations of equity portfolios. *Journal of Financial Economics* 63: 443–94. [CrossRef]

Barnett, Michael, William Brock, and Lars Peter Hansen. 2020. Pricing uncertainty induced by climate change. *Review of Financial Studies* 33: 1024–66. [CrossRef]

Chekhlov, Alexei, Stanislav Uryasev, and Michael Zabaranki. 2003. Portfolio optimization with drawdown constraints. In *Asset and Liability Management Tools*. Edited by Berndt Scherer. London: Risk Books, pp. 263–78.

Chekhlov, Alexei, Stanislav Uryasev, and Michael Zabaranki. 2005. Drawdown measure in portfolio optimization. *International Journal of Theoretical and Applied Finance* 8: 13–58. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56. [CrossRef]

Garcia, Rene. 1998. Asymptotic null distribution of the likelihood ratio test in Markov switching models. *International Economic Review* 39: 763–88. [CrossRef]

Goldberg, Lisa R., and Ola Mahmoud. 2015. *On a Convex Measure of Drawdown Risk*. Working Paper. Berkeley: Center for Risk Management Research. Available online: http://arxiv.org/abs/1404.7493 (accessed on 25 March 2021).

Goldberg, Lisa R., and Ola Mahmoud. 2016. Drawdown: From practice to theory and back again. *Mathematics and Financial Economics* 11: 275–97. [CrossRef]

Gray, Stephen F. 1996. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics* 42: 27–62. [CrossRef]

Grossman, Sanford, and Zhongquan Zhou. 1993. Optimal investment strategies for controlling draw-downs. *Mathematical Finance* 3: 241–276. [CrossRef]

Guidolin, Massimo, and Allan Timmermann. 2004. *Value at Risk and Expected Shortfall Under Regime Switching*. Working Paper. San Diego: University of California. Available online: https://ssrn.com/abstract=557091 (accessed on 1 May 2021).

Guidolin, Massimo, and Allan Timmermann. 2007. Asset allocation under multivariate regime switching. *Journal of Economic Dynamics and Control* 31: 3503–44. [CrossRef]

Guidolin, Massimo, and Allan Timmermann. 2008. International asset allocation under regime switching, skew, and kurtosis preferences. *Review of Financial Studies* 21: 889–935. [CrossRef]

Haas, Markus, and Ji-Chun Liu. 2018. A multivariate regime-switching GARCH model with an application to global stock market and real estate equity returns. *Studies in Nonlinear Dynamics and Econometrics* 22: 1–27. [CrossRef]

Haas, Markus, Stefan Mittnik, and Marc S. Paolella. 2004. A new approach to Markov-switching GARCH models. *Journal of Financial Econometrics* 2: 493–530. [CrossRef]

Haas, Markus, and Marc S. Paolella. 2012. Mixture and regime-switching GARCH models. In *Handbook of Volatility Models and Their Applications*; Edited by Luc Bauwens, Christian M. Hafner, and Sebastien Laurent. Hoboken: John Wiley & Sons, pp. 71–102.

Hamilton, James D. 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57: 357–84. [CrossRef]

Hansen, Bruce E. 1992. The likelihood ratio test under non-standard conditions: Testing the Markov switching model of GNP. *Journal of Applied Econometrics* 7: S61–S82. [CrossRef]

Hansen, Bruce E. 1996. Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica* 64: 413–30. [CrossRef]

Hansen, Peter Reinhard. 2009. *In-Sample Fit and Out-of-Sample Fit: Their Joint Distribution and Its Implications for Model Selection*. Working Paper. Stanford: Department of Economics, Stanford University.

Huang, Wei, Qianqiu Liu, S. Ghon Rhee, and Feng Wu. 2012. Extreme downside risk and expected stock returns. *Journal of Banking and Finance* 36: 1492–502. [CrossRef]

Karydas, Christos, and Anastasios Xepapadeas. 2019. *Climate Change Financial Risks: Pricing and Portfolio Allocation*. Economics Working Paper Series, No. 19/327, 327. Zurich: Center of Economic Research (CER-ETH). [CrossRef]

Klaassen, Franc. 2002. Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* 27: 363–94. [CrossRef]

Pagano, Marco, Christian Wagner, and Josef Zechner. 2023. Disaster resilience and asset prices. *Journal of Financial Economics* 150: 103712. [CrossRef]

Patton, Andrew. 2004. On the out-of-sample importance of skewness and asymmetric dependence for asset allocation. *Journal of Financial Econometrics* 2: 130–68. [CrossRef]

Pelletier, Denis. 2006. Regime switching for dynamic correlations. *Journal of Econometrics* 131: 445–73. [CrossRef]

Peng, Cheng, Young Shin Kim, and Stefan Mittnik. 2022. Portfolio optimization on multivariate regime switching GARCH model with normal tempered stable innovation. *Journal of Risk and Financial Management* 15: 230. [CrossRef]

Perez-Quiros, Gabriel, and Allan Timmermann. 2000. Firm size and cyclical variations in stock returns. *Journal of Finance* 55: 1229–62. [CrossRef]

Reveiz, Alejandro, and Carlos Leon. 2008. Efficient portfolio optimization in the wealth creation and maximum drawdown space. In *Interest Rate Models, Asset Allocation and Quantitative Techniques for Central Banks and Sovereign Wealth Funds*. Edited by Arjan B. Berkelaar, Joachim Coche, and Ken Nyholm. London: Palgrave Macmillan, pp. 134–157.

Article

Fitting the Seven-Parameter Generalized Tempered Stable Distribution to Financial Data

Aubain Nzokem [1,*,†] and Daniel Maposa [2,†]

[1] Department of Mathematics & Statistics, York University, Toronto, ON M3J 1P3, Canada
[2] Department of Statistics and Operations Research, University of Limpopo, Sovenga 0727, South Africa; daniel.maposa@ul.ac.za
[*] Correspondence: hilaire77@gmail.com or hil14@yorku.ca
[†] These authors contributed equally to this work.

Abstract: This paper proposes and implements a methodology to fit a seven-parameter Generalized Tempered Stable (GTS) distribution to financial data. The nonexistence of the mathematical expression of the GTS probability density function makes maximum-likelihood estimation (MLE) inadequate for providing parameter estimations. Based on the function characteristic and the fractional Fourier transform (FRFT), we provide a comprehensive approach to circumvent the problem and yield a good parameter estimation of the GTS probability. The methodology was applied to fit two heavy-tailed data (Bitcoin and Ethereum returns) and two peaked data (S&P 500 and SPY ETF returns). For each historical data, the estimation results show that six-parameter estimations are statistically significant except for the local parameter, μ. The goodness of fit was assessed through Kolmogorov–Smirnov, Anderson–Darling, and Pearson's chi-squared statistics. While the two-parameter geometric Brownian motion (GBM) hypothesis is always rejected, the GTS distribution fits significantly with a very high p-value and outperforms the Kobol, Carr–Geman–Madan–Yor, and bilateral Gamma distributions.

1. Introduction

Modeling high-frequency asset return with the normal distribution is the underlying assumption in many financial tools, such as the Black–Scholes–Merton option pricing model and the risk metric variance–covariance technique for value at risk (VAR). However, substantial empirical evidence rejects the normal distribution for various asset classes and financial markets. The symmetric and rapidly decreasing tail properties of the normal distribution cannot describe the skewed and fat-tailed properties of the asset return distribution.

The α-stable distribution has been proposed (Nolan 2020; Sato 1999) as an alternative to the normal distribution for modeling asset return and many types of physical and economic systems. The theoretical and empirical argument is that the stable distribution generalizes the Central Limit Theorem regardless of the variance nature (finite or infinite) (Nzokem 2024; Rachev et al. 2011). There are two major drawbacks (Borak et al. 2005; Nolan 2020): firstly, the lack of closed formulas for densities and distribution functions, except for the normal distribution ($\alpha = 2$), Cauchy distribution ($\alpha = 1$), and Lévy distribution ($\alpha = \frac{1}{2}$) (Tsallis 1997); secondly, most of the moments of the stable distribution are infinite. An infinite variance of asset return leads to an infinite price for derivative instruments such as options.

The Generalized Tempered Stable (GTS) distribution was developed to overcome the shortcomings of the two distributions, and the tails of the GTS distribution are heavier than

the normal distribution but thinner than the stable distribution (Grabchak and Samorodnitsky 2010; Kim et al. 2009). The general form of the GTS distribution can be defined by the following Lévy measure ($V(dx)$) (1):

$$V(dx) = \left(\frac{\alpha_+ e^{-\lambda_+ x}}{x^{1+\beta_+}} \mathbf{1}_{x>0} + \frac{\alpha_- e^{-\lambda_- |x|}}{|x|^{1+\beta_-}} \mathbf{1}_{x<0} \right) dx \qquad (1)$$

where $0 \leq \beta_+ \leq 1$, $0 \leq \beta_- \leq 1$, $\alpha_+ \geq 0$, $\alpha_- \geq 0$, $\lambda_+ \geq 0$, and $\lambda_- \geq 0$. More details on the Tempered Stable distribution are provided in (Küchler and Tappe 2013; Rachev et al. 2011).

The rich class of the GTS distribution (1) has a myriad of applications ranging from financial to mathematical physics and economic systems. However, few studies (Fallahgoul and Loeper 2021; Massing 2024; Nzokem and Montshiwa 2022) have covered the methods and techniques to estimate the parameters of the GTS distribution. This study aims to contribute to the literature by providing a methodology for fitting the seven-parameter GTS distribution. As illustrations, the study used four historical prices: two heavy-tailed data (Bitcoin and Ethereum returns) and two peaked data (S&P 500 and SPY ETF returns). The GTS distribution is fitted to the underlying distribution of each data index and goodness-of-fit analysis is carried out. The main disadvantage of the GTS distribution is the lack of the closed forms of the density, cumulative, and derivative functions. We use a computational algorithm, called the enhanced fast FRFT scheme (Nzokem 2023a), to circumvent the problem.

The rest of the paper is organized as follows: Section 2 provides some theoretical framework of the GTS distribution. Section 3 presents the multivariate maximum-likelihood (ML) method and the analytic version of the two-parameter normal distribution. Section 4 presents the results of the GTS parameter estimations, along with the associated statistical tests for the heavy-tailed Bitcoin and Ethereum returns. Section 5 fits the GTS distribution to the traditional indices S&P 500 and SPY ETF returns, while Section 6 presents the results of the goodness-of-fit test. Section 7 provides the concluding remarks.

2. Generalized Tempered Stable (GTS) Distribution

The Lévy measure of the GTS distribution ($V(dx)$) is defined in (2) as a product of a tempering function $q(x)$ and a Lévy measure of the α-stable distribution $V_{stable}(dx)$:

$$q(x) = e^{-\lambda_+ x} \mathbf{1}_{x>0} + e^{-\lambda_- |x|} \mathbf{1}_{x<0}$$

$$V_{stable}(dx) = \left(\alpha_+ \frac{1}{x^{1+\beta_+}} \mathbf{1}_{x>0} + \alpha_- \frac{1}{|x|^{1+\beta_-}} \mathbf{1}_{x<0} \right) dx \qquad (2)$$

$$V(dx) = q(x) V_{stable}(dx) = \left(\alpha_+ \frac{e^{-\lambda_+ x}}{x^{1+\beta_+}} \mathbf{1}_{x>0} + \alpha_- \frac{e^{-\lambda_- |x|}}{|x|^{1+\beta_-}} \mathbf{1}_{x<0} \right) dx$$

where $0 \leq \beta_+ \leq 1$, $0 \leq \beta_- \leq 1$, $\alpha_+ \geq 0$, $\alpha_- \geq 0$, $\lambda_+ \geq 0$ and $\lambda_- \geq 0$.

The six parameters that appear have important interpretations. β_+ and β_- are the indexes of stability bounded below by 0 and above by 2 (Borak et al. 2005). They capture the peakedness of the distribution similarly to the β-stable distribution, but the distribution tails are tempered. If β increases (decreases), then the peakedness decreases (increases). α_+ and α_- are the scale parameters, also called the process intensity (Boyarchenko and Levendorskii 2002); they determine the arrival rate of jumps for a given size. λ_+ and λ_- control the decay rate on the positive and negative tails. Additionally, λ_+ and λ_- are also skewness parameters. If $\lambda_+ > \lambda_-$ ($\lambda_+ < \lambda_-$), then the distribution is skewed to the left (right), and if $\lambda_+ = \lambda_-$, then it is symmetric (Fallahgoul et al. 2019; Rachev et al. 2011). α and λ are related to the degree of peakedness and thickness of the distribution. If α increases (decreases), the peakedness and the thickness decrease (increase). Similarly, if λ increases (decreases), then the peakedness increases (decreases) and the thickness decreases (increases) (Bianchi et al. 2019). For more details on the tempering function and Lévy measure of the tempered stable distribution, refer to (Küchler and Tappe 2013; Rachev et al. 2011).

The activity process of the GTS distribution can be studied from the integral (3) of the Lévy measure (2):

$$\int_{-\infty}^{+\infty} V(dx) = \begin{cases} +\infty & \text{if } 0 \leq \beta_+ < 1 \wedge 0 \leq \beta_- < 1 \\ \alpha_+ \lambda_+{}^{\beta_+} \Gamma(-\beta_+) + \alpha_- \lambda_-{}^{\beta_-} \Gamma(-\beta_-) & \text{if } \beta_+ < 0 \wedge \beta_- < 0. \end{cases} \tag{3}$$

As shown in (3), if $\beta_+ < 0$ and $\beta_- < 0$, $GTS(\beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$ is of a finite activity process and can be written as a compound Poisson (Barndorff-Nielsen and Shephard 2002). When $0 \leq \beta_+ < 1$ and $0 \leq \beta_- < 1$, this Lévy density ($V(dx)$) is not integrable as it goes off to infinity too rapidly as x goes to zero (Barndorff-Nielsen and Shephard 2002), which means in practice that there will be a large number of very small jumps. As shown in (3), $GTS(\beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$ is an infinite activity process with infinite jumps in any given time interval.

In addition to the infinite activities process, the variation in the process can be studied through the following integral:

$$\int_{-1}^{1} |x| V(dx) = \int_{-1}^{0} |x| V(dx) + \int_{0}^{1} |x| V(dx)$$
$$= \alpha_- \lambda_-^{\beta_- - 1} \gamma(1 - \beta_-, \lambda_-) + \alpha_+ \lambda_+^{\beta_+ - 1} \gamma(1 - \beta_+, \lambda_+)$$

where $\gamma(s, x) = \int_0^x y^{s-1} e^{-y} dy$ is the lower incomplete gamma function.

And we have:

$$\int_{-1}^{1} |x| V(dx) < +\infty \quad \text{if } 0 < \beta_- \leq 1 \, \& \, 0 < \beta_+ \leq 1. \tag{4}$$

As shown in (4), $GTS(\beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$ generates a finite variance process, which is contrary to the Brownian motion process. $GTS(\beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$ generates a type B Lévy process (Ken-Iti 2001), which is a purely non-Gaussian infinite activity Lévy process of finite variation whose sample paths have an infinite number of small jumps and a finite number of large jumps in any finite time interval.

The GTS distribution can be denoted by $X \sim GTS(\beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$ and $X = X_+ - X_-$ with $X_+ \geq 0$, $X_- \geq 0$. $X_+ \sim TS(\beta_+, \alpha_+, \lambda_+)$ and $X_- \sim TS(\beta_-, \alpha_-, \lambda_-)$. By adding the location parameter, the GTS distribution becomes $GTS(\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$, and we have (5):

$$Y = \mu + X = \mu + X_+ - X_-, \qquad Y \sim GTS(\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-). \tag{5}$$

2.1. GTS Distribution and Characteristic Exponent

Theorem 1. *Consider a variable* $Y \sim GTS(\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$. *The characteristic exponent can be written as:*

$$\Psi(\xi) = \mu \xi i + \alpha_+ \Gamma(-\beta_+) \left((\lambda_+ - i\xi)^{\beta_+} - \lambda_+{}^{\beta_+} \right) + \alpha_- \Gamma(-\beta_-) \left((\lambda_- + i\xi)^{\beta_-} - \lambda_-{}^{\beta_-} \right). \tag{6}$$

Proof. $V(dx)$ in (2) is a Lévy measure. The following relation is satisfied from (4):

$$\int_{-\infty}^{+\infty} Min(1, |x|) V(dx) < +\infty.$$

More details on the proof are provided in (Nzokem and Maposa 2024).

The Lévy–Khintchine representation (Barndorff-Nielsen and Shephard 2002) for non-negative Lévy process is applied on Y. $Y = \mu + X = \mu + X_+ - X_-$ and we have:

$$\Psi(\xi) = Log\left(Ee^{iY\xi}\right) = i\mu\xi + Log\left(Ee^{iX_+\xi}\right) + Log\left(Ee^{-iX_-\xi}\right)$$

$$= i\mu\xi + \int_0^{+\infty}\left(e^{iy\xi}-1\right)\frac{\alpha_+ e^{-\lambda_+ y}}{y^{1+\beta_+}}dy + \int_0^{+\infty}\left(e^{-iy\xi}-1\right)\frac{\alpha_- e^{-\lambda_- y}}{y^{1+\beta_-}}dy, \tag{7}$$

$$\int_0^{+\infty}\left(e^{iy\xi}-1\right)\frac{\alpha_+ e^{-\lambda_+ y}}{y^{1+\beta_+}}dy = \alpha_+\lambda_+^{\beta_+}\Gamma(-\beta_+)\sum_{k=1}^{+\infty}\frac{\Gamma(k-\beta_+)}{\Gamma(-\beta_+)k!}(\frac{i\xi}{\lambda_+})^k$$

$$= \alpha_+\lambda_+^{\beta_+}\Gamma(-\beta_+)\sum_{k=1}^{+\infty}\binom{\beta_+}{k}(-\frac{i\xi}{\lambda_+})^k \tag{8}$$

$$= \alpha_+\Gamma(-\beta_+)\left((\lambda_+ - i\xi)^{\beta_+} - \lambda_+^{\beta_+}\right).$$

Similarly, we have :

$$\int_0^{+\infty}\left(e^{-iy\xi}-1\right)\frac{\alpha_- e^{-\lambda_- y}}{y^{1+\beta_-}}dy = \alpha_-\Gamma(-\beta_-)\left((\lambda_- + i\xi)^{\beta_-} - \lambda_-^{\beta_-}\right). \tag{9}$$

The expression in (7) becomes:

$$\Psi(\xi) = i\mu\xi + \alpha_+\Gamma(-\beta_+)\left((\lambda_+ - i\xi)^{\beta_+} - \lambda_+^{\beta_+}\right) + \alpha_-\Gamma(-\beta_-)\left((\lambda_- + i\xi)^{\beta_-} - \lambda_-^{\beta_-}\right).$$

□

Theorem 2. *Consider a variable* $Y \sim GTS(\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$.

If $(\beta_-, \beta_+) \to (0,0)$, *GTS becomes a bilateral Gamma distribution with the following characteristic exponent:*

$$\Psi(\xi) = \mu\xi i - \alpha_+ \log\left(1 - \frac{1}{\lambda_+}i\xi\right) - \alpha_- \log\left(1 + \frac{1}{\lambda_-}i\xi\right). \tag{10}$$

In addition to $(\beta_-, \beta_+) \to (0,0)$, *if* $\alpha_- = \alpha_+ = \alpha$, *GTS becomes a Variance-Gamma (VG) distribution with parameter* $(\mu, \delta, \sigma, \alpha, \theta)$

$$\delta = \lambda_- - \lambda_+ \quad \sigma = 1 \quad \alpha = \alpha_- = \alpha_+ \quad \theta = \frac{1}{\lambda_-\lambda_+}$$

and the following characteristic exponent:

$$\Psi(\xi) = \mu\xi i - \alpha \log\left(1 - \frac{\lambda_- - \lambda_+}{\lambda_+\lambda_-}i\xi + \frac{1}{\lambda_+\lambda_-}\xi^2\right). \tag{11}$$

Proof.

$$\Gamma(-\beta_+) = -\frac{\Gamma(1-\beta_+)}{\beta_+}$$

$$\lim_{\beta_+ \to 0}\Gamma(-\beta_+)\left((\lambda_+ - i\xi)^{\beta_+} - \lambda_+^{\beta_+}\right) = -\log\left(1 - \frac{1}{\lambda_+}i\xi\right). \tag{12}$$

Similarly, (12) works for $\beta_- \to 0$, and we have the characteristic exponent (10).
In addition, if $\alpha_- = \alpha_+ = \alpha$, from (10), the characteristic exponent becomes:

$$\Psi(\xi) = \mu\xi i - \alpha \log\left(1 - \frac{\lambda_- - \lambda_+}{\lambda_+\lambda_-}i\xi + \frac{1}{\lambda_+\lambda_-}\xi^2\right),$$

which is a Variance-Gamma (VG) distribution with parameter $(\mu, \lambda_- - \lambda_+, 1, \alpha, \frac{1}{\lambda_-\lambda_+})$. For more details on the VG model, refer to (Madan et al. 1998; Nzokem 2023c). □

Theorem 3 (Cumulants κ_k). *Consider a variable* $Y \sim GTS(\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$. *The cumulants* κ_k *of the GTS distribution are defined as follows:*

$$\kappa_0 = 0$$

$$\kappa_1 = \mu + \alpha_+ \frac{\Gamma(1-\beta_+)}{\lambda_+^{1-\beta_+}} - \alpha_- \frac{\Gamma(1-\beta_-)}{\lambda_-^{1-\beta_-}} \tag{13}$$

$$\kappa_k = \alpha_+ \frac{\Gamma(k-\beta_+)}{\lambda_+^{k-\beta_+}} + (-1)^k \alpha_- \frac{\Gamma(k-\beta_-)}{\lambda_-^{k-\beta_-}} \qquad \forall k \in \mathbb{N} \setminus \{0,1\}.$$

Proof. We reconsider the characteristic exponent $\Psi(\xi)$ in (7):

$$
\begin{aligned}
\Psi(\xi) &= i\mu\xi + \int_0^{+\infty} \left(e^{iy\xi} - 1 \right) \frac{\alpha_+ e^{-\lambda_+ y}}{y^{1+\beta_+}} dy + \int_0^{+\infty} \left(e^{-iy\xi} - 1 \right) \frac{\alpha_- e^{\lambda_- y}}{y^{1+\beta_-}} dy \\
&= i\mu\xi + \alpha_+ \sum_{k=1}^{+\infty} \frac{\Gamma(k-\beta_+)}{\lambda_+^{k-\beta_+}} \frac{(i\xi)^k}{k!} + \alpha_- \sum_{k=1}^{+\infty} \frac{\Gamma(k-\beta_-)}{\lambda_-^{k-\beta_-}} \frac{(-i\xi)^k}{k!} \\
&= i\mu\xi + \sum_{k=1}^{+\infty} \frac{1}{k!} \left(\alpha_+ \frac{\Gamma(k-\beta_+)}{\lambda_+^{k-\beta_+}} + \alpha_- \frac{\Gamma(k-\beta_-)}{\lambda_-^{k-\beta_-}} (-1)^k \right) (i\xi)^k \\
&= \sum_{k=0}^{+\infty} \frac{\kappa_k}{k!} (i\xi)^k.
\end{aligned}
\tag{14}
$$

Hence, the k-th order cumulant κ_k is given by comparing the coefficients of both polynomial functions in $i\xi$. For more details on the relationship between the characteristic exponent and cumulant functions, refer to (Feller 1971; Kendall 1945). $\square$

2.2. GTS Distribution and Lévy Process

Corollary 1. *Let* $Y = (Y_t)$ *be a Lévy process on* $\mathbb{R}^+$ *generated by* $GTS(\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$, *and then*

$$Y_t \sim GTS(t\mu, \beta_+, \beta_-, t\alpha_+, t\alpha_-, \lambda_+, \lambda_-) \qquad \forall t \in \mathbb{R}^+. \tag{15}$$

Proof. Let $\Psi(\xi, t)$ be the characteristic exponent of the Lévy process $Y = (Y_t)$. By applying the infinitely divisible property, we have:

$$
\begin{aligned}
\Psi(\xi, t) &= Log\left(E e^{iY_t \xi} \right) = t Log\left(E e^{iX\xi} \right) \\
&= t\mu\xi i + t\alpha_+ \Gamma(-\beta_+)\left((\lambda_+ - i\xi)^{\beta_+} - \lambda_+^{\beta_+} \right) + t\alpha_- \Gamma(-\beta_-)\left((\lambda_- + i\xi)^{\beta_-} - \lambda_-^{\beta_-} \right)
\end{aligned}
$$

and we deduce that $Y_t \sim GTS(t\mu, \beta_+, \beta_-, t\alpha_+, t\alpha_-, \lambda_+, \lambda_-)$. $\square$

Theorem 4 (Asymptotic distribution of Generalized Tempered Stable distribution process). *Let* $Y = Y_t$ *be a Lévy process on* $\mathbb{R}$ *generated by* $GTS(\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$. *Then,* Y_t *converges in distribution to a Lévy process driving by a normal distribution with mean* κ_1 *and variance* κ_2

$$Y_t \xrightarrow{d} N(t\kappa_1, t\kappa_2) \quad as \quad t \to +\infty \tag{16}$$

where

$$\kappa_1 = \mu + \frac{\Gamma(1-\beta_+)}{\lambda_+^{1-\beta_+}} - \alpha_- \frac{\Gamma(1-\beta_-)}{\lambda_-^{1-\beta_-}}$$

$$\kappa_2 = \alpha_+ \frac{\Gamma(2-\beta_+)}{\lambda_+^{2-\beta_+}} + \alpha_- \frac{\Gamma(2-\beta_-)}{\lambda_-^{2-\beta_-}}.$$

Proof. The proof relies on the cumulant-generating function. As in (14), the characteristic exponent ($\Psi(\xi)$) can be written as follows:

$$\Psi(\xi) = Log\left(E e^{iY\xi}\right) = \sum_{j=0}^{+\infty} \kappa_j \frac{(i\xi)^j}{j!}. \tag{17}$$

Let $\phi(\xi, t)$ be the characteristic function of the stochastic process $\frac{Y_t - t\kappa_1}{\sqrt{t\kappa_2}}$ and we have:

$$
\begin{aligned}
\phi(\xi, t) &= E\left(e^{i\frac{Y_t - t\kappa_1}{\sqrt{t\kappa_2}}\xi}\right) = e^{-i\frac{t\kappa_1}{\sqrt{t\kappa_2}}\xi} E\left(e^{i\frac{\xi}{\sqrt{t\kappa_2}}Y_t}\right) \\
&= e^{-i\frac{t\kappa_1}{\sqrt{t\kappa_2}}\xi} e^{t\Psi\left(\frac{\xi}{\sqrt{t\kappa_2}}\right)} = e^{-i\frac{t\kappa_1}{\sqrt{t\kappa_2}}\xi} e^{\sum_{j=0}^{+\infty}\frac{t\kappa_j}{j!}(i\frac{\xi}{\sqrt{t\kappa_2}})^j} \\
&= e^{-\frac{\xi^2}{2} + \sum_{j=3}^{+\infty}\frac{t\kappa_j}{j!}(i\frac{\xi}{\sqrt{t\kappa_2}})^j},
\end{aligned}
\tag{18}
$$

$$
\lim_{t\to+\infty} \sum_{j=3}^{+\infty} \frac{t\kappa_j}{j!}\left(i\frac{\xi}{\sqrt{t\kappa_2}}\right)^j = 0 \quad \lim_{t\to+\infty} \phi(\xi, t) = \lim_{t\to+\infty} e^{-\frac{\xi^2}{2} + \sum_{j=3}^{+\infty}\frac{t\kappa_j}{j!}(i\frac{\xi}{\sqrt{t\kappa_2}})^j}
$$
$$
= e^{-\frac{1}{2}\xi^2}.
\tag{19}
$$

□

3. Multivariate Maximum-Likelihood Method

3.1. Maximum-Likelihood Method: Numerical Approach

From a probability density function $f(x, V)$ with parameter $V = (\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$ and sample data $x = (x_j)_{1 \le j \le m}$, we define the likelihood function and its first and second derivatives as follows:

$$
\begin{aligned}
L_m(x, V) &= \prod_{j=1}^{m} f(x_j, V), \quad l_m(x, V) = \sum_{j=1}^{m} log(f(x_j, V)) \\
\frac{dl_m(x, V)}{dV_j} &= \sum_{i=1}^{m} \frac{\frac{df(x_i, V)}{dV_j}}{f(x_i, V)} \\
\frac{d^2 l_m(x, V)}{dV_k dV_j} &= \sum_{i=1}^{m} \left(\frac{\frac{d^2 f(x_i, V)}{dV_k dV_j}}{f(x_i, V)} - \frac{\frac{df(x_i, V)}{dV_k}}{f(x_i, V)} \frac{\frac{df(x_i, V)}{dV_j}}{f(x_i, V)} \right).
\end{aligned}
\tag{20}
$$

To perform the maximum of the likelihood function ($L_m(x, V)$), we need the gradient of the likelihood function ($\frac{dl_m(x, V)}{dV}$), also known as the score function, and the Hessian matrix ($\frac{d^2 l_m(x, V)}{dV dV'}$), which is the variance–covariance matrix generated by the likelihood function.

Given the parameters $V = (\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$ and the sample data set X, we have the following quantities (21) from the previous development:

$$
I'_m(X, V) = \left(\frac{dl_m(x, V)}{dV_j}\right)_{1 \le j \le p}, \quad I''_m(X, V) = \left(\frac{d^2 l_m(x, V)}{dV_k dV_j}\right)_{\substack{1 \le k \le p \\ 1 \le j \le p}}.
\tag{21}
$$

We use a computational algorithm built as a composite of a standard FRFT to compute the likelihood function and its derivatives (20) in the optimization process. More details on applying the composite of FRFTs for parameter estimations are provided in (Nzokem 2021b, 2021c; Nzokem and Montshiwa 2022, 2023); for other computations (such as probability density and cumulative functions), see (Cherubini et al. 2010; Eberlein 2014; Eberlein et al. 2010; Nzokem 2023b; Nzokem and Maposa 2024).

The computational algorithm yields a local solution, V, and a negative semi-definite matrix, $I''_m(x, V)$, when the following two conditions are satisfied:

$$I'_m(x, V) = 0, \qquad U^T \mathbf{I''_m}(\mathbf{X}, \mathbf{V})U \leq 0 \quad , \forall U \in \mathbb{R}^p. \tag{22}$$

The solutions, V, in (22) are provided by the Newton–Raphson iteration algorithm Formula (23):

$$V^{n+1} = V^n - \left(I''_m(x, V^n)\right)^{-1} I'_m(x, V^n). \tag{23}$$

More details on the maximum-likelihood and Newton–Raphson iteration procedures are provided in (Giudici et al. 2013).

3.2. Asymptotic Distribution of the Maximum-Likelihood Estimator (MLE)

Theorem 5 (Cramer-Rao). *Let $T = T(X_1, \ldots, X_m)$ be a statistic and write $E[T] = k(\theta)$. Then, under suitable (smoothness) assumptions,*

$$Var[T] \geq \frac{(\frac{dE[T]}{d\theta})^2}{mI(\theta)}. \tag{24}$$

For the proof of Theorem 5, refer to (Casella and Berger 2024; Van den Bos 2007).

Theorem 6 (Consistency Estimator). *Let $X_1, \ldots, X_m$ be independent and identically distributed (i.i.d) random variables with density $f(x|\theta)$ satisfying some regularity conditions (Lehmann 1999). Let θ be the true parameter; then, there exists a sequence $\hat{\theta}_m = \theta_m(X_1, \ldots, X_m)$ of local maxima of the likelihood function $L_m(\theta)$ which is consistent, that is, which satisfies*

$$\hat{\theta}_m \overset{a.s.}{\to} \theta \quad as \quad m \to +\infty. \tag{25}$$

More details on the proof of Theorem 6 are provided in (Casella and Berger 2024; Lehmann 1999).

Theorem 7 (Asymptotic Efficiency and Normality). *Let $X_1, \ldots, X_m$ be independent and identically distributed (i.i.d) random variables with density $f(x|\theta)$ satisfying some regularity conditions in (Lehmann 1999). There exists a solution $\hat{\theta}_m = \theta_m(X_1, \ldots, X_m)$ of the likelihood equations which is consistent, and any such solution satisfies:*

$$\hat{\theta}_m - \theta \overset{d}{\to} N\left(0, I_m^{-1}(\theta)\right) \quad as \quad m \to +\infty, \tag{26}$$

where $\theta = (\theta_1, \ldots, \theta_k)$ is the actual parameter and $I_m(\theta)$ is the Fisher information matrix.

More details on the proof of Theorem 7 are provided in (Hall and Oakes 2023; Lehmann 1999; Olive 2014).

Theorem 8 (Likelihood Ratio Test). *Suppose the assumptions of Theorem 7 hold and that $(\hat{\theta}_{1n}, \ldots, \hat{\theta}_{kn})$ are consistent roots of the likelihood equations for $\theta = (\theta_1, \ldots, \theta_k)$. In addition, suppose that the corresponding assumptions hold for the parameter vector $(\theta_{r+1}, \ldots, \theta_k)$ when $r < k$ and that $\hat{\hat{\theta}}_{r+1,n}, \ldots, \hat{\hat{\theta}}_{kn}$ are consistent roots of the likelihood equations for $(\theta_{r+1}, \ldots, \theta_k)$ under the null hypothesis. We consider the likelihood ratio statistic*

$$\frac{l_m(x, \hat{\theta})}{l_m(x, \hat{\hat{\theta}})} \tag{27}$$

where $\hat{\hat{\theta}} = (\theta_1, \ldots, \theta_r, \hat{\hat{\theta}}_{r+1,n}, \ldots, \hat{\hat{\theta}}_{kn})$. Then under the null hypothesis H_0, if

$$\Delta_n = l_m(x, \hat{\theta}) - l_m(x, \hat{\hat{\theta}}), \tag{28}$$

the statistic $2\Delta_n$ has a limiting χ_r^2 distribution.

More details on the proof of Theorem 8 are provided in (Lehmann 1999; Vuong 1989).

3.3. Asymptotic Test and Confidence Interval

The above results allow us to construct an asymptotically efficient estimator $\hat{\theta}_m = (\hat{\theta}_{1m}, \ldots, \hat{\theta}_{km})$ of $\theta = (\theta_1, \ldots, \theta_k)$ such that

$$(\hat{\theta}_{1m} - \theta_1, \ldots, \hat{\theta}_{km} - \theta_k) \tag{29}$$

has a joint multivariate limit distribution with mean $(0, \ldots, 0)$ and covariance matrix $I_m^{-1}(\theta) = (J_{ij})$. In particular, we have:

$$\hat{\theta}_{jm} - \theta_j \xrightarrow{d} N(0, J_{jj}) \quad \text{as} \quad m \to +\infty. \tag{30}$$

One approach to constructing an asymptotically valid confidence interval for the parameters is via the asymptotic distribution of the ML estimator (27). An approximate $(1 - \frac{\alpha}{2})$ confidence interval for $\hat{\theta}_{jm}$ can be written as follows:

$$\hat{\theta}_{jm} \pm z(\frac{\alpha}{2}) * \sqrt{J_{jj}} \quad \text{as} \quad m \to +\infty, \tag{31}$$

where $z(\frac{\alpha}{2})$ is the $\frac{\alpha}{2}$ quantile of the standard normal distribution.

3.4. Applications of the Log-Likelihood Estimator to the Normal Distribution

We suppose the sample data $x = (x_j)_{1 \le j \le m}$ are independent observations and have a normal distribution (Mensah et al. 2023) with parameter $V(\mu, \sigma^2)$, that is, $y \sim \mathcal{N}(\mu, \sigma^2)$; then, the density is

$$f(y|V) = (2\pi\sigma^2)^{-\frac{1}{2}} exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right). \tag{32}$$

The log-likelihood function in (20) becomes

$$l_m(x|V) = \sum_{j=1}^{m} log(f(x_j|V)) = -\frac{m}{2}log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{j=1}^{m}(x_j - \mu)^2. \tag{33}$$

The first-order derivatives of the log-likelihood function with respect to μ and σ^2 in (20) become

$$I_m'(X, V) = \begin{pmatrix} \frac{dl_m(x,V)}{d\mu} \\ \frac{dl_m(x,V)}{d\sigma^2} \end{pmatrix} = \begin{pmatrix} \frac{1}{\sigma^2}\sum_{j=1}^{m}(x_j - \mu) \\ \frac{1}{2\sigma^4}\sum_{j=1}^{m}(x_j - \mu)^2 - \frac{m}{2\sigma^2}. \end{pmatrix} \tag{34}$$

By setting $I_m'(X, V) = 0$, we have

$$\hat{\mu} = \frac{1}{m}\sum_{j=1}^{m} x_j \qquad \hat{\sigma}^2 = \frac{1}{m}\sum_{j=1}^{m}(x_j - \hat{\mu})^2. \tag{35}$$

The second-order derivative of the log-likelihood function with respect to μ and σ^2 in (20) becomes

$$I_m''(X,V) = \begin{pmatrix} \frac{d^2 l_m(x,V)}{d\mu^2} & \frac{d l_m(x,V)}{d\mu d\sigma^2} \\ \frac{d l_m(x,V)}{d\sigma^2 d\mu} & \frac{d^2 l_m(x,V)}{(d\sigma^2)^2} \end{pmatrix}$$

$$= \begin{pmatrix} -\frac{m}{\sigma^2} & -\frac{\sum_{j=1}^{m}(x_j - \mu)}{\sigma^4} \\ -\frac{\sum_{j=1}^{m}(x_j - \mu)}{2\sigma^4} & -\frac{1}{\sigma^6}\sum_{j=1}^{m}(x_j - \mu)^2 + \frac{m}{2\sigma^4}. \end{pmatrix} \tag{36}$$

Refer to (Casella and Berger 2024) for more details.

We have the Fisher information matrix and the inverse:

$$I_m(V) = -E\big(I_m''(X,V)\big) = \begin{pmatrix} \frac{m}{\sigma^2} & 0 \\ 0 & \frac{m}{2\sigma^4}, \end{pmatrix} \qquad I_m^{-1}(V) = \begin{pmatrix} \frac{\sigma^2}{m} & 0 \\ 0 & \frac{2\sigma^4}{m} \end{pmatrix}. \tag{37}$$

Corollary 2. *The limiting distribution of the MLE is given by:*

$$\begin{pmatrix} \hat{\mu} \\ \hat{\sigma}^2 \end{pmatrix} \xrightarrow{d} N\left(\begin{pmatrix} \mu \\ \sigma^2 \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2}{m} & 0 \\ 0 & \frac{2\sigma^4}{m} \end{pmatrix} \right), \quad as \quad m \to +\infty. \tag{38}$$

The proof of Corollary 2 comes from Theorem 7, Equation (26).

4. Fitting Tempered Stable Distribution to Cryptocurrencies: Bitcoin (BTC) and Ethereum

4.1. Data Summaries

Bitcoin was the first cryptocurrency created in 2009 by Satoshi Nakamoto. The idea behind Bitcoin was to create a peer-to-peer electronic payment system that allows online payments to be sent directly from one party to another without going through a financial institution (Nakamoto 2008). Since its inception, Bitcoin has grown in popularity and adoption and is now viewed as a viable legal tender in some countries. Bitcoin is currently used more as an investment tool, a risk-diversified tool, and less as a medium of exchange, a store of value, or a unit of account (Nzokem and Maposa 2024).

Bitcoin (BTC) and Ethereum (ETH) prices were extracted from CoinMarketCap. The period spans from 28 April 2013 to 4 July 2024 for Bitcoin and from 7 August 2015 to 4 July 2024 for Ethereum.

The daily price dynamics are provided in Figure 1. The prices have an increasing trend, even after having major significant increases and decreases over the studied period. Figure 1a,b show that Bitcoin outperforms Ethereum, which is the second-largest cryptocurrency by market capitalization after Bitcoin.

(**a**) Bitcoin (BTC) daily price   (**b**) Ethereum (ETH) daily price
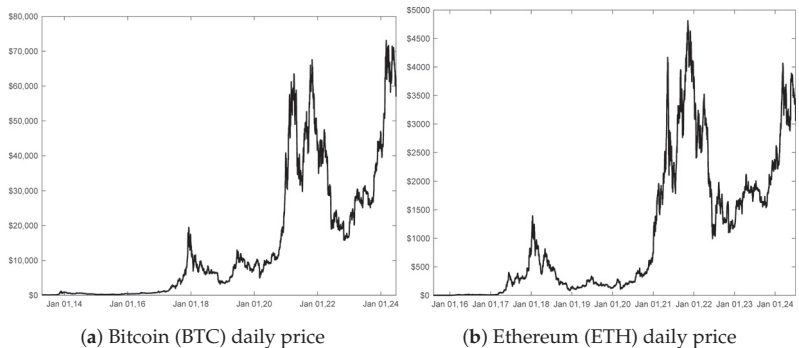
Figure 1. Daily price.

Let m be the number of observations and S_j be the daily observed price on the day t_j with $j = 1, \ldots, m$. The daily return (y_j) is computed as follows:

$$y_j = \log(S_j/S_{j-1}) \qquad j = 2, \dots, m. \tag{39}$$

As shown in Figure 2a,b, the daily return reaches the lowest level (-46% for Bitcoin and -55% for Ethereum) in the first quarter of 2020 amid the coronavirus pandemic and massive disruptions in the global economy. Nine values were identified as outliers and removed from the dataset to avoid a negative impact on the GTS model estimation and the empirical statistics.

(**a**) Daily Bitcoin return   (**b**) Daily Ethereum return
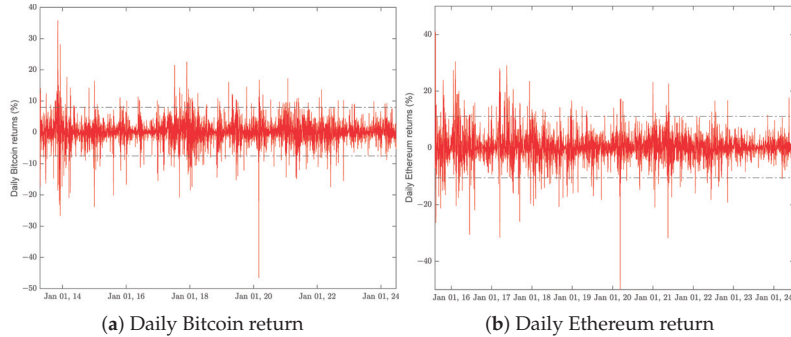
Figure 2. Daily return.

4.2. Multidimensional Estimation Results for Cryptocurrencies

The results of the GTS parameter estimation are summarized in Table 1 for Bitcoin and Table 2 for Ethereum data. The brackets are the asymptotic standard errors computed using the inverse of the Hessian matrix built in (20). The ML estimate of μ is negative for both Bitcoin and Ethereum, while others are positive, as expected in the literature. The asymptotic standard error for μ is quite large and suggests that μ is not statistically significant at 5%.

Table 1. Maximum-likelihood GTS parameter estimation for Bitcoin.

Model	Parameter	Estimate	Std Err	z	$Pr(Z > \|z\|)$	[95% Conf.Interval]	
GTS	μ	-0.121571	(0.375)	-0.32	7.5×10^{-01}	-0.856	0.613
	β_+	0.315548	(0.136)	2.33	2.0×10^{-02}	0.050	0.581
	β_-	0.406563	(0.117)	3.48	4.9×10^{-04}	0.178	0.635
	α_+	0.747714	(0.047)	15.76	6.2×10^{-56}	0.655	0.841
	α_-	0.544565	(0.037)	14.56	4.8×10^{-48}	0.471	0.618
	λ_+	0.246530	(0.036)	6.91	4.9×10^{-12}	0.177	0.316
	λ_-	0.174772	(0.026)	6.69	2.2×10^{-11}	0.124	0.226
	Log(ML)	$-10,606$					
	AIC	21,227					
	BIK	21,271					
GBM	μ	0.151997	(0.060)	2.51	1.2×10^{-02}	0.033	0.271
	σ	3.865132	(0.330)	11.69	7.2×10^{-32}	3.217	4.513
	Log(ML)	$-11,313$					
	AIC	22,630					
	BIK	22,638					

The log-likelihood, Akaike's information Criteria (AIC), and Bayesian information criteria (BIK) statistics show that the GTS distribution with seven parameters performs better than the two-parameter normal distribution (GBM). A comprehensive and detailed examination of the statistical significance of the results will be carried out in Section 6.

Table 1 summarizes the estimation results for Bitcoin returns. The skewness parameters (λ_+, λ_-) are statistically significant at 5%. The difference is positive and statistically

significant, which proves that the Bitcoin return is asymmetric and skewed to the left. The process intensity parameters (α_+, α_-) are statistically significant at 5%. Similarly, the difference is positive and statistically significant, showing that the Bitcoin is more likely to produce positive returns than negative ones. The index of stability parameters (β_+, β_-) are both statistically significant at 5%. However, the difference is positive but not statistically significant.

The GTS distribution with $\beta = \beta_+ = \beta_-$, called the Kobol distribution, was fitted to the Bitcoin data as well, and the estimation results are presented in Appendix B.1. As shown in Table A6, all the parameters are statistically significant at 5%, and have the expected positive sign. However, the likelihood ratio test in Table 6 shows that the GTS distribution in Table 1 is not significantly different from the Kobol distribution as the *p*-value (69.6%) is large. Refer to (Boyarchenko and Levendorskii 2002) for more details on the Kobol distribution.

As shown in Table 2, the parameters for Ethereum returns data are statistically significant at 5%, except μ and β_-. The difference ($\lambda_+ - \lambda_-$) in skewness parameters is negative and not statistically significant, showing that the Ethereum return is asymmetric and skewed to the right. Similarly, the difference ($\alpha_+ - \alpha_-$) in the intensity parameters is positive and not statistically significant, as shown the confidence interval. Contrary to the Bitcoin return, the Ethereum return has a larger process intensity, which provides evidence that Ethereum has a lower level of peakedness and a higher level of thickness.

Table 2. Maximum-likelihood GTS parameter estimation for Ethereum.

Model	Param	Estimate	Std Err	z	$Pr(Z > |z|)$	[95% Conf.Interval]	
GTS	μ	−0.4854	(1.008)	−0.48	6.3×10^{-01}	−2.461	1.491
	β_+	0.3904	(0.164)	2.38	1.7×10^{-02}	0.069	0.712
	β_-	0.4045	(0.210)	1.93	5.4×10^{-02}	−0.007	0.816
	α_+	0.9582	(0.106)	9.01	1.1×10^{-19}	0.750	1.167
	α_-	0.8005	(0.110)	7.25	4.2×10^{-13}	0.584	1.017
	λ_+	0.1667	(0.029)	5.72	1.1×10^{-08}	0.110	0.224
	λ_-	0.1708	(0.036)	4.71	2.5×10^{-06}	0.110	0.242
	Log(ML)	−9552					
	AIC	19,119					
	BIK	19,162					
GBM	μ	0.267284	(0.091)	2.93	3.4×10^{-03}	0.088	0.446
	σ	5.205539	(0.672)	7.74	1.0×10^{-14}	3.887	6.524
	Log(ML)	−9960					
	AIC	19,925					
	BIK	19,933					

We consider the following constraints $\lambda = \lambda_+ = \lambda_-$ and $\beta = \beta_+ = \beta_-$, which are the Carr–Geman–Madan–Yor (CGMY) distribution, also called the Classical Tempered Stable Distribution. The CGMY distribution was fitted as well, and the estimation results are presented in Appendix B.2. As shown in Table A8, all the parameters are statistically significant at 5%, and have the expected positive sign. However, the likelihood ratio test in Table 6 shows, with a high *p*-value (35.3%), that the GTS distribution is not significantly different from the CGMY distribution, and the null hypothesis cannot be rejected. Refer to (Carr et al. 2003; Rachev et al. 2011) for more details on the CGMY distribution.

Tables 1 and 2 summarize the last row of Tables A1 and A2, respectively, in Appendix A.1, which describes the convergence process of the GTS parameter for Bitcoin and Ethereum data. The convergence process was obtained using the Newton–Raphson iteration algorithm (23). Each row has eleven columns made of the iteration number, the seven parameters μ, β_+, β_-, α_+, α_-, λ_+, λ_-, and three statistical indicators, the log-likelihood ($Log(ML)$), the norm of the partial derivatives ($||\frac{dLog(ML)}{dV}||$), and the maximum value of the eigenvalues (*MaxEigenValue*). The statistical indicators aim at checking if the two

necessary and sufficient conditions described in (22) are all met. $Log(ML)$ displays the value of the Naperian logarithm of the likelihood function $L(x, V)$, as described in (20); $||\frac{dLog(ML)}{dV}||$ displays the value of the norm of the first derivatives ($\frac{dl(x,V)}{dV_j}$) described in (21); and $MaxEigenValue$ displays the maximum value of the seven eigenvalues generated by the Hessian matrix ($\frac{d^2l(x,V)}{dV_k dV_j}$), as described in (21).

Similarly, Tables A7 and A9 describe the convergence process of the Kobol distribution parameter for Bitcoin returns and the CGMY distribution parameter for Ethereum returns.

GTS parameter estimations in Tables 1 and 2 are used to evaluate the impact of each parameter on the GTS probability density function. As shown in Figures 3 and 4, the effect of the GTS parameters on the probability density function has the same patterns on Bitcoin and Ethereum returns. However, the magnitudes are different. As shown in Figure 3a,b, β_- (α_-) has a higher effect on the probability density function (pdf) than β_+ (α_+). However, λ_- and λ_+ in both graphs seem symmetric and have the same impact.

(a) $V_j = \beta_+, V_j = \beta_-$ **(b)** $V_j = \alpha_+, V_j = \alpha_-$ **(c)** $V_j = \lambda_+, V_j = \lambda_-$
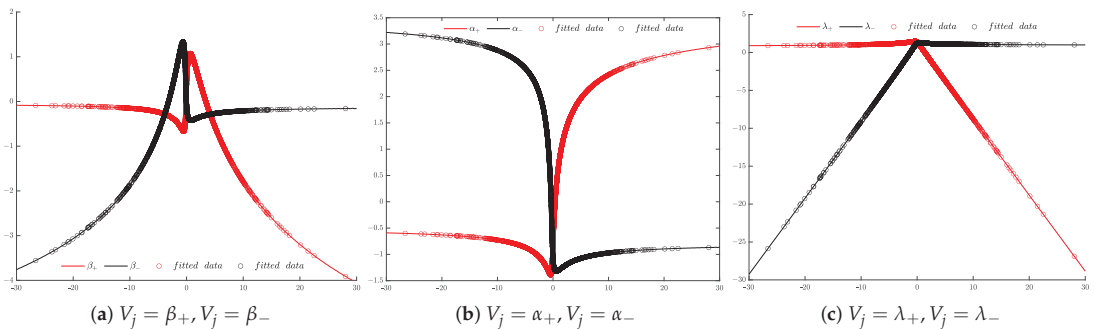
Figure 3. $\frac{\frac{df(x,V)}{dV_j}}{f(x,V)}$: Effect of parameters on the GTS probability density (Bitcoin returns).

(a) $V_j = \beta_+, V_j = \beta_-$ **(b)** $V_j = \alpha_+, V_j = \alpha_-$ **(c)** $V_j = \lambda_+, V_j = \lambda_-$
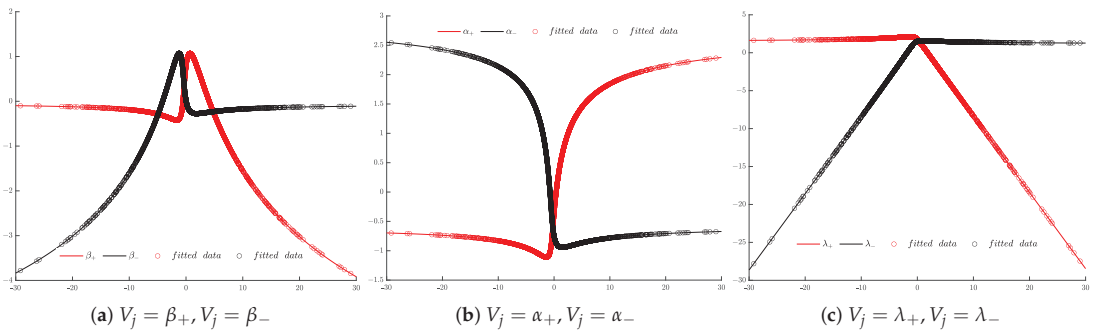
Figure 4. $\frac{\frac{df(x,V)}{dV_j}}{f(x,V)}$: Effect of parameters on the GTS probability density (Ethereum returns).

4.3. Evaluation of the Method of Moments

The method of moments estimates the parameters of the GTS distribution by equating empirical moments and the theoretical moments of the GTS distribution. We empirically estimate the kth moments ($m_k = E(x^k)$), based on sample data $x = (x_j)_{1 \le j \le m}$ as follows:

$$\hat{m}_k = \frac{1}{M} \sum_{j=1}^{m} x_j^k \quad \text{for} \quad k = 1, \ldots, 7. \tag{40}$$

On the other side, the cumulants (κ_k) in Theorem 4 can be related to the moment of the GTS distribution by the following relationship (Poloskov 2021; Rota and Shen 2000; Smith 1995):

$$m_k = E(x^k) = \sum_{j=1}^{k-1} \binom{k-1}{j-1} \kappa_j m_{k-j} + \kappa_k \quad \text{for} \quad k = 1, \ldots, 7. \tag{41}$$

The method of moments estimator for $V = (\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, \lambda_-)$ is defined as the solution to the following system of equations:

$$\hat{m}_k = m_k \quad \text{for} \quad k = 1, \ldots, 7. \tag{42}$$

The system of Equation (42) is often not analytically solvable. For the conditions of existence and uniqueness of the solution, refer to (Küchler and Tappe 2013).

The maximum-likelihood GTS parameter estimations in Tables 1 and 2 are used to evaluate the system of equations in (42). As shown in Table 3, the solution of the maximum-likelihood method satisfies to a certain extent the equations for the first four moments: $\hat{m}_1$, $\hat{m}_2$, $\hat{m}_3$, $\hat{m}_4$ in the system (42). The seventh-moment equation has the highest relative error: 89.9% for Bitcoin (BTC) and 68.3% for Ethereum. Therefore, the maximum likelihood GTS parameter estimation is not the same as the GTS parameter estimation from the method of moments.

In addition to the method of moments estimations, the lower relative errors in Table 3 show that empirical and theoretical standard deviation (σ), skewness, and kurtosis seem to be consistent for Bitcoin and Ethereum. The empirical and theoretical statistics show that the average Ethereum daily return is greater and more volatile than the Bitcoin daily returns. Both assets are thicker than the normal distribution. However, the daily return of Bitcoin is skewed to the left, whereas the daily return of Ethereum is skewed to the right.

Table 3. Evaluation of the method of moments.

	Bitcoin BTC			Ethereum		
	Empirical(1)	Theoretical(2)	$\frac{(1)-(2)}{2}$	Empirical(1)	Theoretical(2)	$\frac{(1)-(2)}{2}$
Sample size	4083			3246		
$\hat{m}_1$	0.152	0.152	0.0%	0.267	0.267	0.0%
$\hat{m}_2$	14.960	15.020	0.4%	27.161	27.388	0.8%
$\hat{m}_3$	−11.320	−15.640	27.6%	55.363	57.867	4.3%
$\hat{m}_4$	2033	2256	9.8%	5267	6307	16.5%
$\hat{m}_5$	−5823	−15,480	62.3%	22,368	32518	31.2%
$\hat{m}_6$	670,695	1,123,215	40.2%	2,114,788	4,361,562	51.5%
$\hat{m}_7$	−1,997,196	−19,777,988	89.9%	12,411,809	39,253,001	68.3%
Standard deviation [1]	3.865	3.873	0.2%	5.206	5.226	0.4%
Skewness [2]	−0.314	−0.387	18.8%	0.238	0.252	5.2%
Kurtosis [3]	9.154	10.082	9.2%	7.112	8.385	15.2%
Max value	28.052			29.013		
Min value	−26.620			−29.174		

[1] $\sigma = \sqrt{\kappa_2}$; [2] Skewness is estimated as $\frac{\kappa_3}{\kappa_2^{3/2}}$; [3] Kurtosis is estimated as $3 + \frac{\kappa_4}{\kappa_2^2}$; κ_1, κ_2 and κ_2 are defined in (13).

5. Fitting Tempered Stable Distribution to Traditional Indices: S&P 500 and SPY EFT

5.1. Data Summaries

The Standard & Poor's 500 Composite Stock Price Index, also known as the S&P 500, is a stock index that tracks the share prices of 500 of the largest public companies with stocks listed on the New York Stock Exchange (NYSE) and the Nasdaq in the United States. It was introduced in 1957 and is often treated as a proxy for describing the overall health of the stock market or the United States (US) economy. The SPDR S&P 500 ETF (SPY), also known as the SPY ETF, is an Exchange-Traded Fund (ETF) that tracks the performance of the S&P 500. SPY ETF provides a mutual fund's diversification, the stock's flexibility, and lower trading fees. The data were extracted from Yahoo Finance. The historical prices span from 4 January 2010 to 22 July 2024 and were adjusted for splits and dividends.

The daily price dynamics are provided in Figure 5. Prices have an increasing trend, even after being temporally disrupted in the first quarter of 2020 by the coronavirus pandemic. The S&P 500 is priced in thousands of US dollars, whereas the SPY ETF is in

hundreds of US dollars. The SPY ETF is cheaper and provides all the attributes of the S&P 500 index, as shown in Figure 5a,b.

(a) S&P 500 Daily Price **(b)** SPY EFT Daily Price
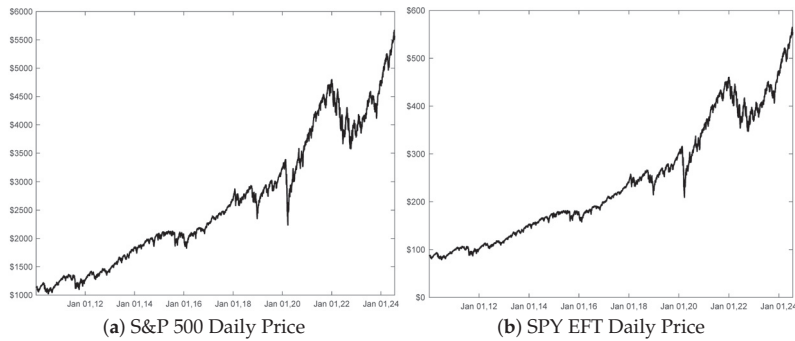
Figure 5. Daily price.

Let the number of observations be m and the daily observed price be S_j on day t_j with $j = 1, \ldots, m$; t_1 is the first observation date (04 January 2010) and t_m is the last observation date (22 July 2024). The daily return, y_j, is computed as in (43):

$$y_j = \log(S_j/S_{j-1}) \qquad j = 2, \ldots, m. \tag{43}$$

The SPY ETF aims to mirror the performance of the S&P 500. Figure 6a,b look similar, which is consistent with the goal of the SPY ETF. As shown in Figure 6a,b, the daily return reaches the lowest level (-12.7% for the S&P 500 and -11.5% for the SPY ETF) in the first quarter of 2020 amid the coronavirus pandemic and massive disruptions in the global economy. Nine values were identified as outliers and removed from the dataset to avoid a negative impact on GTS model estimation and the empirical statistics.

(a) Daily S&P500 return **(b)** Daily SPY ETF return
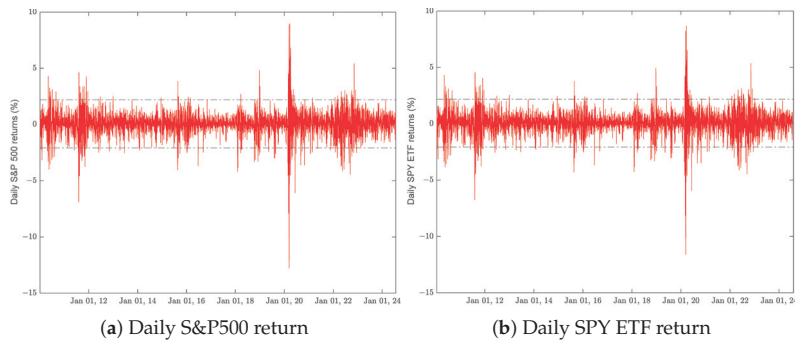
Figure 6. Daily return.

5.2. Multidimensional Estimation Results for Traditional Indices

The estimation results are provided in Table 4 for S&P 500 return data and Table 5 for SPY EFT return data. As previously, the log-likelihood, AIC, and BIK statistics suggest that the GTS distribution with seven parameters performs better than the two-parameter normal distribution (GBM).

As shown in both Tables 4 and 5, the ML estimate of μ is negative, while the others are positive, as expected in the literature. The asymptotic standard error for μ, β_+ and β_- are pretty large and result in μ, β_+ and β_- not being significantly different from zero.

Table 4. Maximum-likelihood GTS parameter estimation for S&P 500 index.

Model	Param	Estimate	Std Err	z	$Pr(Z > |z|)$	[95% Conf.Interval]	
GTS	μ	−0.249408	(0.208)	−1.20	2.3×10^{-01}	−0.658	0.159
	β_+	0.328624	(0.308)	1.07	2.9×10^{-01}	−0.275	0.932
	β_-	0.088640	(0.176)	0.50	6.1×10^{-01}	−0.256	0.433
	α_+	0.792426	(0.350)	2.26	2.4×10^{-02}	0.106	1.479
	α_-	0.542250	(0.107)	5.09	3.6×10^{-07}	0.333	0.751
	λ_+	1.279743	(0.348)	3.68	2.4×10^{-04}	0.597	1.962
	λ_-	0.937133	(0.144)	6.50	8.0×10^{-11}	0.655	1.220
	Log(ML)	−4920					
	AIC	9851					
	BIK	9898					
GBM	μ	0.044875	(0.018)	2.51	1.2×10^{-02}	0.010	0.080
	σ	1.081676	(0.027)	39.53	0.000	1.028	1.135
	Log(ML)	−5330					
	AIC	10,665					
	BIK	10,677					

However, other parameters have larger t-statistics ($|z| > 2$) and are statistically significant at 5%. Except for the index of stability parameters (β_+, β_-), the estimation results for the S&P 500 and SPY ETF indexes show that the difference in skewness parameters (λ_+, λ_-) and intensity parameters (α_+, α_-) are positive but are not statistically significant.

The hypothesis with $\beta_+ = \beta_- = 0$ was considered by fitting the S&P 500 and SPY ETF indexes to the bilateral Gamma distribution. The estimation results are summarized in Appendices C.1 and C.2. As shown in Tables A10 and A12, the skewness parameters (λ_+, λ_-) are positive and statistically significant, and the difference ($\lambda_+ - \lambda_-$) is also positive and statistically significant, which proves that the S&P 500 and SPY ETF returns are skewed to the left. We have the same statistical features for the intensity parameters (α_+, α_-), and both indexes are more likely to produce positive returns than negative returns. Refer to (Küchler and Tappe 2008; Nzokem 2021a) for more details on the bilateral Gamma distribution.

Table 5. Maximum-likelihood GTS parameter estimation for SPY EFT data.

Model	Param	Estimate	Std Err	z	$Pr(Z > |z|)$	[95% Conf.Interval]	
GTS	μ	−0.260643	(0.135)	−1.94	5.3×10^{-02}	−0.524	0.003
	β_+	0.340880	(0.189)	1.80	7.1×10^{-02}	−0.030	0.711
	β_-	0.022212	(0.212)	0.10	9.2×10^{-01}	−0.393	0.437
	α_+	0.787757	(0.225)	3.50	4.6×10^{-04}	0.347	1.229
	α_-	0.597110	(0.141)	4.22	2.4×10^{-05}	0.320	0.874
	λ_+	1.288555	(0.226)	5.70	1.2×10^{-08}	0.846	1.731
	λ_-	1.014353	(0.177)	5.74	9.4×10^{-09}	0.668	1.361
	Log(ML)	−4893					
	AIC	9800					
	BIK	9843					
GBM	μ	0.054344	(0.017)	3.13	1.8×10^{-03}	0.020	0.088
	σ	1.050217	(0.026)	40.71	0.000	1.000	1.101
	Log(ML)	−54,275					
	AIC	10,554					
	BIK	10,566					

The likelihood ratio test in Table 6 shows that, even with non-statistically significant parameters, the GTS distribution fits significantly better than the bilateral Gamma distribution for both the S&P 500 and SPY ETF indexes. Contrary to the AIC statistics, the BIK

statistics do not provide the same information. A comprehensive and detailed examination of the statistical significance of the results is carried out in Section 6.

Tables 4 and 5 summarize the last row of Tables A3 and A4, respectively, in Appendix A.1, which describes the convergence process of the GTS parameter for S&P 500 index, and SPY ETF return data. The convergence process was obtained using the Newton–Raphson iteration algorithm (23). Each row has eleven columns made of the iteration number; the seven parameters μ, β_+, β_-, α_+, α_-, λ_+, λ_-; and three statistical indicators, the log-likelihood ($Log(ML)$), the norm of the partial derivatives ($||\frac{dLog(ML)}{dV}||$), and the maximum value of the eigenvalues ($MaxEigenValue$). The statistical indicators aim at checking if the two necessary and sufficient conditions described in (22) are all met. $Log(ML)$ displays the value of the Naperian logarithm of the likelihood function $L(x, V)$, as described in (20); $||\frac{dLog(ML)}{dV}||$ displays the value of the norm of the first derivatives ($\frac{dl(x,V)}{dV_j}$) described in Equation (21); and $MaxEigenValue$ displays the maximum value of the seven eigenvalues generated by the Hessian matrix ($\frac{d^2l(x,V)}{dV_k dV_j}$), as described in (21).

Similarly, Tables A11 and A13 describe the convergence process of the bilateral Gamma distribution parameter for S&P 500 index and SPY ETF return data.

The GTS parameter estimations in Tables 4 and 5 were used to evaluate the impact of the parameters on the GTS probability density function. As shown in Figures 7 and 8, the effect of the GTS parameters on the probability density function generated by the S&P 500 and SPY ETF have the same patterns. However, the magnitudes are different. As shown in Figure 7a,b on the S&P 500 return data, β_+ (α_+) has a higher effect on the probability density function than β_- (α_-). However, λ_- and λ_+ in Figure 7c are symmetric and have the same impact.

(**a**) $V_j = \beta_+, V_j = \beta_-$ (**b**) $V_j = \alpha_+, V_j = \alpha_-$ (**c**) $V_j = \lambda_+, V_j = \lambda_-$
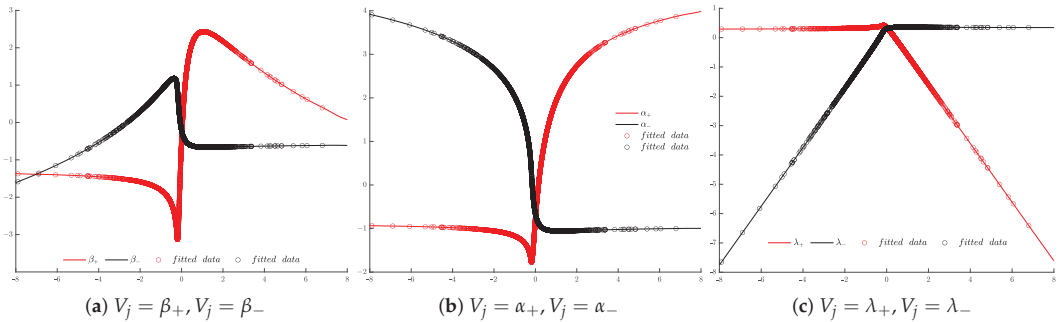
Figure 7. $\frac{\frac{df(x,V)}{dV_j}}{f(x,V)}$: Effect of parameters on the GTS probability density (S&P 500 index).

(**a**) $V_j = \beta_+, V_j = \beta_-$ (**b**) $V_j = \alpha_+, V_j = \alpha_-$ (**c**) $V_j = \lambda_+, V_j = \lambda_-$
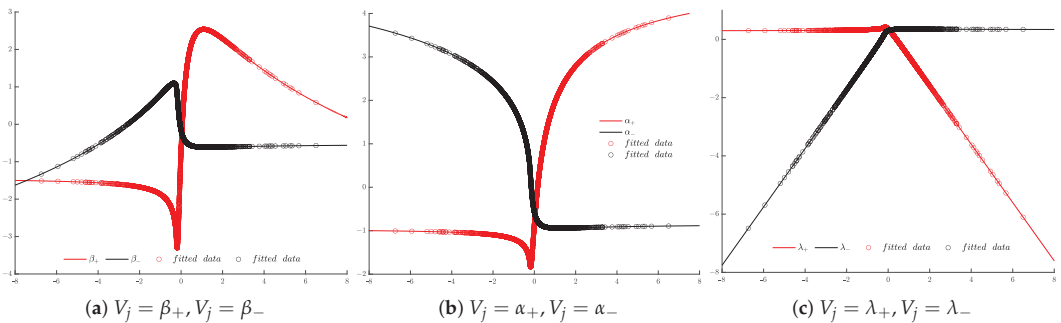
Figure 8. $\frac{\frac{df(x,V)}{dV_j}}{f(x,V)}$: Effect of parameters on the GTS probability density (SPY EFT).

Table 6. Likelihood ratio test statistic and *p*-value.

		GTS	**GTS Variants**	χ^2**-Value**	**df**	**p-Value**
Bitcoin	**Log(ML)**	−10,606.73	−10,606.81	0.1525	1	0.6962
	AIC	21,227.47	21,225.62			
	BIK	21,271.67	21,263.51			
Ethereum	**Log(ML)**	−9552.86	−9553.90	2.0810	2	0.3533
	AIC	19,119.72	19,117.81			
	BIK	19,162.32	19,148.23			
S&P 500	**Log(ML)**	−4920.52	−4924.62	8.1828	2	0.0167
	AIC	9851.06	9859.24			
	BIK	9898.49	9890.26			
SPY ETF	**Log(ML)**	−4893.21	−4898.67	10.9234	2	0.0042
	AIC	9800.42	9807.34			
	BIK	9843.84	9838.36			

5.3. Evaluation of the Methods of Moments

The maximum-likelihood GTS parameter estimations in Tables 4 and 5 are used to evaluate the system of equations in (42). As shown in Table 7, the solution of the maximum-likelihood method satisfies to a certain extent the equations (42) for the following first four moments: $\hat{m}_1, \hat{m}_2, \hat{m}_4, \hat{m}_5$. As for Bitcoin and Ethereum, the seventh-moment equation has the highest relative error: 53.3% for S&P 500 index and −85.9% for SPY ETF. Therefore, the maximum-likelihood GTS parameter estimation is not the GTS parameter estimation from the method of moments.

In addition to the moment estimations in Table 7, the empirical and theoretical standard deviation (σ), skewness, and kurtosis are consistent with lower relative errors for both S&P 500 and SPY ETF. The empirical and theoretical statistics show that both assets are skewed to the left and also thicker than the normal distribution.

Table 7. Evaluation of the methods of moments.

	S&P 500 Index			**SPY ETF**		
	Empirical(1)	**Theoretical(2)**	$\frac{(1)-(2)}{2}$	**Empirical(1)**	**Theoretical(2)**	$\frac{(1)-(2)}{2}$
Sample size	3656			3655		
$\hat{m}_1$	0.045	0.045	−0.5%	0.054	0.054	0.0%
$\hat{m}_2$	1.069	1.083	−1.3%	1.053	1.044	0.8%
$\hat{m}_3$	−0.447	−0.341	31.2%	−0.214	−0.351	−39.0%
$\hat{m}_4$	8.371	9.764	−14.3%	8.197	7.691	6.6%
$\hat{m}_5$	−16.386	−11.128	47.3%	−3.969	−12.717	−68.8%
$\hat{m}_6$	193.563	247.811	−21.9%	157.645	162.048	−2.7%
$\hat{m}_7$	−840.097	−547.882	53.3%	−85.003	−602.447	−85.9%
Standard deviation [1]	1.082	1.033	4.7%	1.050	1.021	2.9%
Skewness [2]	−0.432	−0.535	−19.2%	−0.358	−0.490	−26.9%
Kurtosis [3]	8.413	7.435	13.1%	7.495	7.177	4.4%
Max value	6.797			6.501		
Min value	−7.901			−6.734		

[1] $\sigma = \sqrt{\kappa_2}$; [2] Skewness is estimated as $\frac{\kappa_3}{\kappa_2^{3/2}}$; [3] Kurtosis is estimated as $3 + \frac{\kappa_4}{\kappa_2^2}$; κ_1, κ_2 and κ_2 are defined in (13).

6. Goodness-of-Fit Analysis

6.1. Kolmogorov–Smirnov (KS) Analysis

Given the sample of daily return $\{y_1, y_2 \ldots y_m\}$ of size m and the empirical cumulative distribution function, $F_m(x)$, for each index, the Kolmogorov–Smirnov (KS) test is performed under the null hypothesis, H_0, that the sample $\{y_1, y_2 \ldots y_m\}$ comes from the GTS distribution, $F(x)$. The cumulative distribution function of the theoretical distribution, $F(x)$, needs to be computed. The density function, $f(x)$, does not have a closed form, the same for the cumulative function, $F(x)$, in (45). However, we know the closed form of the Fourier of the density function, $\mathscr{F}[f]$, and the relationship in (46) provides the Fourier

of the cumulative distribution function, $\mathscr{F}[F]$. The GTS distribution function, $F(x)$, was computed from the inverse of the Fourier of the cumulative distribution, $\mathscr{F}[F]$, in (47):

$$Y \sim GTS(\mu, \beta_+, \beta_-, \alpha_+, \alpha_-, \lambda_+, f\lambda_-) \tag{44}$$

$$F(x) = \int_{-\infty}^{x} f(t)\mathrm{d}t \quad f \text{ is the density function of } Y \tag{45}$$

$$\mathscr{F}[F](x) = \frac{\mathscr{F}[f](x)}{ix} + \pi\mathscr{F}[f](0)\delta(x) \tag{46}$$

$$F(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \frac{\mathscr{F}[f](y)}{iy} e^{ixy} \, \mathrm{d}y + \frac{1}{2} \tag{47}$$

See Appendix A in (Nzokem 2021a) for (46) proof.

The two-sided KS goodness-of-fit statistic (D_m) is defined as follows:

$$D_m = \sup_x |F(x) - F_m(x)|, \tag{48}$$

where m is the sample size, $F_m(x)$ denotes the empirical cumulative distribution of $\{y_1, y_2 \ldots y_m\}$.

The distribution of Kolmogorov's goodness-of-fit measure D_m has been studied extensively in the literature. It was shown (Massey 1951) that the D_m distribution is independent of the theoretical distribution, $F(x)$, under the null hypothesis, H_0. The discrete, mixed, and discontinuous distributions case has also been studied (Dimitrova et al. 2020). Under the null hypothesis, H_0, that the sample $\{y_1, y_2 \ldots y_m\}$ of size m comes from the hypothesized continuous distribution, it was shown (An 1933) that the asymptotic statistic $\sqrt{n}D_n$ converges to the Kolmogorov distribution.

The limiting form for the distribution function of Kolmogorov's goodness-of-fit measure D_m is

$$\lim_{m \to +\infty} Pr(\sqrt{m}D_m \leq x) = 1 - 2\sum_{k=1}^{+\infty} (-1)^{k-1} e^{-2k^2x^2} = \frac{\sqrt{2\pi}}{x} \sum_{k=1}^{+\infty} e^{-\frac{(2k-1)^2\pi^2}{8x^2}}. \tag{49}$$

The first representation was given in (An 1933), and the second came from a standard relation for theta functions (Marsaglia et al. 2003).

As shown in Figure 9, the asymptotic statistic, $\sqrt{n}D_n$, is a positively skewed distribution with a mean and a standard deviation (Marsaglia et al. 2003) as follows.

$$\mu = \sqrt{\frac{\pi}{2}} log(2) \sim 0.8687, \qquad \sigma = \sqrt{\frac{\pi^2}{12} - \mu^2} \sim 0.2603. \tag{50}$$

Figure 9. Asymptotic statistic ($\sqrt{m}D_m$) probability density function (PDF).

At a 5% risk level, the risk threshold is $d = 1.3581$ and represents the area in the shaded area under the probability density function.

The p-value of the test statistic, D_m, is computed based on (49) as follows:

$$p\_value = Pr(D_m > \hat{D}_m | H_0) = 1 - Pr(\sqrt{m}D_m \leq \sqrt{m}\hat{D}_m). \tag{51}$$

A p-value is defined as the probability that values are even more extreme or more in the tail than our test statistic. A small p-value leads to a rejection of the null hypothesis, H_0, because the test statistic, D_m, is already extreme. We reject the hypothesis if the p-value is less than our level of significance, which we take to be equal to 0.05.

$\hat{D}_m$ is a realization value of the KS estimator D_m computed from the sample $\{y_1, y_2 \ldots y_m\}$. $\hat{D}_m$ is estimated (Krysicki et al. 1999) as follows:

$$\hat{D}_m = Max(\sup_{0 \leq j \leq P} |F(x_j) - F_m(x_j)|, \sup_{1 \leq j \leq P} |F(x_j) - F_m(x_{j-1})|). \tag{52}$$

The following computations were performed for Bitcoin (BTC) data, and the quantity $\hat{D}_m$ was obtained:

$$\begin{aligned} \sup_{0 \leq j \leq P} |F(x_j) - F_m(x_j)| &= 0.01300 \\ \sup_{1 \leq j \leq P} |F(x_j) - F_m(x_{j-1})| &= 0.00538 \\ \hat{D}_m &= 0.01300 \\ p\_value &= prob(\sqrt{m}D_m > 0.6903 | H_0) = 49.48\%. \end{aligned} \tag{53}$$

For each index, KS statistics ($\hat{D}_m$) and associated p-values were computed and summarized in Table 8, along with the index sample size, m.

Table 8. Kolmogorov–Smirnov statistics and p-values.

Index	GTS			GBN			GTS Variants			Sample Size
	$\hat{D}_m$	$\sqrt{m}\hat{D}_m$	p-Value	$\hat{D}_m$	$\sqrt{m}\hat{D}_m$	p-Value	$\hat{D}_m$	$\sqrt{m}\hat{D}_m$	p-Value	m
Bicoin BTC	0.013	0.830	0.494	0.106	6.803	0.000	0.014 [1]	0.863	0.445	4083
Ethereum	0.012	0.721	0.674	0.092	5.249	0.000	0.013 [2]	0.749	0.627	3246
S&P 500	0.012	0.750	0.627	0.091	5.550	0.000	0.014 [3]	0.897	0.395	3656
SPY ETF	0.014	0.869	0.436	0.089	5.438	0.000	0.016 [3]	1.010	0.258	3655

[1] Kobol distribution ($\beta = \beta_- = \beta_+$); [2] Carr–Geman–Madan–Yor (CGMY) distributions ($\beta = \beta_- = \beta_+$; $\alpha = \alpha_- = \alpha_+$); [3] bilateral Gamma distribution ($\beta_- = \beta_+ = 0$).

The asymptotic statistics, $\sqrt{n}D_n$, produced from the two-parameter geometric Brownian motion (GBM) hypothesis, have high values and show that the GBM hypothesis is always rejected. On the other hand, the high p-values generated by the asymptotic statistics suggest insufficient evidence to reject the assumption that the data were randomly sampled from a GTS. The same observations work for the GTS variants: the Kobol, CGMY, and bilateral Gamma distributions. In addition, as shown the p-value indicator in Table 8, the GTS distribution outperforms the bilateral Gamma distribution for the S&P 500 and SPY ETF indexes. However, the Kobol and CGMY distributions, respectively, for Bitcoin and Ethereum have almost the same performance as the GTS distribution.

6.2. Anderson–Darling Test Analysis

The Anderson–Darling test (Anderson 2008) is a goodness-of-fit test that allows the control of the hypothesis that the distribution of a random variable observed in a sample follows a certain theoretical distribution. The Anderson–Darling statistic belongs to the class of quadratic EDF statistics (Stephens 1974) based on the empirical distribution function.

The quadratic EDF statistics measure the distance between the hypothesized distribution ($F(x)$) and empirical distribution. It is defined as

$$m \int_{-\infty}^{+\infty} (F_m(x) - F(x))^2 w(x)\, dFx, \tag{54}$$

where m is the number of elements in the sample, $w(x)$ is a weighting function, and $F_m(x)$ is the empirical distribution function defined on the sample of size m.

When the weighting function is $w(x) = 1$, the statistic (54) is the Cramér—Von Mises statistic, while the Anderson–Darling statistic is obtained by choosing the weighting function $w(x) = F(x)(1 - F(x))$. Compared with the Cramér–Von Mises statistic, the Anderson–Darling statistic places more weight on the tails of the distribution.

The Anderson–Darling statistic is

$$A_m^2 = m \int_{-\infty}^{+\infty} \frac{(F_m(x) - F(x))}{F(x)(1 - F(x))}\, dF(x). \tag{55}$$

It can be shown that the asymptotic distribution of the Anderson–Darling statistic, A_m^2, is independent of the theoretical distribution under the null hypothesis. The asymptotic distribution (Lewis 1961; Marsaglia and Marsaglia 2004) is defined as follows:

$$G(x) = \lim_{m \to \infty} Pr\left[A_m^2 < x\right] = \sum_{j=0}^{+\infty} a_j (xb_j)^{-\frac{1}{2}} exp\left(-\frac{b_j}{x}\right) \int_0^{+\infty} f_j(y) exp(-y^2)\, dy$$

$$f_j(y) = exp\left(\frac{1}{8} \frac{xb_j}{y^2x + b_j}\right), \qquad a_j = \frac{(-1)^j (2)^{\frac{1}{2}} (4j+1)\Gamma(j+\frac{1}{2})}{j!} \tag{56}$$

$$b_j = \frac{1}{2}(4j+1)^2 \pi^2.$$

As shown in Figure 10, the asymptotic distribution of the Anderson–Darling statistic (A_m^2) is a positively skewed distribution with a mean and a standard deviation (Anderson 2011) as follows

$$\mu = 1, \qquad \sigma = \sqrt{\frac{2}{3}(\pi^2 - 9)} \sim 0.761. \tag{57}$$

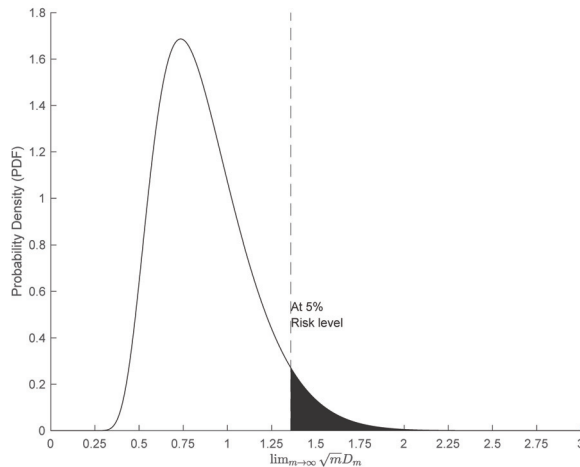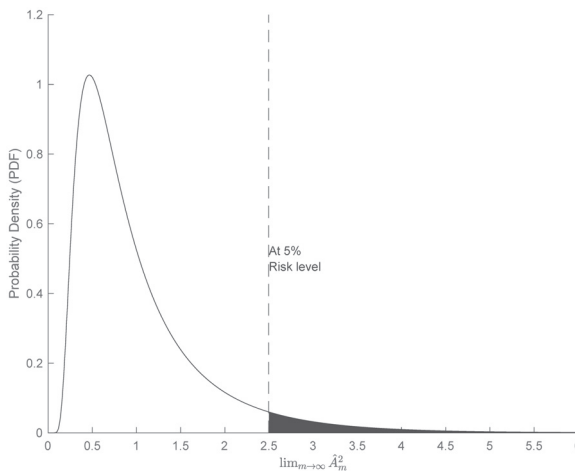**Figure 10.** Asymptotic Anderson–Darling statistic (A_m^2) probability density function (PDF).

At a 5% risk level, the risk threshold is $d = 2.4941$ and represents the area in the shaded area under the probability density function.

The *p*-value of the test statistic, A_m^2, is defined as follows:

$$p\text{-}value = prob(A_m^2 > \hat{A}_m^2 | H_0) = 1 - G(\hat{A}_m^2). \tag{58}$$

In order to compute the Anderson–Darling statistic, A_m^2, in (55), the sample of daily return $\{y_1, y_2 \ldots y_m\}$ of size m is arranged in ascending order: $y_{(1)} < y_{(2)} < \cdots < y_{(m)}$. The Anderson–Darling statistic (Lewis 1961) then becomes

$$A_m^2 = -m - \frac{1}{m} \sum_{j=1}^{m} \Big[(2j-1)log(F(y_{(j)})) + (2(n-j)+1)log(F(y_{(j)})) \Big]. \tag{59}$$

For each index, the Anderson–Darling statistic (59) is computed, along with the *p*-value statistic. Table 9 shows the KS statistics (A_m^2) and *p*-values for the GTS, GBM, and GTS variant distributions. While the two-parameter GBM hypothesis is always rejected, the GTS hypothesis is accepted and yields a very high *p*-value.

Table 9. Anderson–Darling statistics and *p*-values.

	GTS		**GBN**		**GTS Variants**		**Sample Size**
Index	$\hat{A}_m^2$	***p*-Value**	$\hat{A}_m^2$	***p*-Value**	$\hat{A}_m^2$	***p*-Value**	***m***
Bicoin BTC	0.1098	0.9999	99.706	0.0000	0.1105 [1]	0.9999	4083
Ethereum	0.1018	0.9999	59.157	0.0001	0.2123 [2]	0.9866	3246
S&P 500	0.3007	0.9376	54.304	0.0001	0.5010 [3]	0.7458	3656
SPY ETF	0.3017	0.9368	51.516	0.0001	0.6684 [3]	0.5857	3655

[1] Kobol distribution ($\beta = \beta_- = \beta_+$); [2] Carr–Geman–Madan–Yor (CGMY) distributions ($\beta = \beta_- = \beta_+$; $\alpha = \alpha_- = \alpha_+$); [3] bilateral Gamma distribution ($\beta_- = \beta_+ = 0$).

In addition, as shown by the *p*-value indicator in Table 9, the GTS distribution outperforms the bilateral Gamma distribution for the S&P 500 and SPY ETF indexes. However, the Kobol and CGMY distributions for Bitcoin and Ethereum, respectively, have almost the same performance as the GTS distribution.

6.3. Pearson's Chi-Squared Test Analysis

Pearson's chi-squared test (Schoutens 2003) counts the number of sample points falling into certain intervals and compares them with the expected number under the null hypothesis. Under the null hypothesis, H_0, a random sample $\{y_1, y_2 \ldots y_m\}$ comes from the GTS distribution, which has seven parameters estimated in Section 5. Suppose that m observations in the sample from a population are classified into K mutually exclusive classes with respective observed numbers of observations N_j (for $j = 1, 2, \ldots, K$), and a null hypothesis gives the probability $\Pi_j = F(x_j) - F(x_{j-1})$ (47) that an observation falls into the jth class.

The following Pearson statistic calculates the value of the chi-squared goodness-of-fit test:

$$\chi^2(K-1-p) = \sum_{j=1}^{K} \frac{(N_j - m\Pi_j)^2}{m\Pi_j}. \tag{60}$$

Under the null hypothesis assumption, as m goes to $+\infty$, the limiting distribution (Schoutens 2003) of the Pearson statistic (60) follows the $\chi^2(K-1-p)$ distribution with $K-1-p$ degrees of freedom, and p is the number of estimated parameters.

Table 10 shows the Pearson chi-squared statistics ($\hat{\chi}^2(K-1-p)$), *p*-values, and class number for the GTS, GBM, and GTS variant distributions. While the two-parameter GBM hypothesis is always rejected, the GTS hypothesis is accepted and yields a high *p*-value.

Table 10. Pearson statistics and *p*-values.

Index	GTS		GBN		GTS Variants			Class Number
	$\hat{\chi}^2(K-8)$	*p*-Value	$\hat{\chi}^2(K-3)$	*p*-Value	$\hat{\chi}^2(K-p-1)$	*p*	*p*-Value	K
Bicoin BTC	12.234	0.508	1375	0.000	12.549 [1]	6	0.562	21
Ethereum	6.910	0.863	805	0.000	8.618 [2]	5	0.854	20
S&P 500	9.886	0.703	574	0.000	12.844 [3]	5	0.614	21
SPY ETF	13.955	0.377	605	0.000	18.228 [3]	5	0.251	21

[1] Kobol distribution ($\beta = \beta_- = \beta_+$); [2] Carr–Geman–Madan–Yor (CGMY) distributions ($\beta = \beta_- = \beta_+$; $\alpha = \alpha_- = \alpha_+$); [3] bilateral Gamma distribution ($\beta_- = \beta_+ = 0$).

In addition, as shown by the *p*-value indicator in Table 10, the GTS distribution outperforms the bilateral Gamma distribution for the S&P 500 and SPY ETF indexes. However, the Kobol and CGMY distributions for Bitcoin and Ethereum, respectively, have almost the same performance as the GTS distribution. For more details on the estimation of the Pearson statistic inputs under the GTS distribution, refer to Table A5 in Appendix A.2.

7. Conclusions

This study provides a methodology for fitting the rich class of the seven-parameter GTS distribution to financial data. Four historical prices were considered in the methodology application: two heavy-tailed data (Bitcoin and Ethereum returns) and two peaked data (S&P 500 and SPY ETF returns). The study used each historical data to fit the seven-parameter GTS distribution to the underlying data return distribution. The advanced fast FRFT scheme, based on the classic fast FRFT algorithm and the 11-point composite Newton–Cotes rule, was used to perform the maximum-likelihood estimation of seven parameters of the GTS distribution. The maximum likelihood estimate results show that, for each index, the location parameter, μ, is negative, while the others are positive, as expected in the literature. The statistical significance of the parameters was analyzed. The non-statistical significance of the index of stability parameters (β_+, β_-) has led to the fitting of the Kobol, CGMY, and bilateral Gamma distributions. The goodness of fit was assessed through Kolmogorov–Smirnov, Anderson–Darling, and Pearson's chi-squared statistics. While the two-parameter GBM hypothesis is always rejected, the goodness-of-fit analysis shows that the GTS distribution fits significantly the four historical data with a very high *p*-value.

As a main limitation of the study, the applied methodology is computation-intensive, and the researchers need good skills in computer programming. In future work, the estimated parameter of the GTS distribution will be used in the Ornstein–Uhlenbeck-type process to simulate the daily cumulative returns of financial assets.

Appendix A

Appendix A.1. Iterative Maximum-Likelihood Estimation (MLE) Procedure

Table A1. Convergence of the GTS parameter for Bitcoin return data.

Iterations	μ	β_+	β_-	α_+	α_-	λ_+	λ_-	$Log(ML)$	$\|\frac{dLog(ML)}{dV}\|$	MaxEigenValue
1	−0.7369246	0.4613783	0.2671787	0.8100173	0.5173470	0.2156289	0.1919378	−10609.058	282.6765666	3.6240151
2	−0.7977019	0.4654390	0.2169392	0.7846817	0.4905332	0.2164395	0.2049523	−10607.253	26.7522215	−1.6194299
3	−0.4455841	0.3884721	0.3213867	0.7758150	0.5193395	0.2340187	0.1883953	−10607.001	50.1355291	3.0916011
4	−0.7634445	0.4521878	0.2217702	0.7935129	0.4959371	0.2218253	0.2055181	−10607.210	4.8235882	−2.6390063
5	−0.4906746	0.4146531	0.3404176	0.7722729	0.5222110	0.2269202	0.1846457	−10607.059	67.6646338	9.0971871
6	−0.5515834	0.4434827	0.3335905	0.7724484	0.5190619	0.2197566	0.1853686	−10607.022	17.4476962	−0.4021102
7	−0.4914586	0.4327714	0.3503012	0.7686883	0.5235361	0.2216450	0.1826269	−10606.991	16.2838831	−0.1781480
8	−0.2900908	0.3885350	0.3956186	0.7563357	0.5370260	0.2300772	0.1754994	−10606.864	12.0116477	−2.4090216
9	−0.2752698	0.3832660	0.3969704	0.7555456	0.5377367	0.2312224	0.1753571	−10606.847	11.4457840	−2.5487401
10	−0.2609339	0.3780400	0.3982456	0.7547812	0.5384209	0.2323632	0.1752258	−10606.832	10.8628213	−2.6874876
11	−0.2085409	0.3576927	0.4025762	0.7519966	0.5408864	0.2368544	0.1748113	−10606.782	8.3600783	−3.4356438
12	−0.1970109	0.3528575	0.4034002	0.7513923	0.5414138	0.2379362	0.1747455	−10606.772	7.6818408	−3.6954428
13	−0.1761733	0.3436416	0.4046818	0.7503191	0.5423414	0.2400174	0.1746675	−10606.756	6.2516380	−4.2766527
14	−0.1668421	0.3392794	0.4051522	0.7498492	0.5427438	0.2410120	0.1746529	−10606.750	5.5002876	−4.5807361
15	−0.1581860	0.3350854	0.4055256	0.7494209	0.5431090	0.2419740	0.1746517	−10606.745	4.7262048	−4.8824015
16	−0.1501600	0.3310612	0.4058166	0.7490311	0.5434404	0.2429024	0.1746615	−10606.741	3.9306487	−5.1742197
17	−0.1209376	0.3159301	0.4066945	0.7476393	0.5446224	0.2464122	0.1747326	−10606.734	2.8592342	−6.2251311
18	−0.1216487	0.3155707	0.4064438	0.7477179	0.5445608	0.2465247	0.1747753	−10606.734	0.0014787	−6.2014232
19	−0.1215714	0.3155483	0.4064635	0.7477142	0.5445652	0.2465296	0.1747719	−10606.734	1.82×10^{-06}	−6.2026532
20	−0.1215714	0.3155483	0.4064635	0.7477142	0.5445652	0.2465296	0.1747719	−10606.734	9.80×10^{-10}	−6.2026530

Table A2. Convergence of the GTS parameter for Ethereum return data.

Iterations	μ	β_+	β_-	α_+	α_-	λ_+	λ_-	$Log(ML)$	$\|\frac{dLog(ML)}{dV}\|$	MaxEigenValue
1	−0.1215714	0.3155483	0.4064635	0.7477142	0.5445652	0.2465296	0.1747719	−9745.171	2673.428257	206.013602
2	−0.1724835	0.3319505	0.4091022	0.7364129	0.5479094	0.2227870	0.1684568	−9700.715	2388.609394	180.884105
3	−0.2041418	0.3384742	0.4118929	0.7338794	0.5531083	0.2083203	0.1632896	−9669.986	2139.267660	157.699659
4	−0.4006157	0.3530035	0.4393474	0.7513784	0.6172425	0.1135743	0.1221930	−9586.115	1471.570475	32.410140
5	−0.6485551	0.4493817	0.4404508	0.9247887	0.7210031	0.1412949	0.1482307	−9556.026	380.605737	56.584055
6	−0.6290525	0.4371402	0.4359516	0.9780784	0.7824777	0.1582340	0.1608694	−9553.005	24.905322	−0.719221
7	−0.5545412	0.3994778	0.3918188	0.9627486	0.7936571	0.1652438	0.1724287	−9552.866	5.834338	−0.847574
8	−0.4744837	0.3913982	0.4093404	0.9582366	0.8022858	0.1665103	0.1699928	−9552.862	2.963150	−0.933466
9	−0.4825586	0.3902160	0.4051365	0.9580755	0.8007651	0.1667400	0.1706850	−9552.862	0.214871	−0.931142
10	−0.4853678	0.3904369	0.4044899	0.9582486	0.8004799	0.1667119	0.1707853	−9552.862	0.004754	−0.931872
11	−0.4853800	0.3904362	0.4044846	0.9582487	0.8004779	0.1667121	0.1707862	−9552.862	2.96×10^{-07}	−0.931836
12	−0.4853800	0.3904362	0.4044846	0.9582487	0.8004779	0.1667121	0.1707862	−9552.862	1.18×10^{-10}	−0.931836
13	−0.4853800	0.3904362	0.4044846	0.9582487	0.8004779	0.1667121	0.1707862	−9552.862	1.27×10^{-11}	−0.931836

Table A3. Convergence of the GTS parameter for S&P 500 return data.

Iterations	μ	β_+	β_-	α_+	α_-	λ_+	λ_-	$Log(ML)$	$\|\frac{dLog(ML)}{dV}\|$	MaxEigenValue
1	−0.2606426	0.34087979	0.02221141	0.78775729	0.59711061	1.28855513	1.01435308	−4921.0858	147.214541	−0.476265
2	−0.2747887	0.37848567	0.02517846	0.72538248	0.594628	1.22107935	1.01081205	−4920.9765	107.910271	−12.169518
3	−0.2852743	0.34562742	0.01628972	0.78353361	0.58024658	1.27423544	0.9888729	−4920.6236	23.70873	11.9588258
4	−0.2971254	0.37985815	0.05392593	0.74068472	0.55972179	1.22737986	0.96278568	−4920.5493	4.21443356	0.29705471
5	−0.3415082	0.42600675	0.0432239	0.69783497	0.56106365	1.18286494	0.966753	−4920.5722	37.0642417	−1.7903876
6	−0.2995817	0.40315129	0.12236507	0.7168274	0.522172	1.20383351	0.9117451	−4920.574	3.07232514	−0.7101089
7	−0.2944623	0.3977257	0.12218751	0.72174351	0.52260032	1.20899714	0.9121201	−4920.5701	2.63567879	−1.0187469
8	−0.2767429	0.37561063	0.11561097	0.7427615	0.52696799	1.23067384	0.91742165	−4920.5511	1.83311761	−2.1103436
9	−0.274204	0.37177939	0.11355883	0.74659763	0.52814335	1.2345524	0.91893546	−4920.5477	1.75839181	−2.177405
10	−0.2559812	0.34147926	0.09643312	0.77815581	0.53784221	1.26594249	0.93144448	−4920.5308	1.33811298	−2.6594121
11	−0.2496977	0.32954013	0.08928069	0.79125494	0.54186044	1.27868642	0.93662846	−4920.5291	0.79520373	−2.8166517
12	−0.2494237	0.32866495	0.08869445	0.79238161	0.54221561	1.27970094	0.93708759	−4920.5291	0.00166731	−2.6765739
13	−0.2494072	0.32862462	0.08864569	0.79242579	0.54224632	1.27974278	0.93712865	−4920.5291	0.00013552	−2.6768326
14	−0.2494082	0.32862428	0.08864047	0.79242619	0.54224944	1.27974312	0.93713293	−4920.5291	1.47×10^{-05}	−2.6766945
15	−0.2494083	0.32862424	0.08863992	0.79242624	0.54224977	1.27974315	0.93713338	−4920.5291	1.57×10^{-06}	−2.67668
16	−0.2494083	0.32862424	0.08863985	0.79242624	0.54224981	1.27974316	0.93713344	−4920.5291	1.89×10^{-09}	−2.6766783
17	−0.2494083	0.32862424	0.08863985	0.79242624	0.54224981	1.27974316	0.93713344	−4920.5291	2.09×10^{-10}	−2.6766783

Table A4. Convergence of the GTS parameter for SPY EFT return data.

Iterations	μ	β_+	β_-	α_+	α_-	λ_+	λ_-	$Log(ML)$	$\|\|\frac{dLog(ML)}{dV}\|\|$	MaxEigenValue
1	−0.0518661	0.1161846	0.2186548	1.04269292	0.52712574	1.52244991	0.91168779	−4894.2279	14.7801725	−6.5141947
2	−0.1102477	0.18491276	0.17478472	0.94756655	0.52844271	1.4399315	0.91415148	−4893.8278	29.8166141	−1.9290981
3	−0.2094204	0.29377592	0.0891446	0.84029122	0.56054563	1.34797271	0.96500981	−4893.3554	16.9940095	4.33892902
4	−0.1985564	0.29758208	0.13230013	0.83156167	0.53656079	1.33833078	0.93230856	−4893.4206	10.9048744	1.04588745
5	−0.078883	0.25939922	0.39611543	0.84865673	0.40365522	1.35595932	0.7410936	−4895.8806	241.028178	94.6293224
6	−0.0753571	0.26704857	0.33754158	0.84120908	0.45446164	1.3452823	0.80751063	−4894.3899	25.1995505	−2.805571
7	−0.196642	0.31624372	0.20068543	0.80509106	0.50322368	1.30837612	0.88967028	−4893.888	140.257551	34.770691
8	−0.1898283	0.3045047	0.15900291	0.81380451	0.52672075	1.31341912	0.91775259	−4893.4694	6.29433991	−4.6080872
9	−0.2275214	0.32940996	0.10770535	0.79340215	0.55020025	1.29360449	0.95260474	−4893.3049	7.34361008	−8.1891832
10	−0.2726283	0.34972465	0.01601222	0.78061523	0.59736004	1.28153304	1.01757433	−4893.2192	14.0784211	−3.6408772
11	−0.2499816	0.32645286	0.01851524	0.80217301	0.60018154	1.30243672	1.01792703	−4893.2125	6.27794755	−4.6455546
12	−0.2575953	0.33832596	0.02643321	0.79008383	0.59450637	1.29085215	1.01101001	−4893.208	1.23298227	−6.8035318
13	−0.2607071	0.34052644	0.02075252	0.78817075	0.59805376	1.28895161	1.01555438	−4893.2076	0.07363298	−6.71708
14	−0.2606368	0.34088815	0.02227012	0.78774693	0.59707082	1.28854532	1.01430383	−4893.2076	0.00156771	−6.6908109
15	−0.2606432	0.34087911	0.02220633	0.78775813	0.59711397	1.28855593	1.01435731	−4893.2076	0.00010164	−6.6915902
16	−0.2606426	0.34087985	0.02221188	0.78775721	0.5971103	1.28855506	1.01435268	−4893.2076	8.45×10^{-06}	−6.6915177
17	−0.2606426	0.34087979	0.02221141	0.78775729	0.59711061	1.28855513	1.01435308	−4893.2076	7.21×10^{-07}	−6.6915239

Appendix A.2. Pearson Statistic Inputs

Table A5. Observed versus expected statistics under GTS distribution.

k	Bitcoin			Ethereum			sp500			SPY EFT		
	x(k)	$n*\Pi_k$	n(k)	x(k)	$n*\Pi_k$	n(k)	x(k)	$n*\Pi_k$	n(k)	x(k)	$n*\Pi_k$	n(k)
1	−18.988	7.512	8	−20.861	7.531	6	−4.341	10.264	12	−4.405	8.327	11
2	−17.080	4.144	7	−18.321	5.583	6	−3.935	5.456	5	−4.007	4.562	3
3	−15.172	6.603	7	−15.781	10.018	15	−3.529	8.442	7	−3.608	7.116	7
4	−13.264	10.678	9	−13.241	18.331	20	−3.123	13.138	15	−3.210	11.144	12
5	−11.356	17.586	13	−10.700	34.424	29	−2.717	20.588	20	−2.811	17.538	18
6	−9.448	29.661	32	−8.160	66.980	68	−2.311	32.543	30	−2.413	27.775	27
7	−7.540	51.657	47	−5.620	137.268	134	−1.905	52.023	48	−2.015	44.350	41
8	−5.632	94.188	107	−3.080	305.591	305	−1.499	84.479	89	−1.616	71.617	73
9	−3.724	184.486	168	−0.540	769.951	769	−1.093	140.406	147	−1.218	117.552	122
10	−1.816	411.503	419	2.000	965.210	966	−0.687	242.564	244	−0.819	198.101	186
11	0.092	1195.725	1186	4.540	458.955	466	−0.281	455.971	456	−0.421	351.660	348
12	2.000	1150.470	1159	7.080	219.873	222	0.126	896.809	896	−0.023	725.476	730
13	3.908	473.177	469	9.620	111.760	101	0.532	749.106	733	0.376	867.735	867
14	5.816	217.783	227	12.160	59.253	60	0.938	430.841	426	0.774	541.022	522
15	7.724	107.387	102	14.700	32.379	32	1.344	234.692	260	1.173	300.464	325
16	9.632	55.272	51	17.241	18.099	21	1.750	126.820	137	1.571	163.491	189
17	11.540	29.294	39	19.781	10.296	12	2.156	68.688	57	1.969	88.862	82
18	13.448	15.861	14	22.321	5.939	5	2.562	37.387	33	2.368	48.491	36
19	15.356	8.729	9	24.861	3.465	5	2.968	20.460	15	2.766	26.600	23
20	17.264	4.866	4		5.091	4	3.374	11.256	12	3.165	14.669	13
21		6.419	6					14.067	14		18.450	20

Appendix B

Appendix B.1. Bitcoin BTC: Kobol Distribution ($\beta = \beta_- = \beta_+$)

$$V(dx) = \left(\alpha_+ \frac{e^{-\lambda_+ x}}{x^{1+\beta}} \mathbf{1}_{x>0} + \alpha_- \frac{e^{-\lambda_- |x|}}{|x|^{1+\beta}} \mathbf{1}_{x<0} \right) dx \tag{A1}$$

$$\Psi(\xi) = \mu\xi i + \Gamma(-\beta)\left[\alpha_+((\lambda_+ - i\xi)^\beta - \lambda_+{}^\beta) + \alpha_-((\lambda_- + i\xi)^\beta - \lambda_-{}^\beta)\right] \quad \text{(A2)}$$

Table A6. Kobol maximum-likelihood estimation for Bitcoin return data.

Model	Parameter	Estimate	Std Err	z	$Pr(Z > |z|)$	[95% Conf.Interval]	
	μ	-0.292833	(0.126)	-2.32	2.1×10^{-02}	-0.541	-0.045
	β	0.367074	(0.086)	4.27	1.9×10^{-05}	0.199	0.535
	α_+	0.755914	(0.047)	16.02	4.7×10^{-58}	0.663	0.848
	α_-	0.535121	(0.034)	15.68	9.6×10^{-56}	0.468	0.602
GTS	λ_+	0.235266	(0.027)	8.87	3.6×10^{-19}	0.183	0.287
	λ_-	0.181602	(0.023)	7.94	9.8×10^{-16}	0.137	0.226
	Log(ML)	$-10,607$					
	AIC	$21,226$					
	BIK	$21,264$					

Table A7. Convergence of the Kobol parameter for Bitcoin return data.

Iterations	μ	β	α_+	α_-	λ_+	λ_-	$Log(ML)$	$\|\|\frac{dLog(ML)}{dV}\|\|$	*MaxEigenValue*
1	-0.1215714	0.3155483	0.7477142	0.5445652	0.2465296	0.17477186	-10614.93879	450.0556974	25.6678081
2	-0.255172	0.36516958	0.73215119	0.53194253	0.22955186	0.17909072	-10607.01058	51.74982347	-46.893383
3	-0.2912276	0.37096854	0.75410108	0.53529439	0.23387716	0.18070591	-10606.81236	1.484798964	-53.728563
4	-0.2922408	0.36574333	0.7559582	0.53508403	0.23560819	0.18189913	-10606.81041	0.258928464	-53.391237
5	-0.2928641	0.36714311	0.75591147	0.53512239	0.23524801	0.18158588	-10606.81025	0.01286122	-53.406734
6	-0.292837	0.36708319	0.75591382	0.53512107	0.23526357	0.18159941	-10606.81025	0.00174219	-53.40643
7	-0.2928328	0.36707373	0.75591419	0.53512086	0.23526603	0.18160154	-10606.81025	1.18×10^{-05}	-53.406384
8	-0.2928328	0.36707379	0.75591419	0.53512086	0.23526602	0.18160153	-10606.81025	1.60×10^{-06}	-53.406384
9	-0.2928328	0.36707379	0.75591419	0.53512086	0.23526601	0.18160153	-10606.81025	2.18×10^{-07}	-53.406384
10	-0.2928328	0.36707379	0.75591419	0.53512086	0.23526601	0.18160153	-10606.81025	1.09×10^{-08}	-53.406384

Appendix B.2. Ethereum: Carr–Geman–Madan–Yor (CGMY) Distributions

$$V(dx) = \left(\alpha\frac{e^{-\lambda_+ x}}{x^{1+\beta}}\mathbf{1}_{x>0} + \alpha\frac{e^{-\lambda_- |x|}}{|x|^{1+\beta}}\mathbf{1}_{x<0}\right)dx \quad \text{(A3)}$$

$$\Psi(\xi) = \mu\xi i + \alpha\Gamma(-\beta)\left[((\lambda_- - i\xi)^\beta - \lambda_+{}^\beta) + ((\lambda_- + i\xi)^\beta - \lambda_-{}^\beta)\right] \quad \text{(A4)}$$

Table A8. CGMY maximum-likelihood estimation for Ethereum return data.

Model	Parameter	Estimate	Std Err	z	$Pr(Z > |z|)$	[95% Conf.Interval]	
	μ	-0.147089	(0.079)	-1.86	6.3×10^{-02}	-0.302	-0.008
	β	0.398418	(0.127)	3.12	1.8×10^{-03}	0.148	0.649
	α	0.887161	(0.058)	15.22	1.2×10^{-52}	0.773	1.001
GTS	λ_+	0.155369	(0.023)	6.56	5.2×10^{-11}	0.109	0.202
	λ_-	0.185991	(0.025)	7.29	2.9×10^{-13}	0.136	0.236
	Log(ML)	-9554					
	AIC	$19,118$					
	BIK	$19,149$					

Table A9. Convergence of the CGMY parameter for Ethereum return data.

Iterations	μ	β	α	λ_+	λ_-	$Log(ML)$	$||\frac{dLog(ML)}{dV}||$	$MaxEigenValue$
1	-0.48538	0.39043616	0.95824875	0.16671208	0.17078617	-9596.2658	1653.57149	-140.21456
2	-0.0545131	0.40148247	0.88205674	0.15875317	0.18060554	-9554.6834	80.0921993	-19.637869
3	-0.1479632	0.39049434	0.88271998	0.15631408	0.18704084	-9553.9065	3.84112398	-44.76325
4	-0.1465893	0.40345482	0.88868927	0.15450383	0.18507239	-9553.9036	0.4436942	-55.029934
5	-0.1472464	0.39683622	0.88667597	0.15564017	0.18628001	-9553.9026	0.14094418	-51.274906
6	-0.1470247	0.39907668	0.88736036	0.15525581	0.18587143	-9553.9025	0.05563606	-52.523819
7	-0.1471148	0.39816841	0.88708569	0.15541227	0.18603772	-9553.9025	0.02143506	-52.017698
8	-0.1470898	0.39842098	0.88716234	0.15536883	0.18599155	-9553.9025	0.00019543	-52.158334
9	-0.14709	0.39841855	0.88716161	0.15536924	0.18599199	-9553.9025	1.16×10^{-05}	-52.156981
10	-0.14709	0.39841867	0.88716164	0.15536922	0.18599197	-9553.9025	1.78×10^{-06}	-52.157046
11	-0.14709	0.39841869	0.88716165	0.15536922	0.18599197	-9553.9025	2.71×10^{-07}	-52.157055
12	-0.14709	0.39841869	0.88716165	0.15536922	0.18599197	-9553.9025	4.14×10^{-08}	-52.157057

Appendix C

Appendix C.1. S&P 500 Index: Bilateral Gamma (BG) Distribution ($\beta_- = \beta_+ = 0$)

$$V(dx) = \left(\alpha_+ \frac{e^{-\lambda_+ x}}{x} \mathbf{1}_{x>0} + \alpha_- \frac{e^{-\lambda_- |x|}}{|x|} \mathbf{1}_{x<0} \right) dx \tag{A5}$$

$$\Psi(\xi) = \mu \xi i - \alpha_+ \log\left(1 - \frac{1}{\lambda_+} i\xi \right) - \alpha_- \log\left(1 + \frac{1}{\lambda_-} i\xi \right) \tag{A6}$$

Table A10. BG maximum-likelihood estimation for S&P 500 return data.

Model	Parameter	Estimate	Std Err	z	$Pr(Z > |z|)$		[95% Conf.Interval]
	μ	-0.031467	(0.010)	-3.07	2.1×10^{-03}	-0.052	-0.011
	α_+	1.092741	(0.058)	18.98	2.6×10^{-80}	0.980	1.206
	α_-	0.701784	(0.042)	16.80	2.3×10^{-63}	0.620	0.784
GTS	λ_+	1.539690	(0.064)	22.82	3.1×10^{-115}	1.407	1.672
	λ_-	1.110737	(0.050)	22.07	6.6×10^{-108}	1.012	1.209
	Log(ML)	-4925					
	AIC	9859					
	BIK	9890					

Table A11. Convergence of the BG parameter for S&P 500 return data.

Iterations	μ	α_+	α_-	λ_+	λ_-	$Log(ML)$	$||\frac{dLog(ML)}{dV}||$	$MaxEigenValue$
1	0	0.79242624	0.54224981	1.27974316	0.93713344	-4951.1439	1138.53458	-265.251
2	-0.0038447	0.93153413	0.64461254	1.41132138	1.05504868	-4931.7583	549.025405	-171.22804
3	-0.0103214	1.03062846	0.70198868	1.49426667	1.10555794	-4926.8412	286.215345	-126.18156
4	-0.0186317	1.07922912	0.71421996	1.53377391	1.11392285	-4925.393	135.694287	-113.12071
5	-0.0279475	1.09450205	0.70493092	1.54418172	1.10795103	-4924.7065	38.0551545	-116.58686
6	-0.0313951	1.09325663	0.70162581	1.54038766	1.10996346	-4924.621	1.54417452	-120.06271
7	-0.0314671	1.09274119	0.70178365	1.53969027	1.11073682	-4924.6205	0.02788236	-120.35435
8	-0.0314664	1.09276971	0.70182788	1.53971127	1.11079928	-4924.6205	0.00198685	-120.34482
9	-0.0314663	1.09277213	0.70183158	1.5397131	1.11080431	-4924.6205	0.00016039	-120.34394
10	-0.0314663	1.09277232	0.70183188	1.53971325	1.11080472	-4924.6205	1.29×10^{-05}	-120.34387
11	-0.0314662	1.09277234	0.7018319	1.53971326	1.11080475	-4924.6205	1.04×10^{-06}	-120.34387
12	-0.0314662	1.09277234	0.7018319	1.53971326	1.11080476	-4924.6205	8.43×10^{-08}	-120.34387
13	-0.0314662	1.09277234	0.7018319	1.53971326	1.11080476	-4924.6205	6.80×10^{-09}	-120.34387
14	-0.0314662	1.09277234	0.7018319	1.53971326	1.11080476	-4924.6205	5.63×10^{-10}	-120.34387
15	-0.0314662	1.09277234	0.7018319	1.53971326	1.11080476	-4924.6205	5.73×10^{-11}	-120.34387

Appendix C.2. SPY ETF: Bilateral Gamma (BG) Distribution ($\beta_- = \beta_+ = 0$)

$$V(dx) = \left(\alpha_+ \frac{e^{-\lambda_+ x}}{x} \mathbf{1}_{x>0} + \alpha_- \frac{e^{-\lambda_- |x|}}{|x|} \mathbf{1}_{x<0} \right) dx \tag{A7}$$

$$\Psi(\xi) = \mu\xi i - \alpha_+ \log\left(1 - \frac{1}{\lambda_+}i\xi\right) - \alpha_- \log\left(1 + \frac{1}{\lambda_-}i\xi\right) \tag{A8}$$

Table A12. BG maximum-likelihood estimation for SPY EFT return data.

Model	Parameter	Estimate	Std Err	z	$Pr(Z > \lvert z \rvert)$	[95% Conf.Interval]	
	μ	0.015048	(0.012)	1.28	2.0×10^{-01}	-0.008	0.038
	α_+	1.068239	(0.067)	16.02	8.6×10^{-58}	0.938	1.199
GTS	α_-	0.764449	(0.044)	17.33	3.0×10^{-67}	0.678	0.851
	λ_+	1.525718	(0.073)	20.98	1.1×10^{-97}	1.383	1.668
	λ_-	1.156439	(0.052)	22.15	1.1×10^{-108}	1.054	1.259
Log(ML)		-4899					
AIC		9807					
BIK		9838					

Table A13. Convergence of the BG parameter for SPY EFT return data

Iterations	μ	α_+	α_-	λ_+	λ_-	$Log(ML)$	$\lVert \frac{dLog(ML)}{dV} \rVert$	*MaxEigenValue*
1	0	0.78775729	0.59711061	1.28855513	1.01435308	-4918.7331	406.35365	-252.28104
2	0.02867773	0.97562263	0.67572827	1.46822249	0.97596762	-4908.5992	226.190986	-116.11753
3	0.02727407	1.05127517	0.78618306	1.51846501	1.17883595	-4899.0798	45.2281041	-96.788275
4	0.00834089	1.07692251	0.75226045	1.5348003	1.14577232	-4898.9955	131.637516	-107.77617
5	0.01126962	1.07242358	0.7568497	1.53011913	1.1494212	-4898.751	48.0005418	-103.03258
6	0.01386478	1.06933921	0.76167668	1.52688303	1.15363987	-4898.6763	11.3246873	-100.1483
7	0.01492745	1.06823047	0.76397541	1.52573245	1.15588409	-4898.6693	0.98802026	-99.171136
8	0.01504464	1.06821119	0.76439389	1.52569528	1.15636567	-4898.6693	0.02529683	-99.040575
9	0.01504742	1.06823539	0.76444163	1.5257152	1.15642976	-4898.6693	0.00300624	-99.030178
10	0.01504762	1.06823881	0.76444764	1.52571803	1.15643796	-4898.6693	0.00038489	-99.028926
11	0.01504764	1.06823925	0.76444841	1.52571839	1.15643901	-4898.6693	4.92×10^{-05}	-99.028765
12	0.01504765	1.06823931	0.76444851	1.52571844	1.15643915	-4898.6693	6.30×10^{-06}	-99.028745
13	0.01504765	1.06823932	0.76444852	1.52571845	1.15643917	-4898.6693	1.32×10^{-08}	-99.028742
14	0.01504765	1.06823932	0.76444852	1.52571845	1.15643917	-4898.6693	1.69×10^{-09}	-99.028742
15	0.01504765	1.06823932	0.76444852	1.52571845	1.15643917	-4898.6693	2.23×10^{-10}	-99.028742

References

An, Kolmogorov. 1933. Sulla determinazione empirica di una legge didistribuzione. *Giorn Dell'inst Ital Degli Att* 4: 89–91.

Anderson, Theodore W. 2008. Anderson–darling test. In *The Concise Encyclopedia of Statistics*. Edited by Yadolah Dodge. New York: Springer, pp. 12–14. [CrossRef]

Anderson, Theodore W. 2011. Anderson–darling tests of goodness-of-fit. In *International Encyclopedia of Statistical Science*. Edited by Miodrag Lovric. Berlin and Heidelberg: Springer, pp. 52–54. [CrossRef]

Barndorff-Nielsen, Ole E., and Neil Shephard. 2002. Financial Volatility, Lévy Processes and Power Variation. Available online: https://www.olsendata.com/data_products/client_papers/papers/200206-NielsenShephard-FinVolLevyProcessPowerVar.pdf (accessed on 27 August 2024).

Bianchi, Michele Leonardo, Stoyan V. Stoyanov, Gian Luca Tassinari, Frank J. Fabozzi, and Sergio M. Focardi. 2019. *Handbook of Heavy-Tailed Distributions in Asset Management and Risk Management*. Volume 7 of Financial Economics. Singapore: World Scientific Publishing. [CrossRef]

Borak, Szymon, Wolfgang Härdle, and Rafał Weron. 2005. Stable distributions. In *Statistical Tools for Finance and Insurance*. Berlin and Heidelberg: Springer, pp. 21–44. [CrossRef]

Boyarchenko, Svetlana, and Sergei Z. Levendorskii. 2002. *Non-Gaussian Merton-Black-Scholes Theory*. Singapore: World Scientific Publishing, vol. 9.

Carr, Peter, Hélyette Geman, Dilip B. Madan, and Marc Yor. 2003. Stochastic volatility for lévy processes. *Mathematical Finance* 13: 345–82. [CrossRef]

Casella, George, and Roger Berger. 2024. *Statistical Inference*. Chapman & Hall/CRC Texts in Statistical Science. New York: CRC Press.

Cherubini, Umberto, Giovanni Della Lunga, Sabrina Mulinacci, and Pietro Rossi. 2010. *Fourier Transform Methods in Finance*. Hoboken: John Wiley & Sons.

Dimitrova, Dimitrina S., Vladimir K. Kaishev, and Senren Tan. 2020. Computing the kolmogorov-smirnov distribution when the underlying cdf is purely discrete, mixed, or continuous. *Journal of Statistical Software* 95: 1–42. [CrossRef]

Eberlein, Ernst. 2014. Fourier-based valuation methods in mathematical finance. In *Quantitative Energy Finance: Modeling, Pricing, and Hedging in Energy and Commodity Markets*. Edited by Fred Espen Benth, Valery A. Kholodnyi and Peter Laurence. New York: Springer, pp. 85–114. [CrossRef]

Eberlein, Ernst, Kathrin Glau, and Antonis Papapantoleon. 2010. Analysis of fourier transform valuation formulas and applications. *Applied Mathematical Finance* 17: 211–40. [CrossRef]

Fallahgoul, Hasan, and Gregoire Loeper. 2021. Modelling tail risk with tempered stable distributions: An overview. *Annals of Operations Research* 299: 1253–80.

Fallahgoul, Hasan A., David Veredas, and Frank J. Fabozzi. 2019. Quantile-based inference for tempered stable distributions. *Computational Economics* 53: 51–83. [CrossRef]

Feller, William. 1971. *An Introduction to Probability Theory and its Applications*, 2nd ed. New York: John Wiley & Sons, vol. 2.

Giudici, Paolo, Geof H. Givens, and Bani K. Mallick. 2013. *Wiley Series in Computational Statistics*. Hoboken: Wiley Online Library.

Grabchak, Michael, and Gennady Samorodnitsky. 2010. Do financial returns have finite or infinite variance? A paradox and an explanation. *Quantitative Finance* 10: 883–93. [CrossRef]

Hall, W. Jackson, and David Oakes. 2023. *A Course in the Large Sample Theory of Statistical Inference*. New York: CRC Press.

Ken-Iti, Sato. 2001. Basic results on lévy processes. In *Lévy Processes: Theory and Applications*. Edited by Ole E. Barndorff-Nielsen, Thomas Mikosch and Sidney I. Resnick. New York: Springer Science & Business Media, pp. 1–37. [CrossRef]

Kendall, Maurice George. 1945. *The Advanced Theory of Statistics*, 2nd ed. London: Charles Griffin & Co. Ltd., vol. 1.

Kim, Young Shin, Svetlozar T. Rachev, Michele Leonardo Bianchi, and Frank J. Fabozzi. 2009. A new tempered stable distribution and its application to finance. In *Risk Assessment: Decisions in Banking and Finance*. Edited by Georg Bol, Svetlozar T. Rachev and Reinhold Würth. Berlin and Heidelberg: Springer, pp. 77–109.

Krysicki, W., J Bartos, W. Dyczka, K. Królikowska, and M. Wasilewski. 1999. *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*. Cz. II. Statystyka matematyczna. Warszawa: PWN.

Küchler, Uwe, and Stefan Tappe. 2008. Bilateral gamma distributions and processes in financial mathematics. *Stochastic Processes and their Applications* 118: 261–83. [CrossRef]

Küchler, Uwe, and Stefan Tappe. 2013. Tempered stable distributions and processes. *Stochastic Processes and Their Applications* 123: 4256–93. [CrossRef]

Lehmann, Erich Leo. 1999. *Elements of Large-Sample Theory*. New York: Springer.

Lewis, Peter A. W. 1961. Distribution of the anderson-darling statistic. *The Annals of Mathematical Statistics* 32: 1118–24. [CrossRef]

Madan, Dilip B., Peter P. Carr, and Eric C. Chang. 1998. The variance gamma process and option pricing. *Review of Finance* 2: 79–105. [CrossRef]

Marsaglia, George, and John Marsaglia. 2004. Evaluating the anderson-darling distribution. *Journal of Statistical Software* 9: 1–5. [CrossRef]

Marsaglia, George, Wai Wan Tsang, and Jingbo Wang. 2003. Evaluating kolmogorov's distribution. *Journal of Statistical Software* 8: 1–4. [CrossRef]

Massey, Frank J., Jr. 1951. The kolmogorov-smirnov test for goodness of fit. *Journal of the American Statistical Association* 46: 68–78. [CrossRef]

Massing, Till. 2024. Parametric estimation of tempered stable laws. *ALEA Latin American Journal of Probability and Mathematical Statistics* 21: 1567–600. [CrossRef]

Mensah, Eric Teye, Alexander Boateng, Nana Kena Frempong, and Daniel Maposa. 2023. Simulating stock prices using geometric brownian motion model under normal and convoluted distributional assumptions. *Scientific African* 19: e01556. [CrossRef]

Nakamoto, Satoshi. 2008. Bitcoin: A Peer-to-Peer Electronic Cash System. Decentralized Business Review. Available online: https://www.ussc.gov/sites/default/files/pdf/training/annual-national-training-seminar/2018/Emerging_Tech_Bitcoin_Crypto.pdf (accessed on 10 May 2024).

Nolan, John P. 2020. *Modeling with Stable Distributions*. Cham: Springer International Publishing, chp. 2, pp. 25–52. [CrossRef]

Nzokem, Aubain H. 2021a. Fitting infinitely divisible distribution: Case of gamma-variance model. *arXiv*, arXiv:2104.07580.

Nzokem, Aubain H. 2021b. Gamma variance model: Fractional fourier transform (FRFT). *Journal of Physics: Conference Series* 2090: 012094. [CrossRef]

Nzokem, Aubain H. 2021c. Numerical solution of a gamma—Integral equation using a higher order composite newton-cotes formulas. *Journal of Physics: Conference Series* 2084: 012019. [CrossRef]

Nzokem, Aubain H. 2023a. Enhanced the fast fractional fourier transform (frft) scheme using the closed newton-cotes rules. *arXiv*, arXiv:2311.16379.

Nzokem, Aubain H. 2023b. European option pricing under generalized tempered stable process: Empirical analysis. *arXiv*, arXiv:2304.06060.

Nzokem, Aubain H. 2023c. Pricing european options under stochastic volatility models: Case of five-parameter variance-gamma process. *Journal of Risk and Financial Management* 16: 55. [CrossRef]

Nzokem, Aubain H. 2024. Self-decomposable laws associated with general tempered stable (gts) distribution and their simulation applications. *arXiv*, arXiv:2405.16614.

Nzokem, Aubain H., and Daniel Maposa. 2024. Bitcoin versus s&p 500 index: Return and risk analysis. *Mathematical and Computational Applications* 29: 44. [CrossRef]

Nzokem, Aubain H., and V. T. Montshiwa. 2022. Fitting generalized tempered stable distribution: Fractional fourier transform (frft) approach. *arXiv*, arXiv:2205.00586.

Nzokem, Aubain H., and V. T. Montshiwa. 2023. The ornstein–uhlenbeck process and variance gamma process: Parameter estimation and simulations. *Thai Journal of Mathematics*, 160–68. Available online: https://thaijmath2.in.cmu.ac.th/index.php/thaijmath/article/view/1477 (accessed on 17 June 2024).

Olive, David J. 2014. *Statistical Theory and Inference*. New York: Springer.

Poloskov, Igor E. 2021. Relations between cumulants and central moments and their applications. *Journal of Physics: Conference Series* 1794: 012004. [CrossRef]

Rachev, Svetlozar T., Young Shin Kim, Michele L. Bianchi, and Frank J. Fabozzi. 2011. Stable and tempered stable distributions. In *Financial Models with Lévy Processes and Volatility Clustering*. Edited by Svetlozar T. Rachev, Young Shin Kim, Michele L. Bianchi and Frank J. Fabozzi. Volume 187 of The Frank J. Fabozzi Series. Hoboken: John Wiley & Sons, Ltd., chp. 3, pp. 57–85. [CrossRef]

Rota, Gian-Carlo, and Jianhong Shen. 2000. On the combinatorics of cumulants. *Journal of Combinatorial Theory, Series A* 91: 283–304. [CrossRef]

Sato, Ken-Iti. 1999. *Lévy Processes and Infinitely Divisible Distributions*. Cambridge: Cambridge University Press.

Schoutens, Wim. 2003. *Lévy Processes in Finance: Pricing Financial Derivatives*. West Sussex: John Wiley & Sons.

Smith, Peter J. 1995. A recursive formulation of the old problem of obtaining moments from cumulants and vice versa. *The American Statistician* 49: 217–18. [CrossRef]

Stephens, Michael A. 1974. Edf statistics for goodness of fit and some comparisons. *Journal of the American Statistical Association* 69: 730–37. [CrossRef]

Tsallis, Constantino. 1997. Lévy distributions. *Physics World* 10: 42. [CrossRef]

Van den Bos, Adriaan. 2007. *Precision and Accuracy*. Hoboken: John Wiley & Sons, Ltd., chp. 4, pp. 45–97. [CrossRef]

Vuong, Quang H. 1989. Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society* 57: 307–33. [CrossRef]

Article

A Double Optimum New Solution Method Based on EVA and Knapsack

Theofanis Petropoulos [1,*], Paris Patsis [1], Konstantinos Liapis [1] and Evangelos Chytis [2]

[1] Department of Economic & Regional Development, Panteion University, Syngrou Av. 136 176-71,
    48100 Athens, Greece; p_patsis@panteion.gr (P.P.); konstantinos.liapis@panteion.gr (K.L.)
[2] Department of Accounting and Finance, University of Ioannina, Campus Preveza, 48100 Preveza, Greece;
    ehytis@uoi.gr
* Correspondence: f.petropoulos@panteion.gr

Abstract: Optimizing resource allocation often requires a trade-off between multiple objectives. Since projects must be fully implemented or not at all, this issue is modeled as an integer programming problem, precisely a knapsack-type problem, where decision variables are binary (1 or 0). Projects may be complementary/supplementary and competitive/conflicting, meaning some are prerequisites for others, while some prevent others from being implemented. In this paper, a two-objective optimization model in the energy sector is developed, and the Non-dominated Sorting Genetic Algorithm III (NSGA III) is adopted to solve it because the NSGA-III method is capable of handling problems with non-linear characteristics as well as having multiple objectives. The objective is to maximize the overall portfolio's EVA (Economic Value Added). EVA is different from traditional performance measures and is more appropriate because it incorporates the objectives of all stakeholders in a business. Furthermore, because each project generates different kilowatts, maximizing the total production of the portfolio is appropriate. Data from the Greek energy market show optimal solutions on the Pareto efficiency front ranging from (14.7%, 38,000) to (11.91%, 40,750). This paper offers a transparent resource allocation process for similar issues in other sectors.

Keywords: portfolio management; mathematic programming; finance; decision-making; multi-objective optimization; mathematics of quantitative finance; mathematical models in optimal portfolio theory

1. Introduction

The rational allocation of resources is vital in both public and private sectors. A significant subset of the above is the problem of where to implement any project; the whole project must be completed rather than part of it. In this paper, we focus on the knapsack problem type according to Dantzig (1957). Particularly in Greece, due to the decade-long economic crisis and the coronavirus crisis, it is appropriate to allocate the limited budget as rationally and optimally as possible. The knapsack problem is a subclass of integer programming, achieving the optimal solution. Much scientific research has dealt with this issue, but it has yet to receive the appropriate attention in Greece.

We focus on the energy sector in Greece, which is developing rapidly and is attracting increasing interest in investment. In recent years, due to climate change, there has been an urgent need for conversion to alternative energy sources in line with the European policy objective of substantially increasing the share of renewable energy sources in electricity generation and more excellent absorption of funds from the European Recovery Fund.

In Greece, this interest is particularly evident in the case of the exploitation of wind energy due to the favorable legislative framework and the significant wind potential that exists in several Greek regions. Private investors are strongly motivated towards the exploitation of wind energy, mainly because the associated investment costs are not

prohibitive, and a somewhat favorable legislative framework ensures satisfactory rates of return (Mavrotas et al. 2003). Consequently, composing a portfolio with an optimal allocation of resources and an excellent strategic alignment based on future returns is crucial for companies to avoid fines and comply with legislation, as well as to develop scientific and technological capabilities that can help them drive innovation and gain competitive advantage (Bin et al. 2015).

The energy sector satisfies the knapsack problem's conditions and attracts attention to these problems. The challenges of traditional methods, including many runs, can be overcome using evolutionary strategies. In addition to guiding the search towards the Pareto optimal front, the goal of the multi-objective optimization technique is to maintain population diversity within a collection of non-dominated solutions. Among the most significant multi-objective optimization algorithms available today, the NSGA-III (Deb and Jain 2013; Jain and Deb 2013) is a potent strategy to overcome the shortcomings of NSGA-II, including its lack of uniform diversity and lateral diversity preserving operator. NSGA-III is selected among these algorithms because of its uniform diversity in obtaining the Pareto optimal front from a group of non-dominated solutions and its relatively greater capacity to handle many objectives.

Our approach takes into account the fact that some projects depend on others for implementation, such as an intermediate power station, while others cannot proceed without certain prerequisites being met. We categorize these dependencies into two matrices of complementary and conflicting investment projects and then use integer programming to identify the best investment option. Our approach differs from previous research as we utilize the EVA (Economic Value Added) index as an optimization goal. Typically, it is possible to select some investments with negative EVA if necessary to achieve maximum Eva for the portfolio as Sharma and Kumar (2010) found. When projects have different rates of return, cost and production must be considered during the maximization process. Each project's different production in megawatts results from different guaranteed price contracts (feed-in tariffs) or the exchange market, which varies over time.

Our model is further differentiated and idealized as an additional target of total portfolio energy production is added. This helps both to cover the investor in the event of a price decline and simultaneously meet each region's energy needs in the context of sustainable development. Our article first contributes to the debate on the use of EVA against revenue and the use of suitable data in a unique knapsack model for the energy sector of Greece that professionals and academics use. We attempt to use data from the Greek energy market that would be implemented in other countries' energy markets. The literature review and the main features of the energy sector selection are then presented, as well as the conclusions and discussion.

In the Section 2, we analyze the literature review relevant to the problem we consider. In the Section 3, we present the main research question. In the Section 4, we provide our data and the specific form of our model. In the Section 5, we provide the estimates and results. In the Section 6, we summarize all the above with our evaluation and attempt to find the rational answer to the main question.

2. Literature Review

Most research on knapsack problems has dealt with the version with a single constraint, e.g., Balas and Zemel (1980), Horowitz and Sahni (1974), Salkin and Kluyver (1975). Although the single-constraint version of this problem has received a lot of attention, the multi-constraint knapsack problem has not received proper attention in the Greek context (Jaszkiewicz 2004; Erlebach et al. 2002; Zitzler and Thiele 1999; Klamroth and Wiecek 2000). One of the first references to the multiple-constraint knapsack problem is by Lorie and Savage (1955) and Manne and Markowitz (1957) as a capital budgeting model. In Greece, a related study is provided by Florios et al. (2010).

The first accurate algorithms for the multi-constraint knapsack problem started in the 1960s. Gilmore and Gomory (1966) described a dynamic programming algorithm.

Later, Shih (1979) presented a branching and bounding algorithm for the multidimensional knapsack problem (MKP). In recent years, with the development of artificial intelligence, genetic algorithms can solve knapsack-type problems much faster with equally good results as the traditional methods (Kellerer et al. 2004; Khuri et al. 1994).

This paper divides the constraints into two categories: complementary and disjunctive. We use the negative disjunctive constraint as a particular case of a disjunctive constraint (Pferschy and Schauer 2017). Some papers dealing with the disjunctive constraint are Yamada et al. (2002), where a branch-and-bound algorithm is presented, and more recent exact computation algorithms are presented in Hifi and Otmani (2012) and Hifi et al. (2014). The 0-1 knapsack problem can be applied to various financial planning and portfolio management models. It determines which subset of assets provides the highest return under a given budget. The knapsack problem is a well-known combinatorial optimization problem with a wide range of business applications in capital budgeting (Bas 2011) and production planning (Camargo et al. 2012), among others.

Recent advancements in multi-objective optimization for the knapsack problem, particularly in energy portfolios, have introduced innovative methodologies. Notably, the Factored NSGA-II framework enhances exploration through overlapping subpopulations, effectively addressing multiple objectives such as profit maximization and weight minimization (Peerlinck and Sheppard 2022). Additionally, robust optimization algorithms have emerged, focusing on solutions resilient to variable changes, thereby broadening the solution space compared to traditional methods (Miyamoto and Fujiwara 2022).

The integration of quantum computing via the Quantum Approximate Optimization Algorithm (QAOA) presents a novel approach, demonstrating significant improvements in asset allocation within financial portfolios, which can be analogous to energy portfolio optimization (Huot et al. 2024), a great improvement on the previous topic provided by Awasthi et al. (2023). These methods collectively enhance the efficiency and effectiveness of solving complex multi-objective knapsack problems in energy contexts.

Faia et al. (2018) propose a portfolio optimization model using particle swarm optimization. This model addresses multi-objective challenges in energy markets by balancing risk and profit in energy portfolio decisions.

The hybrid algorithm combines k-nearest neighbor with quantum cuckoo search to enhance resource allocation solutions for the multidimensional knapsack problem, outperforming state-of-the-art algorithms in most instances (García and Maureira 2021).

The Harmony Search (HS) algorithm is a prominent heuristic for solving both single- and multi-objective 0-1 knapsack problems (KPs) and effectively solves single and multi-objective 0-1 knapsack problems (Adamuthe et al. 2020). Furthermore, innovations like the hybrid HS with distribution estimation have been proposed to avoid local optima, improving the algorithm's global search capabilities (Liu et al. 2022).

Dynamic Evolutionary Optimization (DEO) is increasingly applied to the multi-objective knapsack problem (MKP), addressing the complexities of real-world scenarios where objectives and constraints can change over time (de Queiroz Lafetá and Oliveira 2020).

Cacchiani et al. (2022) provide an excellent overview of solving techniques for the knapsack problem, especially the quadratic form. A novel optimization algorithm is designed to tackle the multidimensional knapsack problem (MKP), classified as NP-hard. This algorithm is an enhancement of the traditional moth search algorithm (MS), incorporating self-learning mechanisms to improve its efficiency and effectiveness (Feng and Wang 2022).

Lin et al. (2022) propose a single-preference-conditioned model to directly generate approximate Pareto solutions for any trade-off preference and design an efficient multi-objective reinforcement learning algorithm to train this model. Sur et al. (2022) adopted a deep reinforcement learning (DRL)-based approach; the experimental results indicate that the proposed method outperforms the random and greedy methods, particularly when the profits and weights of items have a non-linear relationship, such as quadratic forms. Nomer et al. (2020) introduce a heuristic solver based on neural networks and deep learning for the knapsack problem.

Olivas et al. (2021) conclude that incorporating fuzzy logic into hyperheuristics provides a robust mechanism for improving solutions to the knapsack problem. This approach enhances adaptability and results in better performance across various instances of the problem.

A new class of optimization problems called Mixed Pareto-Lexicographic Multi-objective Optimization Problems (MPL-MOPs) provides a suitable model for scenarios where some objectives have priority over others (Lai et al. 2020).

In the investment area, financial ratios such as the internal rate of return (IRR) and the Weighted Average Cost of Capital (WACC) play a vital role in the selection process of available investments (Kos et al. 2009). The financial measure IRR (internal rate of return) is widely applied in the financial and investment sector and, in many cases, is preferred to the NPV (Net Present Value), although NPV provides the highest accuracy. In this paper, we prefer IRR to reduce the computational cost to produce EVA, a critical investment indicator. EVA is a reliable way to evaluate whether some investments should be completed; it is possible to select some investments with a negative EVA if necessary to achieve maximum EVA for the portfolio (Sharma and Kumar 2010).

By integrating data science and advanced management strategies, organizations can optimize their portfolios to enhance profitability while minimizing environmental impacts. In recent years, scholars have been increasingly focused on EVA in relation to other methods of evaluating business performance. EVA has grown at a CAGR of 9.60%, compared to other business performance evaluation methods, which have only grown at a CAGR of 5.67%, as per publications in Scopus-listed journals (Tripathi et al. 2022).

In general, there are conflicting studies on EVA as a financial metric; (Faiteh and Mohammed 2023) state that EVA demonstrates superiority over traditional metrics for listed companies and can be adapted for unlisted firms using accounting beta, making it a versatile financial metric for value creation. According to Dobrowolski et al. (2022), on the other hand, EVA is not a universal financial metric; it fails to accurately reflect conditions in unstable markets, leading to potential mismanagement and limited shareholder value.

Chen et al. (2023b) found that EVA-related metrics naturally induce long-term, strategic and sustainable decision-making without limiting executives to overly focus on short-term profitability or develop a pseudo environment to illustrate EVA's managerial benefits and potential to cultivate sustainable growth.

A recent study relates to developing a mixed integer non-linear programming (MINLP) model that incorporates financial risk metrics into a robust closed-loop supply chain design, considering the unpredictability of final product demand to maximize EVA (Polo et al. 2019). Multi-objective optimization methods significantly impact optimization outcomes.

One well-known way to deal with this is the Weighted Sum Method. The objectives are often combined into a single objective, and conventional optimization procedures are employed to find the best solution. In this strategy, decision-makers must determine the weights. This method requires correct objective function weights and may not be suitable for nonconvex issues. Di Somma et al. (2018) suggested a stochastic integer programming model that converts the minimization of total energy costs and carbon dioxide emissions into a single objective using the weighted sum approach; then, they employed the branch-and-cut method to solve the researched issue.

In multi-objective evolutionary algorithms, a set of potential non-dominated solutions must be generated first, and the decision-maker selects from these solutions. There have been several reviews on the methods and application of multi-objective optimization (MOO). One of the most used methods is the Pareto method (Ehrgott 2005). The Pareto method is based on the principle of dominance where a dominated and a non-dominated solution emerges constantly from a continuously updated algorithm. The solution using the Pareto method generates a Pareto optimal front where it reflects the amount that must be sacrificed from the optimal solution of one objective to improve another objective.

In finance (Tapia and Coello 2007), to identify critical technical analysis patterns in financial time series, the niched-Pareto genetic algorithm (NPGA) is used. An alternative

method is the e-constraint method (Haimes et al. 1971) where all constraints are transformed into equality by adding or removing the appropriate constant (Mavrotas 2009; Mavrotas and Florios 2013). Mesquita-Cunha et al. (2023) developed a recent improvement of this algorithm for integer programming problems for knapsack-type resource allocation problems.

Genetic algorithms can also be applied to portfolio management (Metawa et al. 2016; Liu and Xiao 2021; Krink and Paterlini 2011). Giagkiozis and Fleming (2015) and Gunantara (2018) provide two comprehensive literature reviews of multi-objective optimization methods.

In the energy sector, optimizing portfolios is a critical concern, as it involves balancing various factors such as cost, risk, and renewable energy integration (Schönberger 2016). One approach to address this challenge is the use of the knapsack problem, a well-known optimization problem in the field of operations research (Ioannou et al. 2017)

The most common type of MOEA mentioned in the literature is dominance-based algorithms, specifically NSGA-II. Li and Qiu (2016) used an improved version of the NSGA-II to optimize a hydro-photovoltaic power system model, considering both power output smoothness and annual power generation. Noorollahi et al. (2017) created NSGA-II to solve a multi-objective problem. Indicator-based algorithms use indicator functions to assess population quality in MOEAs. Keshavarzzadeh and Ahmadi (2019) compared various strategies for optimizing a multi-objective model, including NSGA-II, generalized differential evaluation, indicator-based evolutionary algorithms, speed-constrained multi-objective algorithm, and strength Pareto algorithms.

The population quality in MOEAs is measured using indicator functions on indicator-based algorithms. Keshavarzzadeh and Ahmadi (2019) optimized a well-known multi-objective model using a variety of techniques, including NSGA-II, generalized differential evaluation, an indicator-based evolutionary algorithm, a speed-constrained multi-objective algorithm, and strong Pareto evolutionary algorithms. They then compared the outcomes of these algorithms.

Zhou et al. (2024) addresses the critical need for effective planning in integrated energy systems (IESs) to support energy revolution and sustainability goals. The authors propose a novel planning framework that integrates multi-objective optimization with fuzzy multi-criteria decision-making (MCDM). This framework is designed to tackle the complexities of IES planning by modeling it as a multi-objective optimization problem. The optimization problem is solved using a multi-objective state transition algorithm based on decomposition (MOSTA/D). This method generates a Pareto set that allows for trade-offs among conflicting objectives, which is a common challenge in multi-objective optimization.

The knapsack problem involves selecting a subset of items from a given set, where each item has a weight and a value, and the goal is to maximize the total value while staying within a weight constraint (Göteman et al. 2020).

Recent research has shown a significant increase in the application of optimization techniques, including the knapsack problem, to address energy-related challenges. For example, goal programming has been used to balance the trade-off between the cost per kWh of an electricity generation portfolio and the total risk for an investor-owned utility (Ioannou et al. 2017).

Chen et al. (2023a) reviewed multi-objective optimization in long-term energy systems, emphasizing the need to consider economic, environmental, and energy security objectives to address complex energy demands.

Recent advanced studies have used machine learning or reinforcement learning in portfolio management. Vaish et al. (2024) introduce the use of the Random Forest (RF) model, a popular machine learning algorithm, for optimizing microgrid configurations. The paper compares the RF model's performance with other methodologies, such as particle swarm optimization (PSO) and artificial neural networks (ANNs), indicating a growing trend in applying machine learning techniques to energy optimization problems.

A comprehensive overview of new machine learning techniques in the field of the energy sector is provided by Alazemi et al. (2024).

An alternative approach is the RNN for asset allocation management (Giacomazzi Dantas 2021; Milhomem and Dantas 2021; Tao et al. 2021). deLlano-Paz et al. (2017) provide a comprehensive literature survey on applying Modern Portfolio Theory (MPT) in energy planning.

The model proposed by Roques et al. (2010) seeks to identify the portfolio consisting of those European plants (inter-State12) that minimize the variability of the wind production output for a specific production level as an objective function. They propose an alternative definition for return, referring to it as the mean capacity factor for the different locations. Risk is defined as the hourly variability of production (Rombauts et al. 2011).

Overall, the literature review highlights the growing importance of the knapsack problem in optimizing energy portfolios, as it provides a valuable tool for balancing various objectives and constraints in the energy sector.

3. The Main Question from a Theoretical Point of View—Methodology

Our main question is, "Is there a rational and efficient way to allocate a given amount of budget to a set of projects among alternatives to achieve maximum profitability?". Our variables are Boolean for selected and unselected projects. In this way, the problem is treated as a budget problem (knapsack). A project is selected to be covered by the budget if the total profitability of the selected projects is maximum. We develop our model to solve this problem by answering the main question.

We consider the classical 0-1 knapsack problem, where a subset of n data projects must be allocated to a knapsack of capacity c. Each project has a profit r_j and a weight w_j, and the goal is to select a fraction of projects that maximize the total profit without exceeding the available budget. A binary variable $X_j = 0$ or 1 is defined because each project must be implemented as a whole or not at all. Several variations of the knapsack problem have been proposed to help organizations make sound project selection decisions in different sectors (Martello and Toth 1990).

We use an integer programming formulation (0-1) where there is a binary decision variable for each alternative, and these take values of 0 (alternative not selected) or 1 (alternative selected). In the case of a target, an integer programming (IP) formulation is usually preferred, particularly the knapsack formulation. We construct the model using linear algebra to achieve a practical way to implement several subproblems with different efficiency rates. Another subproblem arises among projects: One project is independent of another; one project is complementary to another; one project is disjoint-conflicting. Thus, the model is defined for the three categories of constraints as follows:

Matrix of Decision Variables X

$$X_{1 \times n} = [x_i] = [x_1, x_2, x_3, \ldots x_n] \; \forall i = 1, 2 \ldots n \; projects \tag{1}$$

where x_i is the binary variable 1 or 0 if the project i is implemented or not.

Matrix cost of projects C

$$C_{1 \times n} = [c_i] = [c_1, c_2, c_3, \ldots c_n] \; \forall i = 1, 2 \ldots n \; projects \tag{2}$$

where c_i is the specific cost for each project construction.

Matrix of Production of Projects

$$P_{1 \times n} = [p_i] = \left[p_1, p_2, p_3, \ldots p_n\right] \; \forall i = 1, 2 \ldots n \; projects \tag{3}$$

where p_i is the specific production for each project at Megawatts (MW).

Matrix of Return of Projects

$$R_{1 \times n} = [r_i] = [r_1, r_2, r_3, \ldots r_n] \; \forall i = 1, 2 \ldots n \; projects \tag{4}$$

where r_i is the specific return for each project as a percentage of the cost.

Matrix ROICC= *Return − OPPEX*

$$R'_{1 \times n} = [r'_i] = [r'_1, r'_2, r'_3, \ldots r'_n] \; \forall i = 1, 2 \ldots n \; projects \tag{5}$$

$$WACC = i_s \frac{E}{E + DTE * E} + (1 - \varphi) i_D \frac{DTE * E}{E + DTE * E} \tag{6}$$

Objective Function (O.F)

$$\textbf{Max O.F 1} \; \mathbf{X_{1 \times n}} \times (\mathbf{ROICC} - \mathbf{WACC}) \; \forall \mathbf{i = 1, 2 \ldots n \; projects} \tag{7}$$

$$\textbf{Max O.F 2} \; \mathbf{X_{1 \times n}} \times P_{1 \times n}{}^T \; \forall \mathbf{i = 1, 2 \ldots n \; projects} \tag{8}$$

Constraints:

Budget restriction

$$\textbf{KNAPSACK} \; \mathbf{X_{1 \times n}} \times \mathbf{C^T_{1 \times n}} = \mathbf{KNAPSACK_{1 \times 1}} \; \forall \mathbf{i = 1, 2 \ldots n \; projects} \tag{9}$$

Nowadays, particularly in Greece because of the prolonged recession after the economic crisis of 2008 and the current coronavirus pandemic, it is imperative to use the available budget in the best possible way. To achieve this, the key role is in the conversion of traditional investments to alternative and modern investments such as renewable sources of energy. In this area, another critical issue emerges, which is the complementarity and competitiveness of projects.

Complementarity is very important because it indicates that implementation of a project is impossible and unprofitable without another project being completed first. This second point reveals a precise need for further research on the subject. It is possible that a project is not only necessary for the realization of an investment but may also meet the needs of other investments that can be carried out with the remaining available budget. A feature example is an energy station that meets the energy needs of multiple energy investments, such as photovoltaics or windmills.

On the other hand, competitiveness means that if an investment is made, one or more available investments cannot be created. It is very important that the available budget is not wasted and that the selection of investments is carried out objectively so that the plurality of all types of investments in the portfolio is appropriate. In the energy sector, this is achieved by not allowing several investments of one type to be made at the same time and by not making more investments in an area where the availability of energy needs is met.

Conflict projects A

$A_{n \times k} = [a_{ij}]$ where i = projects and j = group of conflict projects

So $i = n$ and $j = k$, $a_{ij} = [1, 0]$ where 1 means projects are conflicting and zero in any other cases.

$$X_{1 \times n} \times A_{n \times k} \leq 1_{1 \times k} \; \forall i = 1, 2 \ldots n \; projects \tag{10}$$

Complementary projects S

$\mathbf{S_{n \times m}} = [\mathbf{b_{ij}}]$ where $\mathbf{i}$ = projects and $\mathbf{j}$ = group of projects in combination.

So $\mathbf{i = n}$ and $\mathbf{j = m}$, $\mathbf{b_{ij}} = [\mathbf{1, -1, 0}]$ where 1 and −1, respectively, mean projects are in combination, and 0 in any other cases.

$$X_{1 \times n} \times S_{n \times m} \leq 0_{1 \times m} \; \forall i = 1, 2 \ldots n \; projects \tag{11}$$

To reduce the dimension of the table: If there are z projects in combination with another project, respectively, the table of complementary projects is transformed.

$i = n$ and $j = m$, $b_{ij} = [1, -z, 0]$ where 1 is if the project is combined with another and -z indicates an energy station project to be created if any of the above is carried out, and 0 for all other cases.

$$X_{1 \times n} \times S_{n \times m} \leq 0_{1 \times m} \; \forall i = 1, 2 \ldots n \; projects \tag{12}$$

With the last two equations from the financial position, we ensure that the concentration risk in our portfolio is kept low, and from the ethical point of view, we avoid a biased position in a particular type of investment.

$$WACC = i_s \times \frac{S}{D+S} + i_D \times (1 - \varphi) \times \frac{D}{D+S} \tag{13}$$

According to the INTERNATIONAL VALUATION STANDARDS COUNCIL (IVSC) each investor requires a different rate of return (i_s) depending on the type of investment (photovoltaics or windmill). Similarly, banks vary the mortgage rate (i_D) according to the type of investment and, in addition, the creditworthiness of the investor. For this reason, in our model, each project has a different rate of return. Another important measure is the ratio of investors' capital $S/(D + S)$ and banks' $D/(D + S)$ achieved through the leverage, processing the essential capital to maximize profit, and φ represents tax rate.

The Weighted Average Cost of Capital (WACC) represents the average rate of return a company must pay to finance its assets, calculated as a weighted sum of the costs of equity and debt. It reflects the minimum return required by investors and is crucial for evaluating investment decisions, as it serves as the discount rate in capital budgeting and valuation models. WACC incorporates the firm's capital structure and market risk to provide a comprehensive financing cost.

Another question that arises about the profitability of this project is as follows: We first assume that profitability is equal to the NPV of each project. If the profitability of the rate of return is IRR, then we lose the upper part of our profit using the discount factor of the NPV of each project if we replace the rate of return with the Average Return on Invested Capital (AROIC) of each project. The ROIC substitution must be calculated using present values to turn our problem into a linear algebra problem where a matrix R or X or C must be diagonal, and we choose C because the investment amount of each project cost is not involved with complementarity and discrimination-conflict constraints. Based on Liapis (2010), we would also like to produce an EVA vector of each project (Stewart 2009). The reason for all this transformation is to introduce the EVA theory into the problem.

Economic Value Added (EVA) is a performance measurement tool used to assess a company's ability to generate value beyond the required return on its invested capital. It is calculated by subtracting the cost of capital from the firm's net operating profit after taxes (NOPAT). EVA emphasizes the importance of creating shareholder wealth by ensuring returns exceed the opportunity cost of capital employed.

Subtracting (12) by (5), the O.F gives the following:

$$
\begin{aligned}
&X_{1 \times n} \times C^D_{n \times n} \times R'^T_{n \times 1} \; - \\
&X_{1 \times n} \times C^D_{n \times n} \times WACC_{n \times 1} \Leftrightarrow \\
&X_{1 \times n} \times C^D_{n \times n} \times \left[R'^T - WACC \right]_{n \times 1} \Leftrightarrow \\
&X_{1 \times n} \times C^D_{n \times n} \times EVA_{n \times 1} = Profit_{1 \times 1} \; \forall i = 1, 2 \ldots n \; projects
\end{aligned}
\tag{14}
$$

With the abovementioned approach, an EVA vector is critical for knapsack fulfillment. All the above transformations give the value of a different approach to decision-making for the multiple-investment problem under knapsack constraints.

Many projects may have negative EVA in addition to our constraints so a new filter should be imposed as follows: Eva ≥ 0 or Ri $\geq$ WACC

$C \times X_i$ could be a stepwise product, but for EVA to be accurate for each project, WACC should be a vector with the same number of elements.

Another issue concerning WACC is the combination of different types of capital reflected through leverage, which can vary from project to project. Using the preferred capital, competitiveness and complementarity between projects, in addition to technical and economic reasons, depends on environmental and socio-regulatory factors.

We selected the NSGA III algorithm to address the given problem as the Pareto front (trade-off) between objectives is nonconvex and because the problem has complex non-linear characteristics.

1. NSGA-III is designed to handle more difficult computations and constraints, especially in cases with many decision variables. The algorithm manages to simultaneously optimize multiple dimensions without degrading the quality of solutions and offers better allocation in problems with complex solution sets.
2. In problems with many iterations (looping structures) and complex constraints, NSGA-III can handle complexity better because it searches in multidimensional space and uses reference points to find solutions in each part of the objective space. The constraints are considered through the non-dominated classification process and the distance strategy from the reference points.
3. NSGA-III is known for its ability to explore the multidimensional solution space more fully through the Niche Preservation process and the way it manages benchmarks. This allows it to find solutions that may not be easily identified by NSGA-II. In NSGA-II, solutions close to the Pareto front can be clustered in specific regions, leaving other regions empty. NSGA-III, however, uses a strategy that ensures that solutions are evenly distributed along the Pareto front.
4. A better advantage of NSGA-III is that it does not require additional parameters compared to NSGA-II.

The Non-dominated Sorting Genetic Algorithm III (NSGA-III) is a powerful multi-objective optimization technique designed to handle problems involving many objectives. It extends the concepts of NSGA-II by introducing a reference-point-based approach to maintain a well-distributed set of Pareto optimal solutions. Unlike NSGA-II, which uses crowding distance to promote diversity, NSGA-III uses predefined reference points to guide the search process toward a more uniform spread across the objective space. This makes it particularly useful in high-dimensional objective spaces, where maintaining diversity becomes challenging.

NSGA-III begins by generating an initial population, evaluating it based on the objectives, and sorting individuals into different Pareto fronts. It then associates each solution with the nearest reference point, preserving niche diversity by selecting one solution per reference point. The population evolves through evolutionary operators such as selection, crossover, and mutation, iteratively refining the Pareto front. The algorithm continues until a stopping criterion is met, such as a maximum number of generations or a convergence criterion where the solutions do not significantly improve. NSGA-III's ability to produce diverse solutions makes it suitable for complex, real-world applications with multiple conflicting objectives.

Step-by-step representation of NSGA III pseudo code is shown in Appendix A.

An overview of this paper is shown in Figure 1.
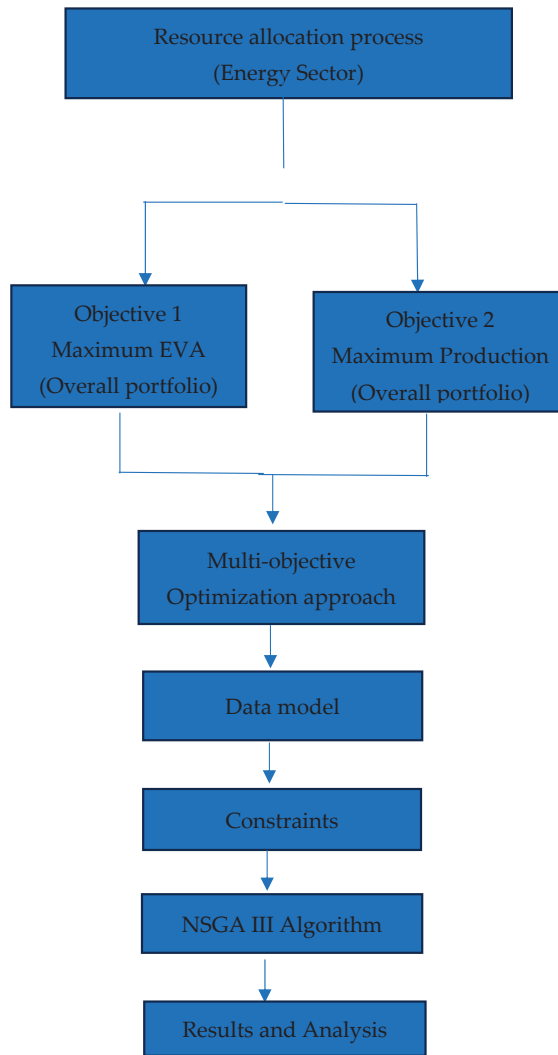
Figure 1. Flow chart of process (Source: Author's calculations).

4. Data and Model

We use data from the Greek green energy market; we have fifteen alternative investment types of projects: four windmills, thirteen photovoltaics, and three power stations. A budget volume analysis is conducted of the windmills, photovoltaics and mandatory connection infrastructure. These types of projects have a mean rate of return, and the infrastructure has returns derived from other projects. Our specific model and data tables are provided below. The budget volume is allocated between equity or equity capital and debt capital using a leverage ratio, a financial metric that measures the proportion of a company's capital that comes from debt. The energy station is expressed at 0.2€ and a return of 6%, which is identical to the WACC because the energy station does not individually generate ROIC but is suitable to generate another project.

Objective Function:

$$O.F_1 = \sum_{i=1}^{20} [X_i \times (ROICC_i - WACC_i)] \qquad \forall\, i = 1, 2\ldots n\ projects \qquad (15)$$

$$O.F_2 = \sum_{i=1}^{20} [X_i \times (P_i^T)] \qquad \forall\, i = 1, 2\ldots n\ projects \qquad (16)$$

$$EVA^T = [EVA_i] = [EVA_1, EVA_2, \ldots EVA_{20}] \qquad \forall\, i = 1, 2\ldots n\ projects \qquad (17)$$

$$EVA_i = r'_i - wacc_i \qquad \forall\, i = 1, 2\ldots n\ projects \qquad (18)$$

$$r'_i = \text{ROIC of each project} \qquad \forall\, i = 1, 2\ldots n\ projects$$

Budget = 15 million €

WACC: A table of WACC for each project

$$i_s = 15\%,\ i_d = 6\%,\ \varphi = 22\%$$

P8, P13, and P15 represent the energy plants where they do not provide a profit on their own but are necessary for the operation of the other types of investments, and therefore, EVA = 0.

Restrictions:

P1 or P3 or P14	(19)
P12 and P5 and P16 then P8	(20)
P11 and P4 then P15	(21)
P7 or P9	(22)
X12 × P12 + X5 × P5 + X16 × P16 ≥ 14000	(23)
X11 × P11 + X4 × P4 ≥ 3000	(24)
$\sum_{i=1}^{20} X_i \times C_i \leq Budget \quad \forall\, i = 1, 2\ldots n\ projects$	(25)

Explanations of the restrictions:

Only one of the investments P1, P2, and P14 can be implemented according to

$$x_1 + x_3 + x_{14} \leq 1. \qquad (26)$$

P8 represents a power station, and P4, P5, and P17 are photovoltaic investments that will be built in the same area. If at least one of the three is implemented, it is appropriate to create P8 because it will meet the energy needs of all 3. The equation that represents this is

$$x_{12} + x_5 + x_{16} - 3x_8 \leq 0. \qquad (27)$$

P15 represents a power station, and P11 and P6 are photovoltaic investments that will be created in the same area. If at least one of the two is implemented, P15 is appropriate to be created because it will meet the energy needs of both. The equation representing this is

$$x_{11} + x_4 - 2x_{15} \leq 0. \qquad (28)$$

Only one of P4 and P6 can be implemented according to the following equation:

$$x_7 + x_9 \leq 1. \qquad (29)$$

Projects under constraint 2 must, if an energy station is built, cover the minimum energy needs of the region.

Projects falling under restriction 3 must, if an energy station is built, cover at least the minimum energy needs of the area.

The total cost must not exceed the available budget:

$$\text{Budget} \leq 15000000\ \text{€} \qquad (30)$$

In Table 1, the cost and ROIC of available investment projects and power plants provide information on the investment projects.

Table 1. Data on investment projects.

	Cost	Production	Price	Revenue	Type	DTE
P1	5.00 €	18,000	50 €	0.9 €	Windmill	4
P2	3.00 €	9750	40 €	0.39 €	Windmill	2
P3	4.00 €	10,000	60 €	0.6 €	Windmill	3
P4	1.00 €	2300	65 €	0.15 €	Photovoltaics	4
P5	0.75 €	1200	68 €	0.081 €	Photovoltaics	2
P6	1.50 €	2800	70 €	0.196 €	Photovoltaics	3
P7	0.60 €	1250	72 €	0.09 €	Photovoltaics	4
P8	0.20 €				Electric Station	
P9	0.40 €	850	75 €	0.063 €	Photovoltaics	4
P10	0.50 €	750	66 €	0.049 €	Photovoltaics	2
P11	1.00 €	1500	74 €	0.111 €	Photovoltaics	2
P12	0.50 €	700	65 €	0.045 €	Photovoltaics	2
P13	0.20 €				Electric Station	
P14	4.00 €	14,000	45 €	0.63 €	Windmill	4
P15	0.20 €				Electric Station	
P16	5.00 €	13,500	63 €	0.85 €	Photovoltaics	4
P17	0.70 €	830	67 €	0.055 €	Photovoltaics	2
P18	0.80 €	770	70 €	0.053 €	Photovoltaics	2
P19	2.00 €	4750	63 €	0.299 €	Photovoltaics	4
P20	1.45 €	2230	65 €	0.15 €	Photovoltaics	2

Source: Data from the Greek energy market.

5. Estimations and Results

Using the NSGA III algorithm and matrix calculations, our findings are provided below:

In Table 2, the appropriate formulation gives the table for constraint Equations (26)–(29) according to a general form of Equations (10)–(12) (e.g., if project 1 and project 2 are conflicting, then the equation representing them is $x_7 + x_9 \leq 1$. If projects 3 and 4 are complementary to project 5 of the energy station, then the corresponding equation is $x_{11} + x_4 - 2 \times x_{15} \leq 0$).

Table 2. Matrixes of complementary and conflict projects.

	Conflict		Complementary	
X1	1	0	0	0
X2	0	0	0	0
X3	1	0	0	0
X4	0	0	0	1
X5	0	0	1	0
X6	0	0	0	0
X7	0	1	0	0
X8	0	0	−3	0
X9	0	1	0	0
X10	0	0	0	0

Table 2. *Cont.*

	Conflict			Complementary	
X11	0	0		0	1
X12	0	0		1	0
X13	0	0		0	0
X14	1	0		0	0
X15	0	0		0	−2
X16	0	0		1	0
X17	0	0		0	0
X18	0	0		0	0
X19	0	0		0	0
X20	0	0		0	0
	RESTRICTIONS			RESTRICTIONS	
X × A	1	0	X × S	−1	0
	≤			≤	
	LIMIT OF RESTRICTIONS			LIMIT OF RESTRICTIONS	
	1	1		0	0

Source: Author's calculations.

In Table 3, the appropriate formulation gives the table for constraint Equations (23) and (24).

Table 3. Matrixes of complementary and conflict projects.

	Production		Conflict
X1	18,000	0	0
X2	9750	0	0
X3	10,000	0	0
X4	2300		0
X5	1200	0	0
X6	2800	0	0
X7	1250	0	1
X8			0
X9	850	0	1
X10	750	0	0
X11	1500		0
X12	700	700	0
X13		0	0
X14	14,000	0	0
X15		0	0
X16	13,500	13,500	0
X17	830	0	0
X18	770	0	0

Table 3. *Cont.*

	Production	Conflict	
X19	4750	0	0
X20	2230	0	0
		RESTRICTIONS	
$X \times P^T$	14,200	38,000	
	$\geq$		
	LIMIT OF RESTRICTIONS		
	14,000	30,000	

Table 4 presents the details of each eligible or ineligible project.

Table 4. Structure of an optimal portfolio.

	EVA = ROIC − WACC	ROIC	WACC	COST ^{D}	X	SELECTED PROJECTS
X1	6%	13%	6.7%	5.00 €	0	WINDMILL
X2	0%	8%	8.1%	3.00 €	1	-
X3	3%	10%	7.3%	4.00 €	0	-
X4	3%	10%	6.7%	1.00 €	1	-
X5	−2%	6%	8.1%	0.75 €	0	PHOTOVOLTAICS
X6	1%	8%	7.3%	1.50 €	0	PHOTOVOLTAICS
X7	3%	10%	6.7%	0.60 €	0	-
X8	0%	0%	0%	0.20 €	1	ENERGY STATION
X9	4%	11%	6.7%	0.40 €	0	-
X10	−3%	5%	8.1%	0.50 €	0	-
X11	−2%	6%	8.1%	1.00 €	1	PHOTOVOLTAICS
X12	−4%	4%	8.1%	0.50 €	1	PHOTOVOLTAICS
X13	0%	0%	0%	0.20 €	0	-
X14	4%	11%	6.7%	4.00 €	1	WINDMILL
X15	0%	0%	0%	0.20 €	1	ENERGY STATION
X16	5%	12%	6.7%	5.00 €	1	PHOTOVOLTAICS
X17	−5%	3%	8.1%	0.70 €	0	-
X18	−4%	4%	8.1%	0.80 €	0	-
X19	0%	3%	6.7%	2.00 €	0	-
X20	1%	−3%	8.1%	1.45 €	0	-

Source: Author's calculations. D referred to diagonal matrix.

The EVA of power stations is 0% because they do not generate a profit, but they are suitable for the creation of another project.

Table 5 presents an estimation of the Objective Functions (OFs)—total EVA or PROFIT, total production, and total budget,

Table 5. Optimal portfolio OF.

MAX EVA OF KNAPSACK	**11.91%**
FULFIL OF KNAPSACK = X × C	**14.9 €**
MAX PRODUCTION	**40,750**

Source: Author's calculations.

Figure 2 reflects the Pareto efficient front, and the optimal solutions (Max EVA, Max Production) range from 14.7% (38,000) to 11.91% (40,750).
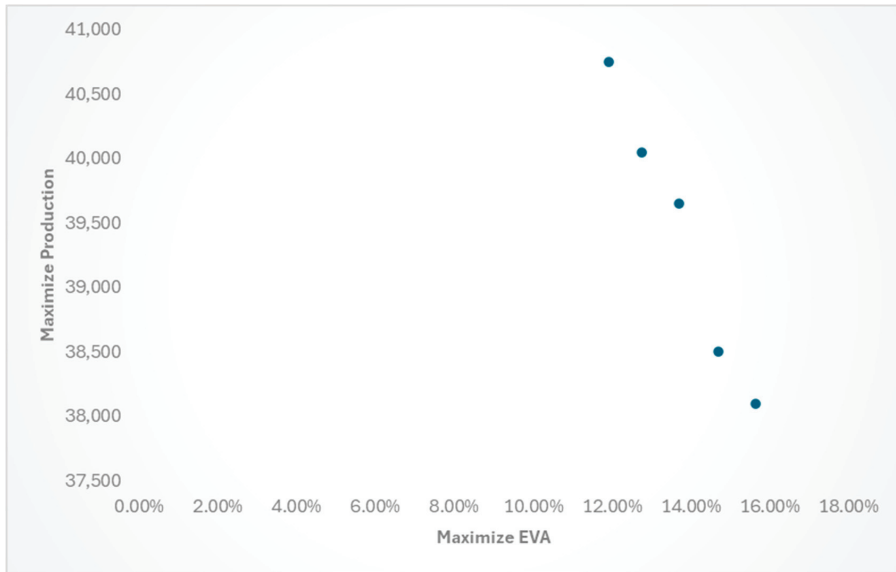

Figure 2. Pareto front (Source: Author's calculations).

Data from any project similar to the data we used from the Greek Green Energy Market, which are taken from a case study, can be used.

From the above estimation, it is concluded that the optimal portfolio structure is as follows:

1. Fulfilling the knapsack—budget: In our example, the optimal budget to be consumed is 14.9 M€.
2. Selection of projects xi: Our optimal portfolio structure includes projects P1, P4, P11, P12 and P19, P16 and power plants P8 and P15.
3. OF 1: The optimal profit is 11.91%.
4. OF 2: Optimal production 40,750.

6. Discussions and Conclusions

6.1. Discussions and Key Findings

The paper aims to extend the existing literature on portfolio optimization in the energy market. From the overall analysis carried out in this case study, its contribution to the optimal allocation of budget resources lies in constructing a specialized model for energy projects based on the data of the Greek energy market.

The paper presents a novel optimization approach to resource allocation in the energy sector, employing a combination of Economic Value Added (EVA) and the knapsack problem. The authors developed a double optimization model within a knapsack-type

framework using the NSGA III algorithm. The research focuses on maximizing EVA alongside production levels to enhance decision-making in the energy sector, particularly in Greece. Applying the model to the Greek energy market offers valuable insights, especially in balancing investments in renewable energy, such as wind power, against overall portfolio returns.

This study has several findings.

First, complementarity and competitiveness between projects are key issues addressed in this paper. Its contribution to this issue is that the creation of alternative constraints to the primary Boolean type of problem of the variable to perform or not to perform a project is always an integer, which does not allow continuous mathematical approximations (Pferschy and Schauer 2017).

Second, in this paper, financial science enters the problem, and mainly the objective function, by maximizing the Economic Value Added (EVA) in the optimal projects for the part of the Economic Value Added that constitutes the knapsack budget (Dluhopolskyi et al. 2021). In this way, the optimal projects that constitute the portfolio knapsack are decided by the firm's management, not by the shareholders or banks.

Third, maximizing EVA while maximizing production using the NSGA III algorithm (Eftekharian et al. 2017) that we carried out in our paper is an innovation that focuses on decision-making according to the objectives of the company's management while meeting the energy needs of the region through maximizing production, ensuring that in a period of price decrease, there will not be a significant impact on the EVA of the portfolio (Vazhayil and Balasubramanian 2014).

Fourth, introducing the constant WACC vector per project allows the introduction of leverage per project and the differentiation of the cost of each project while differentiating the cost of capital per project type in energy from the IVSC studies. Finally, for a given budget, it seems that if you want to increase production, the increase in production must "sacrifice" some of the EVA. This is because elective projects are likely to have higher costs.

The strength of the paper lies in its straightforward methodological approach and the innovative use of EVA as a financial metric alongside a traditional knapsack problem to address the complexities of energy portfolio management. The authors differentiate their approach from previous work by introducing a constant WACC vector per project and incorporating the concept of sacrificing some EVA for production maximization, offering a fresh perspective in multi-objective optimization.

6.2. Theoretical Implication and Practical Implication

This study enriches the theoretical research of optimal allocation and portfolio optimization approach in the energy market. The originality lies in the contribution of a specific methodology to decision-making for selecting specific projects in the energy sector instead of the simple and project-specific financial methods, NPV–IRR. The theoretical implication of the article contributes to the academic debate on whether the EVA index can offer better results than traditional indicators, especially in the energy sector (Dobrowolski et al. 2022).

Also, following the multi-objective optimization approach, it tries to find the balance between the optimization of the financial objective (EVA), ensuring that in a period of price decrease, there will not be a significant impact on the EVA of the portfolio. The second objective is to maximize the production of the portfolio covering, in addition to the above objective, the objective of sustainable development meeting the energy needs of each region. The practical application depends on the fact that the model we develop is not a theoretical model but a model of operational research for market practitioners. Each investor can adjust the model to his needs depending on the specific characteristics of the projects they have to choose from and the regions where they are located. Data from the Greek energy market were used to further enrich scientific approaches in the professional field, helping qualified executives make more rational decisions.

6.3. Restrictions and Future Work

Although this study provides valuable insights, there are some restrictions. First, the data on the Greek energy market we used are limited. A second observation is that the decline in portfolio EVA is disproportionate to the increase in the output provided by this "sacrifice". This is probably due to the specific projects to be selected. Third, the optimal solution to the problem is reached relatively quickly, which explains why the Pareto front of this form is, as in Figure 1, influenced by the above two factors.

The next stage of the research will be to expand the available projects for selection and to see if this problem has been eliminated. However, as these projects belong to the photovoltaic sector, the result is expected to stay the same. A large part of the cost comes from creating energy plants, a sufficient and necessary condition that increases costs without producing a direct economic benefit. The restriction to cover the energy needs of each region has also been introduced, which 'sacrifices' part of the purely financial objective for the benefit of society.

An extension of our research would be the introduction of the theory of representation and the constraints it defines between the management–shareholder–banking relations. According to agent theory and Stewart's EVA, it seeks to build a portfolio that satisfies all three parties (banks, shareholders, and managers) because investors handle money from funds and form portfolios from managers, so it is more appropriate to choose the EVA which managers use to secure their fees, and when EVA is positive, both managers and creditors are satisfied. The specific data refer to Greece; any international investor can use the model by importing data from another economy.

Future research should consider networks' availability regarding loads at different times and other technical specifications. Network inefficiencies, which result from losses for both the type of project and the area of implementation, affect both production and price.

Appendix A

Algorithm A1: Generation t of NSGA-III Procedure
Input: H structured reference points Z* or supplied aspiration points Z*, parent population P_t Output: P_{t+1} 1. $S_t = \varnothing$ 2. Q_t = Recombination + Mutation (P_t)   3. $R_t = P_t \cup Q_t$ 4. $(F_1, F_2, \ldots)$ = Non-dominated-sort (R_t)   5. repeat   6. $S_t = S_t \cup F_i$ and $i = i + 1$ 7. until $|S_t| \geq N$ 8. Last front to be included: $F_l = F_i$ 9. if $|S_t| = N$ then   10.　$P_{t+1} = S_t$, break   11. else

---

Algorithm A1: *Cont.*

---

12. $P_{t+1} = \bigcup\limits_{j=1}^{l-1} F_j$

13. Point to chosen from F_l: K = N − $|P_{t+1}|$

14. Normalize objectives and create reference set Z^*:

Normalize (f^n, S_t, Z^r, Z^s, Z^a)

15. Associate each member s of S_t with a reference point:

[π(s),d(s)] = Associate (S_t, Z^*)

16. Compute niche count of reference point

$j \in Z^*$: $\rho_j = \sum\limits_{s \in s_t/F_i} ((\pi(s) = j)?1:0)$

17. Choose K members one at a time from Fl to construct P_{t+1}:

Niching(K, ρ_j, π, d, Z^r, F_l, P_{t+1})

18. end if

Step-1 Normalize f^n, S_t, Z^r, Z^s/Z^a) procedure

Input: St, Z^s (structured points) or Z^a (supplied points)

Output: f^n, Z^r (reference points on normalized hyper-plane)

1. for j = 1 to M do

2.   Compute ideal point: z_j^{min} = $min_{s \in St}$ $f_j(s)$

3.   Translate objectives: $f_j'(s) = f_j(s) − z_j^{min}$ $\forall s \in$ St

4.   Compute extreme points: z^{jmax} = s:

$argmin_{s \in St}$ ASF(s, w^j) = (ϵ, ..., ϵ)^T),

$\epsilon = 10^{-6}$ and $w_i^j = 1$

5. end for

6. Compute intercepts a_j for j= 1, M

7. Normalize objectives (f^n) using

$f_i^n(X) = \dfrac{f'_i(X)}{a_i − Z_i^{min}}$, for i = 1, 2, ..., M

8. if Z^a is given then

9.   Map each (aspiration) point on normalized hyper-plane

and save the points in the set Z′

10. else

11. $Z^r = Z^s$

12. end if

Step-2 Associate (S_t, Z^r) procedure

Input: S_t, Z^r

Output: π (s $\in S_t$), d(s $\in S_t$)

1. for each reference point Z $\in Z^r$ do

2.   Compute $_{reference}$ line w = z

3. end for

4. for each (s $\in S_t$) do

5.   for each w $\in Z^r$ do

6.   Compute $d^\perp$(s, w) = s − w^Ts / $||w||$

7.   end for

8.   Assign π(s) = w: $argmin_{w \in Z^r}$ $d^\perp$(s, w)

9.   Assign d(s) = $d^\perp$(s, π(s))

10. end for

Step-3 Niching (K, ρ_j, π, d, Z^r, F_l, P_{t+1}) procedure

Input: K, ρ_j, π(s $\in S_t$), d(s $\in S_t$), Z^r, F_l

Output: P_{t+1}

1. k = 1

2. while k $\leq$ K do

3. J_{min} = {j: $argmin_{j \in Zr}$ ρj}

4. $\bar{J}$ = random (j_{min})

5.   Ij = {s: π(s) = j, s $\in$ Fl}

Algorithm A1: *Cont.*

6.   if $I_{\bar{J}} \neq \varnothing$ then
7.   if $\rho_j = 0$ then
8. $P_{t+1} = P_{t+1} \cup \{s: \text{argmin}_{\, s \in I_{\bar{J}}}\, ds\}$
9.   else
10. $P_{t+1} = P_{t+1} \cup \text{random}\left(I_{\bar{J}}\right)$
11.   end if
12. $\rho_j = \rho_j + 1$, $F_l = F_l/s$
13. $k = k + 1$
14. else
15. $Z^r = Z^r / \{\bar{J}\}$
16. end if
17. end while

References

Adamuthe, Amol, Vaishnavi Sale, and Sandeep Mane. 2020. Solving single and multi-objective 01 knapsack problem using harmony search algorithm. *Journal of Scientific Research* 64: 160–67. [CrossRef]

Alazemi, Talal, Mohamed Darwish, and Mohamed Radi. 2024. Renewable energy sources integration via machine learning modelling: A systematic literature review. *Heliyon* 10: e26088. [CrossRef] [PubMed]

Awasthi, Abhishek, Francesco Bär, Joseph Doetsch, Hans Ehm, Marvin Erdmann, Maximilian Hess, Johannes Klepsch, Peter A. Limacher, Andre Luckow, Christoph Niedermeier, and et al. 2023. Quantum computing techniques for multi-knapsack problems. Paper presented at Science and Information Conference, London, UK, July 13–14; Cham: Springer Nature Switzerland, pp. 264–84.

Balas, Egon, and Eitan Zemel. 1980. An algorithm for large zero-one knapsack problems. *Operations Research* 28: 1130–54. [CrossRef]

Bas, Esra. 2011. A capital budgeting problem for preventing workplace mobbing by using analytic hierarchy process and fuzzy 0-1 bidimensional knapsack model. *Expert Systems with Applications* 38: 12415–22. [CrossRef]

Bin, Adriana, Anibal Azevedo, Leonardo Duarte, Sérgio Salles-Filho, and Pedro Massaguer. 2015. R&D and innovation project selection: Can optimization methods be adequate? *Procedia Computer Science* 55: 613–21.

Cacchiani, Valentina, Manuel Iori, Alberto Locatelli, and Silvano Martello. 2022. Knapsack problems—An overview of recent advances. Part II: Multiple, multidimensional, and quadratic knapsack problems. *Computers & Operations Research* 143: 105693.

Camargo, Victor, Leandro Mattiolli, and Franklina Toledo. 2012. A knapsack problem as a tool to solve the production planning problem in small foundries. *Computers & Operations Research* 39: 86–92.

Chen, Wenxin, Hongtao Ren, and Wenji Zhou. 2023a. Review of multi-objective optimization in long-term energy system models. *Global Energy Interconnection* 6: 645–60. [CrossRef]

Chen, Yuanzhan, Zhuo Jin, and Bo Qin. 2023b. Economic Value Added in performance measurement: A simulation approach and empirical evidence. *Accounting & Finance* 63: 109–40.

Dantzig, George. 1957. Discrete-variable extremum problems. *Operations Research* 5: 266–88. [CrossRef]

Deb, Kalyanmoy, and Himanshu Jain. 2013. An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, part I: Solving problems with box constraints. *IEEE Transactions on Evolutionary Computation* 18: 577–601. [CrossRef]

deLlano-Paz, Fernando, Anxo Calvo-Silvosa, Susana Inglesias Antelo, and Isabel Soares. 2017. Energy planning and modern portfolio theory: A review. *Renewable and Sustainable Energy Reviews* 77: 636–51. [CrossRef]

de Queiroz Lafetá, Thiago Fialho, and Gina Maira Barbosa de Oliveira. 2020. Applying Dynamic Evolutionary Optimization to the Multiobjective Knapsack Problem. In *Intelligent Systems: 9th Brazilian Conference, BRACIS 2020, Rio Grande, Brazil, October 20–23, 2020, Proceedings, Part I 9*. Berlin/Heidelberg: Springer International Publishing, pp. 49–63.

Di Somma, Marialaura, Giorgio Graditi, Ehsan Heydarian-Forushani, Miadreza Shafie-khah, and Pierluigi Siano. 2018. Stochastic optimal scheduling of distributed energy resources with renewables considering economic and environmental aspects. *Renewable Energy* 116: 272–87. [CrossRef]

Dluhopolskyi, Oleksandr, Vasyl Brych, Olena Borysiak, Mykhailo Fedirko, Nataliya Dziubanovska, and Nataliya Halysh. 2021. Modeling the environmental and economic effect of value added created in the energy service market. *Polityka Energetyczna* 24: 153–64. [CrossRef]

Dobrowolski, Zbysław, Grzegorz Drozdowski, Mirela Panait, and Arkadiusz Babczuk. 2022. Can the economic value added Be used as the universal financial metric? *Sustainability* 14: 2967. [CrossRef]

Eftekharian, Seyedeh Elham, Mohammad Shojafar, and Shahaboddin Shamshirband. 2017. 2-phase NSGA II: An optimized reward and risk measurements algorithm in portfolio optimization. *Algorithms* 10: 130. [CrossRef]

Ehrgott, Matthias. 2005. *Multicriteria Optimization*. Berlin/Heidelberg: Springer Science & Business Media, vol. 491.

Erlebach, Thomas, Hans Kellerer, and Ulrich Pferschy. 2002. Approximating multiobjective knapsack problems. *Management Science* 48: 1603–12. [CrossRef]

Faia, Ricardo, Tiago Pinto, Zita Vale, and Juan Manuel Corchado. 2018. Multi-objective portfolio optimization of electricity markets participation. Paper presented at 2018 Power Systems Computation Conference (PSCC), Dublin, Ireland, June 11–15; Piscataway: IEEE, pp. 1–6.

Faiteh, Anouar, and Rachid Aasri Mohammed. 2023. Economic value added: The best indicator for measuring value creation or just an illusion? *Investment Management & Financial Innovations* 20: 138.

Feng, Yanhong, and Gaige Wang. 2022. A binary moth search algorithm based on self-learning for multidimensional knapsack problems. *Future Generation Computer Systems* 126: 48–64. [CrossRef]

Florios, Kostas, George Mavrotas, and Danae Diakoulaki. 2010. Solving multiobjective, multiconstraint knapsack problems using mathematical programming and evolutionary algorithms. *European Journal of Operational Research* 203: 14–21. [CrossRef]

García, Jose, and Carlos Maureira. 2021. A KNN quantum cuckoo search algorithm applied to the multidimensional knapsack problem. *Applied Soft Computing* 102: 107077. [CrossRef]

Giacomazzi Dantas, Stefano. 2021. Asset Allocation Using Reinforcement Learning. Master's thesis, McGill University, Montreal, QC, Canada.

Giagkiozis, Ioannis, and Peter Fleming. 2015. Methods for multi-objective optimization: An analysis. *Information Sciences* 293: 338–50. [CrossRef]

Gilmore, Paul, and Ralph Gomory. 1966. Theory and computation of knapsack problems. *Operations Research* 14: 1045–74. [CrossRef]

Göteman, Malin, Marianna Giassi, Jens Engström, and Jan Isberg. 2020. Advances and challenges in wave energy park optimization—A review. *Frontiers in Energy Research* 8: 26. [CrossRef]

Gunantara, Nyoman. 2018. A review of multi-objective optimization: Methods and its applications. *Cogent Engineering* 5: 1502242. [CrossRef]

Haimes, Yacov, Leon Lasdon, and Dang Da. 1971. On a bicriterion formulation of the problems of integrated system identification and system optimization. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-1* 3: 296–97. [CrossRef]

Hifi, Mhand, and Nabil Otmani. 2012. An algorithm for the disjunctively constrained knapsack problem. *International Journal of Operational Research* 13: 22. [CrossRef]

Hifi, Mhand, Sagvan Saleh, and Lei Wu. 2014. A fast large neighborhood search for disjunctively constrained knapsack problems. Paper presented at 3rd International Symposium on Combinatorial Optimization, ISCO 2014, Lisbon, Portugal, March 5–7; Lecture Notes in Computer Science. Cham: Springer, vol. 8596, pp. 396–407.

Horowitz, Ellis, and Sartaj Sahni. 1974. Computing partitions with applications to the knapsack problem. *Journal of the Association or Computing Machinery* 21: 277–92. [CrossRef]

Huot, Chansreynich, Kimleang Kea, Tae-Kyung Kim, and Youngsun Han. 2024. Empirical Analysis of Quantum Approximate Optimization Algorithm for Knapsack-based Financial Portfolio Optimization. *arXiv* arXiv:2402.07123.

Ioannou, Anastasia, Andrew Angus, and Feargal Brennan. 2017. Risk-based methods for sustainable energy system planning: A review. *Renewable and Sustainable Energy Reviews* 74: 602–15. [CrossRef]

Jain, Himanshu, and Kalyanmoy Deb. 2013. An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: Handling constraints and extending to an adaptive approach. *IEEE Transactions on Evolutionary Computation* 18: 602–22. [CrossRef]

Jaszkiewicz, Andrzej. 2004. On the computational efficiency of multiple objective metaheuristics. The knapsack problem case study. *European Journal of Operational Research* 158: 418–33. [CrossRef]

Kellerer, Hans, Ulrich Pferschy, and David Pisinger. 2004. Introduction to NP-Completeness of knapsack problems. In *Knapsack Problems*. Berlin/Heidelberg: Springer, pp. 483–93.

Keshavarzzadeh, Amir, and Pouria Ahmadi. 2019. Multi-objective technoeconomic optimization of a solar based integrated energy system using various optimization methods. *Energy Conversion and Management* 196: 196–210. [CrossRef]

Khuri, Sami, Thomas Bäck, and Jorg Heitkotter. 1994. The zero/one multiple knapsack problem and genetic algorithms. Paper presented at 1994 ACM Symposium on Applied Computing, ACM, Phoenix, AZ, USA, March 6–8; pp. 188–93.

Klamroth, Kathrin, and Margaret Wiecek. 2000. Dynamic programming approaches to the multiple criteria knapsack problem. *Naval Research Logistics* 47: 57–76. [CrossRef]

Kos, Ali, David Morton, Elmira Popova, Stephen Hess, Ernie Kee, and Drew Richards. 2009. Prioritizing Project Selection. *The Engineering Economist* 54: 267–97.

Krink, Thiemo, and Sandra Paterlini. 2011. Multiobjective optimization using differential evolution for real-world portfolio optimization. *Computational Management Science* 8: 157–79. [CrossRef]

Lai, Leonardo, Lorenzo Fiaschi, and Marco Cococcioni. 2020. Solving mixed Pareto-lexicographic multi-objective optimization problems: The case of priority chains. *Swarm and Evolutionary Computation* 55: 100687. [CrossRef]

Li, Fang-Fang, and Jun Qiu. 2016. Multi-objective optimization for integrated hydro-photovoltaic power system. *Applied Energy* 167: 377–84. [CrossRef]

Liapis, Konstantinos J. 2010. The Residual Value Models: A Framework for Business Administration. *European Research Studies* 13: 83–102.

Lin, Xi, Zhiyuan Yang, and Qingfu Zhang. 2022. Pareto set learning for neural multi-objective combinatorial optimization. *arXiv* arXiv:2203.15386.

Liu, Kang, Haibin Ouyang, Steven Li, and Liqun Gao. 2022. A Hybrid Harmony Search Algorithm with Distribution Estimation for Solving the 0-1 Knapsack Problem. *Mathematical Problems in Engineering* 2022: 8440165. [CrossRef]

Liu, Shuai, and Chenglin Xiao. 2021. Application and comparative study of optimization algorithms in financial investment portfolio problems. *Mobile Information Systems* 2021: 3462715. [CrossRef]

Lorie, James, and Leonard Savage. 1955. Three problems in capital rationing. *Journal of Business* 28: 229–39. [CrossRef]

Manne, Alan, and Harry Markowitz. 1957. On the solution of discrete programming problems. *Econometrica* 25: 84–110.

Martello, Silvano, and Paolo Toth. 1990. *Knapsack Problems*. New York: Wiley.

Mavrotas, George. 2009. Effective implementation of the ε-constraint method in multi-objective mathematical programming problems. *Applied Mathematics and Computation* 213: 455–65. [CrossRef]

Mavrotas, George, and Kostas Florios. 2013. An improved version of the augmented ε-constraint method (AUGMECON2) for finding the exact Pareto set in multi-objective integer programming problems. *Applied Mathematics and Computation* 219: 9652–69. [CrossRef]

Mavrotas, George, Pantelis Capros, and Danae Diakoulaki. 2003. Combined MCDA–IP Approach for Project Selection in the Electricity Market. *Annals of Operations Research Volume* 120: 159–70. [CrossRef]

Mesquita-Cunha, Mariana, José Rui Figueira, and Ana Barbosa-Póvoa. 2023. New ε-constraint methods for multi-objective integer linear programming: A Pareto front representation approach. *European Journal of Operational Research* 306: 286–307. [CrossRef]

Milhomem, Danilo, and Maria Jose Pereira Dantas. 2021. Analysis of New Approaches Used in Portfolio Optimization: A Systematic. *Evolutionary and Memetic Computing for Project Portfolio Selection and Scheduling* 26: 125.

Miyamoto, Takuya, and Akihiro Fujiwara. 2022. Robust optimization algorithms for multi-objective knapsack problem. Paper presented at 2022 Tenth International Symposium on Computing and Networking Workshops (CANDARW), Himeji, Japan, November 21–24; Piscataway: IEEE, pp. 430–32.

Metawa, Noura, Mohamed Elhoseny, Kabir Hassan, and Aboul Ella Hassanien. 2016. Loan portfolio optimization using Genetic Algorithm: A case of credit constraints. Paper presented at 12th International Computer Engineering Conference. Boundless Smart Societies, Cairo, Egypt, December 28–29; pp. 59–64.

Nomer, Hazem, Khalid Abdulahzz Alnowibet, Ashraf Elsayed, and Ali Wagdy Mohamed. 2020. Neural knapsack: A neural network based solver for the knapsack problem. *IEEE Access* 8: 224200–10. [CrossRef]

Noorollahi, Ehsan, Dawud Fadai, Seyed Hassan Ghodsipour, and Mohsen Akbarpour Shirazi. 2017. Developing a new optimization framework for power generation expansion planning with the inclusion of renewable energy—A case study of Iran. *Journal of Renewable and Sustainable Energy* 9: 015901. [CrossRef]

Olivas, Frumencio, Ivan Amaya, José Carlos Ortiz-Bayliss, Santiago Enrique Conant-Pablos, and Hugo Terashima-Marín. 2021. Enhancing hyperheuristics for the knapsack problem through fuzzy logic. *Computational Intelligence and Neuroscience* 2021: 8834324. [CrossRef] [PubMed]

Peerlinck, Amy, and John Sheppard. 2022. Multi-objective factored evolutionary optimization and the multi-objective knapsack problem. Paper presented at 2022 IEEE Congress on Evolutionary Computation (CEC), Padua, Italy, July 18–23; Piscataway: IEEE, pp. 1–8.

Pferschy, Ulrich, and Joachim Schauer. 2017. Approximation of knapsack problems with conflict and forcing graphs. *Journal of Combinatorial Optimization* 33: 1300–23. [CrossRef]

Polo, Andrés, Dairo Muñoz Numar Peña, Adrian Cañón, and John Willmer Escobar. 2019. Robust design of a closed-loop supply chain under uncertainty conditions integrating financial criteria. *Omega* 88: 110–32. [CrossRef]

Rombauts, Yannick, Erik Delarue, and William D'haeseleer. 2011. Optimal portfolio-theory-based allocation of wind power: Taking into account cross-border transmission-capacity constraints. *Renew Energy* 36: 2374–87. [CrossRef]

Roques, Fabien, Celine Hiroux, and Marcelo Saguan. 2010. Optimal wind power deployment in Europe—A portfolio approach. *Energy Policy* 38: 3245–56. [CrossRef]

Salkin, Harvey, and Corrnelis De Kluyver. 1975. The knapsack problem: A survey. *Naval Research Logistics Quarterly* 22: 127–44. [CrossRef]

Schönberger, Jörn. 2016. Multi-period vehicle routing with limited period load. *IFAC-PapersOnLine* 49: 24–29. [CrossRef]

Sharma, Anil K., and Satish Kumar. 2010. Economic value added (EVA)-literature review and relevant issues. *International Journal of Economics and Finance* 2: 200–20. [CrossRef]

Shih, Wei. 1979. A Branch and Bound Method for the Multiconstraint Zero-One Knapsack Problem. *Journal of the Operational Research Society* 30: 369–78. [CrossRef]

Stewart, Bennett. 2009. EVA Momentum: The One Ratio That Tells the Whole Story 21.2. *Journal of Applied Corporate Finance* 21: 74–86. [CrossRef]

Sur, Giwon, Shun Yuel Ryu, JongWon Kim, and Hyuk Lim. 2022. A deep reinforcement learning-based scheme for solving multiple knapsack problems. *Applied Sciences* 12: 3068. [CrossRef]

Tao, Yuechuan, Jing Qiu, and Shuying Lai. 2021. Deep reinforcement learning-based bidding strategy for EVAs in local energy market considering information asymmetry. *IEEE Transactions on Industrial Informatics* 18: 3831–42. [CrossRef]

Tapia, Ma Guadalupe Castillo, and Carlos Coello Coello. 2007. Applications of multi-objective evolutionary algorithms in economics and finance: A survey. Paper presented at 2007 IEEE Congress on Evolutionary Computation, CEC, Singapore, September 25–28.

Tripathi, Prasoon, Varun Chotia, Umesh Solanki, Rahul Meena, and Vinay Khandelwal. 2022. Economic Value-Added Research: Mapping Thematic Structure and Research Trends. *Risks* 11: 9. [CrossRef]

Vaish, Jayati, Anil Kumar Tiwari, and Seethalekshmi Kaimal. 2024. Multi-objective optimization of distributed energy resources based microgrid using random forest model. *Bulletin of Electrical Engineering and Informatics* 13: 67–75. [CrossRef]

Vazhayil, Joy, and Rajasekhar Balasubramanian. 2014. Optimization of India's electricity generation portfolio using intelligent Pareto-search genetic algorithm. *International Journal of Electrical Power & Energy Systems* 55: 13–20.

Yamada, Takeo, Seiji Kataoka, and Kohtaro Watanabe. 2002. Heuristic and exact algorithms for the disjunctively constrained knapsack problem. *IPSJ Journal* 43: 9.

Zhou, Xiaojun, Wan Tan, Yan Sun, Tingwen Huang, and Chunhua Yang. 2024. Multi-objective optimization and decision making for integrated energy system using STA and fuzzy TOPSIS. *Expert Systems with Applications* 240: 122539. [CrossRef]

Zitzler, Eckart, and Lothar Thiele. 1999. Multiobjective evolutionary algorithms: A comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation* 3: 257–71. [CrossRef]

Article
Rainbow Step Barrier Options

Tristan Guillaume

Laboratoire Thema, CYU Cergy Paris Université, 33 Boulevard du Port, F-95011 Cergy-Pontoise Cedex, France; tristan.guillaume@cyu.fr; Tel.: +33-6-12-22-45-88

Abstract: This article provides exact analytical formulae for various kinds of rainbow step barrier options. These are highly flexible and sophisticated multi-asset barrier options based on the following principle: the option life is divided into several time intervals on which different barriers are monitored w.r.t. different underlying assets. From a mathematical point of view, new results are provided for the first passage time of a multidimensional geometric Brownian motion to a boundary defined as a step function. The article shows how to implement the obtained option valuation formulae in a simple and very efficient manner. Numerical results highlight a strong sensitivity of rainbow step barrier options to the correlations between the underlying assets.

Keywords: rainbow step barrier option; rainbow option; step barrier option; barrier option; multiasset option; multiasset barrier option; first passage time; boundary crossing probability; multidimensional Brownian motion

1. Introduction

Barrier options are characterised by the introduction in the option contract of a parameter called the "barrier", which, in the most standard form, is a predefined reference value of the underlying asset S that may be located above the spot value of the underlying (upward barrier or "up-barrier") or below it (downward barrier or "down-barrier"). The barrier may be of a "knock-out" type, i.e., the option expires worthless if the barrier is hit by S at any time during the option life (in which case the barrier is called "continuous") or at one or a few predefined times (in which case the barrier is called "discrete"). The first passage time of an underlying to a knock-out barrier, before the option expiry, triggers the "deactivation" of the option. Alternatively, the barrier may be of a "knock-in" type, i.e., the option expires worthless unless the barrier has been hit at least once before expiry, an event called "activation".

Barrier options are the oldest and the most widely traded non-vanilla options. They are embedded in a lot of popular structured derivatives in stock and interest rate markets (see, e.g., Bouzoubaa and Osseiran 2010). They are also extensively used as analytical tools in financial modelling, for instance, in the so-called "structural models" of default risk (see, e.g., Bielecki and Rutkowski 2004) or in the valuation of investments (theory of "real options"). Since their first appearance in the financial markets during the 1970s, there have been a huge number of variations in their original payoff, leading to an extraordinary variety of non-standard barrier options. Among the most well-known of them are the step barrier options, which divide the option lifetime into several time intervals on which the barrier takes on different values. In its standard form, a step barrier option features a piecewise constant barrier, i.e., a barrier defined as a step function. This allows modulation of the barrier level during the option life, thus offering increased flexibility and enhanced risk management capabilities, relative to a traditional barrier option. For instance, up-and-out barrier levels can be raised and down-and-out barrier levels can be lowered during time intervals when more protection is required, thus reducing the risk of deactivation. Likewise, up-and-in barrier levels can be lowered and down-and-in barrier levels can be

raised in time intervals during which the implicit volatility of the underlying rises, thus, increasing the chances of activation.

The exact analytical valuation of step barrier functions was first achieved by Guillaume (2001). Further mathematical details on how closed-form valuation can be achieved, as well as exact results for more general deterministic step barriers, are provided in Guillaume (2015, 2016). In the last couple of years, there has been a renewed interest in step barrier options, as they stand out as an essential component of innovative forms of investment, such as autocallable structured products and other equity-linked products. This has led to a new series of academic contributions, such as Lee et al. (2019a, 2019b), Lee et al. (2021), Lee et al. (2022). Unlike previously cited references, these articles do not provide explicit formulae, except for a few simple and already known cases, nor do they handle outside step barriers as in Guillaume (2001), alternatively upward and downward steps as in Guillaume (2015) and exponentially moving step barriers as in Guillaume (2016). These recent contributions ignore previous results given in references they do not cite, such as Guillaume (2001), that actually solve the problems they discuss. They also claim to be able to analytically value a step barrier option with an arbitrary number of steps, but without explaining how they intend to solve the difficult problem known in numerical integration as the "curse of dimensionality", nor even beginning to discuss the numerical implementation of their approach, which constitutes the main issue, though.

There are still a number of unsolved problems related to the valuation of step barrier options. In particular, multi-asset step barrier options are barely touched upon in the existing literature, apart from an isolated formula for an "outside" step barrier option given in Guillaume (2001), also called an "external" step barrier option, featuring one underlying asset w.r.t. which barrier crossing is monitored and another underlying asset w.r.t. which the moneyness of the option is measured at expiry (the reader may refer to Heynen and Kat 1994, or to Kwok et al. 1998, for background on outside barrier options in general). Yet, multi-asset contracts with step barriers are actively traded in today's financial markets, as they allow investors to benefit from the advantages of diversification in terms of risk control and expansion of investment opportunities. A particularly important subset of these contracts is the so-called "rainbow" step barrier option. Broadly speaking, in the literature on options, the denomination "rainbow" applies to payoffs linked to the performances of two or more underlying assets (Chang et al. 2005; Gao and Wu 2022); metaphorically, each underlying represents a different color, so that the association of all of these factors makes up a rainbow. In the realm of barrier options, the rainbow step barrier option is characterised by the property that, at each time interval, the barrier is monitored w.r.t. a different underlying asset. A contract featuring a number $n \in \mathbb{N}$ of underlying assets associated with n time intervals $[t_0 = 0, t_1], \ldots, [t_{n-1}, t_n]$, on which n steps of a piecewise constant barrier are monitored, is called an $n-$the colour rainbow barrier option. Each time interval is matched with a specific step of the barrier and a specific underlying asset. In the standard form of the contract, it is the $n-$th asset associated with the $n-$th last step of the barrier that is used to determine the moneyness of the contract at expiry. Rainbow step barrier options are typically priced by Monte Carlo simulation, even in a standard Black–Scholes model, because of the difficulties of the entailed analytical calculations and also because the dimension of the valuation problem quickly increases with the number of "colors", leading to non-trivial issues of numerical evaluation of high-dimensional integrals. Due to these obstacles, the present article is restricted to two-colour rainbow step barrier options. Closed-form valuation is achieved not only for standard two-colour contracts but also for two-colour outside step barrier options involving a third correlated asset at expiry, and for two-colour contracts featuring a two-sided barrier (also known as a double barrier), i.e., both an upward and a downward barrier on each time interval. Numerical evaluation of the obtained analytical solutions is dealt with so that the valuation formulae derived in this paper can be immediately implemented and yield extremely accurate results in a few tenths of one second. Numerical results are provided, which reveal a strong and stable

dependency of rainbow step barrier options on the correlations between the underlying assets, as well as the importance of the volatility of the asset used to measure moneyness when dealing with rainbow outside step barrier options. These findings suggest clear application ideas to traders and investors, whether for a hedging or speculative purposes. They highlight the specificity of rainbow step barrier options as instruments highly sensitive to correlation, in contrast to standard step barrier options, which are sensitive to volatility but, by constuction, cannot be sensitive to correlation.

This article is organised as follows: Section 2 states the main analytical results, provides numerical results, and discusses their implications; Section 3 gives the mathematical proofs of the analytical results presented in Section 2.

2. Formulae and Numerical Results

Let us begin with a few definitions. Let S_1 and S_2 be two GBMs (geometric Brownian motions) modelling two asset prices, whose differentials, under a given probability measure P, are given by:

$$dS_1(t) = v_1 S_1(t)dt + \sigma_1 S_1(t)dB_1(t) \tag{1}$$

$$dS_2(t) = v_2 S_2(t)dt + \sigma_2 S_2(t)dB_2(t) \tag{2}$$

where $v_1, v_2 \in \mathbb{R}$, $\sigma_1, \sigma_2 \in \mathbb{R}_+$, and B_1 and B_2 are two standard Brownian motions whose correlation coefficient is denoted by $\rho_{1.2}$.

The measure P is characterised by the pair (v_1, v_2) or, equivalently, by the pair:

$$\left(\mu_1 = v_1 - \sigma_1^2/2, \mu_2 = v_2 - \sigma_2^2/2 \right) \tag{3}$$

If we refer to the log-return processes $X_i(t) = \ln(S_i(t)/S_i(0))$, $i \in \{1, 2\}$, whose differentials under P are given by:

$$dX_i(t) = \mu_i dt + \sigma_i dB_i(t) \tag{4}$$

Let H_1, H_2, K_1, K_2, K_3 be positive real numbers. The numbers H_1, H_2 are the values of two knock-out continuous barriers. H_1 is monitored w.r.t. S_1 on a time interval $[t_0 = 0, t_1]$, while H_2 is monitored w.r.t. S_2 on a time interval $[t_1, t_2]$. The numbers K_1, K_2, K_3 are the values of three discrete knock-out barriers; K_1 is monitored w.r.t. S_1 at time t_1, while K_2 and K_3 are monitored w.r.t. S_2 at times t_1 and t_2, respectively.

We can now begin to value two-colour step barrier options in the following order:

- both steps either upward or downward (Section 2.1);
- one upward step and one downward step (Section 2.2);
- reverse-type contract (Section 2.3);
- outside or external two-colour step barrier (Section 2.4);
- two-colour step double barrier (Section 2.5).

2.1. Valuation of Two-Colour Step Barrier Options When the Steps of the Barrier Are on the Same Side in Each Time Interval

Section 2.1 deals with the valuation of two-colour step barrier options when the steps of the barrier are either both upward or both downward. Our objective is to find the value of the joint cumulative distribution function $P_{[RUU]}(\mu_1, \mu_2)$ defined by:

$$P_{[RUU]}(\mu_1, \mu_2) \triangleq P\left(\sup_{0 \leq t \leq t_1} S_1(t) \leq H_1, S_1(t_1) \leq K_1, S_2(t_1) \leq K_2, \sup_{t_1 \leq t \leq t_2} S_2(t) \leq H_2, S_2(t_2) \leq K_3 \right) \tag{5}$$

where the acronym "[RUU]" stands for "Rainbow Up and Up".

The main result of Section 2.1 is given by the following Proposition 1.

Proposition 1. *The exact value of* $P_{[RUU]}(\mu_1, \mu_2)$ *is given by:*

$$P_{[RUU]}(\mu_1, \mu_2) = N_3\left[\frac{\min(k_1, h_1) - \mu_1 t_1}{\sigma_1\sqrt{t_1}}, \frac{\min(k_2, h_2) - \mu_2 t_1}{\sigma_2\sqrt{t_1}}, \frac{\min(k_3, h_2) - \mu_2 t_2}{\sigma_2\sqrt{t_2}}; \theta_{1.2}, \theta_{1.3}, \theta_{2.3}\right] \tag{6}$$

$$-\exp\left(\frac{2\mu_1 h_1}{\sigma_1^2}\right)$$
$$\times N_3\left[\begin{array}{l}\frac{\min(k_1,h_1)-2h_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}, \frac{\min(k_2,h_2)-\mu_2 t_1}{\sigma_2\sqrt{t_1}} - \frac{2\theta_{1.2}h_1}{\sigma_1\sqrt{t_1}}, \frac{\min(k_3,h_2)-\mu_2 t_2}{\sigma_2\sqrt{t_2}} - \frac{2\theta_{1.2}h_1}{\sigma_1\sqrt{t_2}};\\ \theta_{1.2}, \theta_{1.3}, \theta_{2.3}\end{array}\right] \tag{7}$$

$$-\exp\left(\frac{2\mu_2 h_2}{\sigma_2^2}\right)$$
$$\times N_3\left[\begin{array}{l}\frac{\min(k_1,h_1)-\mu_1 t_1}{\sigma_1\sqrt{t_1}} + \frac{2\theta_{1.2}\mu_2\sqrt{t_1}}{\sigma_2}, \frac{\min(k_2,h_2)+\mu_2 t_1}{\sigma_2\sqrt{t_1}}, \frac{\min(k_3,h_2)-2h_2-\mu_2 t_2}{\sigma_2\sqrt{t_2}};\\ \theta_{1.2}, -\theta_{1.3}, -\theta_{2.3}\end{array}\right] \tag{8}$$

$$+\exp\left(\left(\frac{2\mu_1}{\sigma_1^2} - \frac{4\mu_2\theta_{1.2}}{\sigma_1\sigma_2}\right)h_1 + \frac{2\mu_2 h_2}{\sigma_2^2}\right)$$
$$\times N_3\left[\begin{array}{l}\frac{\min(k_1,h_1)-2h_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}} + \frac{2\theta_{1.2}\mu_2\sqrt{t_1}}{\sigma_2}, \frac{\min(k_2,h_2)+\mu_2 t_1}{\sigma_2\sqrt{t_1}} - \frac{2\theta_{1.2}h_1}{\sigma_1\sqrt{t_1}},\\ \frac{\min(k_3,h_2)-2h_2-\mu_2 t_2}{\sigma_2\sqrt{t_2}} + \frac{2\theta_{1.2}h_1}{\sigma_1\sqrt{t_2}}; \theta_{1.2}, -\theta_{1.3}, -\theta_{2.3}\end{array}\right] \tag{9}$$

where the μ_i's are given by (3) and:

- $N_3[b_1, b_2, b_3; c_{12}, c_{13}, c_{23}]$ *is the trivariate standard normal cumulative distribution function with correlation coefficients c_{12}, c_{13}, c_{23}*

$$h_1 = \ln\left(\frac{H_1}{S_1(0)}\right), h_2 = \ln\left(\frac{H_2}{S_2(0)}\right), k_1 = \ln\left(\frac{K_1}{S_1(0)}\right), k_2 = \ln\left(\frac{K_2}{S_2(0)}\right), k_3 = \ln\left(\frac{K_3}{S_2(0)}\right) \tag{10}$$

$$\theta_{1.2} = \rho_{1.2}, \theta_{1.3} = \sqrt{\frac{t_1}{t_2}}\rho_{1.2}, \theta_{2.3} = \sqrt{\frac{t_1}{t_2}} \tag{11}$$

Corollary 1. *It suffices to multiply by (-1) all the first three arguments of each $N_3[., ., .; ., ., .]$ function and to substitute each* min *operator by a* max *operator in Proposition 1 to obtain an exact formula for* $P_{[RDD]}(\mu_1, \mu_2)$ *defined as:*

$$P_{[RDD]}(\mu_1, \mu_2) \triangleq P\left(\inf_{0 \le t \le t_1} S_1(t) \ge H_1, S_1(t_1) \ge K_1, S_2(t_1) \ge K_2, \inf_{t_1 \le t \le t_2} S_2(t) \ge H_2, S_2(t_2) \ge K_3\right) \tag{12}$$

where the acronym "[RDD]" stands for "Rainbow Down and Down".

Corollary 2. *The term numbered (9) in Proposition 1 gives the value of the corresponding knock-in probability denoted by* $P_{[RUU]}^{(I)}(\mu_1, \mu_2)$ *and defined by:*

$$P_{[RUU]}^{(I)}(\mu_1, \mu_2) \triangleq P\left(\sup_{0 \le t \le t_1} S_1(t) \ge H_1, S_1(t_1) \le K_1, S_2(t_1) \le K_2, \sup_{t_1 \le t \le t_2} S_2(t) \ge H_2, S_2(t_2) \le K_3\right) \tag{13}$$

Corollary 3. *It suffices to substitute each argument $\theta_{2.3}$ in each $N_3[., ., .; ., ., .]$ function of Proposition 1 by $\sqrt{\frac{t_2}{t_3}}$, $\forall t_3 \ge t_2$, to obtain an exact formula for the early-ending variant* $P_{[EERUU]}(\mu_1, \mu_2)$ *defined by:*

$$P_{[EERUU]}(\mu_1, \mu_2) \triangleq P\left(\sup_{0 \le t \le t_1} S_1(t) \le H_1, S_1(t_1) \le K_1, S_2(t_1) \le K_2, \sup_{t_1 \le t \le t_2} S_2(t) \le H_2, S_2(t_3) \le K_3\right) \tag{14}$$

Corollary 4. *Let $\hat{p}$ be the value of $\mathrm{P}_{[\mathrm{RUU}]}(\mu_1, \mu_2)$ when the value of K_3 becomes "very high", i.e., high enough for the probability $P(S_2(t_2) \leq K_3)$ to tend to zero; then, the difference $\hat{p} - \mathrm{P}_{[\mathrm{RUU}]}(\mu_1, \mu_2)$ provides the value of the following minor variant:*

$$\hat{p} - \mathrm{P}_{[\mathrm{RUU}]}(\mu_1, \mu_2) = P\left(\sup_{0 \leq t \leq t_1} S_1(t) \leq H_1, S_1(t_1) \leq K_1, S_2(t_1) \leq K_2, \sup_{t_1 \leq t \leq t_2} S_2(t) \leq H_2, S_2(t_2) > K_3 \right) \quad (15)$$

End of Proposition 1.

Equipped with Proposition 1, one can value in closed form a two-colour step barrier option with two successive upward or two successive downward steps. Applying the theory of non-arbitrage pricing in a complete market (Harrison and Kreps 1979; Harrison and Pliska 1981), the value of a two-colour up-and-up knock-out put, denoted by $V_{[\mathrm{RUU}]}$, is given by:

$$V_{[\mathrm{RUU}]} = e^{-rt_2} \left(E_Q \left[K_3 \mathbf{1}_{\{A\}} - S_2(t_2) \mathbf{1}_{\{A\}} \right] \right) \quad (16)$$

where r is the riskless interest rate assumed to be constant, $\mathbf{1}_{\{\cdot\}}$ is the indicator function and A is the set constructed by the intersection of elements of the $\sigma-$algebra generated by the pair of processes $(S_1(t), S_2(t))$ that characterises the probability $\mathrm{P}_{[\mathrm{RUU}]}(\mu_1, \mu_2)$ as given by the arguments of the probability operator in (5).

A simple application of the Cameron–Martin–Girsanov theorem yields:

$$V_{[\mathrm{RUU}]} = e^{-rt_2} K_3 \mathrm{P}_{[\mathrm{RUU}]}\left(\mu_1^{(Q)}, \mu_2^{(Q)} \right) - S_2(0) \mathrm{P}_{[\mathrm{RUU}]}\left(\mu_1^{(P_2)}, \mu_2^{(P_2)} \right) \quad (17)$$

where

$$\mu_i^{(Q)} = r - \frac{\sigma_i^2}{2}, \ \mu_1^{(P_2)} = r - \frac{\sigma_1^2}{2} + \sigma_1 \sigma_2 \rho_{1.2}, \ \mu_2^{(P_2)} = r + \frac{\sigma_2^2}{2}, \quad (18)$$

Q is the measure under which $\left\{ B_i(t) + \frac{\mu_i - r}{\sigma_i} t, t \geq 0 \right\}$ is a standard Brownian motion (the classical so-called risk-neutral measure), while P_2 is the measure under which $\left\{ B_1(t) - \sigma_2 \rho_{1.2} t, t \geq 0 \right\}$ and $\left\{ B_2(t) - \sigma_2 \sqrt{1 - \rho_{1.2}^2} t, t \geq 0 \right\}$ are two independent standard Brownian motions.

To factor in a continuous dividend rate δ_i associated with each asset S_i, simply replace r by $r - \delta_i$.

All the other two-colour rainbow barrier options subsequently mentioned in Sections 2.1 and 2.2, whether they be knock-in or feature a mixture of a downward and an upward barrier, are identically valued, by taking the relevant $\mathrm{P}_{[\cdot]}$ probability along with the pairs $\left(\mu_1^{(Q)}, \mu_2^{(Q)} \right)$ and $\left(\mu_1^{(P_2)}, \mu_2^{(P_2)} \right)$.

The numerical implementation of Proposition 1 is easy. Using Genz's (2004) algorithm to evaluate the trivariate standard normal cumulative distribution function, the accuracy and efficiency required for all practical purposes can be achieved in computational times in the order of 0.1 s. Table 1 provides the prices of a few two-colour up-and-up knock-out put options, for various levels of the volatility and correlation parameters of the underlying assets S_1 and S_2, and different values of the knock-out barriers. All the initial values of the underlying assets $S_i(0)$ and the strike prices K_i are set at 100. Expiry is 1 year. The two time intervals $[t_0, t_1]$ and $[t_1, t_2]$ have equal length, i.e., $t_1 =$ 6 months, but unequal time lengths can be handled just as well by the formulae. The riskless interest rate is assumed to be 2.5%.

In each cell, four prices are reported: the first one is the exact analytical value as obtained by implementing Proposition 1, while the prices in brackets are three successive approximations obtained by performing increasingly large Monte Carlo simulations. More specifically, these approximations rely on the conditional Monte Carlo method, which is well known for its accuracy and efficiency (Glasserman 2003). The number of simulations performed is 500,000 for the first approximation, 2,000,000 for the second approximation,

and 10,000,000 for the third approximation. The pseudo-random numbers are drawn from the reliable Mersenne Twister generator (Matsumoto and Nishimura 1998).

Table 1. Two-colour up-and-up knock-out put.

	$\rho_{1.2} = -0.6$	$\rho_{1.2} = -0.2$	$\rho_{1.2} = 0.2$	$\rho_{1.2} = 0.6$
$\sigma_1 = \sigma_2 = 20\%$ $H_1 = H_2 = 115$	1.189 (1.123,1.168, 1.182)	2.163 (2.237, 2.141, 2.161)	3.134 (3.082, 3.116, 3.132)	4.141 (4.174, 4.127, 4.140)
$\sigma_1 = \sigma_2 = 60\%$ $H_1 = H_2 = 125$	4.065 (4.109, 4.072, 4.061)	6.440 (6.392, 6.449, 6.442)	8.769 (8.734, 8.755, 8.763)	11.213 (11.278, 11.196, 11.218)
$\sigma_1 = 20\%, \sigma_2 = 60\%$ $H_1 = 115, H_2 = 125$	4.454 (4.411, 4.443, 4.454)	7.260 (7.355, 7.301, 7.264)	9.979 (9.912, 9.101, 9.975)	12.844 (12.957, 12.892, 12.848)
$\sigma_1 = 60\%, \sigma_2 = 20\%$ $H_1 = 125, H_2 = 115$	1.096 (1.082, 1.108, 1.091)	1.921 (1.107, 1.953, 1.928)	2.753 (2.796, 2.745, 2.752)	3.617 (3.692, 3.599, 3.614)

In purely numerical terms, it can be clearly observed that the conditional Monte Carlo approximations gradually converge to the analytical values as more and more simulations are performed. A minimum of 10,000,000 simulations are necessary to guarantee a modest 10^{-3} convergence. This requires a computational time of approximately 35 s on a computer equipped with a Core i7 CPU. Much more accurate values can be obtained by means of Proposition 1 in only two-tenths of a second. This gap in accuracy and efficiency makes a particularly valuable difference when pricing large portfolios of options.

From a financial point of view, the most striking phenomenon observed in Table 1 is that the option price regularly and significantly increases with the value of the correlation coefficient between assets S_1 and S_2, whatever the volatilities and the levels of the barriers. Roughly speaking, the price of an at-the-money two-colour up-and-up knock-out put option when $\rho_{1.2} = 0.6$ is three times greater than when $\rho_{1.2} = -0.6$. This property can be exploited by traders who take positions on correlation, as the prices of these options will substantially increase if implicit correlation turns out to be underestimated by the markets. This property can also be harnessed by traders to construct hedges on sold derivatives that are sensitive to pairwise correlation. From an investor's perspective, the observed phenomenon allows to define effective strategies to reduce the cost of hedging by tapping into negative correlation. Such a significant functional relation w.r.t. correlation is a major attraction of rainbow step barrier options relative to non-rainbow step barrier options, as the latter can only handle volatility effects.

Another noticeable fact in Table 1 is that lowering the up-and-out barriers seems much more effective in reducing the option's price than lowering the volatilities of assets S_1 and S_2, regardless of the sign and the magnitude of correlation. Indeed, looking at row 1 in Table 1, one can see that the options are relatively cheap, although the volatilities of both assets S_1 and S_2 are low, because the knock-out barriers are located quite near the spot prices of the underlying assets; and looking at row 2 in Table 1, one can see that the options are relatively expensive, although the volatilities of both assets S_1 and S_2 are high because the knock-out barriers are more distant. This shows that the barrier effect, which drives prices down as up-and-out barriers become lower and conversely drives prices up as the up-and-out barrier becomes higher, and prevails over the volatility effect, which exerts its influence in the opposite direction, i.e., a lower volatility pushes prices up by decreasing the probability of knocking out before expiry and a higher volatility pushes prices down by increasing the latter probability. This phenomenon can be explained by the ambivalent nature of volatility: on the one hand, less volatility means less risk of being deactivated before expiry, but on the other hand, it also means fewer chances of ending in-the-money

at expiry; whichever of this positive and this negative effect weighs more on the option price depends on the relative values of barrier, strike, volatility and expiry parameters in a complex manner.

2.2. Valuation of Two-Colour Step Barrier Options Involving One Upward Step and One Downward Step

Section 2.2 deals with the case when the steps of the barrier are not on the same side in each time interval, i.e., either first downward, then upward, or first upward, then downward.

The main result of Section 2.2 is given by the following Proposition 2.

Proposition 2. *Let* $P_{[RUD]}(\mu_1, \mu_2)$ *denote the joint cumulative distribution function defined by:*

$$P_{[RUD]}(\mu_1, \mu_2) \triangleq P\left(\sup_{0 \leq t \leq t_1} S_1(t) \leq H_1, S_1(t_1) \leq K_1, S_2(t_1) \geq K_2, \inf_{t_1 \leq t \leq t_2} S_2(t) \geq H_2, S_2(t_2) \geq K_3 \right) \tag{19}$$

where the acronym "[RUD]" stands for "Rainbow Up and Down".

Then, the exact value of $P_{[RUD]}(\mu_1, \mu_2)$ *is given by:*

$$P_{[RUD]}(\mu_1, \mu_2)$$
$$= N_3 \left[\frac{\min(k_1, h_1) - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}, \frac{-\max(k_2, h_2) + \mu_2 t_1}{\sigma_2 \sqrt{t_1}}, \frac{-\max(k_3, h_2) + \mu_2 t_2}{\sigma_2 \sqrt{t_2}}; -\theta_{1.2}, -\theta_{1.3}, \theta_{2.3} \right] \tag{20}$$

$$- \exp\left(\frac{2\mu_1 h_1}{\sigma_1^2} \right)$$
$$\times N_3 \left[\begin{array}{l} \frac{\min(k_1, h_1) - 2h_1 - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}, \frac{-\max(k_2, h_2) + \mu_2 t_1}{\sigma_2 \sqrt{t_1}} + \frac{2\theta_{1.2} h_1}{\sigma_1 \sqrt{t_1}}, \\ \frac{-\max(k_3, h_2) + \mu_2 t_2}{\sigma_2 \sqrt{t_2}} + \frac{2\theta_{1.2} h_1}{\sigma_1 \sqrt{t_2}}; -\theta_{1.2}, -\theta_{1.3}, \theta_{2.3} \end{array} \right] \tag{21}$$

$$- \exp\left(\frac{2\mu_2 h_2}{\sigma_2^2} \right)$$
$$\times N_3 \left[\begin{array}{l} \frac{\min(k_1, h_1) - \mu_1 t_1}{\sigma_1 \sqrt{t_1}} + \frac{2\theta_{1.2} \mu_2 \sqrt{t_1}}{\sigma_2}, \frac{-\max(k_2, h_2) - \mu_2 t_1}{\sigma_2 \sqrt{t_1}}, \frac{-\max(k_3, h_2) + 2h_2 + \mu_2 t_2}{\sigma_2 \sqrt{t_2}}; \\ -\theta_{1.2}, \theta_{1.3}, -\theta_{2.3} \end{array} \right] \tag{22}$$

$$+ \exp\left(\left(\frac{2\mu_1}{\sigma_1^2} - \frac{4\mu_2 \theta_{1.2}}{\sigma_1 \sigma_2} \right) h_1 + \frac{2\mu_2 h_2}{\sigma_2^2} \right)$$
$$\times N_3 \left[\begin{array}{l} \frac{\min(k_1, h_1) - 2h_1 - \mu_1 t_1}{\sigma_1 \sqrt{t_1}} + \frac{2\theta_{1.2} \mu_2 \sqrt{t_1}}{\sigma_2}, \frac{-\max(k_2, h_2) - \mu_2 t_1}{\sigma_2 \sqrt{t_1}} + \frac{2\theta_{1.2} h_1}{\sigma_1 \sqrt{t_1}}, \\ \frac{-\max(k_3, h_2) + 2h_2 + \mu_2 t_2}{\sigma_2 \sqrt{t_2}} - \frac{2\theta_{1.2} h_1}{\sigma_1 \sqrt{t_2}}; \theta_{1.2}, -\theta_{1.3}, -\theta_{2.3} \end{array} \right] \tag{23}$$

where all the notations are identical, as in Proposition 1.

Corollary 1. *It suffices to multiply by* (-1) *all the first three arguments of each* $N_3[., ., .; ., ., .]$ *function and substitute each* min *operator by a* max *operator as well as each* max *operator by a* min *operator in Proposition 2 to obtain an exact formula for* $P_{[RDU]}(\mu_1, \mu_2)$ *defined as:*

$$P_{[RDU]}(\mu_1, \mu_2) \triangleq P\left(\inf_{0 \leq t \leq t_1} S_1(t) \geq H_1, S_1(t_1) \geq K_1, S_2(t_1) \leq K_2, \sup_{t_1 \leq t \leq t_2} S_2(t) \leq H_2, S_2(t_2) \leq K_3 \right) \tag{24}$$

Corollary 2. *The term numbered (23) in Proposition 2 provides the value of the corrresponding up-and-in, then down-and-in probability, denoted as* $P_{[RUD]}^{(I)}(\mu_1, \mu_2)$ *and defined by:*

$$P_{[RUD]}^{(I)}(\mu_1, \mu_2) \triangleq P\left(\sup_{0 \leq t \leq t_1} S_1(t) \geq H_1, S_1(t_1) \leq K_1, S_2(t_1) \geq K_2, \inf_{t_1 \leq t \leq t_2} S_2(t) \leq H_2, S_2(t_2) \geq K_3 \right) \tag{25}$$

End of Proposition 2.

Equipped with Proposition 2, one can value in closed form a two-colour step barrier option with one upward step and one downward step, by taking the relevant $P_{[.]}$ or $P_{[.]}^{(I)}$ probability along with the pairs $\left(\mu_1^{(Q)}, \mu_2^{(Q)}\right)$ and $\left(\mu_1^{(P_2)}, \mu_2^{(P_2)}\right)$ defined in (18), as explained in Section 2.1. Table 2 reports the prices of a few down-and-up two-colour knock-out put options by implementing Proposition 2 to obtain exact analytical values and by computing three successive conditional Monte Carlo approximations in the same way, as in Table 1.

Table 2. Two-colour down-and-up knock-out put.

	$\rho_{1.2} = -0.6$	$\rho_{1.2} = -0.2$	$\rho_{1.2} = 0.2$	$\rho_{1.2} = 0.6$
$\sigma_1 = \sigma_2 = 20\%$ $H_1 = 85, H_2 = 115$	4.299 (4.287, 4.305, 4.297)	3.293 (3.318, 3.286, 3.292)	2.307 (2.282, 2.298, 2.305)	1.303 (1.284, 1.309, 1.304)
$\sigma_1 = \sigma_2 = 60\%$ $H_1 = 75, H_2 = 125$	10.221 (10.142, 10.228, 10.224)	7.610 (7.563, 7.597, 7.612)	5.212 (5.255, 5.204, 5.214)	2.862 (2.834, 2.854, 2.865)
$\sigma_1 = 20\%, \sigma_2 = 60\%$ $H_1 = 85, H_2 = 125$	13.377 (13.385, 13.391, 13.376)	10.496 (10.472, 10.482, 10.948)	7.735 (7.783, 7.717, 7.731)	4.857 (4.894, 4.866, 4.857)
$\sigma_1 = 60\%, \sigma_2 = 20\%$ $H_1 = 75, H_2 = 115$	3.344 (3.387, 3.358, 3.346)	2.401 (2.383, 2.413, 2.402)	1.545 (1.596, 1.530, 1.542)	0.748 (0.884, 0.787, 0.752)

In Table 2, the most salient feature is still the functional dependency of the option's price on the correlation between assets S_1 and S_2, but, this time, the direction is opposite to that in Table 1, i.e., the two-colour down-and-up knock-out put prices steadily decrease as $\rho_{1.2}$ goes from -60% to 60%. In a trader's perspective, one could sum up the argument by saying that two-colour rainbow barrier options are a bet on a positive correlation when both barriers are on the same side (upward or downward), while they are a bet on a negative correlation when the barriers stand on opposite sides (up-and-down or down-and-up).

The barrier effect also prevails over the volatility effect in Table 2. Overall, two-colour down-and-up knock-out puts display maximum values that are a little higher, and minimum values that are a little lower than two-colour up-and-up knock-out puts, although up-and-out barriers and down-and-out barriers are designed with the exact same distance to the spot prices of S_1 and S_2.

2.3. Valuation of Reverse Two-Colour Step Barrier Options

A two-colour rainbow barrier option is said to be reverse when the moneyness of the option is defined w.r.t. the first and former "colour" (i.e., asset S_1) instead of the second and last one (asset S_2): the option, so to speak, reverts back to asset one at expiry, hence the denomination. From a computational standpoint, this is not a trivial difference since it adds an additional dimension to the integral formulation of the problem. Let us define as $P_{[RRUU]}(\mu_1, \mu_2)$ the following cumulative joint distribution at the core of reverse rainbow option valuation:

$$P_{[RRUU]}^{(Rev)}(\mu_1, \mu_2) = P\left(\sup_{0 \le t \le t_1} S_1(t) \le H_1, S_1(t_1) \le K_1, S_2(t_1) \le K_2, \sup_{t_1 \le t \le t_2} S_2(t) \le H_2, S_1(t_2) \le K_3 \right) \quad (26)$$

where the acronym "[RRUU]" stands for "Reverse Rainbow Up and Up".

Then, Proposition 3 provides the exact value of $P_{[RRUU]}(\mu_1, \mu_2)$ in the form of a triple integral.

Proposition 3.

$$P_{[RRUU]}(\mu_1, \mu_2) = \frac{1}{(2\pi)^{3/2}\sigma_{2|1}\sigma_{3|1.2}\sigma_1^2\sigma_2 t_1\sqrt{t_2}} \int\limits_{-\infty}^{\min(k_1,h_1)} \int\limits_{-\infty}^{h_2} \int\limits_{-\infty}^{\min(k_2,h_2)} \varphi_2(x_1)\varphi_3(x_2,x_3) \quad (27)$$

$$e^{-\frac{1}{2}\left(\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)^2 - \frac{1}{2\sigma_{2|1}^2}\left(\frac{x_2-\mu_2 t_1}{\sigma_2\sqrt{t_1}}-\theta_{1.2}\left(\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)\right)^2 - \frac{1}{2\sigma_{3|1.2}^2}\left(\frac{x_3-\mu_2 t_2}{\sigma_2\sqrt{t_2}}-\theta_{1.3}\left(\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)-\frac{\theta_{2.3|1}}{\sigma_{2|1}}\left(\frac{x_2-\mu_2 t_1}{\sigma_2\sqrt{t_1}}-\theta_{1.2}\left(\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)\right)\right)}$$

$$N\left[\frac{1}{\sigma_{4|1.2.3}}\left(\begin{array}{l}\frac{k_3-\mu_1 t_2}{\sigma_1\sqrt{t_2}}-\theta_{1.4}\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}-\frac{\theta_{2.4|1}}{\sigma_{2|1}}\left(\frac{x_2-\mu_2 t_1}{\sigma_2\sqrt{t_1}}-\theta_{1.2}\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)\\ -\frac{\theta_{3.4|1.2}}{\sigma_{3|1.2}}\left(\frac{x_3-\mu_2 t_2}{\sigma_2\sqrt{t_2}}-\theta_{1.3}\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}-\frac{\theta_{2.3|1}}{\sigma_{2|1}}\left(\frac{x_2-\mu_2 t_1}{\sigma_2\sqrt{t_1}}-\theta_{1.2}\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)\right)\end{array}\right)\right] dx_3 dx_2 dx_1$$

where:

$$\theta_{2.4|1} = \frac{\theta_{2.4}-\theta_{1.2}\theta_{1.4}}{\sqrt{1-\theta_{1.2}^2}}, \theta_{3.4|1.2} = \frac{\theta_{3.4}-\theta_{1.3}\theta_{1.4}-\theta_{2.3|1}\theta_{2.4|1}}{\sigma_{3|1.2}}, \sigma_{4|1.2.3} = \sqrt{1-\theta_{1.4}^2-\theta_{2.4|1}^2-\theta_{3.4|1.2}^2} \quad (28)$$

-   $N[.]$ *is the univariate standard normal cumulative distribution function;*
-   *the functions φ_2 and φ_3 are defined by (80) and (81) in Section 3.*

    All the other notations in Proposition 3 have been previously defined.

Remark 1. *Other types of reverse two-colour knock-out or knock-in barrier probability distributions are handled similarly by modifying the upper bounds of the integral and, possibly, the φ_i functions, according to the considered combination of events.*

Remark 2. $\theta_{2.4|1}$ *is the partial correlation between $X_2(t_1)$ and $X_1(t_2)$ conditional on $X_1(t_1)$, while $\theta_{3.4|1.2}$ is the partial correlation between $X_2(t_2)$ and $X_1(t_2)$ conditional on $X_1(t_1)$ and $X_2(t_1)$, and $\sigma_{4|1.2.3}$ is the conditional standard deviation of $X_1(t_2)$ given $X_1(t_1)$, $X_2(t_1)$ and $X_2(t_2)$.*
    End of Proposition 3.

The application of Proposition 3 to value a reverse two-colour step barrier option is now discussed. The no-arbitrage price of a reverse two-colour rainbow up-and-up knock-out put, denoted by $V_{[RRUU]}$, is given by:

$$V_{[RRUU]}^{(R)} = e^{-rt_2}\left(E_Q\left[K_3 \mathbf{1}_{\{A\}}-S_1(t_2)\mathbf{1}_{\{A\}}\right]\right) = e^{-rt_2}K_3 P_{[RRUU]}^{(R)}\left(\mu_1^{(Q)}, \mu_2^{(Q)}\right) - S_2(0)P_{[RRUU]}^{(R)}\left(\mu_1^{(P_1)}, \mu_2^{(P_1)}\right) \quad (29)$$

where

$$\mu_1^{(P_1)} = r + \frac{\sigma_1^2}{2}, \ \mu_2^{(P_1)} = r - \frac{\sigma_2^2}{2} + \sigma_1\sigma_2\rho_{1.2} \quad (30)$$

-   A *is the set constructed by the intersection of elements of the $\sigma-$algebra generated by the pair of processes $(S_1(t), S_2(t))$ that characterises the probability $P_{[RRUU]}(\mu_1, \mu_2)$ as given by the arguments of the probability operator in (26);*
-   P_1 *is the measure under which $B_1(t) - \sigma_1 t$ is a standard Brownian motion.*

However, it is less easy to evaluate Proposition 3 than to evaluate Proposition 1 and Proposition 2. The problem at hand has two "nice" features from the standpoint of numerical integration: first, the dimension, equal to 3, is moderate; second, the integrand is continuous. The snag is the large number of parameters in each evaluation of the integrand in a quadrature process, especially the various conditional standard deviations at the denominators of the fractions, that may hinder fast convergence when they take on absolute values that become smaller and smaller. That is why it is recommended to use a subregion adaptive algorithm of numerical integration, as explained by Berntsen et al.

(1991), that adapts the number of integrand evaluations in each subregion according to the rate of change of the integrand. Although more time-consuming than a fixed degree rule, it is more accurate to control the approximation error, as the subdivision of the integration domain stops only when the sum of the local error deterministic estimates becomes smaller than some prespecified requested accuracy. Adaptive integration can be enhanced by a Kronrod rule to reduce the number of required iterations (see, e.g., Davis and Rabinowitz 2007). These techniques are widely used in numerical integration, and it is easy to find available code or built-in functions in the usual scientific computing software.

2.4. Valuation of Two-Colour Outside Step Barrier Options

In this section, a third correlated asset S_3 is introduced, w.r.t. which the option's moneyness is measured at expiry, while the assets S_1 and S_2 serve exclusively as the underlyings w.r.t. which barrier crossing is monitored. This is an important extension, as outside barrier options allow to manage volatility more consistently than standard (non-outside) barrier options, as explained, e.g., by Das (2006).

Let us consider a third asset S_3 with the following differential:

$$dS_3(t) = v_3 S_3(t)dt + \sigma_3 S_3(t)dB_3(t) \tag{31}$$

The instantaneous pairwise correlations between the Brownian motions B'_is are denoted as $\rho_{i.j}$.

The objective is to compute the probabilities $p_m(\mu_1, \mu_2, \mu_3), m \in \{1, 2, 3, 4\}$ defined by:

$$p_1(\mu_1, \mu_2, \mu_3) = P \left(\begin{array}{l} \sup_{0 \leq t \leq t_1} S_1(t) \leq H_1, S_1(t_1) \leq K_1, S_2(t_1) \leq H_2, \\ \sup_{t_1 \leq t \leq t_2} S_2(t) \leq H_2, S_2(t_2) \leq K_2, S_3(t_2) \leq K_3 \end{array} \right) \tag{32}$$

$$p_2(\mu_1, \mu_2, \mu_3) = P \left(\begin{array}{l} \inf_{0 \leq t \leq t_1} S_1(t) \geq H_1, S_1(t_1) \geq K_1, S_2(t_1) \geq H_2, \\ \inf_{t_1 \leq t \leq t_2} S_2(t) \geq H_2, S_2(t_2) \geq K_2, S_3(t_2) \geq K_3 \end{array} \right) \tag{33}$$

$$p_3(\mu_1, \mu_2, \mu_3) = P \left(\begin{array}{l} \sup_{0 \leq t \leq t_1} S_1(t) \leq H_1, S_1(t_1) \leq K_1, S_2(t_1) \geq H_2, \\ \inf_{t_1 \leq t \leq t_2} S_2(t) \geq H_2, S_2(t_2) \geq K_2, S_3(t_2) \geq K_3 \end{array} \right) \tag{34}$$

$$p_4(\mu_1, \mu_2, \mu_3) = P \left(\begin{array}{l} \inf_{0 \leq t \leq t_1} S_1(t) \geq H_1, S_1(t_1) \geq K_1, S_2(t_1) \leq H_2, \\ \sup_{t_1 \leq t \leq t_2} S_2(t) \leq H_2, S_2(t_2) \leq K_2, S_3(t_2) \leq K_3 \end{array} \right) \tag{35}$$

Let $\mathrm{x} = [c_1, c_2, c_3, c_4, c_5]$ be a vector of five coordinates where each $c_i \in\,]-1, 1[$, $\forall i \in \{1, \ldots, 5\}$.

Let the function $\Psi_4[b_1, b_2, b_3, b_4; \mathrm{x}], \forall b_1, b_2, b_3, b_4 \in \mathbb{R}$, be defined by:

$$\Psi_4[b_1, b_2, b_3, b_4; \mathrm{x}]$$

$$= \int_{x_1=-\infty}^{b_1} \int_{x_2=-\infty}^{\frac{b_2 - c_1 x_1}{\sqrt{1-c_1^2}}} \int_{x_3=-\infty}^{\frac{b_3 - c_4 x_2\sqrt{1-c_1^2} - c_4 c_1 x_1}{\sqrt{1-c_2^2}}} \frac{1}{(2\pi)^{3/2}} \exp\left(-\frac{x_1^2}{2} - \frac{x_2^2}{2} - \frac{x_3^2}{2} \right) \tag{36}$$

$$N\left[\frac{b_4 - \frac{c_5 - c_2 c_3}{1 - c_2^2}\left(x_3\sqrt{1 - c_4^2} + x_2 c_4\sqrt{1 - c_1^2}\right) - x_1\left(c_3 + \frac{c_5 - c_2 c_3}{1 - c_2^2}(c_1 c_4 - c_2)\right)}{\sqrt{1 - c_3^2 - \left(\frac{c_5 - c_2 c_3}{1 - c_2^2}\right)^2}}\right] dx_3 dx_2 dx_1$$

The following Proposition 4 combines all the probabilities defined in (32)–(35) into a single formula.

Proposition 4. *The exact values of the probabilities* $p_m(\mu_1, \mu_2, \mu_3), m \in \{1, 2, 3, 4\}$, *written in shorter notation as* p_m, *are given by:*

$$p_m = \Psi_4\left[\delta_1\left(\frac{G_1(k_1, h_1) - \mu_1 t_1}{\sigma_1\sqrt{t_1}}\right), \delta_2\left(\frac{h_2 - \mu_2 t_1}{\sigma_2\sqrt{t_1}}\right), \delta_2\left(\frac{G_2(k_2, h_2) - \mu_2 t_2}{\sigma_2\sqrt{t_2}}\right), \delta_2\left(\frac{k_3 - \mu_3 t_2}{\sigma_3\sqrt{t_2}}\right); x_1\right] \tag{37}$$

$$-\exp\left(\frac{2\mu_1 h_1}{\sigma_1^2}\right) \times \Psi_4\left[\begin{array}{l} \delta_1\left(\frac{G_1(k_1, h_1) - 2h_1 - \mu_1 t_1}{\sigma_1\sqrt{t_1}}\right), \delta_2\left(\frac{h_2 - \mu_2 t_1}{\sigma_2\sqrt{t_1}} - \theta_{1.2}\frac{2h_1}{\sigma_1\sqrt{t_1}}\right), \\ \delta_2\left(\frac{G_2(k_2, h_2) - \mu_2 t_2}{\sigma_2\sqrt{t_2}} - \theta_{1.2}\frac{2h_1}{\sigma_1\sqrt{t_2}}\right), \\ \delta_2\left(\frac{k_3 - \mu_3 t_2}{\sigma_3\sqrt{t_2}} - \theta_{1.4}\frac{2h_1}{\sigma_1\sqrt{t_1}} - \theta_{3.4|1}\left(\theta_{1.2}\frac{2h_1}{\sigma_1\sqrt{t_2}} - \theta_{1.3}\frac{2h_1}{\sigma_1\sqrt{t_1}}\right)\right); x_1 \end{array}\right] \tag{38}$$

$$-\exp\left(\frac{2\mu_2 h_2}{\sigma_2^2}\right) \times \Psi_4\left[\begin{array}{l} \delta_1\left(\frac{G_1(k_1, h_1) - \mu_1 t_1}{\sigma_1\sqrt{t_1}} + \theta_{1.2}\frac{2\mu_2 t_1}{\sigma_2\sqrt{t_1}}\right), \delta_2\left(\frac{h_2 + \mu_2 t_1}{\sigma_2\sqrt{t_1}}\right), \\ \delta_2\left(\frac{G_2(k_2, h_2) - 2h_2 - \mu_2 t_2}{\sigma_2\sqrt{t_2}}\right), \\ \delta_2\left(\frac{k_3 - \mu_3 t_2}{\sigma_3\sqrt{t_2}} - \theta_{3.4|1}\left(\frac{2h_2}{\sigma_2\sqrt{t_2}} + \theta_{1.2}\theta_{1.3}\frac{2\mu_2 t_1}{\sigma_2\sqrt{t_1}}\right) + \theta_{1.2}\theta_{1.4}\frac{2\mu_2 t_1}{\sigma_2\sqrt{t_1}}\right); x_2 \end{array}\right] \tag{39}$$

$$+ \exp\left(\left(\frac{2\mu_1}{\sigma_1^2} - \frac{4\mu_2\rho_{1.2}}{\sigma_1\sigma_2}\right)h_1 + \frac{2\mu_2 h_2}{\sigma_2^2}\right)$$

$$\times \Psi_4\left[\begin{array}{l} \delta_1\left(\frac{G_1(k_1, h_1) - 2h_1 - \mu_1 t_1}{\sigma_1\sqrt{t_1}} + \theta_{1.2}\frac{2\mu_2 t_1}{\sigma_2\sqrt{t_1}}\right), \\ \delta_2\left(\frac{h_2 + \mu_2 t_1}{\sigma_2\sqrt{t_1}} - \theta_{1.2}\frac{2h_1}{\sigma_1\sqrt{t_1}}\right), \\ \delta_2\left(\frac{G_2(k_2, h_2) - 2h_2 - \mu_2 t_2}{\sigma_2\sqrt{t_2}} + \theta_{1.2}\frac{2h_1}{\sigma_1\sqrt{t_2}}\right), \\ \delta_2\left(\begin{array}{l}\frac{k_3 - \mu_3 t_2}{\sigma_3\sqrt{t_2}} + \theta_{1.4}\left(\theta_{1.2}\frac{2\mu_2\sqrt{t_1}}{\sigma_2} - \frac{2h_1}{\sigma_1\sqrt{t_1}}\right) \\ -\theta_{3.4|1}\left(\frac{2h_2}{\sigma_2\sqrt{t_2}} - \theta_{1.2}\frac{2h_1}{\sigma_1\sqrt{t_2}} + \theta_{1.3}\left(\theta_{1.2}\frac{2\mu_2\sqrt{t_1}}{\sigma_2} - \frac{2h_1}{\sigma_1\sqrt{t_1}}\right)\right)\end{array}\right); x_2 \end{array}\right] \tag{40}$$

where k_1, k_2, h_1, h_2 *are as in Proposition 2,* $k_3 = \ln\left(\frac{K_3}{S_3(0)}\right)$, *and we have:*

$$\theta_{1.2} = \rho_{1.2}, \theta_{1.3} = \sqrt{\frac{t_1}{t_2}}\rho_{1.2}, \theta_{1.4} = \sqrt{\frac{t_1}{t_2}}\rho_{1.3}, \theta_{2.3} = \sqrt{\frac{t_1}{t_2}}, \theta_{3.4} = \rho_{2.3}, \ \theta_{3.4|1} = \frac{\theta_{3.4} - \theta_{1.3}\theta_{1.4}}{\sqrt{1 - \theta_{1.3}^2}} \tag{41}$$

$$\delta_1 = \begin{cases} 1 \text{ if } p_m = p_1 \text{ or } p_m = p_3 \\ -1 \text{ if } p_m = p_2 \text{ or } p_m = p_4 \end{cases}, \ \delta_2 = \begin{cases} 1 \text{ if } p_m = p_1 \text{ or } p_m = p_4 \\ -1 \text{ if } p_m = p_2 \text{ or } p_m = p_3 \end{cases} \tag{42}$$

$$G_1(.,.) = \begin{cases} \min(.,.) \text{ if } p_m = p_1 \text{ or } p_m = p_3 \\ \max(.,.) \text{ if } p_m = p_2 \text{ or } p_m = p_4 \end{cases}, \ G_2(.,.) = \begin{cases} \max(.,.) \text{ if } p_m = p_1 \text{ or } p_m = p_3 \\ \min(.,.) \text{ if } p_m = p_2 \text{ or } p_m = p_4 \end{cases} \tag{43}$$

$$x_1 = \begin{cases} [\theta_{1.2}, \theta_{1.3}, \theta_{1.4}, \theta_{2.3}, \theta_{3.4}] \text{ if } p_m = p_1 \text{ or } p_m = p_2 \\ [-\theta_{1.2}, -\theta_{1.3}, -\theta_{1.4}, \theta_{2.3}, \theta_{3.4}] \text{ if } p_m = p_3 \text{ or } p_m = p_4 \end{cases} \tag{44}$$

$$x_2 = \begin{cases} [\theta_{1.2}, \theta_{1.3}, \theta_{1.4}, -\theta_{2.3}, \theta_{3.4}] \text{ if } p_m = p_1 \text{ or } p_m = p_2 \\ [-\theta_{1.2}, -\theta_{1.3}, -\theta_{1.4}, -\theta_{2.3}, \theta_{3.4}] \text{ if } p_m = p_3 \text{ or } p_m = p_4 \end{cases} \tag{45}$$

Corollary 1. *The corresponding knock-in probabilities can be inferred in the same way as in Proposition 1 and Proposition 2. Let the probabilities* $p_m^{(I)}(\mu_1, \mu_2, \mu_3), m \in \{1, 2, 3, 4\}$ *be defined by:*

$$p_1^{(I)}(\mu_1, \mu_2, \mu_3) = P \left(\begin{array}{l} \sup\limits_{0 \le t \le t_1} S_1(t) \ge H_1, S_1(t_1) \le K_1, S_2(t_1) \le H_2, \\ \sup\limits_{t_1 \le t \le t_2} S_2(t) \ge H_2, S_2(t_2) \le K_2, S_3(t_2) \le K_3 \end{array} \right) \tag{46}$$

$$p_2^{(I)}(\mu_1, \mu_2, \mu_3) = P \left(\begin{array}{l} \inf\limits_{0 \le t \le t_1} S_1(t) \le H_1, S_1(t_1) \ge K_1, S_2(t_1) \ge H_2, \\ \inf\limits_{t_1 \le t \le t_2} S_2(t) \le H_2, S_2(t_2) \ge K_2, S_3(t_2) \ge K_3 \end{array} \right) \tag{47}$$

$$p_3^{(I)}(\mu_1, \mu_2, \mu_3) = P \left(\begin{array}{l} \sup\limits_{0 \le t \le t_1} S_1(t) \ge H_1, S_1(t_1) \le K_1, S_2(t_1) \ge H_2, \\ \inf\limits_{t_1 \le t \le t_2} S_2(t) \le H_2, S_2(t_2) \ge K_2, S_3(t_2) \ge K_3 \end{array} \right) \tag{48}$$

$$p_4^{(I)}(\mu_1, \mu_2, \mu_3) = P \left(\begin{array}{l} \inf\limits_{0 \le t \le t_1} S_1(t) \le H_1, S_1(t_1) \ge K_1, S_2(t_1) \le H_2, \\ \sup\limits_{t_1 \le t \le t_2} S_2(t) \ge H_2, S_2(t_2) \le K_2, S_3(t_2) \le K_3 \end{array} \right) \tag{49}$$

Then, $p_m^{(I)}(\mu_1, \mu_2, \mu_3)$ *is given by (40).*

Corollary 2. *It suffices to substitute each argument* $\theta_{3.4}$ *in each* $\Psi_4[., ., ., ., ; ., ., ., ., .]$ *function of Proposition 4 by* $\rho_{2.3}\sqrt{\frac{t_2}{t_3}}$, $\forall t_3 \ge t_2$, *to obtain an exact formula for the early-ending variant of* $p_m(\mu_1, \mu_2, \mu_3)$.
End of Proposition 4.

Equipped with Proposition 4, one can value in closed form a two-colour outside step barrier option. More precisely, the value of a two-colour outside up-and-out put, denoted by $V_{[ORUU]}$, is given by:

$$V_{[ORUU]} = e^{-rt_2} \left(E_Q \left[K_3 \mathbf{1}_{\{A\}} - S_3(t_2) \mathbf{1}_{\{A\}} \right] \right) \tag{50}$$

where A is the set constructed by the intersection of elements of the $\sigma-$algebra generated by the pair of processes $(S_1(t), S_2(t))$ that characterises the probability $p_1(\mu_1, \mu_2)$ as given by the arguments of the probability operator in (32), and the acronym "[ORUU]" stands for "Outside Rainbow Up and Up".

Using the following orthogonal decomposition of Brownian motion $B_3(t)$:

$$B_3(t) = \rho_{1.3}W_1(t) + \rho_{2.3|1}W_2(t) + \sigma_{3|1.2}W_3(t) \tag{51}$$

where:

$$\rho_{2.3|1} = \frac{\rho_{2.3} - \rho_{1.2}\rho_{1.3}}{\sqrt{1 - \rho_{1.2}^2}}, \; \sigma_{3|1.2} = \sqrt{1 - \rho_{1.3}^2 - \rho_{2.3|1}^2} \tag{52}$$

and $(W_1(t), W_2(t), W_3(t))$ is a basis of three independent Brownian motions (Guillaume 2018), the multidimensional Cameron-Martin-Girsanov theorem yields:

$$V_{[ORUU]} = e^{-rt_2}K_3 p_1 \left(\mu_1^{(Q)}, \mu_2^{(Q)}, \mu_3^{(Q)} \right) - S_3(0) p_1 \left(\mu_1^{(P_3)}, \mu_2^{(P_3)}, \mu_3^{(P_3)} \right) \tag{53}$$

where:

$$\mu_1^{(P_3)} = r - \frac{\sigma_1^2}{2} + \sigma_1\sigma_3\rho_{1.3}, \; \mu_2^{(P_3)} = r - \frac{\sigma_2^2}{2} + \sigma_2\sigma_3\rho_{2.3}, \; \mu_3^{(P_3)} = r + \frac{\sigma_3^2}{2} \tag{54}$$

The measure P_3 is the measure under which $B_1(t) - \sigma_3\rho_{1.3}t$, $B_2(t) - \sigma_3\rho_{2.3|1}t$ and $B_3(t) - \sigma_3\sigma_{3|1.2}t$ are three independent standard Brownian motions.

A simple and robust numerical evaluation of the function Ψ_4 consists in selecting an appropriate cutoff value for the negative infinity lower bounds and then applying a fixed-degree quadrature rule. Given the smoothness of the integrand, even a low-degree rule will perform well. Table 3 provides the prices of a few two-colour outside up-and-down knock-out call options for various levels of the volatility and correlation parameters of the underlying assets S_1, S_2, and S_3, and different values of the knock-out barriers. The parameters $S_i(0)$, K_i, t_1, t_2, and r are identical as those as in Tables 1 and 2. In each cell, the first reported value is the exact analytical price, as obtained by implementing Proposition 4 by means of a classical 16-point Gauss–Legendre quadrature, while the numbers in the brackets are three successive Monte Carlo approximations, as explained in Section 2.1.

From a purely numerical standpoint, the pattern of convergence of conditional Monte Carlo approximations to the analytical values is as clear in Table 3 as in Tables 1 and 2. This illustrates the robustness of our numerical integration scheme for the Ψ_4 function. The efficiency gap between Monte Carlo pricing and analytical pricing is even more pronounced than for non-outside rainbow step barrier options due to the presence of an additional stochastic process to simulate: the average computational time required by simulation is 42 s, whereas the evaluation of the analytical formula based on Proposition 4 only takes a few tenths of a second.

Table 3. Outside two-colour up-and-down knock-out call.

	$\rho_{1.2} = -0.6$, $\rho_{1.3} = \rho_{2.3} = -0.4$	$\rho_{1.2} = -0.6$, $\rho_{1.3} = \rho_{2.3} = 0.4$	$\rho_{1.2} = 0.6$ $\rho_{1.3} = \rho_{2.3} = 0.4$	$\rho_{1.2} = 0.6$ $\rho_{1.3} = \rho_{2.3} = -0.4$
$\sigma_1 = \sigma_2 = 20\%$ $\sigma_3 = 20\%$ $H_1 = 115, H_2 = 85$	2.378 (2.452, 2.361, 2.375)	2.772 (2.914, 2.812, 2.779)	1.717 (1.585, 1.731, 1.720)	1.182 (1.193, 1.178, 1.180)
$\sigma_1 = \sigma_2 = 20\%$ $\sigma_3 = 60\%$ $H_1 = 115, H_2 = 85$	5.769 (5.728, 5.781, 5.764)	7.053 (7.137, 7.036, 7.054)	4.522 (4.534, 4.541, 4.524)	2.897 (2.852, 2.923, 2.893)
$\sigma_1 = \sigma_2 = 60\%$ $\sigma_3 = 20\%$ $H_1 = 125, H_2 = 75$	1.627 (1.592, 1.614, 1.628)	2.783 (2.848, 2.767, 2.788)	1.351 (1.320, 1.365, 1.353)	0.849 (0.915, 0.828, 0.842)
$\sigma_1 = \sigma_2 = 60\%$ $\sigma_3 = 60\%$ $H_1 = 125, H_2 = 75$	3.864 (3.814, 3.872, 3.865)	7.554 (7.518, 7.535, 7.558)	3.823 (3.856, 3.829, 3.827)	2.076 (2.011, 2.091, 2.072)
$\sigma_1 = 20\%, \sigma_2 = 60\%$ $\sigma_3 = 20\%$ $H_1 = 115, H_2 = 75$	1.534 (1.502, 1.526, 1.535)	2.573 (2.495, 2.556, 2.577)	1.188 (1.207, 1.179, 1.185)	0.621 (0.774, 0.684, 0.613)
$\sigma_1 = 20\%, \sigma_2 = 60\%$ $\sigma_3 = 60\%$ $H_1 = 115, H_2 = 75$	3.629 (3.787, 3.662, 3.621)	6.697 (6.724, 6.684, 6.692)	3.217 (3.051, 3.252, 3.221)	1.518 (1.586, 1.476, 1.511)
$\sigma_1 = 60\%, \sigma_2 = 20\%$ $\sigma_3 = 20\%$ $H_1 = 125, H_2 = 85$	2.572 (2.734, 2.548, 2.567)	2.989 (3.125, 3.016, 2.994)	1.931 (2.071, 1.965, 1.938)	1.496 (1.634, 1.454, 1.489)
$\sigma_1 = 60\%, \sigma_2 = 20\%$ $\sigma_3 = 60\%$ $H_1 = 125, H_2 = 85$	6.248 (6.304, 6.237, 6.241)	7.953 (8.060, 7.984, 7.961)	5.323 (5.212, 5.348, 5.324)	3.707 (3.569, 3.726, 3.702)

From a financial point of view, the prices in Table 3 display a very different pattern from those in Tables 1 and 2. With regard to correlation, the highest option values attained are when the correlation between S_1 and S_2 is negative and the correlation between S_3 and both S_1 and S_2 is positive. The lowest option values are when the correlation between S_1 and S_2 is positive and the correlation between S_3 and both S_1 and S_2 is negative. On average across all volatilities and barrier levels in Table 3, options are approximately three

times more expensive under the former correlation structure than under the latter one. In terms of volatility, the highest option values attained are when the volatility of asset S_3 is high. This remains true under very different combinations of values for all the other parameters (volatilities of S_1 and S_2, barrier levels and correlation structure). Such an observation highlights the prominent role of the volatility of the asset chosen to determine the moneyness of the option at expiry. In particular, the value of a rainbow outside step barrier option is a monotonically increasing function of σ_3, whereas the value of a rainbow step barrier option is not a monotonically increasing function of σ_2, just like the value of a reverse rainbow step barrier option is not a monotonically increasing function of σ_1. This is because a rainbow outside step barrier option allows to make a clear distinction between the functions of each underlying asset: two of them, S_1 and S_2, are only concerned with barrier crossing during the option life, and the third one, S_3, is only concerned with moneyness testing at the option expiry. That distinction is impossible to make when it comes to non-outside rainbow step barrier options, so that the impact of volatility becomes ambiguous and difficult to handle. It should be emphasised that, for the vast majority of parameters, the sensitivities of the rainbow outside step knock-out barrier options to σ_1 and σ_2 is negative, reflecting an increased risk of being deactivated before expiry. Only for quite specific correlation structures between the underlying assets and quite specific combinations of barrier values can these sensitivities be positive. A major advantage of closed form formulae such as those derived in this article is precisely to allow measurement of such sensitivities with high precision by mere differentiation of the formulae w.r.t. the relevant parameters.

One more noticeable difference in the reported numerical results between outside and non-outside rainbow step barrier options is that the volatility effect prevails over the barrier effect in Table 3, in contrast to Tables 1 and 2. Indeed, in row 2 of Table 3, tight barrier levels do not preclude relatively high option prices thanks to the volatility of asset S_3 set at 60%. Likewise, in row 3 of Table 3, wider barrier levels do not preclude relatively low option prices due to the volatility of asset S_3 set at only 20%.

2.5. Valuation of Two-Sided, Two-Colour Step Barrier Options

In this section, a two-sided barrier is introduced in each time interval, i.e., the valuation of rainbow step double barrier options is handled.

Let H_1 and H_2 denote an upward and a downward barrier, respectively, on the time interval $[t_0 = 0, t_1]$. Similarly, H_3 and H_4 represent an upward and a downward barrier, respectively, on the time interval $[t_1, t_2]$. As in the previous sections, barrier crossing is monitored w.r.t. process S_1 following Equation (1) on $[t_0 = 0, t_1]$ and w.r.t. process S_2 following Equation (2) on $[t_1, t_2]$. Our objective now is to find the value of the joint cumulative distribution function $P_{[\text{RDKO}]}(\mu_1, \mu_2)$ defined by:

$$P_{[\text{RDKO}]}(\mu_1, \mu_2) \tag{55}$$

$$= P \left(\begin{array}{c} \sup\limits_{0 \leq t \leq t_1} S_1(t) \leq H_1, \inf\limits_{0 \leq t \leq t_1} S_1(t) \geq H_2, S_1(t_1) \leq \min(K_1, H_1), \\ \sup\limits_{t_1 \leq t \leq t_2} S_2(t) \leq H_3, \inf\limits_{t_1 \leq t \leq t_2} S_2(t) \geq H_4, S_2(t_2) \leq \min(H_3, K_2) \end{array} \right)$$

where the acronym "[RDKO]" stands for "Rainbow Double Knock Out".

The main result of Section 2.5 is given by the following Proposition 5.

Proposition 5. *The exact value of* $\mathrm{P}_{[\mathrm{RDKO}]}(\mu_1, \mu_2)$ *is given by:*

$$\mathrm{P}_{[\mathrm{RDKO}]}(\mu_1, \mu_2) = \sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \exp\left(\frac{2\mu_1}{\sigma_1^2} n_1 a_1 + \frac{2\mu_2}{\sigma_2^2} n_2 a_2 \right)$$

$$\{N_3[A_1(\min(k_1,h_1)), A_2(h_3), A_3(\min(h_3,k_2)); x_1] - N_3\left[\begin{array}{l} A_1(h_2), A_2(h_3), \\ A_3(\min(h_3,k_2)); x_1 \end{array} \right]$$

$$-N_3[A_1(\min(k_1,h_1)), A_2(h_4), A_3(\min(h_3,k_2)); x_1] + N_3\left[\begin{array}{l} A_1(h_2), A_2(h_4), \\ A_3(\min(h_3,k_2)); x_1 \end{array} \right]$$

$$-N_3[A_1(\min(k_1,h_1)), A_2(h_3), A_3(h_4); x_1] + N_3[A_1(h_2), A_2(h_3), A_3(h_4); x_1]$$

$$+N_3[A_1(\min(k_1,h_1)), A_2(h_4), A_3(h_4); x_1] - N_3[A_1(h_2), A_2(h_4), A_3(h_4); x_1]\}$$

(56)

$$-\sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \exp\left(n_1 a_1 \left(\frac{2\mu_1}{\sigma_1^2} - \frac{4\rho_{1.2}\mu_2}{\sigma_1\sigma_2} \right) + \frac{2\mu_2}{\sigma_2^2}(h_4 - n_2 a_2) \right)$$

$$\{N_3[A_4(\min(k_1,h_1)), A_5(h_3), A_6(\min(h_3,k_2)); x_2] - N_3\left[\begin{array}{l} A_4(h_2), A_5(h_3), \\ A_6(\min(h_3,k_2)); x_2 \end{array} \right]$$

$$-N_3[A_4(\min(k_1,h_1)), A_5(h_4), A_6(\min(h_3,k_2)); x_2] + N_3\left[\begin{array}{l} A_4(h_2), A_5(h_4), \\ A_6(\min(h_3,k_2)); x_2 \end{array} \right]$$

$$-N_3[A_4(\min(k_1,h_1)), A_5(h_3), A_6(h_4); x_2] + N_3[A_4(h_2), A_5(h_3), A_6(h_4); x_2]$$

$$+N_3[A_4(\min(k_1,h_1)), A_5(h_4), A_6(h_4); x_2] - N_3[A_4(h_2), A_5(h_4), A_6(h_4); x_2]\}$$

(57)

$$-\sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \exp\left(\frac{2\mu_1}{\sigma_1^2}(h_2 - n_1 a_1) + \frac{2\mu_2}{\sigma_2^2} n_2(h_3 - a_2) \right)$$

$$\{N_3[A_7(\min(k_1,h_1)), A_8(h_3), A_9(\min(h_3,k_2)); x_1] - N_3\left[\begin{array}{l} A_7(h_2), A_8(h_3), \\ A_9(\min(h_3,k_2)); x_1 \end{array} \right]$$

$$-N_3[A_7(\min(k_1,h_1)), A_8(h_4), A_9(\min(h_3,k_2)); x_1] + N_3\left[\begin{array}{l} A_7(h_2), A_8(h_4), \\ A_9(\min(h_3,k_2)); x_1 \end{array} \right]$$

$$-N_3[A_7(\min(k_1,h_1)), A_8(h_3), A_9(h_4); x_1] + N_3[A_7(h_2), A_8(h_3), A_9(h_4); x_1]$$

$$+N_3[A_7(\min(k_1,h_1)), A_8(h_4), A_9(d_2); x_1] - N_3[A_7(h_2), A_8(h_4), A_9(d_2); x_1]\}$$

(58)

$$+\sum_{n_1=-\infty}^{\infty} \sum_{n_2=-\infty}^{\infty} \exp\left((h_2 - n_1 a_1)\left(\frac{2\mu_1}{\sigma_1^2} - \frac{4\rho_{1.2}\mu_2}{\sigma_1\sigma_2} \right) + \frac{2\mu_2}{\sigma_2^2}(h_4 - n_2 a_2) \right)$$

$$\{N_3[A_{10}(\min(k_1,h_1)), A_{11}(h_3), A_{12}(\min(h_3,k_2)); x_2] - N_3\left[\begin{array}{l} A_{10}(h_2), A_{11}(h_3), \\ A_{12}(\min(h_3,k_2)); x_2 \end{array} \right]$$

$$-N_3[A_{10}(\min(k_1,h_1)), A_{11}(h_4), A_{12}(\min(h_3,k_2)); x_2] + N_3\left[\begin{array}{l} A_{10}(h_2), A_{11}(h_4), \\ A_{12}(\min(h_3,k_2)); x_2 \end{array} \right]$$

$$-N_3[A_{10}(\min(k_1,h_1)), A_{11}(h_3), A_{12}(h_4); x_2] + N_3[A_{10}(h_2), A_{11}(h_3), A_3(h_4); x_2]$$

$$+N_3[A_{10}(\min(k_1,h_1)), A_{11}(h_4), A_{12}(h_4); x_2] - N_3[A_{10}(h_2), A_{11}(h_4), A_{12}(h_4); x_2]\}$$

(59)

where:

$$- h_1 = \ln\left(\frac{H_1}{S_1(0)} \right) > 0, h_2 = \ln\left(\frac{H_2}{S_1(0)} \right) < 0, h_3 = \ln\left(\frac{H_3}{S_2(0)} \right), h_4 = \ln\left(\frac{H_4}{S_2(0)} \right) \quad (60)$$

$$- a_1 = h_1 - h_2, a_2 = h_3 - h_4 \quad (61)$$

$$- A_1(x) = \frac{x - 2n_1 a_1 - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}, A_2(x) = \frac{x - \mu_2 t_1}{\sigma_2 \sqrt{t_1}} - \frac{2\rho_{1.2} n_1 a_1}{\sigma_1 \sqrt{t_1}} \quad (62)$$

$$- A_3(x) = \frac{x - 2n_2 a_2 - \mu_2 t_2}{\sigma_2 \sqrt{t_2}} - \frac{2\rho_{1.2} n_1 a_1}{\sigma_1 \sqrt{t_2}} \quad (63)$$

$$- A_4(x) = A_1(x) - \frac{2\rho_{1.2}\mu_2\sqrt{t_1}}{\sigma_2}, A_5(x) = \frac{x + \mu_2 t_1}{\sigma_2\sqrt{t_1}} - \frac{2\rho_{1.2}n_1 a_1}{\sigma_1\sqrt{t_1}} \tag{64}$$

$$- A_6(x) = \frac{x - 2h_4 + 2n_2 a_2 - \mu_2 t_2}{\sigma_2\sqrt{t_2}} + \frac{2\rho_{1.2}n_1 a_1}{\sigma_1\sqrt{t_2}} \tag{65}$$

$$- A_7(x) = \frac{x - 2h_2 + 2n_1 a_1 - \mu_1 t_1}{\sigma_1\sqrt{t_1}}, A_8(x) = \frac{x - \mu_2 t_1}{\sigma_2\sqrt{t_1}} - \frac{2\rho_{1.2}(h_2 - n_1 a_1)}{\sigma_1\sqrt{t_1}} \tag{66}$$

$$- A_9(x) = \frac{x - 2n_2 a_2 - \mu_2 t_2}{\sigma_2\sqrt{t_2}} - \frac{2\rho_{1.2}(h_2 - n_1 a_1)}{\sigma_1\sqrt{t_2}} \tag{67}$$

$$- A_{10}(x) = A_7(x) - \frac{2\rho_{1.2}\mu_2\sqrt{t_1}}{\sigma_2}, A_{11}(x) = \frac{x + \mu_2 t_1}{\sigma_2\sqrt{t_1}} - \frac{2\rho_{1.2}(h_2 - n_1 a_1)}{\sigma_1\sqrt{t_1}} \tag{68}$$

$$- A_{12}(x) = \frac{x - 2h_4 + 2n_2 a_2 - \mu_2 t_2}{\sigma_2\sqrt{t_2}} + \frac{2\rho_{1.2}(h_2 - n_1 a_1)}{\sigma_1\sqrt{t_2}} \tag{69}$$

$$- x_1 = \left\{\rho_{1.2}, \rho_{1.2}\sqrt{\frac{t_1}{t_2}}, \sqrt{\frac{t_1}{t_2}}\right\}, x_2 = \left\{\rho_{1.2}, -\rho_{1.2}\sqrt{\frac{t_1}{t_2}}, -\sqrt{\frac{t_1}{t_2}}\right\} \tag{70}$$

All other notations have been defined in the previous sections.
End of Proposition 5.

Pricing two-colour double knock-out barrier options can be achieved through the same changes of probability measures as those applicable to two-colour single knock-out barrier options, i.e., the value of a two-colour double knock-out put, denoted as $V_{[RDKO]}$, is given by:

$$V_{[RDKO]} = e^{-rt_2}K_3 P_{[RDKO]}\left(\mu_1^{(Q)}, \mu_2^{(Q)}\right) - S_2(0)P_{[RDKO]}\left(\mu_1^{(P_2)}, \mu_2^{(P_2)}\right) \tag{71}$$

where the parameters $\mu_i^{(Q)}$ and $\mu_i^{(P_2)}$ are given by Equation (18).

Table 4 provides the prices of a few two-colour knock-out double barrier puts for various levels of the volatility and correlation parameters of the underlying assets S_1 and S_2, and different values of the knock-out barriers. Expiry is 6 months and t_1 is one quarter of a year. The parameters $S_i(0)$, K_i, and r are identical to those in Tables 1–3. In each cell, the first number is the exact analytical value as derived from (71), while the numbers in the brackets are three successive Monte Carlo approximations, as explained in Section 2.1.

Table 4. Two-colour rainbow double knock-out put.

	$\rho_{1.2} = -0.6$	$\rho_{1.2} = -0.2$	$\rho_{1.2} = 0.2$	$\rho_{1.2} = 0.6$
$\sigma_1 = \sigma_2 = 15\%$ $H_1 = H_3 = 120$ $H_2 = H_4 = 80$	2.919 (2.985, 2.956, 2.916)	2.938 (2.792, 2.883, 2.932)	2.936 (2.974, 2.918, 2.931)	3.027 (3.191, 3.088, 3.032)
$\sigma_1 = \sigma_2 = 30\%$ $H_1 = H_3 = 130$ $H_2 = H_4 = 70$	4.370 (4.529, 4.281, 4.365)	5.791 (5.978, 5.697, 5.786)	4.427 (4.196, 4.443, 4.426)	5.175 (5.329, 5.092, 5.158)
$\sigma_1 = 15\%, \sigma_2 = 30\%$ $H_1 = 120, H_3 = 130$ $H_2 = 80, H_4 = 70$	4.726 (4.594, 4.752, 4.728)	4.744 (4.868, 4.785, 4.749)	4.724 (4.574, 4.771, 4.728)	5.042 (5.226, 5.018, 5.045)
$\sigma_1 = 30\%, \sigma_2 = 15\%$ $H_1 = 130, H_3 = 120$ $H_2 = 70, H_4 = 80$	2.614 (2.429, 2.576, 2.610)	2.683 (2.872, 2.612, 2.679)	2.771 (2.602, 2.742, 2.773)	2.942 (3.165, 3.036, 2.953)

Thanks to the rapidly decaying exponential functions in the integrands, a level of 10^{-7} convergence is attained by stopping at 8, the number of iterations controlled by the absolute values of n_1 and n_2 in the double sum operators, which results in a total computational time of less than 1 s. For higher values of the volatility parameters than those in Table 4, however, the uniform convergence of the double sums in (56)–(59) may require a greater number of iterations and thus take more time. The implementation of Proposition 5 using the Φ_3 function introduced in Section 3 is slightly faster than the one using the trivariate standard normal cumulative distribution function N_3, although the difference is relatively negligible for most practical purposes. Both methods of implementation yield prices equal to at least 4 decimals.

From a financial standpoint, a striking contrast between the numerical results in Table 4 and those of the previous sections is the much weaker dependency of the option value on the correlation structure, as illustrated by the smaller differences between the four option prices associated with each combination of volatility and barrier parameters. It seems that, the more volatility, the more dependency on the correlation structure, as suggested by the comparison between row 1 and row 2. Another noticeable difference is that the functional relation with the correlation structure is not monotonic. This is particularly clear in row 2 where a relatively significant increase in value from $\rho_{1.2} = -0.6$ to $\rho_{1.2} = -0.2$ is followed by a relatively significant decrease in value from $\rho_{1.2} = -0.2$ to $\rho_{1.2} = 0.2$, before a new increase in value from $\rho_{1.2} = 0.2$ to $\rho_{1.2} = 0.6$. This more complex and unstable dependency on correlation structure suggests that two-colour knock-out double barrier options are a less suitable instrument for correlation trading than two-colour knock-out single barrier options. However, one should remain wary of drawing hasty conclusions from the comparison of the results in Table 4 and those in the previous sections, as the option parameters are not identical, especially regarding volatility and expiry.

3. Proofs of Formulae

The proofs of Propositions 2 and 3 are only outlined as they essentially follow the same steps as the proof of Proposition 1.

Proof of Proposition 1. Since the log function is strictly increasing, we have:

$$P_{[RUU]}(\mu_1, \mu_2) = P\left(\sup_{0 \leq t \leq t_1} X_1(t) \leq h_1, X_1(t_1) \leq k_1, X_2(t_1) \leq k_2, \sup_{t_1 \leq t \leq t_2} X_2(t) \leq h_2, X_2(t_2) \leq k_3 \right) \tag{72}$$

Next, it can be noticed that, despite the non-zero correlation between $X_1(t)$ and $X_2(t)$, the law of $\sup_{0 \leq t \leq t_1} X_1(t)$ conditional on $X_1(t_1)$ and $X_2(t_1)$ is equal to the law of $\sup_{0 \leq t \leq t_1} X_1(t_1)$ conditional on $X_1(t_1)$.

Indeed, denoting the density function operator as $f(.)$ and making use of the Markov property of $X_2(t)$ we have:

$$f\left(\sup_{0 \leq t \leq t_1} X_1(t)|X_1(t_1), X_2(t_1) \right) = \frac{f\left(X_2(t_1) \Big| \sup_{0 \leq t \leq t_1} X_1(t), X_1(t_1) \right) f\left(\sup_{0 \leq t \leq t_1} X_1(t), X_1(t_1) \right)}{f(X_1(t_1), X_2(t_1))} \tag{73}$$

$$= \frac{f(X_2(t_1)|X_1(t_1))f\left(\sup_{0 \leq t \leq t_1} X_1(t), X_1(t_1) \right)}{f(X_1(t_1), X_2(t_1))} = \frac{f(X_1(t_1), X_2(t_1))}{f(X_1(t_1))} \frac{f\left(\sup_{0 \leq t \leq t_1} X_1(t), X_1(t_1) \right)}{f(X_1(t_1), X_2(t_1))} \tag{74}$$

$$= \frac{f\left(\sup_{0 \leq t \leq t_1} X_1(t), X_1(t_1) \right)}{f(X_1(t_1))} = f\left(\sup_{0 \leq t \leq t_1} X_1(t)|X_1(t_1) \right) \tag{75}$$

A translation from the time interval $[t_0 = 0, t_1]$ to the time interval $[t_1, t_2]$, through the substitution of $X_1(0)$ with $X_2(t_1)$, of $X_1(t_1)$ with $X_2(t_2)$ and of $X_2(t_1)$ with $X_1(t_2)$, shows similarly that the law of $\sup_{t_1 \leq t \leq t_2} X_2(t)$ conditional on $X_2(t_1)$, $X_2(t_2)$ and $X_1(t_2)$ is equal to the law of $\sup_{t_1 \leq t \leq t_2} X_2(t)$ conditional on $X_2(t_1)$ and $X_2(t_2)$.

Thus, by conditioning w.r.t. the absolutely continuous random variables $X_1(t_1)$, $X_2(t_1)$ and $X_2(t_2)$, we can express the problem as the following integral:

$$
\begin{aligned}
&P_{[\text{RUU}]}(\mu_1, \mu_2) \\
&= \int_{-\infty}^{\min(k_1, h_1)} \int_{-\infty}^{\min(k_2, h_2)} \int_{-\infty}^{\min(k_3, h_2)} \varphi_1(x_1, x_2, x_3) \varphi_2(x_1) \varphi_3(x_2, x_3) dx_3 dx_2 dx_1
\end{aligned}
\tag{76}
$$

where

$$
\varphi_1(x_1, x_2, x_3) = P(X_1(t_1) \in dx_1, X_2(t_1) \in dx_2, X_2(t_2) \in dx_3) dx_3 dx_2 dx_1
\tag{77}
$$

$$
\varphi_2(x_1) = P\left(\sup_{0 \leq t \leq t_1} X_1(t) \leq h_1 | X_1(t_1) \in dx_1 \right) dx_1
\tag{78}
$$

$$
\varphi_3(x_2, x_3) = P\left(\sup_{t_1 \leq t \leq t_2} X_2(t) \leq h_2 | X_2(t_1) \in dx_2, X_2(t_2) \in dx_3 \right) dx_2 dx_3
\tag{79}
$$

The functions φ_2 and φ_3 in (78) and (79) can be expanded by applying known formulae that can be found in Wang and Pötzelberger (1997):

$$
\varphi_2(x_1) = 1 - \exp\left(\frac{2 h_1 (x_1 - h_1)}{\sigma_1^2 t_1} \right)
\tag{80}
$$

$$
\varphi_3(x_2, x_3) = 1 - \exp\left(\frac{2(h_2 - x_2)(x_3 - h_2)}{\sigma_2^2 (t_2 - t_1)} \right)
\tag{81}
$$

The function φ_1 derives from the trivariate normality of the triple $(X_1(t_1), X_2(t_1), X_2(t_2))$. It is elementary to obtain the marginal distributions:

$$
X_1(t_1) \sim \mathcal{N}\left(\mu_1 t_1, \sigma_1^2 t_1 \right), X_2(t_1) \sim \mathcal{N}\left(\mu_2 t_1, \sigma_2^2 t_1 \right), X_2(t_2) \sim \mathcal{N}\left(\mu_2 t_1, \sigma_2^2 t_1 \right)
\tag{82}
$$

where $\mathcal{N}(a, b^2)$ refers to the normal distribution with expectation a and variance b^2.

Denoting by Z_1, Z_2, Z_3 three independent standard normal random variables, the pairwise covariances can be written as follows:

$$
\text{cov}[X_1(t_1), X_2(t_1)] = \text{cov}\left[\mu_1 t_1 + \sigma_1 \sqrt{t_1} Z_1, \mu_2 t_1 + \sigma_2 \sqrt{t_1}\left(\rho_{1.2} Z_1 + \sqrt{1 - \rho_{1.2}^2} Z_2 \right) \right] = \sigma_1 \sigma_2 \rho_{1.2} t_1
\tag{83}
$$

$$
\text{cov}[X_2(t_1), X_2(t_2)] = \text{cov}\left[\begin{array}{l} \mu_2 t_1 + \sigma_2 \sqrt{t_1}\left(\rho_{1.2} Z_1 + \sqrt{1 - \rho_{1.2}^2} Z_2 \right), \\ \mu_2 t_2 + \sigma_2 \sqrt{t_1}\left(\rho_{1.2} Z_1 + \sqrt{1 - \rho_{1.2}^2} Z_2 \right) + \sigma_2 \sqrt{t_2 - t_1} Z_3 \end{array} \right] = \sigma_2^2 t_1
\tag{84}
$$

$$
\text{cov}[X_1(t_1), X_2(t_2)] = \text{cov}\left[\mu_1 t_1 + \sigma_1 \sqrt{t_1} Z_1, \mu_2 t_2 + \sigma_2 \sqrt{t_1}\left(\rho_{1.2} Z_1 + \sqrt{1 - \rho_{1.2}^2} Z_2 \right) + \sigma_2 \sqrt{t_2 - t_1} Z_3 \right]
$$
$$
= \sigma_1 \sigma_2 \rho_{1.2} t_1
\tag{85}
$$

where we have applied the bilinearity of the covariance operator, the independence of increments of Brownian motion, and the orthogonal decomposition of two-dimensional correlated Brownian motion. The correlation coefficients $\theta_{1.2}, \theta_{1.3}, \theta_{2.3}$ in Proposition 1

ensue. Expanding the trivariate normal density function $\varphi_1(x_1, x_2, x_3)$ as a product of normal conditional densities (Guillaume 2018), we obtain:

$$\varphi_1(x_1, x_2, x_3)$$

$$= \frac{e^{-\frac{1}{2}\left(\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)^2 - \frac{1}{2\sigma_{2|1}^2}\left(\frac{x_2-\mu_2 t_1}{\sigma_2\sqrt{t_1}} - \theta_{1.2}\left(\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)\right)^2 - \frac{1}{2\sigma_{3|1.2}^2}\left(\frac{x_3-\mu_2 t_2}{\sigma_2\sqrt{t_2}} - \theta_{1.3}\left(\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right) - \frac{\theta_{2.3|1}}{\sigma_{2|1}}\left(\frac{x_2-\mu_2 t_1}{\sigma_2\sqrt{t_1}} - \theta_{1.2}\left(\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)\right)\right)^2}}{(2\pi)^{3/2}\sigma_{2|1}\sigma_{3|1.2}\sigma_2^2\sigma_1 t_1\sqrt{t_2}} \quad (86)$$

where:

$$\sigma_{2|1} = \sqrt{1 - \theta_{1.2}^2}, \theta_{2.3|1} = \frac{\theta_{2.3} - \theta_{1.2}\theta_{1.3}}{\sqrt{1 - \theta_{1.2}^2}}, \sigma_{3|1.2} = \sqrt{1 - \theta_{1.3}^2 - \theta_{2.3|1}^2} \quad (87)$$

The terms $\sigma_{2|1}$, $\theta_{2.3|1}$ and $\sigma_{3|1.2}$ have the following precise meanings:

- $\sigma_{2|1}$ is the conditional standard deviation of $X_2(t_1)$ given $X_1(t_1)$;
- $\theta_{2.3|1}$ is the conditional correlation between $X_2(t_1)$ and $X_2(t_2)$ given $X_1(t_1)$;
- $\sigma_{3|1.2}$ is the conditional standard deviation of $X_2(t_2)$ given $X_1(t_1)$ and $X_2(t_1)$.

The rest of the proof, whose cumbersome details are omitted, then consists in solving the four integrals implied by (76). The final result takes the form of the linear combination of four N_3 functions written in Proposition 1.

Corollary 1 comes from the property of symmetry of Brownian paths.

Corollary 2 is a consequence of the fact that:

$$P_{[RUU]}^{(I)}(\mu_1, \mu_2)$$

$$= \int_{-\infty}^{\min(k_1, h_1)} \int_{-\infty}^{\min(k_2, h_2)} \int_{-\infty}^{\min(k_3, h_2)} \varphi_1(x_1, x_2, x_3) \exp\left(\frac{2h_1(x_1 - h_1)}{\sigma_1^2 t_1}\right) \exp\left(\frac{2(h_2 - x_1)(x_2 - h_2)}{\sigma_2^2(t_2 - t_1)}\right) dx_3 dx_2 dx_1 \quad (88)$$

Corollary 3 comes from the fact that the correlation coefficient between the random variables $S_2(t_2)$ and $S_2(t_3)$ is equal to $\sqrt{\frac{t_2}{t_3}}$.

Corollary 4 is a straightforward application of the law of total probability. $\square$

Proof of Proposition 2. Using similar steps as in the proof of Proposition 1, one can express the problem at hand as the following integral:

$$P_{[RUD]}(\mu_1, \mu_2) = \int_{-\infty}^{\min(k_1, h_1)} \int_{\max(k_2, h_2)}^{\infty} \int_{\max(k_3, h_2)}^{\infty} \varphi_1(x_1, x_2, x_3) \varphi_2(x_1) \varphi_4(x_2, x_3) dx_3 dx_2 dx_1 \quad (89)$$

where the functions φ_1 and φ_2 are given by (86) and (80), respectively, and:

$$\varphi_4(x_2, x_3) = P\left(\inf_{t_1 \leq t \leq t_2} X_2(t) \geq h_2 | X_2(t_1) \in dx_2, X_2(t_2) \in dx_3\right) dx_2 dx_3 = \varphi_3(x_2, x_3) \quad (90)$$

where the function φ_3 is given by (81).

Performing the necessary calculations, one can obtain the linear combination of four N_3 functions given in Proposition 2.

As in Proposition 1, Corollary 1 comes from the property of symmetry of Brownian paths.

Corollary 2 is a consequence of the fact that:

$$P_{[RUD]}^{(I)}(\mu_1, \mu_2)$$

$$= \int_{-\infty}^{\min(k_1, h_1)} \int_{\max(k_2, h_2)}^{\infty} \int_{\max(k_3, h_2)}^{\infty} \varphi_1(x_1, x_2, x_3) e^{\frac{2\mu_1}{\sigma_1^2} h_1} \frac{e^{-\frac{1}{2}\left(\frac{x_1 - 2h_1 - \mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)^2}}{\sigma_1\sqrt{2\pi t_1}} e^{\frac{2\mu_2}{\sigma_2^2}(h_2 - x_2)} \frac{e^{-\frac{1}{2}\left(\frac{-x_3 - x_2 + 2h_2 + \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right)^2}}{\sigma_2\sqrt{2\pi(t_2 - t_1)}} dx_1 dx_2 dx_3 \quad (91)$$

$\square$

Proof of Proposition 3. One can express the problem at hand as the following integral:

$$P_{[\text{RRUU}]}(\mu_1, \mu_2)$$

$$= \int_{-\infty}^{\min(k_1,h_1)} \int_{-\infty}^{h_2} \int_{-\infty}^{\min(k_2,h_2)} \int_{-\infty}^{k_3} \varphi_2(x_1)\varphi_3(x_2,x_3)\varphi_5(x_1,x_2,x_3,x_4)dx_4dx_3dx_2dx_1 \tag{92}$$

where

$$\varphi_5(x_1,x_2,x_3,x_4) = P(X_1(t_1) \in dx_1, X_2(t_1) \in dx_2, X_2(t_2) \in dx_3, X_1(t_2) \in dx_4)dx_4dx_3dx_2dx_1 \tag{93}$$

Plugging the quadrivariate normal joint density function of the set of random variables $X_1(t_1)$, $X_2(t_1)$, $X_1(t_2)$ and $X_2(t_2)$, as a product of conditional density functions as explained in Guillaume (2018), and then factoring in the conditional cumulative distribution function of $X_1(t_2)$ given the triple $(X_1(t_1), X_2(t_1), X_2(t_2))$, Proposition 3 ensues. □

Proof of Proposition 4. Proof is given only for $p_1(\mu_1, \mu_2, \mu_3)$ and $p_3(\mu_1, \mu_2, \mu_3)$, as $p_2(\mu_1, \mu_2, \mu_3)$ and $p_4(\mu_1, \mu_2, \mu_3)$ can then be deduced by the same symmetry argument as that already used in Corollary 1 of Proposition 1.

Following steps similar to the beginning of the proof of Proposition 1, one can express the problem at hand as the following two integrals:

$$p_1(\mu_1, \mu_2, \mu_3)$$

$$= \int_{x_1=-\infty}^{\min(k_1,h_1)} \int_{x_2=-\infty}^{\min(k_2,h_2)} \int_{x_3=-\infty}^{\min(k_3,h_2)} \int_{x_4=-\infty}^{k_4} \varphi_6(x_1)\varphi_7(x_1,x_2)\varphi_8(x_2,x_3)\varphi_9(x_1,x_3,x_4)dx_4dx_3dx_2dx_1 \tag{94}$$

$$p_3(\mu_1, \mu_2, \mu_3)$$

$$= \int_{x_1=-\infty}^{\min(k_1,h_1)} \int_{x_2=\max(k_2,h_2)}^{\infty} \int_{x_3=\max(k_3,h_2)}^{\infty} \int_{x_4=k_4}^{\infty} \varphi_6(x_1)\varphi_7(x_1,x_2)\varphi_{10}(x_2,x_3)\varphi_9(x_1,x_3,x_4)dx_4dx_3dx_2dx_1 \tag{95}$$

where

$$\varphi_6(x_1) = P\left(\sup_{0 \leq t \leq t_1} X_1(t) \leq h_1, X_1(t_1) \in dx_1\right)dx_1 \tag{96}$$

$$\varphi_7(x_1,x_2) = P(X_2(t_1) \in dx_2 | X_1(t_1) \in dx_1)dx_1dx_2 \tag{97}$$

$$\varphi_8(x_2,x_3) = P\left(\sup_{t_1 \leq t \leq t_2} X_2(t) \leq h_2, X_2(t_2) \in dx_3 | X_2(t_1) \in dx_2\right)dx_2dx_3 \tag{98}$$

$$\varphi_{10}(x_2,x_3) = P\left(\inf_{t_1 \leq t \leq t_2} X_2(t) \geq h_2, X_2(t_2) \in dx_3 | X_2(t_1) \in dx_2\right)dx_2dx_3 \tag{99}$$

$$\varphi_9(x_1,x_3,x_4) = P(X_3(t_2) \in dx_4 | X_2(t_2) \in dx_3, X_1(t_1) \in dx_1) \tag{100}$$

The function φ_6 is obtained by differentiating the classical formula for the joint cumulative distribution of the maximum of a Brownian motion with drift and its endpoint over a closed time interval (see, e.g., Karatzas and Shreve 2000):

$$\varphi_6(x_1) = \frac{e^{-\frac{1}{2}\left(\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)^2}}{\sigma_1\sqrt{2\pi t_1}} - e^{\frac{2\mu_1}{\sigma_1^2}h_1}\frac{e^{-\frac{1}{2}\left(\frac{x_1-2h_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)^2}}{\sigma_1\sqrt{2\pi t_1}}dx_1 \tag{101}$$

The function φ_7 is easily derived from the bivariate normality of the pair $(X_1(t_1), X_2(t_1))$:

$$\varphi_7(x_1,x_2) = \frac{e^{-\frac{1}{2(1-\rho_{1.2}^2)}\left(\frac{x_2-\mu_2 t_1}{\sigma_2\sqrt{t_1}} - \rho_{1.2}\frac{x_1-\mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)^2}}{\sigma_2\sqrt{2\pi t_1(1-\rho_{1.2}^2)}}dx_1dx_2 \tag{102}$$

To handle the function φ_8, we notice that, by conditioning w.r.t. the filtration at time t_1 the same classical formula as the one used to derive φ_6, we can obtain:

$$P\left(\sup_{t_1 \leq t \leq t_2} S_2(t) \leq H_2, S_2(t_2) \leq S_2(0)e^{x_3} | S_2(t_1) = S_2(0)e^{x_2}\right)$$

$$= N\left[\frac{\ln\left(\frac{S_2(0)e^{x_3}}{S_2(0)e^{x_2}}\right) - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] - \left(\frac{H_2}{S_2(0)e^{x_2}}\right)^{\frac{2\mu_2}{\sigma_2^2}} N\left[\frac{\ln\left(\frac{S_2(0)e^{x_3}}{S_2(0)e^{x_2}}\right) - 2\ln\left(\frac{H_2}{S_2(0)e^{x_2}}\right)}{-\mu_2(t_2 - t_1)}\frac{}{\sigma_2\sqrt{t_2 - t_1}}\right] \quad (103)$$

for any given $(x_2, x_3) \in \mathbb{R}^2$ and $H_2 > S_2(0)e^{x_3}$.

Equation (103) can be rewritten as follows:

$$P\left(\sup_{t_1 \leq t \leq t_2} X_2(t) \leq h_2, X_2(t_2) \leq x_3 | X_2(t_1) \in dx_2\right) \quad (104)$$

$$= N\left[\frac{x_3 - x_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] - \exp\left(\frac{2\mu_2}{\sigma_2^2}(h_2 - x_2)\right) N\left[\frac{x_3 - x_2 - 2(h_2 - x_2) - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right]$$

Therefore, by differentiating (104) w.r.t. x_3, we obtain:

$$\varphi_8(x_2, x_3) = \frac{e^{-\frac{1}{2}\left(\frac{x_3 - x_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right)^2}}{\sigma_2\sqrt{2\pi(t_2 - t_1)}} - e^{\frac{2\mu_2}{\sigma_2^2}(h_2 - x_2)}\frac{e^{-\frac{1}{2}\left(\frac{x_3 + x_2 - 2h_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right)^2}}{\sigma_2\sqrt{2\pi(t_2 - t_1)}} dx_2 dx_3 \quad (105)$$

By the symmetry of paths of Brownian motion, we have:

$$P\left(\inf_{t_1 \leq t \leq t_2} X_2(t) \geq h_2, X_2(t_2) \geq x_3 | X_2(t_1) \in dx_2\right) \quad (106)$$

$$= N\left[\frac{-x_3 + x_2 + \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] - \exp\left(\frac{2\mu_2}{\sigma_2^2}(h_2 - x_2)\right) N\left[\frac{-x_3 + x_2 + 2(h_2 - x_2) + \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right]$$

Hence,

$$\varphi_{10}(x_2, x_3) = \varphi_8(x_2, x_3) \quad (107)$$

The function φ_9 derives from the joint trivariate normality of the triple $(X_1(t_1), X_2(t_2), X_3(t_2))$. The marginal distributions of the elements of this triple come from the known marginal distributions of $S_1(t_1)$, $S_2(t_2)$ and $S_3(t_2)$. The pairwise correlations, as given by $\theta_{1.3}, \theta_{1.4}$ and $\theta_{3.4}$ in Proposition 4 can be easily determined using the same method as in the proof of Proposition 3. We obtain:

$$\varphi_9(x_1, x_3, x_4) = \frac{e^{-\frac{1}{2\phi_{4|1.3}^2}\left(\frac{x_4 - \mu_3 t_2}{\sigma_3\sqrt{t_2}} - \theta_{1.4}\left(\frac{x_1 - \mu_1 t_1}{\sigma_1\sqrt{t_1}}\right) - \theta_{3.4|1}\left(\frac{x_3 - \mu_2 t_2}{\sigma_2\sqrt{t_2}} - \theta_{1.3}\left(\frac{x_1 - \mu_1 t_1}{\sigma_1\sqrt{t_1}}\right)\right)\right)^2}}{\phi_{4|1.3}\sigma_3\sqrt{2\pi t_2}} \quad (108)$$

where $\sigma_{3|1} = \sqrt{1 - \theta_{1.3}^2}$.

The rest of the proof, whose cumbersome details are omitted, then consists in solving the integrals implied by (94) and (95). The final result can be expressed as the linear combination of four Ψ_4 functions written in Proposition 4. The origin of the function Ψ_4, which is a special form of quadrivariate normal cumulative distribution, lies in the FDD

(Finite Dimensional Distribution) of the quadruple $[S_1(t_1), S_2(t_1), S_2(t_2), S_3(t_2)]$. Indeed, a little algebra shows that, $\forall D_1, D_2, D_3, D_4 \in \mathbb{R}^+$, we have:

$$P(S_1(t_1) \leq D_1, S_2(t_1) \leq D_2, S_2(t_2) \leq D_3, S_3(t_2) \leq D_4)$$

$$= \Psi_4 \begin{bmatrix} \frac{\ln\left(\frac{D_1}{S_1(0)}\right) - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}, \frac{\ln\left(\frac{D_2}{S_2(0)}\right) - \mu_2 t_1}{\sigma_2 \sqrt{t_1}}, \frac{\ln\left(\frac{D_3}{S_2(0)}\right) - \mu_2 t_2}{\sigma_2 \sqrt{t_2}}, \frac{\ln\left(\frac{D_4}{S_3(0)}\right) - \mu_3 t_2}{\sigma_3 \sqrt{t_2}}; \\ \theta_{1.2}, \theta_{1.3}, \theta_{1.4}, \theta_{2.3}, \theta_{3.4} \end{bmatrix} \tag{109}$$

Corollary 1 is a consequence of the fact that:

$$p_1^{(I)}(\mu_1, \mu_2) = \int_{x_1=-\infty}^{\min(k_1, h_1)} \int_{x_2=-\infty}^{\min(k_2, h_2)} \int_{x_3=-\infty}^{\min(k_3, h_2)} \int_{x_4=-\infty}^{k_4} e^{\frac{2\mu_1}{\sigma_1^2} h_1} \frac{e^{-\frac{1}{2}\left(\frac{x_1 - 2h_1 - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}\right)^2}}{\sigma_1 \sqrt{2\pi t_1}} \varphi_7(x_1, x_2)$$

$$e^{\frac{2\mu_2}{\sigma_2^2}(h_2 - x_2)} \frac{e^{-\frac{1}{2}\left(\frac{x_3 + x_2 - 2h_2 - \mu_2(t_2 - t_1)}{\sigma_2 \sqrt{t_2 - t_1}}\right)^2}}{\sigma_2 \sqrt{2\pi(t_2 - t_1)}} \varphi_9(x_1, x_3, x_4) dx_4 dx_3 dx_2 dx_1 \tag{110}$$

and:

$$p_3^{(I)}(\mu_1, \mu_2) = \int_{x_1=-\infty}^{\min(k_1, h_1)} \int_{x_2=\max(k_2, h_2)}^{\infty} \int_{x_3=\max(k_3, h_2)}^{\infty} \int_{x_4=-\infty}^{k_4} e^{\frac{2\mu_1}{\sigma_1^2} h_1} \frac{e^{-\frac{1}{2}\left(\frac{x_1 - 2h_1 - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}\right)^2}}{\sigma_1 \sqrt{2\pi t_1}} \varphi_7(x_1, x_2)$$

$$e^{\frac{2\mu_2}{\sigma_2^2}(h_2 - x_2)} \frac{e^{-\frac{1}{2}\left(\frac{-x_3 - x_2 + 2h_2 + \mu_2(t_2 - t_1)}{\sigma_2 \sqrt{t_2 - t_1}}\right)^2}}{\sigma_2 \sqrt{2\pi(t_2 - t_1)}} \varphi_9(x_1, x_3, x_4) dx_4 dx_3 dx_2 dx_1 \tag{111}$$

Corollary 2 comes from the fact that the correlation coefficient between the random variables $S_2(t_2)$ and $S_3(t_3)$ is equal to $\rho_{2.3} \sqrt{\frac{t_2}{t_3}}$. $\square$

Proof of Proposition 5. Following steps similar to the beginnings of the previous proofs, one can express the problem at hand as the following integral:

$$P_{[RDKO]}(\mu_1, \mu_2) \tag{112}$$

$$= \int_{x_1=h_2}^{\min(k_1, h_1)} \int_{x_2=h_4}^{h_3} \int_{x_3=h_4}^{\min(k_2, h_3)} \varphi_1(x_1, x_2, x_3) \varphi_{11}(x_1, x_2) \varphi_{12}(x_2, x_3) dx_3 dx_2 dx_1$$

where the function φ_1 is given by Equation (86) and:

$$\varphi_{11}(x_1, x_2) = P\left(\sup_{0 \leq t \leq t_1} X_1(t) \leq h_1, \inf_{0 \leq t \leq t_1} X_1(t) \geq h_2 | X_1(t_1) \in dx_1 \right) dx_2 dx_1 \tag{113}$$

$$\varphi_{12}(x_2, x_3) \tag{114}$$

$$= P\left(\sup_{t_1 \leq t \leq t_2} X_2(t) \leq h_3, \inf_{t_1 \leq t \leq t_2} X_2(t) \geq h_4 | X_2(t_1) \in dx_2, X_2(t_2) \in dx_3 \right) dx_3 dx_2$$

From Pötzelberger and Wang (2001), one can plug:

$$\varphi_{11}(x_1, x_2) = \sum_{n=-\infty}^{\infty} e^{\frac{2na_1(x_1 - na_1)}{\sigma_1^2 t_1}} - e^{\frac{2(h_1 - na_1)(x_1 - h_1 + na_1)}{\sigma_1^2 t_1}} \tag{115}$$

$$\varphi_{12}(x_2, x_3) = \sum_{n=-\infty}^{\infty} e^{\dfrac{2na_2(x_3 - x_2 - na_2)}{\sigma_2^2(t_2 - t_1)}} - e^{\dfrac{2(h_3 - x_2 - na_2)(x_3 - h_3 + na_2)}{\sigma_2^2(t_2 - t_1)}} \tag{116}$$

The bulk of the proof, whose cumbersome details are omitted, then consists of solving the sixteen integrals implied by (112). The final result takes the form of the linear combinations of double sums of N_3 functions in Proposition 5.

An elementary adjustment identical to the one in Corollary 3 of Proposition 1 allows to value an early-ending variant of $P_{[RDKO]}(\mu_1, \mu_2)$.

Alternatively, one can also expand the problem as the following integral:

$$P_{[RDKO]}(\mu_1, \mu_2) = \int_{x_1=h_2}^{\min(k_1,h_1)} \int_{x_2=h_4}^{h_3} \int_{x_3=h_4}^{\min(k_2,h_3)} \varphi_{13}(x_1)\varphi_7(x_1, x_2)\varphi_{14}(x_2, x_3)\,dx_3\,dx_2\,dx_1 \tag{117}$$

where the function φ_7 is given by Equation (102) and:

$$\varphi_{13}(x_1) = P\left(\sup_{0 \le t \le t_1} X_1(t) \le h_1, \inf_{0 \le t \le t_1} X_1(t) \ge h_2, X_1(t_1) \in dx_1\right)dx_1 \tag{118}$$

$$\varphi_{14}(x_2, x_3) \tag{119}$$

$$= P\left(\sup_{t_1 \le t \le t_2} X_2(t) \le h_3, \inf_{t_1 \le t \le t_2} X_2(t) \ge h_4, X_2(t_2) \in dx_3 | X_2(t_1) \in dx_2\right)dx_3\,dx_2$$

According to the classical formula for the distribution of the maximum, the minimum and the endpoint of a Brownian motion with drift over a closed time interval, which can be traced back to Anderson (1960), we have:

$$P\left(\sup_{0 \le t \le t_1} S_1(t) < H_1, \inf_{0 \le t \le t_1} S_1(t) > H_2, S_1(t_1) < S_1(0)e^{x_1} | S_1(0)\right)$$

$$= \sum_{n_1=-\infty}^{\infty} \exp\left(\frac{2\mu_1}{\sigma_1^2} n_1 a_1\right) \left\{ N\left[\frac{x_1 - 2n_1 a_1 - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}\right] - N\left[\frac{h_2 - 2n_1 a_1 - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}\right] \right\} \tag{120}$$

$$- \sum_{n_1=-\infty}^{\infty} \exp\left(\frac{2\mu_1}{\sigma_1^2}(h_2 - n_1 a_1)\right) \left\{ N\left[\frac{x_1 - 2h_2 + 2n_1 a_1 - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}\right] - N\left[\frac{-h_2 + 2n_1 a_1 - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}\right] \right\}$$

for a given $x_1 \in \mathbb{R}$, and $\forall H_1 > S_1(0)e^{x_1}$.

Mere differentiation of (120) w.r.t. x_1 yields the function φ_{13}:

$$\varphi_{13}(x_1) = \sum_{n_1=-\infty}^{\infty} \frac{e^{\frac{2\mu_1}{\sigma_1^2}n_1 a_1 - \frac{1}{2\sigma^2 t_1}(x_1 - \mu_1 t_1 - 2n_1 a_1)^2}}{\sigma_1 \sqrt{2\pi t_1}} - \sum_{n_1=-\infty}^{\infty} \frac{e^{\frac{2\mu}{\sigma_1^2}(h_2 - n_1 a_1) - \frac{1}{2\sigma_1^2 t_1}(x_1 - 2h_2 - \mu_1 t_1 + 2n_1 a_1)^2}}{\sigma_1 \sqrt{2\pi t_1}} \tag{121}$$

To handle the function φ_{14}, we notice that, by conditioning w.r.t. the filtration at time t_1 the same classical formula as the one used to derive φ_{13}, we can obtain:

$$P\left(\sup_{t_1 \le t \le t_2} S_2(t) < H_3, \inf_{t_1 \le t \le t_2} S_2(t) > H_4, S_2(t_2) \le S_2(0)e^{x_3} | S_2(t_1) = S_2(0)e^{x_2}\right)$$

$$
= \sum_{n_2=-\infty}^{\infty} \exp\left(\frac{2\mu_2}{\sigma_2^2} n_2 a_2\right) \left\{ \begin{array}{l} N\left[\dfrac{\ln\left(\dfrac{S_2(0)e^{x_3}}{S_2(0)e^{x_2}}\right) - 2n_2 a_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] \\[2em] -N\left[\dfrac{\ln\left(\dfrac{H_4}{S_2(0)e^{x_2}}\right) - 2n_2 a_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] \end{array} \right\} \tag{122}
$$

$$
- \sum_{n_2=-\infty}^{\infty} \exp\left(\frac{2\mu_2}{\sigma_2^2}\left(\ln\left(\frac{H_4}{S_2(0)e^{x_2}}\right) - n_2 a_2\right)\right)
$$
$$
\left\{ \begin{array}{l} N\left[\dfrac{\ln\left(\dfrac{S_2(0)e^{x_3}}{S_2(0)e^{x_2}}\right) - 2\ln\left(\dfrac{H_4}{S_2(0)e^{x_2}}\right) + 2n_2 a_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] \\[2em] -N\left[\dfrac{-\ln\left(\dfrac{H_4}{S_2(0)e^{x_2}}\right) + 2n_2 a_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] \end{array} \right\} \tag{123}
$$

for any given $(x_2, x_3) \in \mathbb{R}^2$ and $H_3 > S_2(0)e^{x_3}$.

Equations (122) and (123) can be rewritten as follows:

$$
P\left(\sup_{t_1\leq t\leq t_2} X_2(t) < h_3, \inf_{t_1\leq t\leq t_2} X_2(t) > h_4, X_2(t_2) \leq x_3 | X_2(t_1) \in dx_2\right)
$$

$$
= \sum_{n_2=-\infty}^{\infty} \exp\left(\frac{2\mu_2}{\sigma_2^2} n_2 a_2\right) \left\{ \begin{array}{l} N\left[\dfrac{x_3 - x_2 - 2n_2 a_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] \\[1.5em] -N\left[\dfrac{h_4 - x_2 - 2n_2 a_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] \end{array} \right\} \tag{124}
$$

$$
- \sum_{n_2=-\infty}^{\infty} \exp\left(\frac{2\mu_2}{\sigma_2^2}((h_4 - x_2) - n_2 a_2)\right) \left\{ \begin{array}{l} N\left[\dfrac{x_3 - x_2 - 2(h_4 - x_2) + 2n_2 a_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] \\[1.5em] -N\left[\dfrac{-(h_4 - x_2) + 2n_2 a_2 - \mu_2(t_2 - t_1)}{\sigma_2\sqrt{t_2 - t_1}}\right] \end{array} \right\} \tag{125}
$$

Therefore, by differentiating (124) and (125) w.r.t. x_3, we obtain:

$$
\varphi_{14}(x_2, x_3) = \sum_{n_2=-\infty}^{\infty} e^{\left(\frac{2\mu_2}{\sigma_2^2} n_2 a_2\right)} \frac{e^{-\frac{1}{2\sigma_2^2(t_2 - t_1)}(x_3 - x_2 - 2n_2 a_2 - \mu_2(t_2 - t_1))^2}}{\sigma_2\sqrt{2\pi(t_2 - t_1)}} \tag{126}
$$

$$
- \sum_{n_2=-\infty}^{\infty} e^{\frac{2\mu_2}{\sigma_2^2}(h_4 - x_2 - n_2 a_2)} \frac{e^{-\frac{1}{2\sigma_2^2(t_2 - t_1)}(x_3 + x_2 - 2h_4 + 2n_2 a_2 - \mu_2(t_2 - t_1))^2}}{\sigma_2\sqrt{2\pi(t_2 - t_1)}} \tag{127}
$$

This second formulation leads to a formula identical to Proposition 5 except for the fact that the N_3 functions are replaced by Φ_3 functions defined as follows:

$$
\Phi_3[b_1, b_2, b_3; x] = \int_{x=-\infty}^{b_2} \frac{\exp\left(-x^2/2\right)}{\sqrt{2\pi}} N\left[\frac{b_1 - c_1 x}{\sqrt{1 - c_1^2}}\right] N\left[\frac{b_3 - c_2 x}{\sqrt{1 - c_2^2}}\right] dx \tag{128}
$$

where $(b_1, b_2, b_3) \in \mathbb{R}^3$ and x is a vector with two real coordinates $c_1, c_2 \in]-1, 1[$.

The vectors of correlation coefficients x_1 and x_2 become:

$$x_1 = \left\{ \rho_{1.2}, \sqrt{\frac{t_1}{t_2}} \right\}, x_2 = \left\{ \rho_{1.2}, -\sqrt{\frac{t_1}{t_2}} \right\} \tag{129}$$

The origin of the function Φ_3, which is a special form of trivariate normal cumulative distribution, lies in the FDD (finite dimensional distribution) of the triple $[S_1(t_1), S_2(t_1), S_2(t_2)]$. Indeed, a little algebra shows that, $\forall D_1, D_2, D_3 \in \mathbb{R}^+$, we have:

$$P(S_1(t_1) \leq D_1, S_2(t_1) \leq D_2, S_2(t_2) \leq D_3)$$

$$= \Phi_3 \left[\frac{\ln\left(\frac{D_1}{S_1(0)}\right) - \mu_1 t_1}{\sigma_1 \sqrt{t_1}}, \frac{\ln\left(\frac{D_2}{S_2(0)}\right) - \mu_2 t_1}{\sigma_2 \sqrt{t_1}}, \frac{\ln\left(\frac{D_3}{S_2(0)}\right) - \mu_2 t_2}{\sigma_2 \sqrt{t_2}}; \rho_{1.2}, \sqrt{\frac{t_1}{t_2}} \right] \tag{130}$$

Notice that the two-colour probability distributions of Sections 2.1 and 2.2 can also be written as linear combinations of functions Φ_3. $\square$

4. Conclusions

This article has shown how to value in closed form an important kind of multi-asset step barrier option known as a rainbow step barrier option, under the condition that the number of "colours" is restricted to two, along with widespread variants such as a two-colour outside step barrier and a two-colour step double barrier. It may be feasible, albeit tedious, to find an analytical solution to an extended valuation problem with three or four colours, but the expected benefits, compared with a conditional Monte Carlo approximation method, would greatly depend on the degree of the quadrature required to numerically evaluate the resulting multidimensional integrals. It should be emphasised that, even if more sophisticated models (allowing, e.g., for stochastic volatility) or a greater number of underlying assets are needed, closed form solutions obtained in a low-dimensional Black–Scholes framework remain useful as fast and accurate benchmarks that can: (i) serve as control variates in a simulation; (ii) speed up the calibration process; (iii) facilitate the analysis and the understanding of the interactions between the variables, as well as of the sensitivities of the option value w.r.t. its main parameters, which is instrumental in devising appropriate hedging techniques or trading strategies.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflict of interest.

References

Anderson, Todd W. 1960. A modification of the sequential probability ratio test to reduce the sample size. *Annals of Mathematical Statistics* 31: 165–97. [CrossRef]

Berntsen, Jarle, Terje O. Espelid, and Alan Genz. 1991. An adaptive algorithm for the approximate calculation of multiple integrals. *ACM Transactions on Mathematical Software* 17: 437–51. [CrossRef]

Bielecki, Tomasz R., and Marek Rutkowski. 2004. *Credit Risk: Modeling, Valuation and Hedging*. Berlin/Heidelberg: Springer.

Bouzoubaa, Mohamed, and Adel Osseiran. 2010. *Exotic Options and Hybrids*. Hoboken: Wiley Finance.

Chang, Der-Chen, Eric C. Chang, Haitao Fan, and Duy-Minh Nhieu. 2005. Mathematical analysis of the two-color partial rainbow options. *Applicable Analysis* 84: 737–57. [CrossRef]

Das, Satyajit. 2006. *Structured Products: Volume 1*. Hoboken: Wiley Finance, John Wiley and Sons (Asia).

Davis, Philip J., and Philip Rabinowitz. 2007. *Methods of Numerical Integration*. New York: Dover Publications.

Gao, Rong, and Xiaoli Wu. 2022. Pricing rainbow option for uncertain financial market. *RAIRO Operations Research* 56: 3973–89. [CrossRef]

Genz, Alan. 2004. Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and Computing* 14: 151–60. [CrossRef]

Glasserman, Paul. 2003. *Monte Carlo Methods in Financial Engineering*. New York: Springer.

Guillaume, Tristan. 2001. Analytical valuation of options on joint minima and maxima. *Applied Mathematical Finance* 8: 209–35. [CrossRef]

Guillaume, Tristan. 2015. On the computation of the survival probability of Brownian motion with drift in a closed time interval when the absorbing boundary is a step function. *Journal of Probability and Statistics* 2015: 391681. [CrossRef]

Guillaume, Tristan. 2016. Computation of the survival probability of Brownian motion with drift in a closed time interval when the absorbing boundary is an affine or an exponential function of time. *International Journal of Statistics and Probability* 5: 119–38. [CrossRef]

Guillaume, Tristan. 2018. Computation of the Quadrivariate and Pentavariate normal cumulative distribution functions. *Communications in Statistics—Simulation and Computation* 47: 839–51. [CrossRef]

Harrison, J. Michael, and David M. Kreps. 1979. Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory* 20: 381–408. [CrossRef]

Harrison, J. Michael, and Stanley R. Pliska. 1981. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and Their Applications* 11: 312–16. [CrossRef]

Heynen, Ronald C., and Harry M. Kat. 1994. Crossing barriers. *Risk Magazine* 7: 46–50.

Karatzas, Ioannis, and Steven E. Shreve. 2000. *Brownian Motion and Stochastic Calculus*. New York: Springer.

Kwok, Yue-Kuen, Lixin Wu, and Hong Yu. 1998. Pricing Multi-Asset Options with an External Barrier. *International Journal of Theoretical and Applied Finance* 1: 523–41. [CrossRef]

Lee, Hangsuck, Bangwon Ko, and Seongjoo Song. 2019a. Valuing step barrier options and their icicled variations. *The North American Journal of Economics and Finance* 49: 396–411. [CrossRef]

Lee, Hangsuck, Gaeun Lee, and Seongjoo Song. 2022. Multi-step reflection principle and barrier options. *The Journal of Futures Markets* 42: 692–721. [CrossRef]

Lee, Hangsuck, Hongjun Ha, and Minha Lee. 2021. Valuation of piecewise linear barrier options. *The North American Journal of Economics and Finance* 58: 101470. [CrossRef]

Lee, Hangsuck, Soohan Ahn, and Bangwon Ko. 2019b. Generalizing the reflection principle of Brownian motion and closed form pricing of barrier options and autocallable investments. *The North American Journal of Economics and Finance* 50: 1–13. [CrossRef]

Matsumoto, Makoto, and Takuji Nishimura. 1998. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Transactions on Modeling and Computer Simulation* 8: 3–30. [CrossRef]

Pötzelberger, Klaus, and Liqun Wang. 2001. Boundary crossing probability for Brownian motion. *Journal of Applied Probability* 38: 152–64. [CrossRef]

Wang, Liqun, and Klaus Pötzelberger. 1997. Boundary crossing probability for Brownian motion and general boundaries. *Journal of Applied Probability* 34: 54–65. [CrossRef]

Article

Enhancing Model Selection by Obtaining Optimal Tuning Parameters in Elastic-Net Quantile Regression, Application to Crude Oil Prices

Abdullah S. Al-Jawarneh [1], Ahmed R. M. Alsayed [2,3,*], Heba N. Ayyoub [4], Mohd Tahir Ismail [5], Siok Kun Sek [5], Kivanç Halil Ariç [6] and Giancarlo Manzi [2]

[1] Department of Mathematics, Faculty of Science, Jerash University, Jerash 26150, Jordan
[2] Department of Economics, Quantitative Methods, and Data Mining Centre, University of Milan, 20122 Milan, Italy
[3] Department of Economics, University of Bergamo, 24127 Bergamo, Italy
[4] Department of Mathematics, Faculty of Science, Philadelphia University, Amman 19392, Jordan
[5] School of Mathematical Sciences, Universiti Sains Malaysia, Gelugor 11700, Malaysia
[6] Faculty of Economics and Administrative Sciences, Sivas Cumhuriyet University, 58070 Sivas, Turkey
*   Correspondence: ahmed.alsayed@unimi.it

Abstract: Recently, there has been an increased focus on enhancing the accuracy of machine learning techniques. However, there is the possibility to improve it by selecting the optimal tuning parameters, especially when data heterogeneity and multicollinearity exist. Therefore, this study proposed a statistical model to study the importance of changing the crude oil prices in the European Union, in which it should meet state-of-the-art developments on economic, political, environmental, and social challenges. The proposed model is Elastic-net quantile regression, which provides more accurate estimations to tackle multicollinearity, heavy-tailed distributions, heterogeneity, and selecting the most significant variables. The performance has been verified by several statistical criteria. The main findings of numerical simulation and real data application confirm the superiority of the proposed Elastic-net quantile regression at the optimal tuning parameters, as it provided significant information in detecting changes in oil prices. Accordingly, based on the significant selected variables; the exchange rate has the highest influence on oil price changes at high frequencies, followed by retail trade, interest rates, and the consumer price index. The importance of this research is that policymakers take advantage of the vital importance of developing energy policies and decisions in their planning.

Keywords: quantile regression; tuning parameters; penalized regression; multicollinearity; heterogeneity; cross-validation; crude oil price

1. Introduction

In most research fields that have large time series data, like environmental, medical, marketing, etc., the data have great importance as information that need a development tool to reach the right decision. To detect more information around these data like patterns and trends, advanced machine learning (ML) has been used. ML is classified into two parts, namely supervised and unsupervised methods. Supervised ML algorithms build mathematical models to predict outcomes in the future. One of the main applications of supervised ML is regression analysis (Kassambara 2018; Ray 2019).

Regression analysis faces several challenges and affect prediction accuracy. For example, heterogeneity and the multicollinearity problem exist among the predictor variables; consequently, such a model is difficult to interpret (Qin et al. 2016; Alsayed et al. 2018; Al-Jawarneh et al. 2022). Many researchers continue to develop hybrid regression models to deal with these issues by improving the ordinary least squares method (OLS) method, such as the penalized regularization method; namely, Ridge regression (RR) and the least absolute

shrinkage and selection operator (LASSO) method. However, the RR method still cannot deal with the reduction of the predictor numbers; hence, the unnecessary predictor variables will still exist in the final model (Tibshirani 1996; Zou and Hastie 2005). Meanwhile, the LASSO method is inconsistent for variable selection and dealing with multicollinearity (Fan and Li 2001; Zou and Hastie 2005). Elastic-net (ELNET) methods (Zou and Hastie 2005) were proposed. This method represents a newly developed penalized regularization method for improving the model's interpretability and identifying relevant variables, considering that the procedures with the initial coefficient estimator used to compute the adaptive weights need not be consistent. In addition to that, quantile regression (QR) seeks to search for a model that minimizes the sum of the absolute residuals rather than the sum of the squared residuals. QR measures the effects of unobserved heterogeneity in the included variables. If the dependent variable distribution changes together with the independent variables, then the result is misleading when using the OLS regression, whereas QR shows how such changes in the independent variables affect the distribution shape of the dependent variable. Therefore, it will provide significant estimators for the changing of the heterogeneous distribution of the dependent variable (Alsayed et al. 2020). Moreover, there is a study that proposed the idea of penalized LASSO quantile regression that used the sum of the absolute values of the coefficients as the penalty (Li and Zhu 2008). In addition, recent research proposed an elastic net penalized quantile regression model approach that combines the strengths of the quantile loss and the Elastic net (Su and Wang 2021).

In penalized regression, the tuning parameters play a serious role in improving the penalty to realize the optimal estimation and consistent selection (Xiao and Sun 2019) when the tuning parameters have control of the coefficient shrinkage rate. For instance, if λ has too high a value, which leads to more shrinkage to be a small value or exactly equal zero, that means the model will be under-fitting (i.e., high bias and low variance). As the tuning parameter value increases, the bias increases, the variance decreases, and vice versa, in addition to the choosing of tuning parameter alpha (α) in the elastic net method which belongs to a value between zero and one, several of the studies work to fix this value at alpha equal to 0.5 or 0.75. So, choosing tuning parameter values is a difficult business and very sensitive (Fan and Tang 2013; Desboulets 2018).

To choose the tuning parameter, the literature advised some frequently used methods. These include the minimizing information criterion (IC), namely the Akaike information criterion (AIC) (Akaike et al. 1973), Bayes information criterion (BIC) (Schwarz 1978), Mallow Cp (Efron et al. 2004), and cross-validation (CV) (Stone 1974). The CV method is the simplest and most commonly used method in the literature for estimating and choosing the tuning parameter that has the minimizing CV sum of squared residuals (Chand 2012; Desboulets 2018), where the principle of the CV method presents a grid of λ values and computing the CV error for each λ after choosing the optimal λ, which has the smallest CV error (Gareth et al. 2013).

On the other hand, the importance of predicting and forecasting the energy market is highly needed to reach the optimal balanced point between energy, economics and environmental quality. Crude oil is an ingredient for sustainable economic growth, while the supply and demand are inelastic, and crude oil prices often experience sharp and sustained fluctuations (Alsayed and Manzi 2019). Therefore, this study examines the reaction of crude oil prices during the recent period, which has several economic shocks, particularly in the European Union; these include the global economic crisis in 2008, COVID-19 in 2020, and the recent wars in 2022–2024, which led to challenges for the global economy. Several studies have shown interest in examining the crude oil prices with several global and local factors (Aastveit et al. 2023; He et al. 2021; Kartal 2020; Baumeister and Kilian 2015; Doğrul and Soytas 2010; Amano and Van Norden 1998).

The significant contribution of this research is twofold: First, on the statistical aspect to deal with the heterogeneity, and second, to improve the accuracy of model selection by selecting the predictors that have the most effect on the response variable; the ELNET. QR regression at $\tau = 0.25, 0.5, 0.75$ will be used based on the D-fold C.V. method to select

the optimal tuning parameters (α_{opt} and λ_{opt}). Then, it is evaluated and compared with recently developed methods using both simulations and real applications. Additionally, regarding the novelty of the econometrics aspect, we model and predict the crude oil price using local and global variables—namely, the exchange rate and retail trade, interest rates, and consumer price index—to tackle the time series data that suffer from heterogeneity.

The advantage of using this approach is that this model is distinguished by its superior ability to deal with time series issues compared to the old models used and its ability to keep pace with developments in time series. However, there is the disadvantage that it does not have the oracle property, but this can be treated using the adaptive elastic net method in the future.

This research is organized as follows: Section 1 provides the introduction and literature review, Section 2 presents the methods, quantile regression, elastic net regression, D-fold cross-validation, and the proposed method. Section 3 explains the data and variables. Section 4 presents the empirical findings and discussion, and the Section 5 is the conclusion.

2. Methodology

This section briefly describes the applied methods. The first method is the QR regression method, which deals with heterogeneity problems. The second method pertains to the penalized regularization method by the Elastic-net (ELNET) method and D-fold cross-validation. Finally, this section discusses the proposed method provided, ELNET.QRα_{opt} regression.

2.1. Quantile Regression

QR regression is broadly applied, covering wide research areas. Koenker suggested a general approach of QR for longitudinal data (Koenker 2004). QR is used to estimate the conditional median and any other quantiles of the dependent predictor variables, and it could tackle the unobserved heterogeneity effects. The QR could describe that relationship at different points in the conditional median or quantiles distribution of dependent variable $Q_{y/X}(\tau)$, where τ is the quantiles or percentiles and takes value from $0 < \tau < 1$ (Ambark et al. 2023). The model structure of the multiple linear regression is

$$y = X^T \beta^\tau + \varepsilon \tag{1}$$

where $[y]_{n \times 1}$ is a vector of the response variable, $[X]_{n \times p}$ is a matrix of the predictor variables, $[\beta^\tau]_{p \times 1}$ is the unknown vector of the regression coefficients associated with the τth quantile, $[\varepsilon]_{n \times 1}$ is a vector of the random observation errors that are supposed to be a normal distribution with zero mean error term and variance $E(\varepsilon) = \sigma^2 I_n$.

The linear quantile regression model assumes:

$$Q_{y/X}(\tau) = X^T \beta^\tau \; ; \; \beta^\tau = \left\{ \beta_0^\tau, \beta_1^\tau, \dots, \beta_p^\tau \right\}' \tag{2}$$

where β^τ is the quantile coefficient. Then, the τth quantile regression estimator minimizes the objective function β^τ (Davino et al. 2013), which is given by:

$$\hat{\beta}_\tau = \min_{\beta} \sum_{i=1}^n \rho_\tau \left(y_i - x_i^T \beta \right) \tag{3}$$

where x_i^T is the ith row of X and $\rho_\tau(v)$ is a loss function defined as follows:

$$\rho_\tau(v) = v \left(\tau - I_{\{v<0\}} \right)); \; 0 < \tau < 1 \text{ and } v \in R \tag{4}$$

To improve quantile regression and regularization, Koenker suggested a penalized version, as follows:

$$\hat{\beta}_\tau = \min_{\beta} \sum_{i=1}^n \rho_\tau \left(y_i - x_i^T \beta \right) + P_\lambda(\beta) \tag{5}$$

where $P(\beta)$ is the penalty function and λ is the tuning, which is greater than zero.

2.2. Elastic Net Regression

ELNET regression is proposed to deal with the limitations of LASSO and improve the interpretability model and accuracy prediction by combining two penalties; namely, L_1 penalty (LASSO) and L_2 penalty (RR) (Zou and Hastie 2005; Zou and Zhang 2009; Friedman et al. 2010; Lee et al. 2016). The ELNET estimator is given as follows:

$$\hat{\beta}^{ELNET} = \min_{\beta}\sum_{i=1}^{n} \left(y_i - x_i^T\beta \right) + \lambda_1\|\beta\|_1 + \lambda_2\|\beta\|_2^2 \qquad (6)$$

where $\|\beta\|_1 = \sum_{j=1}^{p}|\beta_j|$ is L_1-norm of β, $\|\beta\|_2^2 = \sum_{j=1}^{p}(\beta_j)^2$ is L_2-norm of β, and λ_1 and λ_2 are the tuning λ_1, $\lambda_2 > 0$. These functions control the strength of shrinkage of the predictor variables. The values of λ_1 and λ_2 are dependent on the dataset, and they are automatically selected using CV (Zou and Hastie 2005; Melkumova and Shatskikh 2017; Masselot et al. 2018; Al-Jawarneh and Ismail 2024). The best values of tuning parameters λ_1 and λ_2 can be defined as the minimum mean squared error (*MSE*) (Friedman et al. 2010; Lee et al. 2016).

Equation (7) becomes equivalent to the following by denoting $\lambda_1 = 2n\lambda\alpha$ and $\lambda_2 = n\lambda(1-\alpha)$ (Haws et al. 2015; Al-Jawarneh et al. 2022):

$$\hat{\beta}^{ELNET} = \min_{\beta}\sum_{i=1}^{n} \left(y_i - x_i^T\beta \right) + \lambda\left(\alpha\|\beta\|_1 + \frac{(1-\alpha)}{2}\|\beta\|_2^2 \right) \qquad (7)$$

where α is a regularization parameter between zero and one. The ELNET estimation undergoes the RR estimator when $\alpha = 0$, whereas it is subject to the LASSO estimator when $\alpha = 1$.

2.3. D-Fold Cross-Validation

The *D*-fold cross-validation (*D*-CV) method was proposed by (Geisser 1975). *D*-CV idea is to split the dataset into *D* folds nearly equal in size. After that, the $D-1$ folds are used as the training set for the estimation of the model, while the dth fold is used as a test set to assess the predictive performance of the model. This process repeats until every *D* fold serves as the test set. Then, the mean prediction error over all folds is calculated (van Houwelingen and Sauerbrei 2013; Gareth et al. 2013). The *D*-CV algorithm is explained in Algorithm 1 (Hastie et al. 2009; Melkumova and Shatskikh 2017; Hastie et al. 2015) and is repeated for each fold as shown in Figure 1.
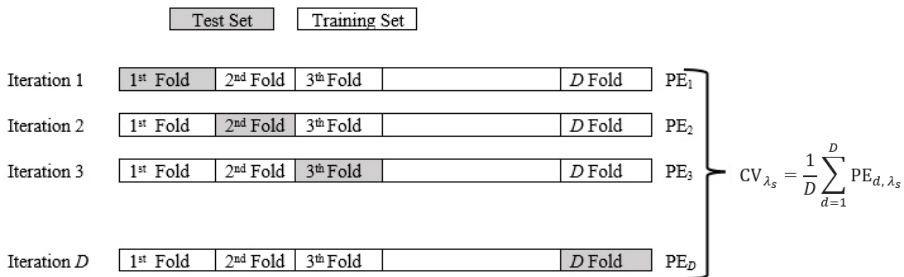

Figure 1. Diagram of *D*-fold cross-validation (*D*-CV).

Algorithm 1: *D*-fold Cross-Validation

1- Randomly, split the whole dataset of size *n* into *D* folds of roughly equal size.

2- ● For a grid of *S* values of λ: *s* in $\underline{1:S}$.

  ● For *d* in $\underline{1:D}$.

    ● Consider $D − 1$ folds as a training set, and d^{th} fold as a test set.
    ● By the training set, using an estimation method to estimate the regression coefficient at λ_s value, and denoted by the fitted function $\hat{f}_{d,\lambda_s}(z)$.
    ● Calculate the prediction error (PE) on the test set.

$$PE_{d,\lambda_s} = \sum_{i=1}^{n_d} \left(V_i - \hat{f}_{d,\lambda_s}(z_i) \right)^2 \tag{8}$$

    ● End for *d*.

  ● Repeat for $d = 1, 2, \ldots, D$
  ● For each λ_s overall fold, calculate the average of *D* prediction errors.

$$CV_{\lambda_s} = \frac{1}{D} \sum_{d=1}^{D} PE_{d,\lambda_s} \tag{9}$$

    ● End for s.
  ● Repeat for *s* in $\overline{1:S}$

3- Choose the optimal λ that gives a minimum average CV.

$$\lambda_{opt} = argmin_{s=1,2,S} \left\{ CV_{\lambda_s} \right\}$$

End.

Figure 1 describes the process of *D*-CV where the dataset is split for *D* folds. Each row represents one iteration and each column represents one fold. For example, in the first iteration, the 1st fold represents the test set, whereas the remaining $D − 1$ folds represent the training set. In the second iteration, the 2nd fold represents the test set, and the other folds are the training set. Thus, repeat this until every *D* fold serves as the test set.

The size of each training set is equal to $[(D − 1)n]/D$ observations, as the *D* value increase leads to a decrease in the bias of the fit model, whereas the variance will increase, as will the correlation among the fitted model because of the overlap among the training sets (Gareth et al. 2013). Usually, the *D* value is chosen at 5 or 10, where these values yield estimates that achieve an intermediate level of bias and are not excessively biased nor from high variance; Thus, *D* = 5 or 10 involve the bias-variance trade-off (Gareth et al. 2013; Kuhn and Johnson 2013; Al-Jawarneh and Ismail 2024).

2.4. Proposed Penalized Quantile Regression Method

The Elastic-net quantile regression method based on D-CV (ELNET.QR) is presented to explain the significance of the predictor variables on the response variable and enhance the prediction error of the final model based on the optimal tuning parameters as follows:

1. Apply the QR method at τ = (0.25, 0.50, 0.75) using all the variables:

$$\hat{\beta}_{\tau}^{QR} = \min_{\beta} \sum_{i=1}^{n} \rho_\tau \left(y_i - x_i^T \beta \right) \tag{10}$$

2. Using the training set only, select the optimal parameters via the *D*-CV method at *D* = 10 as follows:

  i. The regularization parameter α_{opt} value of the sequence $0 < \alpha < 1$, where α_{opt} represents the relative contribution of the L_1 penalty versus L_2 penalty.

$$\alpha_{opt} = argmin_{k=\overline{1:K}} \left\{ CV_{\alpha_k} \right\} ; \tag{11}$$

$$CV_{\alpha_k} = \frac{1}{10}\sum_{d=1}^{10} PE_{\alpha_k} \; ; \; \alpha_k \in (0,1)$$

where k represents the number of α values between one and zero and will be chosen. In this study, we choose $K = 50$.

ii.   The tuning parameter λ_{opt} value is at α_{opt}

$$\lambda_{opt} = argmin_{s=\overline{1:5}}\left\{ CV_{\alpha_{opt}, \lambda_s} \right\}; \tag{12}$$

$$CV_{\alpha_{opt}, \lambda_s} = \frac{1}{10}\sum_{d=1}^{10} MSE_{\alpha_{opt}, \lambda_s}$$

3.   Based on the Equations (7) and (10) at α_{opt} and λ_{opt}, the ELNET penalized regression is used as the following formula:

$$\hat{\beta}_\tau^{ELNET. QR} = \min_\beta \sum_{i=1}^{n} \rho_\tau\left(y_i - x_i^T\beta \right) + \lambda_{opt}\left(\alpha_{opt}\|\hat{\beta}\|_1 + \frac{(1 - \alpha_{opt})}{2}\|\hat{\beta}\|_2^2 \right); \tag{13}$$

$$\rho_\tau(v) = v\left(\tau - I_{\{v<0\}} \right)$$

Finally, a comparison was made between the proposed methods with traditional methods. The performance of the proposed estimated method has been tested by using several well-known criteria, namely; Residual Sum of Squares (*RSS* (Equation (14)), root mean square error (*RMSE*; Equation (15)), mean absolute error (*MAE*; Equation (16)), mean absolute percentage error (*MAPE*; Equation (17)), and mean absolute scaled error (*MASE*; Equation (18)).

$$RSS = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{14}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{l=1}^{n}(y_l - \hat{y}_l)^2} \tag{15}$$

$$MAE = \frac{1}{n}\sum_{l=1}^{n}|y_l - \hat{y}_l| \tag{16}$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left| \frac{y_i - \hat{y}_i}{y_i} \right| \tag{17}$$

$$MASE = \frac{1}{n}\sum_{l=1}^{n} \frac{|y_l - \hat{y}_l|}{\frac{1}{n-1}\sum_{l=2}^{n}|y_l - y_{l-1}|} \tag{18}$$

3. Application

This section implemented the numerical simulation experiment and a real dataset application to show the capacity of the proposed methods. The analyses are performed using open-source R 4.3.1 software by using the *hqreg* package and our developed code to calculate the function of obtaining the best tuning parameters value for ELNET.QR regression.

3.1. Simulation Study

In this section, we present the results of the numerical simulation for the eight methods: namely; RR.QR, LASSO.QR, ELNET.QR at the best α_{opt} value, ELNET.QR at $\alpha = 0.25$, ELNET.QR at $\alpha = 0.5$, ELNET.QR at $\alpha = 0.75$, AdLASSO.QR method based on the minimum *MSE* (λ_{min}), and minimum *MSE* with one standard error (λ_{1se}) of weighted RR. We evaluate and illustrate these eight methods of variable selection and prediction

performance under a normal distribution. Simulation scenarios considered three QR levels at $\tau = 0.25$, 0.5, 0.75, a sample size of n = 150, and iteration= 1000. The 10-CV was applied to select the best tuning parameter values. The simulated data are split into two parts: 70% for training, and 30% for testing the estimated models then evaluated by using the performance criteria.

3.2. Application Datasets

The European Union (EU) is an economic and political union of 27 countries. It operates an internal (or single) market, which allows the free movement of goods, capital, services, and people between member states. In recent years, the European Union has made significant and noticeable progress in implementing new policies in energy consumption to shift to low carbon emissions, while crude oil market prices have experienced important changes and is more volatile than the price of other tradable commodities that have negative effects on investment.

The dataset of energy oil prices with the relevant affected factors is included to evaluate the performance of the penalized regression methods. The explanatory variable is crude oil prices (y) measured by local currency per barrel, while the predictor variables are the consumer price index, retail trade, the foreign exchange rate (Euro/USD), and interest rates represented by $x_1, \ldots, x_4$ respectively. The data are collected on a monthly basis from the beginning of the year 2000 to the end of the year 2022 to include the highest affected period, such as the financial crisis in 2008, COVID-19 in 2020, and the Russian-Ukraine war in 2022. The data for all variables are gathered from Bloomberg (2023). There are several variables used to examine oil price changes, but we included the variables that are available every quarter, and those variables have effects on changes in the oil price according to economic theory, policy, and literature.

Recent literature included those variables as independent variables to measure the effects on crude oil prices (Kirikkaleli and Doğan 2021; Yılmaz and Altay 2016; Alsayed 2023; Amano and Van Norden 1998). Their findings showed that it significantly affects oil prices at various levels. Other studies used advanced statistical methods such as a multivariate adaptive regression splines model to detect the effect of foreign exchange (USD-TRY), credit default swap spread, global uncertainty, and global volatility on local currency oil prices at a local economy level in Turkey during the COVID-19 pandemic using daily data from July 2019 to October 2020 (Kartal 2020). Doğrul and Soytas (2010) detected the relationship between oil prices, interest rate, economic activity, and unemployment in Turkey by applying the Toda–Yamamoto technique. Their findings support that the volatility index is the most important factor influencing crude oil prices.

Descriptive statistics of our dataset are presented in Figure 2a–c. We can observe that crude oil prices were facing three serious shocks with an exponential increase in 2008, the spread of the COVID-19 pandemic in 2020, and the recent oil price slump in 2022, which have substantially raised the economic uncertainty and geopolitical risk levels. The combination of those economic degradations will likely initiate a long-term economic downturn and drive the European Union economy into the next recession.

The estimated model consists of four independent variables $x_1, \ldots, x_4$, and the terms of interaction between variables are $x_5, \ldots, x_{10}$. The dataset is divided into two parts: 70% (192 cases) which used for the training dataset and the remaining 30% of the dataset (83 cases) is used for testing. The whole dataset has been made stationarity at first difference and then standardized before doing the analysis. The interested estimated model is as follows.

$$y_t = \alpha_t + \beta_{1t}x_1 + \beta_{2t}x_2 + \beta_{3t}x_3 + \beta_{4t}x_4 + \beta_{5t}x_5 + \beta_{6t}x_6 + \beta_{7t}x_7 + \beta_{8t}x_8 + \beta_{9t}x_9 + \beta_{10t}x_{10} + \varepsilon_t$$

where y represents the oil price at day t in Europe, x_1 is the consumer price index, x_2 is retail trade, x_3 is exchange rate, x_4 is interest rates, $x_5 = x_1x_2$, $x_6 = x_1x_3$, $x_7 = x_1x_4$, $x_8 = x_2x_3$, $x_9 = x_2x_4$, $x_{10} = x_3x_4$ and ε_t is the error term.

(**a**)

(**b**)

(**c**)
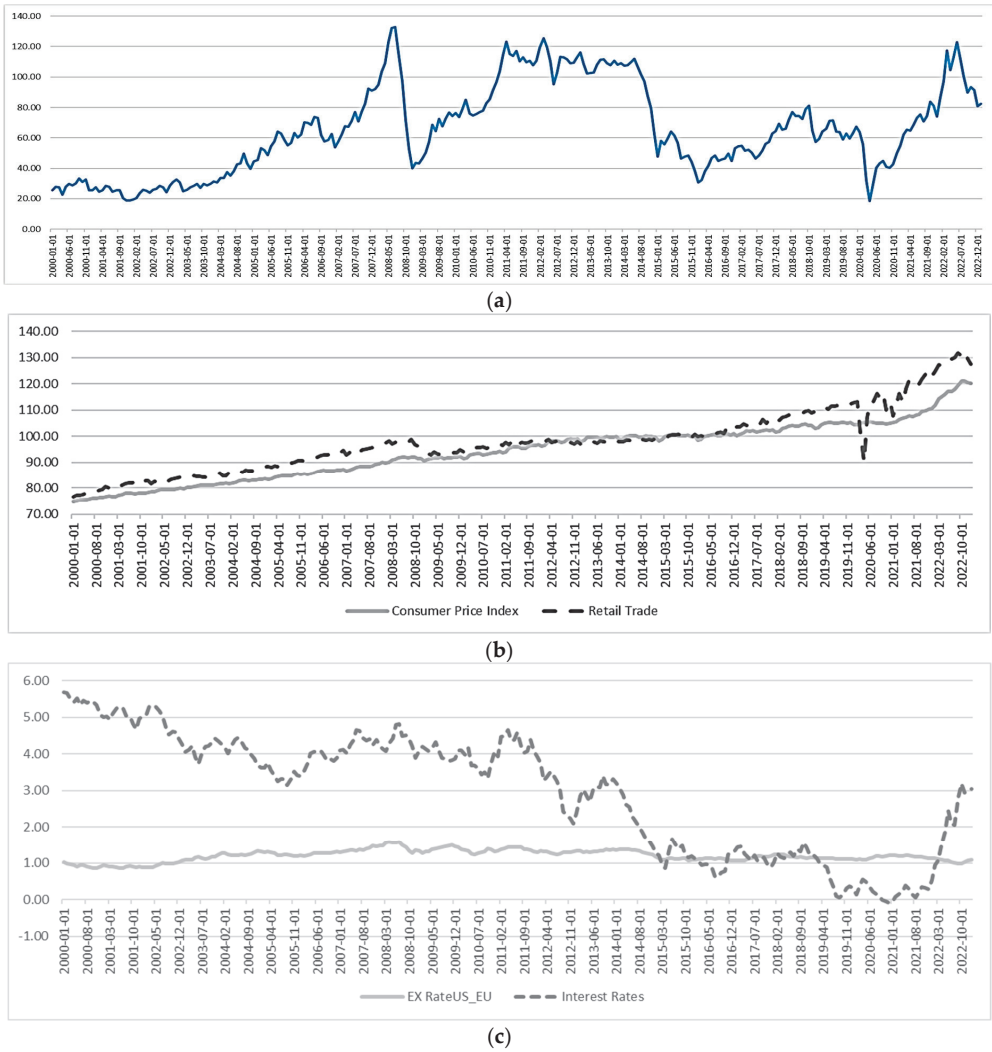
Figure 2. (**a**) The monthly crude oil prices in the European Union. (**b**) The consumer price index and retail trade in the European Union. (**c**) The monthly exchange rate and interest rates in the European Union.

4. Results and Discussion

In this section, we provide the result of the numerical experiment and the application based on the real dataset.

4.1. Simulation Results

Table 1 describes the average of the performance criteria in terms of *RSS*, *RMSE*, *MAE*, *MASE*, and *MAPE* for all the regression methods used in this study. In cases of study at τ = (0.25, 0.50, 0.75), the results show that the proposed regression method ELNET.QR $\alpha_{opt}, \lambda_{min}$ (by determining the best optimal α value) has the smallest error value in these criteria tests. Therefore, ELNET.QR $\alpha_{opt}, \lambda_{min}$ improves the prediction accuracy by producing the smallest error values in terms of *RSS*, *RMSE*, *MAE*, *MASE*, and *MAPE*.

Table 1. Performance criteria of the simulation scenarios.

Method	λ	RSS	RMSE	MAE	MAPE	MASE
		$\tau = 0.25$				
RR.QR	λ_{min}	35.5827	0.88624	0.71977	3.657265	0.794615
	λ_{1se}	39.4718	0.93395	0.76183	4.024955	0.841051
LASSO.QR	λ_{min}	14.1830	0.56093	0.40053	**1.746325**	0.442176
	λ_{1se}	15.3082	0.58278	0.43724	1.999109	0.482706
AdLASSO.QR (RR.W. λ_{min})	λ_{min}	14.5157	0.56762	0.41579	2.082011	0.459031
	λ_{1se}	16.0379	0.59633	0.45133	2.030926	0.498263
AdLASSO.QR (RR.W. λ_{1se})	λ_{min}	14.7190	0.57034	0.41830	2.092005	0.461806
	λ_{1se}	16.2224	0.59878	0.45343	2.044566	0.500584
ELNET.QR $\alpha = 0.25$	λ_{min}	19.0338	0.64701	0.50269	2.272781	0.554958
	λ_{1se}	22.9179	0.71031	0.56536	2.504381	0.624143
ELNET.QR $\alpha = 0.5$	λ_{min}	14.7433	0.57177	0.41665	1.89485	0.459972
	λ_{1se}	17.4372	0.62082	0.47837	2.131715	0.528107
ELNET.QR $\alpha = 0.75$	λ_{min}	14.2895	0.56306	0.40377	1.780631	0.445752
	λ_{1se}	15.7379	0.59067	0.44563	2.044958	0.491972
ELNET.QR α_{opt}	λ_{min}	**13.9615**	**0.55671**	**0.39654**	**1.776101**	**0.437778**
	λ_{1se}	15.7195	0.59049	0.44696	2.020386	0.493431
		$\tau = 0.5$				
RR.QR	λ_{min}	30.8563	0.82567	0.69494	2.127862	0.76720
	λ_{1se}	33.5667	0.86207	0.72893	**2.019126**	0.80472
LASSO.QR	λ_{min}	11.7178	0.50823	0.39718	2.592074	0.438483
	λ_{1se}	13.6771	0.54948	0.45072	2.456135	0.497588
AdLASSO.QR (RR.W. λ_{min})	λ_{min}	13.385	0.53733	0.44051	2.604391	0.486318
	λ_{1se}	14.9137	0.56951	0.46914	2.488478	0.517925
AdLASSO.QR (RR.W. λ_{1se})	λ_{min}	13.4163	0.53805	0.44168	2.602395	0.487607
	λ_{1se}	14.9282	0.56976	0.46944	2.48853	0.518257
ELNET.QR $\alpha = 0.25$	λ_{min}	16.1670	0.59738	0.48819	2.668777	0.538961
	λ_{1se}	19.3541	0.65371	0.53790	2.602162	0.593840
ELNET.QR $\alpha = 0.5$	λ_{min}	13.1976	0.53957	0.44388	2.665215	0.490041
	λ_{1se}	15.5554	0.58596	0.47915	2.659132	0.528980
ELNET.QR $\alpha = 0.75$	λ_{min}	12.3296	0.52141	0.42639	2.639871	0.470737
	λ_{1se}	14.3324	0.56247	0.46268	2.544796	0.510792
ELNET.QR α_{opt}	λ_{min}	**10.3090**	**0.47689**	**0.37236**	2.564913	**0.411081**
	λ_{1se}	13.6800	0.54953	0.45087	2.515077	0.497758
		$\tau = 0.75$				
RR.QR	λ_{min}	42.7458	0.96993	0.79680	6.288212	0.879659
	λ_{1se}	46.8397	1.01652	0.83617	**6.47382**	0.923126
LASSO.QR	λ_{min}	12.9671	0.53623	0.45390	4.716719	0.501104
	λ_{1se}	14.3718	0.5644	0.46955	4.696969	0.518382
AdLASSO.QR (RR.W. λ_{min})	λ_{min}	14.4291	0.55778	0.47079	4.889332	0.519747
	λ_{1se}	15.8043	0.58392	0.48687	4.910642	0.537499

Table 1. *Cont.*

Method	λ	RSS	RMSE	MAE	MAPE	MASE
AdLASSO.QR (RR.W. λ_{1se})	λ_{min}	14.5222	0.55917	0.47169	4.892964	0.520746
	λ_{1se}	15.8926	0.58537	0.48803	4.916813	0.538782
ELNET.QR $\alpha = 0.25$	λ_{min}	20.6517	0.67191	0.55133	5.073862	0.608665
	λ_{1se}	23.6645	0.71873	0.58854	5.305833	0.649741
ELNET.QR $\alpha = 0.5$	λ_{min}	13.7992	0.55295	0.46196	4.620298	0.510000
	λ_{1se}	15.1615	0.57941	0.47896	4.653668	0.528763
ELNET.QR $\alpha = 0.75$	λ_{min}	13.0885	0.53868	0.45467	4.68214	0.501955
	λ_{1se}	14.5195	0.56725	0.47102	4.675978	0.520007
ELNET.QR α_{opt}	λ_{min}	**12.8468**	**0.53376**	**0.45244**	4.714046	**0.499487**
	λ_{1se}	14.3309	0.56360	0.46864	4.683324	0.517377

Note: The bold number indicates the lowest values in favor of the superior performance method compared to the other methods.

4.2. Application Results and Discussion

Figure 3 shows the curve of RSS to choose the optimal alpha (α_{opt}) in three cases of $\tau = 0.25,\ 0.50,\ 0.75$ he y-axis represents the estimation of the RSS and the x-axis represents the alpha values. At $\tau = 0.25$, the minimum RSS value appears at $\alpha_{opt} = 0.38$, whereas at $\tau = 0.50, 0.75$, the minimum RSS values are at 0.02 in two cases. These results indicate that the optimal alpha can reduce the RSS value more than the traditional methods for chosen alpha at fixed points like $\alpha = 0.25,\ 0.5,\ 0.75$, or other methods like lasso and ridge methods.
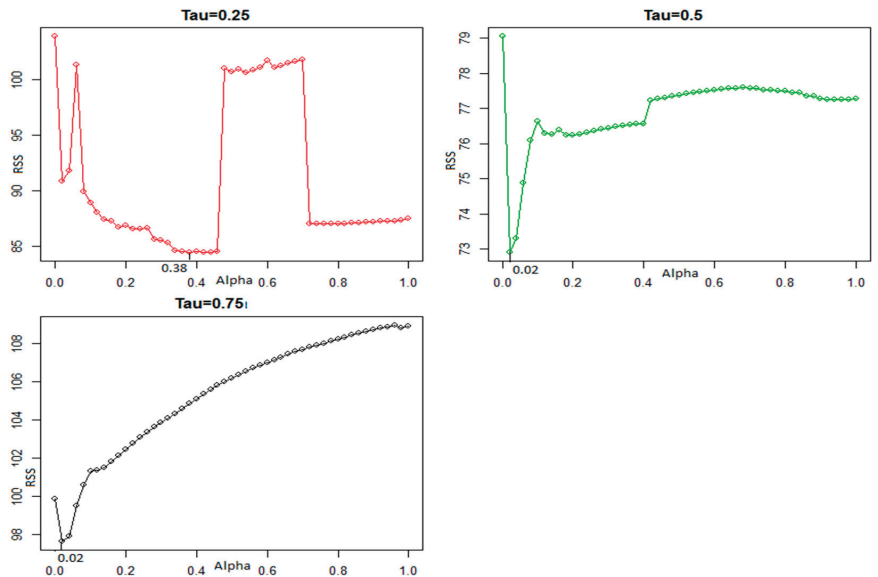

Figure 3. 10-CV estimation for choosing the α_{opt} of ELNET.QR ($\tau = 0.25,\ 0.50,\ 0.75$).

Figure 4 shows the 10-CV estimate plot of the ELNET.Qr α_{opt}; ($\alpha_{opt} = 0.38,\ 0.02,\ \text{and } 0.02$) and $\tau = 0.25, 0.50, 0.75$, respectively. In each plot, the red dotted line is the mean square error (MSE) curve with one standard error band along the error bars. The y-axis represents the MSE and the x-axis denotes the $log(\lambda)$ function. The upper horizontal line of the plot represents the numbers of nonzero regression coefficients in the model at $log(\lambda)$ value. The

first vertical dotted lines from the right represent the point selected at a minimum of the MSE (λ_{min}) rule, while the second vertical line denotes the location of the point selected at a minimum of MSE with the one standard error (λ_{1se}) rule. These two lines show the numbers of nonzero regression coefficients selected at λ_{1se} and λ_{min} rules. The increase in λ value leads to a decrease in the number of non-zero coefficients in the model. Therefore, the selection of the λ is based on the optimal minimum MSE value.
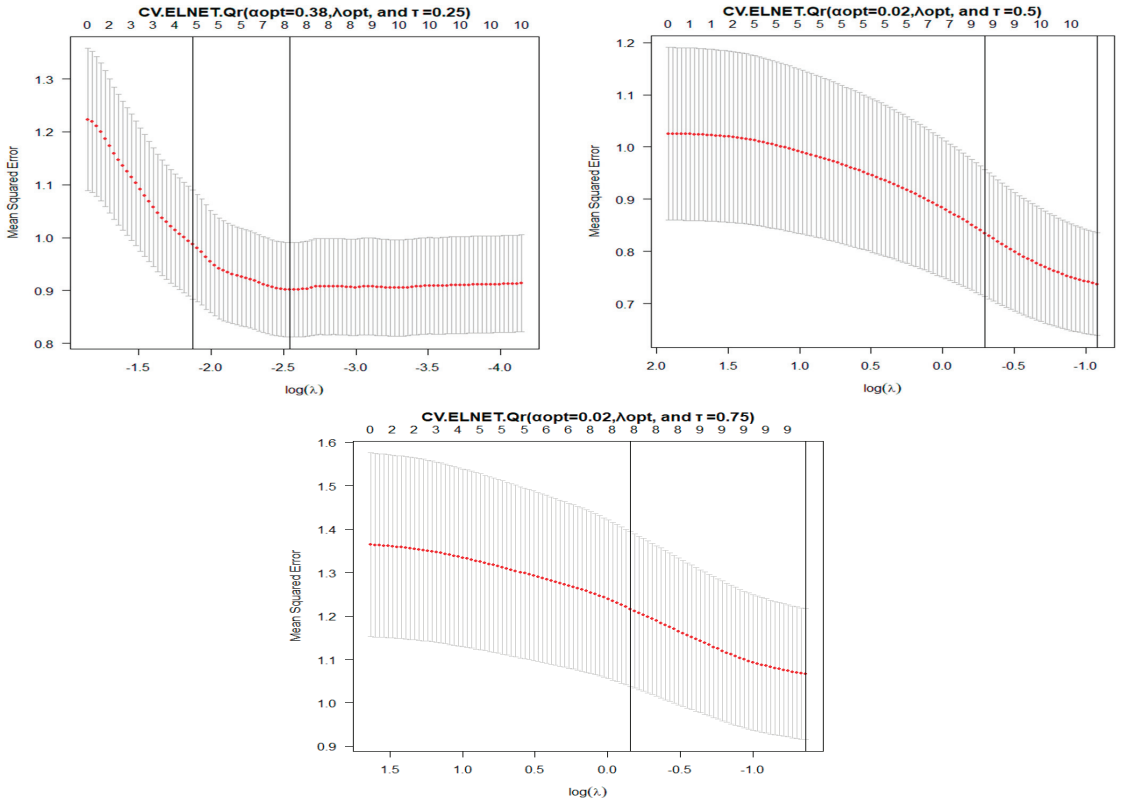

Figure 4. 10-CV estimation of the ELNET.QR α_{opt} at $\tau = 0.25, 0.50, 0.75$.

Figure 5 illustrates the relationship between $log(\lambda)$ and the selected nonzero coefficient estimation in the ELNET.QR $\alpha_{opt}(0.38, 0.02, \text{and } 0.02)$ at current λ, which represents the actual degrees of freedom. All methods have regularization and variable selection. In each one of the figures from right to left, the estimation of the coefficients decreases toward zero with the increase in the λ value and forces it to become zero for the unnecessary coefficient estimation (i.e., $if \ \lambda \to \infty$, then the estimated coefficients $\to 0$). For instance, at $\tau = 0.25$, the ELNET.QR as the $\alpha_{opt} = 0.38$ method selected eight nonzero coefficients at λ_{min} and five nonzero coefficients at λ_{1se} with different significant strengths.

Table 2 display the RSS values and the number of variable selections (Num. of V.S.) of the proposed method compared to those of the previous methods in the testing datasets. Based on the RSS values, the order of the proposed methods among all used methods in this study is as follows: The first-order method among all methods is achieved by ELNET.QR $\alpha_{opt} = 0.38$, $\tau = 0.25$ $(\lambda_{min} = 0.0787; RSS = 84.426;$ Num. of V.S. = 8), c, $\tau = 0.5$ $(\lambda_{min} = 0.3392; RSS = 72.9008;$ Num. of V.S. = 10). At $\tau = 0.75$ $(\lambda_{min} = 0.2555;$ $RSS = 97.6474;$ Num. of V.S. = 9), it has the smallest RSS value.
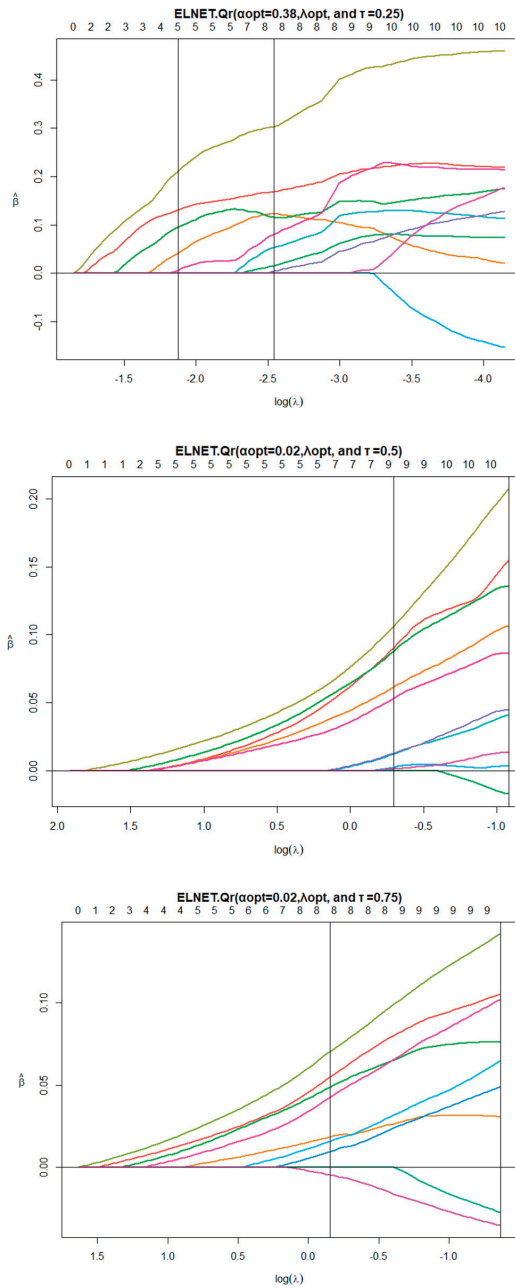
Figure 5. Coefficient estimation of the ELNET.QR α_{opt}.

Table 2. Selected number of variables and RSS error values.

Method	λ	*RSS*	Num. of V.S.	V.S.
RR	$\lambda_{min} = 0.041839$	109.177	10	$x_1, \ldots, x_{10}$
	$\lambda_{1se} = 0.74835$	77.0141	10	$x_1, \ldots, x_{10}$
LASSO	$\lambda_{min} = 0.003994$	120.091	9	$x_1, \ldots, x_4, x_6, \ldots, x_{10}$
	$\lambda_{1se} = 0.078399$	82.3651	7	$x_1, \ldots, x_4, x_6, x_8, x_{10}$
ELNET $\alpha = 0.25$	$\lambda_{min} = 0.013263$	116.129	10	$x_1, \ldots, x_{10}$
	$\lambda_{1se} = 0.260352$	76.9288	7	$x_1, \ldots, x_4, x_6, x_8, x_{10}$
ELNET $\alpha = 0.5$	$\lambda_{min} = 0.007987$	117.951	9	$x_1, \ldots, x_4, x_6, \ldots, x_{10}$
	$\lambda_{1se} = 0.142868$	79.5524	7	$x_1, \ldots, x_4, x_6, x_8, x_{10}$
ELNET $\alpha = 0.75$	$\lambda_{min} = 0.005844$	118.623	9	$x_1, \ldots, x_4, x_6, \ldots, x_{10}$
	$\lambda_{1se} = 0.104532$	80.8656	7	$x_1, \ldots, x_4, x_6, x_8, x_{10}$
	$\tau = 0.25$			
RR.QR	$\lambda_{min} = 0.1322$	103.907	10	$x_1, \ldots, x_{10}$
	$\lambda_{1se} = 0.5480$	92.046	10	$x_1, \ldots, x_{10}$
LASSO.QR	$\lambda_{min} = 0.0471$	87.492	4	$x_1, \ldots, x_4$
	$\lambda_{1se} = 0.0720$	93.659	3	x_1, x_2, x_4
AdLASSO.QR (RR.W. λ_{min})	$\lambda_{min} = 0.0008$	100.139	4	x_1, x_3, x_4, x_{10}
	$\lambda_{1se} = 0.0028$	100.797	3	x_1, x_3, x_4
AdLASSO.QR (RR.W. λ_{1se})	$\lambda_{min} = 0.0003$	96.542	4	$x_1, \ldots, x_4$
	$\lambda_{1se} = 0.0012$	99.369	3	x_1, x_3, x_4
ELNET.QR $\alpha = 0.25$	$\lambda_{min} = 0.0911$	86.377	8	$x_1, \ldots, x_6, x_8, x_{10}$
	$\lambda_{1se} = 0.2002$	87.614	5	$x_1, \ldots, x_4, x_{10}$
ELNET.QR $\alpha = 0.5$	$\lambda_{min} = 0.0317$	100.762	9	$x_1, \ldots, x_6, x_8, x_9, x_{10}$
	$\lambda_{1se} = 0.1275$	90.6738	4	$x_1, \ldots, x_4$
ELNET.QR $\alpha = 0.75$	$\lambda_{min} = 0.054$	87.045	5	$x_1, \ldots, x_4, x_6$
	$\lambda_{1se} = 0.0931$	92.967	4	$x_1, \ldots, x_4$
ELNET.QR $\alpha_{opt} = 0.38$	$\lambda_{min} = 0.0787$	**84.426**	8	$x_1, \ldots, x_6, x_8, x_{10}$
	$\lambda_{1se} = 0.1532$	89.458	5	$x_1, \ldots, x_4, x_{10}$
	$\tau = 0.5$			
RR.QR	$\lambda_{min} = 0.0678$	79.0412	10	$x_1, \ldots, x_{10}$
	$\lambda_{1se} = 0.6563$	74.8137	10	$x_1, \ldots, x_{10}$
LASSO.QR	$\lambda_{min} = 0.0228$	77.2748	8	$x_1, \ldots, x_4, x_6, x_8 \ldots, x_{10}$
	$\lambda_{1se} = 0.0531$	75.0064	5	$x_1, \ldots, x_4, x_8$
AdLASSO.QR (RR.W. λ_{min})	$\lambda_{min} = 0.0006$	79.3498	3	x_1, x_3, x_4
	$\lambda_{1se} = 0.0035$	77.0606	2	x_1, x_3
AdLASSO.QR (RR.W. λ_{1se})	$\lambda_{min} = 0.0001$	73.2189	5	$x_1, \ldots, x_4, x_{10}$
	$\lambda_{1se} = 0.0009$	80.3982	3	x_1, x_3, x_4
ELNET.QR $\alpha = 0.25$	$\lambda_{min} = 0.0390$	76.283	9	$x_1, \ldots, x_6, x_8, x_9, x_{10}$
	$\lambda_{1se} = 0.1719$	74.4287	8	$x_1, \ldots, x_4, x_6, x_8 \ldots, x_{10}$
ELNET.QR $\alpha = 0.5$	$\lambda_{min} = 0.0357$	77.3792	9	$x_1, \ldots, x_6, x_8, x_9, x_{10}$
	$\lambda_{1se} = 0.1000$	74.9740	6	$x_1, \ldots, x_4, x_8, x_{10}$
ELNET.QR $\alpha = 0.75$	$\lambda_{min} = 0.0277$	77.5181	8	$x_1, \ldots, x_4, x_6, x_8, x_9, x_{10}$
	$\lambda_{1se} = 0.0687$	74.9286	6	$x_1, \ldots, x_4, x_8, x_{10}$

Table 2. *Cont.*

Method	λ	*RSS*	Num. of V.S.	V.S.
ELNET.QR $\alpha_{opt}=0.02$	$\lambda_{min}=0.3392$	**72.9008**	10	$x_1,\ldots,x_{10}$
	$\lambda_{1se}=0.7450$	74.9729	9	$x_1,\ldots,x_4,x_6,x_7\ldots,x_{10}$
	$\tau=0.75$			
RR.QR	$\lambda_{min}=0.0511$	99.8799	10	$x_1,\ldots,x_{10}$
	$\lambda_{1se}=0.6298$	99.3733	10	$x_1,\ldots,x_{10}$
LASSO.QR	$\lambda_{min}=0.0053$	108.9124	8	$x_1,\ldots,x_4,x_6,x_8,\ldots,x_{10}$
	$\lambda_{1se}=0.0354$	101.6289	5	x_1,x_3,x_4,x_6,x_{10}
AdLASSO.QR (RR.W. λ_{min})	$\lambda_{min}=0.0004$	113.2957	5	x_1,x_3,x_4,x_6,x_{10}
	$\lambda_{1se}=0.0011$	118.7315	4	x_1,x_3,x_6,x_{10}
AdLASSO.QR (RR.W. λ_{1se})	$\lambda_{min}=0.0001$	100.7581	5	x_1,x_3,x_4,x_6,x_{10}
	$\lambda_{1se}=0.0003$	105.5192	3	x_1,x_3,x_4
ELNET.QR $\alpha=0.25$	$\lambda_{min}=0.0204$	103.2303	8	$x_1,\ldots x_4,x_6,x_8\ldots,x_{10}$
	$\lambda_{1se}=0.1334$	102.9762	6	$x_1 x_3,x_4,\ x_6,x_8,x_{10}$
ELNET.QR $\alpha=0.5$	$\lambda_{min}=0.0102$	106.1962	8	$x_1,\ldots x_4,x_6,x_8\ldots,x_{10}$
	$\lambda_{1se}=0.0688$	101.7987	5	$x_1 x_3,x_4,\ x_6,x_{10}$
ELNET.QR $\alpha=0.75$	$\lambda_{min}=0.0068$	107.9645	8	$x_1,\ldots x_4,x_6,x_8\ldots,x_{10}$
	$\lambda_{1se}=0.0458$	101.4287	5	$x_1 x_3,x_4,\ x_6,x_{10}$
ELNET.QR $\alpha_{opt}=0.02$	$\lambda_{min}=0.2555$	**97.6474**	9	$x_1,\ldots,x_6,x_8\ldots,x_{10}$
	$\lambda_{1se}=0.8571$	105.4114	8	$x_1,\ldots x_4,x_6,x_8\ldots,x_{10}$

Note: The bold number indicates the lowest values in favor of the superior performance method compared to the other methods.

Table 3 shows the performance criteria of the prediction accuracy for comparing the penalized QR regressions methods by using *RMSE*, *MAE*, *MAPE*, and *MASE*. The proposed method is ELNET.QR α_{opt}. In three cases of study at $\tau=$ 0.25, 0.5, 0.75, it provides the smallest error value in terms of *RMSE*, *MAE*, *MASE*, and *MAPE*. For instance, $\tau=0.25$ at $\lambda_{min}=0.0787$ provides the first-order method with the smallest error value, and it remains the same for $\tau=0.5$ at $\lambda_{min}=0.3392$ and $\tau=0.75$ at $\lambda_{min}=0.2555$. However, $\tau=0.25$ and 0.5 provides a different order in terms of MAPE, and at $\tau=0.75$, it provides a second order in terms of MAE. Therefore, ELNET.QR α_{opt} improves prediction accuracy by producing the smallest error values in terms of *RSS*, *RMSE*, *MAE*, and *MAPE*.

Generally, in this application, the proposed ELNET.QR α_{opt} at $\tau=$ 0.25, 0.50, 0.75, and λ_{min} is better in three cases of τ. Moreover, it proved that these predictors have a great significance on the response variable. The ELNET.QR α_{opt} method has achieved the best method for reducing the number of components and selecting predictor variables with high prediction accuracy. The ELNET.QR α_{opt} method deals with multicollinearity in three cases of τ by choosing some of these variables and forcing the other to be zero in the final model, whereas the RR.QR method lost reliability and accuracy for the selection. This method has been selected for all the predictor variables. The rest of the methods are LASSO.QR and AdLASSO.QR methods to deal with multicollinearity.

Based on the findings in the previous section, we will rely on the ELNET.QR α_{opt} estimated coefficients to interpret the oil price model as shown in Table 4, as it is more consistent in terms of *RMSE*, *MAE*, *MAPE*, and *MASE*. These results imply that the changes in the oil prices could be explained with the variables included in the analysis with different significant strengths, particularly the exchange rate (x_3), as it has the highest effect on oil prices. As expected, there is a positive relationship between crude oil prices and exchange rate, retail trade, interest rates and the consumer price index. In other words, crude oil

prices increase when the exchange rate and other variables increase. In addition, the results reveal that the interaction team has a very important role and a mixed effect on the changes in oil prices. The results obtained are consistent with the literature. The oil prices in Europe have an increasing trend during several periods, but are low in comparison during the COVID-19 pandemic until the period at the start of the Russian-Ukraine war; they have been quite high since the beginning of the war (Krozer 2013; Borowski 2020; Balashova and Serletis 2021).

Table 3. Performance criteria.

Method	λ	*RMSE*	*MAE*	*MAPE*	*MASE*
RR	$\lambda_{min} = 0.041839$	1.1469	0.7806	1.7260	0.8729
	$\lambda_{1se} = 0.74835$	0.9633	0.7105	1.2235	0.7945
LASSO	$\lambda_{min} = 0.003994$	1.2029	0.8069	1.8265	0.9022
	$\lambda_{1se} = 0.078399$	0.9962	0.7161	1.3324	0.8008
ELNET $\alpha = 0.25$	$\lambda_{min} = 0.013263$	1.1829	0.7974	1.7923	0.8916
	$\lambda_{1se} = 0.260352$	0.9627	0.7046	1.2343	0.7879
ELNET $\alpha = 0.5$	$\lambda_{min} = 0.007987$	1.1921	0.8018	1.8080	0.8966
	$\lambda_{1se} = 0.142868$	0.9790	0.7096	1.2885	0.7935
ELNET $\alpha = 0.75$	$\lambda_{min} = 0.005844$	1.1955	0.8035	1.8136	0.8985
	$\lambda_{1se} = 0.104532$	0.9871	0.7129	1.3112	0.7972
$\tau = 0.25$					
RR.QR	$\lambda_{min} = 0.1322$	1.1189	0.8876	2.537	0.9924
	$\lambda_{1se} = 0.5480$	1.0531	0.8610	2.727	0.9628
LASSO.QR	$\lambda_{min} = 0.0471$	1.0267	0.8217	2.514	0.9187
	$\lambda_{1se} = 0.0720$	1.0623	0.8520	2.571	0.9527
AdLASSO.QR (RR.W. λ_{min})	$\lambda_{min} = 0.0008$	1.0984	0.8644	2.618	0.9667
	$\lambda_{1se} = 0.0028$	1.1020	0.8742	2.639	0.9776
AdLASSO.QR (RR.W. λ_{1se})	$\lambda_{min} = 0.0003$	1.0785	0.8527	2.704	0.9534
	$\lambda_{1se} = 0.0012$	1.0942	0.8661	2.671	0.9685
ELNET.QR $\alpha = 0.25$	$\lambda_{min} = 0.0911$	1.0201	0.8259	**2.4273**	0.9236
	$\lambda_{1se} = 0.2002$	1.0274	0.8387	2.5879	0.9378
ELNET.QR $\alpha = 0.5$	$\lambda_{min} = 0.0317$	1.1018	0.8779	2.5650	0.9817
	$\lambda_{1se} = 0.1275$	1.0452	0.8432	2.5592	0.9429
ELNET.QR $\alpha = 0.75$	$\lambda_{min} = 0.054$	1.0241	0.8195	2.5257	0.9164
	$\lambda_{1se} = 0.0931$	1.0583	0.8500	2.5636	0.9505
ELNET.QR $\alpha_{opt} = 0.38$	$\lambda_{min} = 0.0787$	**1.0086**	**0.8180**	**2.444**	**0.9146**
	$\lambda_{1se} = 0.1532$	1.0382	0.8404	2.565	0.9398
$\tau = 0.5$					
RR.QR	$\lambda_{min} = 0.0678$	0.9759	0.7058	1.6486	0.7893
	$\lambda_{1se} = 0.6563$	0.9494	0.6876	1.2275	0.7689
LASSO.QR	$\lambda_{min} = 0.0228$	0.9649	0.7037	1.6693	0.7869
	$\lambda_{1se} = 0.0531$	0.9506	0.6907	1.4299	0.7724
AdLASSO.QR (RR.W. λ_{min})	$\lambda_{min} = 0.0006$	0.9778	0.7109	1.7108	0.7949
	$\lambda_{1se} = 0.0035$	0.9636	0.7151	1.6833	0.7996

Table 3. *Cont*.

Method	λ	RMSE	MAE	MAPE	MASE
AdLASSO.QR (RR.W. λ_{1se})	$\lambda_{min} = 0.0001$	0.9392	0.6872	1.5984	0.7685
	$\lambda_{1se} = 0.0009$	0.9842	0.7099	1.3913	0.7938
ELNET.QR $\alpha = 0.25$	$\lambda_{min} = 0.0390$	0.9590	0.6993	1.6403	0.7820
	$\lambda_{1se} = 0.1719$	0.9470	0.6876	**1.3001**	0.7689
ELNET.QR $\alpha = 0.5$	$\lambda_{min} = 0.0357$	0.9655	0.7031	1.6494	0.7863
	$\lambda_{1se} = 0.1000$	0.9504	0.6895	**1.3276**	0.7710
ELNET.QR $\alpha = 0.75$	$\lambda_{min} = 0.0277$	0.9664	0.7040	1.6612	0.7872
	$\lambda_{1se} = 0.0687$	0.9501	0.6901	1.3927	0.7717
ELNET.QR $\alpha_{opt} = 0.02$	$\lambda_{min} = 0.3392$	**0.9372**	**0.6827**	1.3390	**0.7634**
	$\lambda_{1se} = 0.7450$	0.9504	0.6914	**1.1955**	0.7731
	$\tau = 0.75$				
RR.QR	$\lambda_{min} = 0.0511$	1.0970	0.7913	2.9823	0.8848
	$\lambda_{1se} = 0.6298$	1.0942	0.7899	2.9192	0.8833
LASSO.QR	$\lambda_{min} = 0.0053$	1.1455	0.8323	3.1377	0.9307
	$\lambda_{1se} = 0.0354$	1.1065	**0.7785**	2.9182	**0.8705**
AdLASSO.QR (RR.W. λ_{min})	$\lambda_{min} = 0.0004$	1.1683	0.8644	3.1510	0.9665
	$\lambda_{1se} = 0.0011$	1.1960	0.9016	3.4228	1.0082
AdLASSO.QR (RR.W. λ_{1se})	$\lambda_{min} = 0.0001$	1.1018	0.7837	2.909	0.8764
	$\lambda_{1se} = 0.0003$	1.1275	0.8137	3.1028	0.9099
ELNET.QR $\alpha = 0.25$	$\lambda_{min} = 0.0204$	1.1152	0.8057	3.0121	0.9010
	$\lambda_{1se} = 0.1334$	1.1139	0.7947	2.9758	0.8886
ELNET.QR $\alpha = 0.5$	$\lambda_{min} = 0.0102$	1.1311	0.8191	3.0654	0.9159
	$\lambda_{1se} = 0.0688$	1.1075	0.7847	2.9390	0.8775
ELNET.QR $\alpha = 0.75$	$\lambda_{min} = 0.0068$	1.1405	0.8278	3.1141	0.9256
	$\lambda_{1se} = 0.0458$	1.1055	**0.7801**	2.9212	**0.8723**
ELNET.QR $\alpha_{opt} = 0.02$	$\lambda_{min} = 0.2555$	**1.0847**	**0.7812**	**2.8845**	**0.8736**
	$\lambda_{1se} = 0.8571$	1.1270	0.8156	3.0404	0.9120

Note: The bold number indicates the lowest values in favor of the superior performance method compared to the other methods.

Table 4. Coefficients estimation for the predictor variables by ELNET.QR method.

	$\tau = 0.25$, $\alpha_{opt} = 0.38$		$\tau = 0.5$, $\alpha_{opt} = 0.02$		$\tau = 0.75$, $\alpha_{opt} = 0.02$	
	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}	λ_{min}	λ_{1se}
$\hat{\beta}_1$	0.1686	0.1313	0.1545	0.0904	0.1054	0.0548
$\hat{\beta}_2$	0.1225	0.0415	0.1066	0.0614	0.0306	0.0184
$\hat{\beta}_3$	0.3038	0.2110	0.2072	0.1061	0.1421	0.0703
$\hat{\beta}_4$	0.1151	0.0948	0.1360	0.0879	0.0761	0.0487
$\hat{\beta}_5$	0.0150	0	-0.0172	0	-0.0277	0
$\hat{\beta}_6$	0.0529	0	0.0413	0.0123	0.0648	0.0158
$\hat{\beta}_7$	0	0	0.0033	0.0026	0	0
$\hat{\beta}_8$	0.0037	0	0.0448	0.0128	0.0489	0.0093
$\hat{\beta}_9$	0	0	0.0133	0.0013	-0.0357	-0.0048
$\hat{\beta}_{10}$	0.0802	0.0062	0.0862	0.0531	0.1022	0.0425

5. Conclusions

In this study, we proposed ELNET.QR based on selecting the best alpha value α_{opt} using a cross-validation method. The method was used to identify the relationship between the predictor variables and the response variable to improve the accuracy of model selection and to deal with heavy-tailed distributions, heterogeneity, and multicollinearity between the predictor variables by determining the best alpha value. Numerical experiments and actual time-series datasets were carried out. The results showed that the ELNET.QR α_{opt} method effectively selected the actual predictor variables that were most significant for the response variable with a reduced prediction error at $\tau = 0.25$, 0.5, 0.75. The ELNET.QR α_{opt} method selected the best-fitting model with high prediction accuracy, compared to the other methods. It also proved that not all of the alpha values can be used to represent the Elastic net. Thus, the cross-validation represents the best way to choose an alpha value and use it as the optimal value for building the final model.

This method offers additional insights into the behaviour of crude oil prices, as it provides evidence of how oil prices have changed by the exchange rate, retail trade, interest rates, and consumer price index during the economic shocks in the European Union, while the findings reveal that the exchange rate has the highest effect on changing the crude oil price. In conclusion, this research indicates that the selected penalized QR method is well suited for modelling crude oil prices while considering the dynamic effect on the factors that influence local currency and the global economy in the European Union. By considering these results, the European Union should make the exchange rate stable so that the effect of these variables on oil prices are less significant. The main reason behind this recommendation is the focus on local factors, as it is either mostly or partially under the control of European policy.

References

Aastveit, Knut Are, Jamie L. Cross, and Herman K. van Dijk. 2023. Quantifying time-varying forecast uncertainty and risk for the real price of oil. *Journal of Business & Economic Statistics* 41: 523–37.

Akaike, Hirotsugu, Boris Nikolaevich Petrov, and F. Csaki. 1973. *Second International Symposium on Information Theory*. Budapest: Akademiai Kiado, pp. 267–81.

Al-Jawarneh, Abdullah S., and Mohd Tahir Ismail. 2024. The adaptive LASSO regression and empirical mode decomposition algorithm for enhancing modelling accuracy. *Communications in Statistics-Simulation and Computation* 53: 714–26. [CrossRef]

Al-Jawarneh, Abdullah S., Mohd Tahir Ismail, Ahmad M. Awajan, and Ahmed R. M. Alsayed. 2022. Improving accuracy models using elastic net regression approach based on empirical mode decomposition. *Communications in Statistics-Simulation and Computation* 51: 4006–25. [CrossRef]

Alsayed, Ahmed R. M. 2023. Turkish stock market from pandemic to Russian invasion, evidence from developed machine learning algorithm. *Computational Economics* 62: 1107–23. [CrossRef]

Alsayed, Ahmed R. M., and Giancarlo Manzi. 2019. A comparison of monotonic correlation measures with outliers. *WSEAS Transactions on Computers* 18: 223–30.

Alsayed, Ahmed R. M., Zaidi Isa, and Sek Siok Kun. 2018. Outliers detection methods in panel data regression: An application to environment science. *International Journal of Ecological Economics & Statistics* 39: 73–86.

Alsayed, Ahmed R. M., Zaidi Isa, Sek Siok Kun, and Giancarlo Manzi. 2020. Quantile regression to tackle the heterogeneity on the relationship between economic growth, energy consumption, and CO_2 emissions. *Environmental Modeling & Assessment* 25: 251–58.

Amano, Robert A., and Simon Van Norden. 1998. Exchange rates and oil prices. *Review of International Economics* 6: 683–94. [CrossRef]

Ambark, Ali S. A., Mohd Tahir Ismail, Abdullah S. Al-Jawarneh, and Samsul Ariffin Abdul Karim. 2023. Elastic Net Penalized Quantile Regression Model and Empirical Mode Decomposition for Improving the Accuracy of the Model Selection. *IEEE Access* 11: 26152–62. [CrossRef]

Balashova, Svetlana, and Apostolos Serletis. 2021. Oil price uncertainty, globalization, and total factor productivity: Evidence from the European Union. *Energies* 14: 3429. [CrossRef]

Baumeister, Christiane, and Lutz Kilian. 2015. Forecasting the real price of oil in a changing world: A forecast combination approach. *Journal of Business & Economic Statistics* 33: 338–51.

Bloomberg. 2023. Bloomberg Terminal. Available online: https://www.bloomberg.com/professional/solution/bloomberg-terminal/ (accessed on 1 January 2023).

Borowski, Piotr F. 2020. Zonal and Nodal Models of energy market in European Union. *Energies* 13: 4182. [CrossRef]

Chand, Sohail. 2012. On tuning parameter selection of lasso-type methods—A monte carlo study. Paper presented at 2012 9th International Bhurban Conference on Applied Sciences & Technology (IBCAST), Islamabad, Pakistan, January 9–12; pp. 120–29.

Davino, Cristina, Marilena Furno, and Domenico Vistocco. 2013. *Quantile Regression: Theory and Applications*. Hoboken: John Wiley & Sons.

Desboulets, Loann David Denis. 2018. A review on variable selection in regression analysis. *Econometrics* 6: 45. [CrossRef]

Doğrul, H. Günsel, and Ugur Soytas. 2010. Relationship between oil prices, interest rate, and unemployment: Evidence from an emerging market. *Energy Economics* 32: 1523–28.

Efron, Bradley, Trevor Hastie, Iain Johnstone, and Robert Tibshirani. 2004. Least angle regression. *The Annals of Statistics* 32: 407–99. [CrossRef]

Fan, Jianqing, and Runze Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association* 96: 1348–60. [CrossRef]

Fan, Yingying, and Cheng Yong Tang. 2013. Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 75: 531–52. [CrossRef]

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software* 33: 1–22. [CrossRef]

Gareth, James, Witten Daniela, Hastie Trevor, and Tibshirani Robert. 2013. *An Introduction to Statistical Learning with Applications in R*. New York: Springer.

Geisser, Seymour. 1975. The predictive sample reuse method with applications. *Journal of the American statistical Association* 70: 320–28. [CrossRef]

Hastie, Trevor, Robert Tibshirani, Jerome H. Friedman, and Jerome H. Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer Science & Business Media.

Hastie, Trevor, Robert Tibshirani, and Martin Wainwright. 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations*. Boca Raton: CRC Press.

Haws, David C., Irina Rish, Simon Teyssedre, Dan He, Aurelie C. Lozano, Prabhanjan Kambadur, Zivan Karaman, and Laxmi Parida. 2015. Variable-selection emerges on top in empirical comparison of whole-genome complex-trait prediction methods. *PLoS ONE* 10: e0138903. [CrossRef]

He, Yanan, Ai Han, Yongmiao Hong, Yuying Sun, and Shouyang Wang. 2021. Forecasting crude oil price intervals and return volatility via autoregressive conditional interval models. *Econometric Reviews* 40: 584–606. [CrossRef]

Kartal, Mustafa Tevfik. 2020. The Effect of the COVID-19 Pandemic on Oil Prices: Evidence from Turkey. *Energy Research Letters*. Available online: https://erl.scholasticahq.com/article/18723-the-effect-of-the-covid-19-pandemic-on-oil-prices-evidence-from-turkey (accessed on 29 May 2024).

Kassambara, Alboukadel. 2018. *Machine Learning Essentials: Practical Guide in R*. Scotts Valley: CreateSpace Independent Publishing Platform.

Kirikkaleli, Dervis, and Nezahat Doğan. 2021. Energy consumption and refugee migration in Turkey. *Utilities Policy* 68: 101144. [CrossRef]

Koenker, Roger. 2004. Quantile regression for longitudinal data. *Journal of Multivariate Analysis* 91: 74–89. [CrossRef]

Krozer, Yoram. 2013. Cost and benefit of renewable energy in the European Union. *Renewable Energy* 50: 68–73.

Kuhn, Max, and Kjell Johnson. 2013. *Applied Predictive Modeling*. New York: Springer.

Lee, Young, Vy Nguyen, and Duzhe Wang. 2016. On Variable and Grouped Selections of the Elastic Net. *Report CS532* 961: 1–24.

Li, Youjuan, and Ji Zhu. 2008. L 1-norm quantile regression. *Journal of Computational and Graphical Statistics* 17: 163–85. [CrossRef]

Masselot, Pierre, Fateh Chebana, Diane Bélanger, André St-Hilaire, Belkacem Abdous, Pierre Gosselin, and Taha BMJ Ouarda. 2018. EMD-regression for modelling multi-scale relationships, and application to weather-related cardiovascular mortality. *Science of The Total Environment* 612: 1018–29. [CrossRef]

Melkumova, L. E., and S. Ya Shatskikh. 2017. Comparing Ridge and LASSO estimators for data analysis. *Procedia Engineering* 201: 746–55. [CrossRef]

Qin, Lei, Shuangge Ma, Jung-Chen Lin, and Ben-Chang Shia. 2016. Lasso Regression Based on Empirical Mode Decomposition. *Communications in Statistics-Simulation and Computation* 45: 1281–94. [CrossRef]

Ray, Susmita. 2019. A quick review of machine learning algorithms. Paper presented at 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, February 14–16.

Schwarz, Gideon. 1978. Estimating the dimension of a model. *The Annals of Statistics* 6: 461–64. [CrossRef]

Stone, Mervyn. 1974. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B (Methodological)* 36: 111–33.

Su, Meihong, and Wenjian Wang. 2021. Elastic net penalized quantile regression model. *Journal of Computational and Applied Mathematics* 392: 113462. [CrossRef]

Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* 58: 267–88. [CrossRef]

van Houwelingen, Hans C., and Willi Sauerbrei. 2013. Cross-validation, shrinkage and variable selection in linear regression revisited. *Open Journal of Statistics* 3: 79–102. [CrossRef]

Xiao, Hui, and Yiguo Sun. 2019. On Tuning Parameter Selection in Model Selection and Model Averaging: A Monte Carlo Study. *Journal of Risk and Financial Management* 12: 109. [CrossRef]

Yılmaz, Alper, and Hüseyin Altay. 2016. Examining the cointegration relationship and volatility spillover between imported crude oil prices and exchange rate: The Turkish case. *Ege Academic Review* 16: 655–71.

Zou, Hui, and Hao Helen Zhang. 2009. On the adaptive elastic-net with a diverging number of parameters. *Annals of statistics* 37: 1733. [CrossRef]

Zou, Hui, and Trevor Hastie. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 67: 301–20. [CrossRef]

Article

COVID-19 and Non-Performing Loans in Europe

John Hlias Plikas *, Dimitrios Kenourgios and Georgios A. Savvakis

Department of Economics & UoA Center for Financial Studies, National and Kapodistrian University of Athens, 10559 Athens, Greece; dkenourg@econ.uoa.gr (D.K.); gsavvakis@econ.uoa.gr (G.A.S.)
* Correspondence: johnp1@econ.uoa.gr

Abstract: This study investigates the impact of COVID-19 on the non-performing loans (NPLs) in Europe, distinguishing by European subregion, country-level prosperity, NPL type, and NPL economic sector. We utilized panel data analysis covering the period 2015Q1–2021Q4 while controlling for macro, bank-specific, and regulatory indicators. We derived that the COVID-19 deaths and the strictness of lockdown measures positively affected the NPLs, while the economic support policies exerted a negative effect. Profitable, capitalized banks fared better. The strictness of lockdown measures hindered the ability of SMEs to repay their loans, increasing their NPLs. Sectors involving physical work-related activities also experienced an increase in their NPLs. We also deduced that bank securitization and national culture significantly contributed to NPL reduction.

Keywords: non-performing loans; COVID-19; policy responses; European banking system; cultural dimensions

JEL Classification: G21; G28; I18; C33; E58

1. Introduction

The 21st century has been marked by a series of external shocks (Shehzad et al. 2020), with the COVID-19 pandemic standing out as a health-oriented shock, causing unprecedented cross-sector variances, rapid dissemination rates, and a high degree of economic uncertainty (Žunić et al. 2021; Yi et al. 2022). In response to the spread of the COVID-19 pandemic, nations implemented strict social distancing and lockdown policies (Yang and Yang 2021), which resulted in sharp declines in economic growth and enterprise earnings, particularly in the services, travel, and tourism industries (Zheng and Zhang 2021; Ceylan et al. 2020; Bassani 2021).

The economic spillovers of COVID-19 have reverberated globally, significantly impacting businesses, jobs, and incomes (Zheng and Zhang 2021; Banks et al. 2020). The banking industry, unable to evade the negative financial spillovers (Demir and Danisman 2021; Foglia et al. 2022; Shehzad et al. 2020), witnessed high levels of debt and economic imbalances, which have reduced the debtors' ability to repay their loan obligations, resulting in a potential increase in the banks' non-performing loans (NPLs) (Ari et al. 2021; Demir and Danisman 2021; Banks et al. 2020; Park and Shin 2021; Ho et al. 2023).

Preliminary research hinted that the pandemic would resemble the negative repercussions of a banking crisis (Özlem Dursun-de Neef and Schandlbauer 2021; Žunić et al. 2021), potentially resulting in a significant surge in NPLs (Colak and Öztekin 2021). Businesses with lower economic turnover were more likely to be affected due to the lockdown and closure policies, implying a potential surge in the SME NPLs (Cowling et al. 2022; Wellalage et al. 2022). While there is extensive scientific research on the effects of the COVID-19 pandemic on NPLs, predominantly focusing on the European Union due to its peculiar reaction to the European sovereign debt crisis (Foglia et al. 2022; Demir and Danisman 2021; Duan et al. 2021; Rizwan et al. 2020; Ari et al. 2020; Ari et al. 2021; Colak and Öztekin

2021; Park and Shin 2021; Özlem Dursun-de Neef and Schandlbauer 2021; Apergis 2022), there still exist unexplored avenues in this area of study.

Examining the impact of COVID-19 on NPLs within Europe is of intrinsic significance due to the continent's diverse subregions, each with distinct economic structures, levels of prosperity, and policy responses. These subregional variations suggest that the impact of COVID-19 on NPLs could differ markedly across Europe, with peripheral economies experiencing more severe impacts due to their economic vulnerabilities (Foglia et al. 2022; Apergis 2022). For instance, in 2015, while the average NPL ratio across the Eurozone was approximately 12%, Germany had an NPL ratio of less than 2%, whereas Greece faced a staggering 35%, and Italy and Ireland had ratios of around 20% (Rinaldi and Sanchis-Arellano 2021). This highlights the stark contrast between core economies, such as Germany, and peripheral economies, such as Greece and Italy (Jameaba 2020). Moreover, the interconnected economies underscore the potential for cross-border repercussions, emphasizing the need for a localized study. Europe's historical responses to crises, such as the Global Financial Crisis (GFC) of 2008, and unique policy measures ever since, emphasize the importance of understanding the effects of COVID-19 within this context.

As we delve into the investigation of the impact of COVID-19 on NPLs, it becomes essential to consider the diverse countries' cultural backgrounds and policy responses. The diverse cultural fabric of the European economies influences how bank managers, debtors, and the nations as a whole perceive and navigate the challenges posed by the pandemic (Kostis et al. 2018; Petrakis et al. 2015; Petrakis and Kostis 2013; Boubakri et al. 2017; Ashraf et al. 2016; Gaganis et al. 2020; Giannetti and Yafeh 2012; Boubakri et al. 2023). The cultural variations of European nations can shape both borrowers' and banks' attitudes toward risk management strategies and financial decisions, consequently impacting loan repayments and defaults (Kostis et al. 2018; Petrakis et al. 2015; Petrakis and Kostis 2013). For instance, a culture that highly values tradition and security may lead to more conservative financial behaviors, thereby reducing the likelihood of loan defaults. Conversely, a culture that emphasizes innovation and competitiveness might encourage economic activities that enhance loan repayment capabilities (Kostis et al. 2018; Petrakis et al. 2015). Therefore, considering the unique cultural values of European economies is pivotal for conducting a holistic investigation of the pandemic's impact on NPLs in the European landscape.

Motivated by the work of Ari et al. (2021) on the dynamics of non-performing loans during banking crises and Duan et al. (2021) on bank systemic risk around COVID-19, this study examines the influence of COVID-19 on the NPLs of the European Banking System. Specifically, Ari et al. (2021) utilized data on past banking crises to identify pre-crisis predictors of NPLs and provide insights into post-COVID-19 NPL vulnerabilities using the IMF's GDP growth forecasts. However, they did not consider the pandemic period or the heterogeneity within European subregions. Furthermore, while Duan et al. (2021) conducted a comprehensive study on the impact of the pandemic on bank systemic risk, it focused solely on the effect of initial government policy responses on systemic risk and did not consider the influence of quantitative easing (QE) policy responses. Although they employed Hofstede's five cultural dimensions to assess national culture (Hofstede 2001), they did not incorporate crucial cultural factors, such as tradition and security, as outlined by Schwartz (1994). These factors can shape both borrowers' and banks' attitudes toward financial decisions and, consequently, impact loan repayments and defaults. Moreover, Schwartz's (1994) cultural dimensions framework also included data obtained from diverse regions, including socialist countries. Another advantage of Schwartz's (1994) framework is that it delves deeper into the intricacies of national culture, allowing us to capture a broader range of cultural variations that may influence loan defaults.

Utilizing panel data analysis with country-fixed effects, we conducted a comprehensive comparison between the pre-pandemic period (2015Q1–2019Q4), the post-pandemic period (2020Q1–2021Q4), as well as the entire period of analysis (2015Q1–2021Q4). For this purpose, we utilized a unique quarterly dataset of aggregated data spanning from 2015 to 2021. We chose to commence our analysis from 2015Q1 due to several considerations.

First, it allowed us to provide a holistic view of the European banking landscape before the pandemic, reducing potential biases associated with shorter observation periods. Second, in 2014, the European Banking Authority (EBA) introduced a harmonized NPL definition of NPLs across European countries (EBA 2019). We chose to begin our analysis from 2015Q1, since this period coincides with the harmonized NPL definition introduced by the EBA, leading to consistent and comparable data and minimizing biases arising from varying international NPL definitions. We chose to end our analysis in 2021Q4 to focus on the period during which the pandemic's effects on NPLs were most pronounced. Additionally, we chose this period to avoid exogenous disruptions stemming from the war between Russia and Ukraine and to ensure our results remained specific to the pandemic period.

We formulated several questions to be answered: (1) How did COVID-19 impact the NPLs of the European economies? (2) Did it differ between core and peripheral economies? (3) What were the primary factors of COVID-19 that affected the change in NPLs? (4) Did the government's economic support policies to mitigate the pandemic manage to absorb the impact of the pandemic on NPLs? (5) Did central bank QE economic support measures aid in minimizing the risk of a new wave of NPLs? (6) Did national culture shape banking institutions and borrowers' behavior in preventing the rise of NPLs? (7) Did bank securitization strategies contribute to NPL reduction?

Our research contributes to the existing literature (Ari et al. 2020; Žunić et al. 2021; Loang et al. 2023; Apergis 2022; Ari et al. 2021; Duan et al. 2021) by being the first to conduct a comprehensive analysis of the effects of COVID-19 on the European Union's NPLs, distinguishing by European subregion and country-level prosperity. Second, our research also explores the impact of NPL types and sectoral NPLs. Third, it considers the bank securitization strategies as a means of NPL reduction, emphasizing their effectiveness in reducing NPLs. By discerning the impact of COVID-19 on NPLs across various sectors and loan types, we may provide granular insights for effectively managing NPL risks and promoting economic resilience in the aftermath of the pandemic. While Žunić et al. (2021) addressed the factors influencing NPLs during the COVID-19 period, providing useful insights, they lacked a broader European context. Moreover, while Ari et al. (2021) provided insights regarding NPL vulnerabilities for the post-COVID-19 period, they based their analysis on past banking crises, lacking the incorporation of actual post-COVID-19 period data. Furthermore, they did not comprehensively explore the pandemic's impact on various types of NPLs and sectoral NPLs. While Apergis (2022) provided insights into the existence of NPL homogeneity amongst EU countries, he did not consider policy responses, cultural intricacies, or diverse NPLs types/sectoral NPLs in his analysis. Our study endeavors to fill these unexplored territories by conducting a detailed analysis in this area, while also encompassing a broader spectrum of dimensions to foster a more comprehensive understanding.

The remainder of the paper is laid out as follows. In Section 2, we provide the theoretical and conceptual framework. In Section 3, we provide the literature review. The data, variables, econometric models, and empirical methodology used are all described in Section 4. The empirical results and the robustness checks are presented in Sections 5 and 6, respectively. The conclusions and future research are presented in Section 7.

2. Theoretical and Conceptual Framework

This section aims to identify key theories and elaborate the conceptual model of our investigation on the impact of the COVID-19 pandemic on NPLs across European economies. By integrating relevant theoretical perspectives, we provide a holistic framework that explains how cultural dimensions, economic shocks, and policy responses interact and influence NPL dynamics. The next paragraph underpins key theoretical perspectives, and the third paragraph outlines the conceptual framework.

The exploration of the impact of the COVID-19 pandemic on NPLs in European economies can be rooted in utilizing Schwartz's (1994) theory of cultural values, Minsky's (1992), Kindleberger and Aliber's (2011) theories on financial stability and banking crises,

as well as Bernanke's (2009) theory on policy responses to economic crises. Schwartz (1994) identified ten universal values, including stimulation, hedonism, achievement, and benevolence, that shape individual and organizational behavior. Minsky (1992) implied that financial systems are inherently unstable and prone to cycles of boom and bust, often triggered by economic shocks. Kindleberger and Aliber (2011) complemented this by focusing on historical patterns of financial crises, where speculative bubbles and crashes are central themes. Bernanke (2009) emphasized the importance of proactive monetary and fiscal policies in mitigating the impacts of economic downturns.

In our conceptual framework, we integrated these theoretical perspectives to provide a holistic understanding of the dynamics related to the effect of the COVID-19 pandemic on NPLs. The framework encompasses three key components: (1) cultural values, as measured by Schwartz's (1994) framework, which influences borrower behavior and bank risk management; (2) the economic spillovers of the COVID-19 pandemic, which disrupted economic activities and increased financial uncertainties, in line with the theories of Minsky's (1992), Kindleberger and Aliber (2011); (3) government and central bank policy responses, such as economic support and quantitative easing measures, aimed at stabilizing the economy and supporting businesses and households, in line with Bernanke (2009). The significance of this theoretical framework lies in its ability to capture the multifaceted nature of NPL dynamics in the context of the COVID-19 pandemic. Specifically, it captures the complex interplay of cultural values, economic shocks, and policy responses. By doing so, it provides a holistic view of how these factors collectively influenced NPL ratios in the European economies during the pandemic.

3. Literature Review

This section aims to identify the pre- and post-COVID-19 pandemic literature and to identify key studies that will aid in our selection of the appropriate candidate predictor variables for our analysis. The pre-pandemic literature is investigated in Section 3.1 and the post-pandemic literature is investigated in Section 3.2.

3.1. Pre-COVID-19 Pandemic Literature

The initial scientific literature regarding the impact of COVID-19 on European banks' NPLs anticipated that the pandemic would lead to a surge of NPLs. This assumption was primarily based on historical research and speculative reasoning.

Ari et al. (2020) stated that the COVID-19 pandemic would most likely lead to an increase in NPLs. However, they also mentioned that banks have a modest advantage, as they had already undertaken initiatives to raise their capital ratios after the GFC. Laeven and Valencia (2018) agreed, indicating that elevated NPL ratios are a recurrent feature of banking crises and are often assessed in the aftermath of such events. Brunnermeier and Krishnamurthy (2020) disagreed, however, stating that the lessened regulatory stance of the European Central Bank (ECB) and the EBA, along with government loan guarantees, would assist financial institutions in managing the COVID-19 crisis. Bitar and Tarazi (2022) also stated that before the pandemic, banks maintained sufficient funds. However, they also argued that by releasing cash reserves and implementing additional measures, such as easing the management of NPLs, banks' earning potential could be jeopardized, potentially leading to a prolonged downturn. From a regulatory standpoint, the EBA and ECB forecasted a reduction in 2020 and an upsurge in 2021 (Couppey-Soubeyran et al. 2020).

It is clear from the above analysis that in the pre-COVID-19 pandemic literature, there were contradictions regarding the impact of COVID-19 on NPLs. Although one literature branch expected a new wave of NPLs, another literature branch expected that banks would have a modest advantage due to the concentration of capital reserves after the GFC. We also noticed a third literature branch implying that quantitative easing measures might mitigate the impact on NPLs.

3.2. Post-COVID-19 Pandemic Literature

After the COVID-19-related data started to crystalize, scientific studies on the impact of the COVID-19 pandemic on NPLs revealed additional and unexpected outcomes. In our research, we identified the most pertinent literature on the relationship between the COVID-19 pandemic and NPLs.

Sharif et al. (2020) stated that the risk associated with the COVID-19 pandemic is perceived differently in the short and the long term and may be viewed as an economic crisis. Rizwan et al. (2020) derived that the COVID-19 pandemic significantly contributes to a country's systemic risk, due to successive COVID-19 lockdowns. As systemic risk increases, NPLs tend to increase as well. Duan et al. (2021) stated that the COVID-19 pandemic exacerbates preexisting financial vulnerabilities. Apergis (2022) added that those vulnerabilities are not uniform across the EU countries. More specifically, they stated that banks that are deleveraged, undercapitalized, and have low profitability indices are susceptible to the pandemic. On the other hand, banks with elevated capital and profitability indices can withstand the adverse impacts of the pandemic. Dunbar (2022) suggested that although COVID-19 poses a significant risk to the bank's financial soundness, the relaxed regulatory stance of central banks might enable the release of capital buffers, thereby facilitating lending and enhancing overall financial stability. According to Kozak (2021), larger banks, being more profitable, exhibited increased stability during the pandemic, compared to smaller banks. Xie et al. (2024) emphasized that during the pandemic, financially strong banks, offering competitive products and services, can play a significant role in facilitating economic growth. Demir and Danisman (2021) argued that well-capitalized banks that have low NPL levels and are larger in size are more resilient to the pandemic. They added that government financial aid initiatives considerably assisted banks in dealing with the financial and capital losses incurred as a result of the economic spillovers caused by the pandemic. Ari et al. (2021) added that monetary and prudential policies may mitigate the rapid credit expansion, leading to a decrease in NPLs. Yi et al. (2022) emphasized the significance of enhanced regulations in preventing excessive credit expansion.

Foglia et al. (2022) found that the pandemic had a heterogeneous effect on the Eurozone banking system. However, due to the intricate interconnections among European banks, the European banking system may be "too interconnected to fail." This implies that the higher profitability levels of high-income economies may hinder the ability of banks in low-income economies to incur losses. Kryzanowski et al. (2022) added that banks with high-quality capital were more resilient to the crisis and were able to effectively control their NPL ratios. Mateev et al. (2022) argued that bank performance strongly depends on both banks' efficiency and market power. Moreover, Cowling et al. (2022) stated that small business firms are particularly vulnerable to the economic spillovers stemming from COVID-19. They added that COVID-19 increases the possibility of SME bankruptcy, which translates into increased NPL ratios related to those firms. On the other hand, Wellalage et al. (2022) added that although the adverse effect of COVID-19 is anticipated to inflict substantial and enduring damage on SMEs, a firm's access to external finance can mitigate the negative impacts of the pandemic. This was further supported by Naili and Lahrichi (2022), who highlighted that the pandemic posed disproportionate effects. Regarding the impacts of the pandemic on banking performance, Salazar et al. (2023) stated that the COVID-19 pandemic has introduced high uncertainty and economic downturn, ultimately affecting banks' performance. Alnabulsi et al. (2023a) also contributed to this discussion by underscoring the complex relationships between NPLs and bank performance. In another study, Alnabulsi et al. (2023b) also emphasized that NPLs can significantly destabilize banks, particularly in times of economic crisis, stressing the importance of robust risk management practices to mitigate these effects.

It is clear from the post-COVID-19 pandemic literature that the successive lockdowns to mitigate the pandemic were expected to affect the banks' systemic risk and lead to a rise in NPLs. Additionally, the literature highlighted that the pandemic would worsen preexisting financial vulnerabilities, with heterogeneous impacts on European economies.

The literature suggested that SMEs are particularly vulnerable and may experience a rise in their NPLs. Banks that are financially strong were expected to withstand the adverse effects of the pandemic. Furthermore, the post-COVID-19 pandemic literature supplements the pre-COVID-19 pandemic literature, suggesting that government and central bank policy responses might assist banks in dealing with the economic spillovers of the pandemic.

From the above analysis, we derived several gaps in the existing literature. First, the impact of cultural dimensions on NPLs was largely overlooked. Second, there was inadequate attention to the heterogeneous impacts across European subregions. Third, existing studies often did not comprehensively integrate government and central bank policies with other economic indicators. Additionally, many studies focused on the immediate impact of the pandemic without extending the analysis to the post-pandemic period. Our study endeavors to fill these unexplored territories by conducting a detailed analysis, while encompassing a broader spectrum of dimensions to foster a more comprehensive understanding.

4. Data and Specification Model—Empirical Methodology

This section aims to describe the collection of the data used in this study (Section 4.1) and define the candidate predictors and their expected impacts on the NPLs (Section 4.2); furthermore, it outlines the empirical methodology and the empirical models employed (Section 4.3).

4.1. Data Construction

Our dataset consisted of quarterly aggregate country data spanning from 2015Q1 to 2021Q4 for 28 European economies.[1] The final sample consisted of an unbalanced panel of 28 countries with 784 observations (the distribution of observations by country can be found in Table A1 of Appendix A).

The primary dependent variable was the ratio of NPLs to total (gross) loans (NPLs), consisting of 684 observations, while we alternated between various types and sectoral NPLs. The reduction in the NPL sample size was primarily due to data reporting disparities and, in some cases, missing or incomplete data in certain quarters or countries. We deliberately opted not to employ data imputation methods, preserving the data integrity of the dependent variable. The primary dependent variables' data were obtained from the ECB data portal. The data for NPL types and sectoral NPLs were all from the EBA. We chose our sample period to coincide with the establishment of a harmonized NPL approach (EBA 2019) to eliminate international NPL definition inconsistencies. Regarding the candidate predictors, we categorized our data into three variable groups: one group included the COVID-19 variables, another group included the QE-related variables, and the third group contained the COVID-19 government response variables. To effectively capture the impact of our candidate predictors on the NPLs, we incorporated a range of control factors, such as macroeconomic, bank-specific, regulatory, and national culture-related factors. All the variables employed in our analysis are expressed at an aggregate level. We did not need to convert bank-specific variables to the country level by using standardization strategies since the EBA had already aggregated these variables at the country level during the data collection process. Notably, the European banks fill out the required information in the reporting templates following a harmonized methodology approach. The regulatory authority then compiles and aggregates the data contained in those templates by using strategies for addressing variations in reporting practices, ensuring consistency and comparability across European countries.

Ten national cultural dimensions based on Schwartz's (1994) theory of cultural values were included, with data collected from the European Social Survey (ESS). Data from four ESS questionnaires were evaluated: ESS Round 7 (2014), Round 8 (2016), Round 9 (2018), and Round 10 (2020). For each nation, a percentage of positive responses was calculated, and biennial figures were assigned to quarters according to the questionnaire timeframes. We imputed the missing data with values corresponding to the nearest pre-

ceding questionnaire period, assuming that cultural values remained relatively stable in the short term. The missing values from earlier questionnaire periods were left unaltered. Following this approach, we ended up with a small number of missing data[2], while increasing data accuracy. Principal component analysis (PCA) was applied for dimensionality reduction. The synthetic "Culture_PCA" variable contained information from the first two principal components, explaining 73% of the total variation. The generated variable was primarily and positively driven by the cultural values of IMPTRAD ("Tradition"), IMPENV ("Universalism"), IPEQOPT ("Benevolence"), and IPRULE ("Power").

A detailed description of all variables employed, and their respective data sources, is presented in detail in Tables A2–A4 of Appendix A (although Table A4 lists the cultural dimensions and the synthetic "CULTURE_PCA" variable, the tests performed, the eigenvalues, as well as the primary drivers of the synthetic variable are not reported due to space limitations but are available upon request).

4.2. Expected Channels of Impact

Table 1 lists both the candidate predictors and the control variables, along with their projected relationship to the main dependent variable (NPLs), based on the respective literature. Table 1 focuses on presenting the projected relationship of candidate predictors to the output variable NPLs and, therefore, does not include the dependent variables used in this research. A positive projected relationship is indicated by the '+' sign, while a negative projected relationship is indicated by the '−' sign. A detailed presentation of the primary dependent variable and secondary dependent variables is presented in Table A2 of Appendix A.

Table 1. Variables and expected channels of impact.

Variable Group	Variable Symbol	Parameter Shown	Explanation	Related Literature	Expected Sign
Macroeconomic Variables	UNEMP	Percentage (%)	% of unemployment	(Makri et al. 2014; Ceylan et al. 2020; Bassani 2021)	(+)
	CPI	No.	Quarterly Consumer Price Index	(Makri et al. 2014)	(+)
	R_GDP_Q2Q	Percentage (%)	Quarterly percentage growth rate of real GDP	(Makri et al. 2014)	(−)
	GDP_MARKET	No.	Quarterly gross domestic product at market prices	(Makri et al. 2014)	(−)
Bank-specific Variables	NPLS (-1)	Percentage (%)	Previous quarter aggregate non-performing loans to total gross loans	(Makri et al. 2014)	(+)
	ROA	Percentage (%)	Return on assets: profit or loss for the year/total assets	(Makri et al. 2014; Colak and Öztekin 2021)	(−)
	CAP	Percentage (%)	Bank capital and reserves to total assets	(Makri et al. 2014; Colak and Öztekin 2021; Bitar and Tarazi 2022)	(−)/(+)
	LOAN_DISBRS	Percentage (%)	Loan disbursments to customers	(Naili and Lahrichi 2022)	(+)
	FINANCIAL_ASSETS	No.	Total financial instruments on the asset side	(Alessi et al. 2022)	(−)
	PROVISIONS	Percentage (%)	Impairments (credit risk losses)/equity	(Ozili and Outa 2017)	(−)
	RISK_CAPITAL	Percentage (%)	Total risk exposure amount for position, foreign exchange, and commodities risks/total risk exposure amount	(Bitar and Tarazi 2022)	(−)

Table 1. *Cont.*

Variable Group	Variable Symbol	Parameter Shown	Explanation	Related Literature	Expected Sign
Bank-specific Variables	OPER_RISK	No.	Total risk exposure amount for OpePercentage (%)ns/total risk exposure amount	(Bitar and Tarazi 2022)	(−)
	LIABILITIES	No.	Total deposits other than from banks/total liabilities	(Ozili and Outa 2017)	(−)
	CASH_BALANCES	Percentage (%)	Cash positions/total assets	(Alessi et al. 2022)	(−)
	FINANCIAL_ASSETS	No.	Total financial instruments on the asset side	(Alessi et al. 2022)	(−)
	EQUITY	Percentage (%)	Equity instruments/total assets	(Durand and Le Quang 2022)	(−)
	TOTAL_ASSETS	No.	Total assets	(Alessi et al. 2022)	(−)
	RETAINED_EARNINGS	Percentage (%)	Retained earnings/Tier 1 capital volume	(Ahmed et al. 2021)	(−)
	DERIVATIVES	Percentage (%)	Derivatives/total assets	(Mayordomo et al. 2014)	(−)
	CRED_DEPOSITS	Percentage (%)	Deposits from credit institutions/total liabilities	(Ozili 2019)	(−)
Regulatory Variables	TIER1_CAP	No.	Additional Tier 1 capital	(Bitar and Tarazi 2022)	(−)
	COVER_PERCENTAGE (%)	Percentage (%)	Accumulated impairment, accumulated negative changes in fair value due to credit risk for non-performing loans and advances/total gross non-performing loans and advances	(Bitar and Tarazi 2022; Alessi et al. 2022)	(−)
	RWA_VOLUME	No.	RWA volume	(Bitar and Tarazi 2022)	(−)
	OWN_FUNDS_TIER1	No.	Tier 1 capital volume	(Bitar and Tarazi 2022)	(−)
	SECURITIZATION	Percentage (%)	Securitization positions/risk-weighted exposure amounts for credit, counterparty credit, and dilution risks and free deliveries	(Di Tommaso and Pacelli 2022)	(−)
Quantitative Easing Variables	PEPP_PURCHASES	No.	Net purchases at book value	(Rizwan et al. 2020; Ari et al. 2021; (Hoang et al. 2021)	(−)
	ASSET_TO_GDP	Percentage (%)	Total assets/quarterly gross domestic product at market prices	(Rizwan et al. 2020; Ari et al. 2021; Hoang et al. 2021)	(−)
	QE_ANNOUNCEMENT	Binary (1/0)	Quantitative Easing (QE) Announcement: 1 Corresponding to dates: 18/03/2020 and 04/06/2020.	(Rizwan et al. 2020; Ari et al. 2021; Hoang et al. 2021)	(−)
	EXP_ASSET_PURC	No.	Expanded Asset Purchase Program (APP)	(Rizwan et al. 2020; Ari et al. 2021; Hoang et al. 2021)	(−)
	BOND_PURC	No.	Covered bonds purchases at book value (CBPP3)	(Rizwan et al. 2020; Ari et al. 2021; Hoang et al. 2021)	(−)

Table 1. *Cont.*

Variable Group	Variable Symbol	Parameter Shown	Explanation	Related Literature	Expected Sign
COVID-19 Variables	COVID19_DUMMY	Binary (1/0)	COVID-19 pandemic existence	(Demir and Danisman 2021; Laeven and Valencia 2018; Laeven and Valencia 2020, 2021)	(+)
	COVID19_VACCINATED	No.	COVID-19 vaccinated	(Demir and Danisman 2021; Laeven and Valencia 2018; Laeven and Valencia 2020, 2021)	(−)
	COVID19_DEATHS	No.	COVID-19 deaths	(Demir and Danisman 2021; Laeven and Valencia 2018; Laeven and Valencia 2020, 2021)	(+)
Cultural Dimension Variables	CULTURE_PCA	Percentage (%)	Cultural identity	Author's Calculations	(−)
COVID-19 Government Response Variables	CONTNMN	Index	Government response containment index	(Hoang et al. 2021; Couppey-Soubeyran et al. 2020; (Bassani 2021)	(+)
	GOVT_RESP_STR	Index	Government response stringency index	(Hoang et al. 2021; Couppey-Soubeyran et al. 2020; Bassani 2021)	(+)
	GOVT_ECON_SUP	Index	Government response economic support index	(Hoang et al. 2021)	(−)

A positive projected relationship is indicated by the '+' sign, while a negative projected relationship is indicated by the '−' sign.

4.3. Methodology and Econometric Models

We investigated the impact of COVID-19 on the NPLs in the European Union (EU28) during the period 2015–2021. OLS methodology for panel data was utilized to analyze and quantify the impact of the candidate predictor on the NPLs. Panel data analysis was conducted by utilizing fixed and random effects. Panel data leverage both the time-series and cross-sectional dimensions, enabling a comprehensive analysis. Although cointegration techniques are effective for identifying long-term equilibrium relationships between variables, we did not employ these methods in our analysis. Our study focused on the short- to medium-term impacts of COVID-19 on NPLs, making cointegration less relevant for our research objectives. We performed all the requirements for the whole sample period and then used the Hausman test to check the suitability of the random effects over the fixed-effects method. Most of the models developed are estimated using country-fixed effects, allowing for the management of time-constant unobserved country heterogeneity. Our research did not delve into company-specific characteristics. Therefore, we opted for country-fixed effects over individual fixed effects in our analysis, since our study focused on aggregated data at the country level. Several stationarity tests were performed to evaluate if the values were (trend)stationary.[3] We transformed the non-stationary variables to stationary by applying first and second differences accordingly.[4] Notably, the primary dependent variable, NPL, was identified to contain a unit root. To achieve stationarity in the NPL series, we applied the first differences. This step was essential to avoid producing biased results. The newly created NPL series effectively captured the variations in the NPL ratio over time. After differentiating when appropriate, the final dataset consisted only of stationary variables.[5]

Next, we also applied the Durbin–Watson statistic to recognize potential autocorrelation in the residuals. Based on the indication of the Durbin–Watson statistic, we

incorporated one and two lag periods of the dependent variable into our analysis.[6] This strategic inclusion mitigated the autocorrelation in the residuals, ultimately enhancing the robustness of our regression estimates. Moreover, we used the Akaike Information Criterion (AIC) to choose the appropriate lag length. Considering that the dynamic panel may yield biased results, as stated by Roodman (2009), and that an alternative model, such as GMM, could more effectively address issues such as reverse causality and omitted variable bias, we proceeded by comparing the two estimators before continuing with the analysis. Specifically, based on Bettinger (2010), we employed the Hausman test to compare the OLS and Generalized Method of Moments (GMM) method for dynamic panels (Hansen 1982). The Hausman test revealed that OLS yielded consistent estimates. Furthermore, when heteroscedasticity was present, either across cross-sectional units or across time segments, we applied the white cross-section or white period coefficient covariance method, respectively. Finally, we included autoregressive (AR) terms where appropriate, to mitigate potential autocorrelation in the error terms and to capture temporal dependencies between the data. To validate this choice, we performed additional analysis using both robust standard errors and AR components. The comparative results yielded that while robust standard errors adequately addressed autocorrelation, the inclusion of AR terms provided a more comprehensive model fit, capturing the temporal dynamics inherent in the NPL data more effectively.

In line with Xie et al. (2024), our strategy involved facilitating a comparative approach. While Xie et al. (2024) examined the pre- and post-COVID-19 periods, we expanded by analyzing and comparing three sample periods: (1) one related to the pre-COVID-19 period (Q1:2015 to Q4:2019), (2) one related to the post-COVID-19 period (Q1:2020 to Q4:2021), and finally, (3) one related to the entire period of analysis (Q1:2015 to Q4:2021). Both pre- and post-COVID-19 samples were a byproduct of the total sample period. While we included relevant COVID-19 variables in the models related to the post-COVID-19 period, data limitations and their unavailability before the pandemic[7] restricted their inclusion in pre-COVID-19 models. Instead, we incorporated the COVID19_DUMMY variable, a dummy variable with values 1/0 (1, corresponding to the pandemic's existence, and 0 otherwise), in the model related to the entire period of analysis. This variable effectively captures the pandemic occurrence. This strategy allowed us to assess the impacts of COVID-19 across the entire period and facilitate comparison, while also acknowledging the data constraints. While we included the COVID-19 variables in the post-COVID sample period, we included the COVID19_DUMMY variable only in the sample related to the entire sample period.

Based on the above, we formulated the following baseline estimation models:
Pre-COVID-19 period:

$$DNPL_{i,t} = \beta_0 + \beta_1 \times DNPL_{i(t-1),1} + \beta_2 \times DB_{it,2} + \beta_3 \times DM_{it,3} + \beta_4 \times DR_{it,4} + \beta_5 \times CULTURE\_PCA_{it,5} + u_{it} \quad (1)$$

Post-COVID-19 period:

$$DNPL_{i,t} = \beta_0 + \beta_1 \times DNPL_{i(t-1),1} + \beta_2 \times DB_{it,2} + \beta_3 \times DM_{it,3} + \beta_4 \times DR_{it,4} + \beta_5 \times CULTURE\_PCA_{it,5} + \beta_6 \times DQE_{it,6} + \beta_7 \times DG_{it,7} + \beta_8 \times DC_{it,8} + u_{it} \quad (2)$$

Total period:

$$DNPL_{i,t} = \beta_0 + \beta_1 \times DNPL_{i(t-1),1} + \beta_2 \times DB_{it,2} + \beta_3 \times DM_{it,3} + \beta_4 \times DCOVID19\_DUMMY_{it,4} + \beta_5 \times CULTURE\_PCA_{it,5} + \beta_6 \times DQE_{it,6} + u_{it} \quad (3)$$

where $DNPL_{i,t}$ denotes the aggregate non-performing loans to total gross loans, $DNPL_{i(t-1),1}$ corresponds to the NPLs of the prior quarter, $DB_{it,2}$ denotes the bank-specific variables, $DM_{it,3}$ represents the macroeconomic factors, $DR_{it,4}$ denotes the regulatory variables, the $DCOVID19\_DUMMY_{it,4}$ denotes the dummy variable related to COVID-19 existence,[8] the $CULTURE\_PCA_{it,5}$ denotes the control variable representing each nation's cultural identity, $DQE_{it,6}$ denotes the QE policy response variables, $DG_{it,7}$ denotes the government economic policy response variables, and finally, $DC_{it,8}$ denotes the COVID-19-related factors. Note that i corresponds to the examined country of the sample and t to the year. We used one

lag for selected bank-specific and macroeconomic regressors to achieve optimum model fit based on the indication of the Durbin–Watson statistic and to capture the dynamics of explanatory variables over the previous quarter.

We followed a top-down approach by breaking down further and analyzing each period into additional subsamples with alternative characteristics. Specifically, we explored the pandemic's impact on the European subregion, on country-level prosperity, distinguishing by NPL type and NPL economic sector/activity. We distinguish the core and peripheral economies based on the Phillips and Sul (2007) approach. Table A1 of Appendix A displays the classification of countries based on subregions and core/peripheral economies.

To obtain deeper insight into the relevance of the explanatory variables and to account for multicollinearity, we first controlled only by bank-specific and macro variables. We then included the regulatory variables, and next we included the government response variables; finally, we included the QE policy response variables (Table A3 of Appendix A). Moreover, we regressed by different NPL types and NPL economic sectors, by consecutively employing the dependent variables presented in Table A2 of Appendix A.

The empirical analysis was divided into baseline and subsample estimations. Although we chose to include the CULTURE_PCA factor in all baseline estimation models (Section 5.2), in the subsample analysis (Section 5.3), the CULTURE_PCA variable was included only in the models related to the post-COVID-19 period.

5. Results and Discussion

This section presents the main regression estimates, followed by a relevant discussion. Section 5.1 delves into the results of descriptive statistics and the correlation matrix, Section 5.2 presents the results of the baseline estimations, and Section 5.3 presents additional empirical results distinguishing by European subregion, core and peripheral European economies, NPL type, and NPL economic sector.

5.1. Descriptive Statistics

Before proceeding with our regression results, we generated descriptive statistics and correlation matrices. Table 2 depicts the descriptive statistics (individual samples) of both the primary dependent variable and the candidate predictors employed in the current research for the period 2015Q1 until 2021Q4 (descriptive statistics of secondary dependent variables and control variables are not shown due to space constraints). The fixed-effects method effectively mitigated the influence of high values of both non-normality distribution, as derived from the Jarque–Bera statistic, as well as kurtosis and skewness.[9] Additionally, to deal with highly correlated variables, we incorporated them in alternative empirical models.

Table 2. Descriptive statistics.

	NPLS	PEPP_PU RCHASES	ASSET_T O_GDP	QE_ANNOU NCEMENT	EXP_ASSET _PURC	BOND_P URC	COVID19_D UMMY	COVID19_ VACCINATED	COVID19_ DEATHS	CONTNMN	GOVT_ RESP_STR	GOVT_ ECON_SUP
Mean	6.818929	5.705593	0.006400	0.071429	8.883617	1.006164	0.253827	2.11×10^8	1.069482	3.235154	3.152849	3.848498
Median	3.701882	0.000000	0.006056	0.000000	6.361949	8.921000	0.000000	2.568623	2.325710	3.624000	3.454800	4.125000
Maximum	4.774785	3.395420	0.025949	1.000000	2.019962	3.416300	1.000000	3.49×10^9	8.680039	5.402000	5.516100	6.600000
Minimum	0.208018	0.000000	0.000700	0.000000	0.000000	−6.250000	0.000000	0.000000	1.000000	5.450000	5.044000	0.000000
Std. Dev.	8.858129	1.026505	0.004267	0.257704	9.174503	9.572955	0.435477	5.48×10^8	1.847337	1.305790	1.389047	1.986842
Skewness	2.744479	1.405314	0.489372	3.328201	0.073791	1.157496	1.131313	3.807389	2.416297	−0.576944	−0.365652	−0.442150
Kurtosis	1.064595	3.465069	2.403187	1.207692	1.080539	3.403692	2.279869	1.837132	8.485477	2.054388	1.840860	1.989601
Jarque–Bera	2.524793	9.468567	3.756203	4.138809	4.323796	6.442512	1.841768	2.550274	4.297823	1.928888	1.627956	1.562507
Probability	0.000000	0.008789	0.000000	0.000000	0.115106	0.039905	0.000000	0.000000	0.000000	0.000065	0.000292	0.000405
Sum	4.664148	1.597566	4.390062	5.600000	2.487413	2.817260	1.990000	4.39×10^{10}	2.06×10^8	6.729120	6.557925	8.004875
Sum Sq. Dev.	53,592.59	2.85×10^{11}	0.012472	5.200000	2.272630	2.47×10^9	1.484885	6.22×10^{19}	6.55×10^{14}	3.53×10^8	3.99×10^8	8.17×10^8
Observations	684	208	686	784	208	208	784	208	193	208	208	208

Note(s): NPLS stands for aggregate non-performing loans to total gross loans; PEPP_PURCHASES represents the net purchases at book value; ASSET_TO_GDP stands for the total assets/quarterly gross domestic product at market prices; QE_ANNOUNCEMENT denotes the Quantitative Easing (QE) Announcement: 1 Corresponding to Dates: 18 March 2020 and 4 June 2020; EXP_ASSET_PURC represents the Expanded Asset Purchase Program (APP); BOND_PURC stands for covered bonds purchases at book value (CBPP3); COVID19_DUMMY stands for COVID-19 pandemic existence; COVID19_VACCINATED represents the COVID-19 vaccinated population; COVID19_DEATHS represents the COVID-19 deaths; CONTNMN stands for government response containment index; GOVT_RESP_STR stands for government response stringency index; finally, GOVT_ECON_SUP represents the government response economic support index.

5.2. Baseline Estimations

Table 3 summarizes the results of the econometric estimation model related to the three periods of analysis.

Table 3. Main empirical findings.

Empirical Model:	Empirical Model 1	Empirical Model 2	Empirical Model 3	Empirical Model 4	Empirical Model 5	Empirical Model 6	Empirical Model 7
Period Examined:	Total	Before COVID-19	After COVID-19	After COVID-19	After COVID-19	After COVID-19	After COVID-19
Variable Symbol	Dependent Variable: D(NPLS)						
D(NPLS(-1))	−0.342359	−0.183453 **	−1.020644 ***	−1.206490 ***	−1.203173 ***	−1.168123 ***	−1.145578 ***
D(UNEMP(-1))	−0.035229 ***	0.015602 ***	−0.045157 ***	−0.019558	−0.023334	−0.028111	−0.026744
D(ROA(-1))	−0.291491 ***	−0.109478	−0.175687	−0.170662	−0.216442	−0.206493	−0.204913
D(CPI(-1))	−0.000515	0.002510	−0.001905	−0.005244 ***	−0.001308	−0.001966	−0.002714
D(CAP(-1))	0.149963	−0.254603 ***	0.377711 *	0.318489	0.140830	0.230260	0.263534
D(LOAN_DISBRS(-1))	−0.003586	0.014573 *	−0.021015	−0.036377	−0.028291	−0.049109	−0.056952
R_GDP_Q2Q(-1)	−0.001371	0.022920	−0.002708	0.002934	0.000898	−0.000133	−0.000636
COVID19_DUMMY	0.032328						
COVID19_VACCINATED			-33.38×10^{-9} ***		−0.036204 ***	−0.027599 **	−0.027225 **
GOVT_RESP_STR			0.000140 **				572.9648 **
CONTNMN						578.1687 **	
GOVT_ECON_SUP			0.000058	6.64×10^{-5}			
D(ASSET_TO_GDP)				−0.002029 *			−2.775.504
COVID19_DEATHS					−3.305.095		
C	−0.411011	−0.381554	−0.657638	−0.772071	−0.493559	−0.677555	−0.709040
Observations:	402	305	97	90	85	97	90
R-squared:	0.529274	0.656643	0.762686	0.730969	0.745125	0.738525	0.740121
F-statistic:	6.573298	6.731721	8.652614	7.942145	0.637128	0.641406	0.632869
Prob(F-stat):	0.000000	0.000000	0.000000	0.000001	0.000000	0.000000	0.000000

Note(s): (1.) Model 1 refers to the total sample period: 2015Q1–2021Q4, aiming to explore the COVID-19 impact on the change in NPLs of the European banks. Model 2 refers to the period before the COVID-19 pandemic: 2015Q1–2019Q4, aiming to examine the macro and bank-specific variables' effect on the change in NPLs. Models 3 to 7 refer to the period after the pandemic: 2020Q1–2021Q4, aiming to examine the effect of both COVID-19 and policy response variables on the change in NPLs, as well as the effect of the central bank and government policy support measures. (2.) OLS methodology was employed for the regression model estimation. More specifically, Fixed Corrected Panel Effects estimations with country-fixed effects were utilized for all models because of the Hausman test. The table presents the values of the coefficients, while the significance of the p-value is presented with an asterisk: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. (3.) NPLS stands for aggregate non-performing loans to total gross loans; UNEMP represents the % of unemployment; ROA stands for return on assets; CPI stands for quarterly consumer price index; CAP stands for the bank capital and reserves to total assets; LOAN_DISBRS represents the loan disbursements to customers; R_GDP_Q2Q stands for the quarterly percentage growth rate of the real GDP; COVID19_DUMMY stands for the COVID-19 pandemic existence; COVID19_VACCINATED represents the vaccinated population against the COVID-19 pandemic; GOVT_RESP_STR stands for government response stringency index; CONTNMN stands for government response containment index; GOVT_ECON_SUP represents the government response economic support index; ASSET_TO_GDP stands for the total assets/quarterly gross domestic product at market prices; finally, COVID19_DEATHS represents the COVID-19 deaths. (4.) The (-1) denotes one period lag. This note also applies to Tables 3 and 4. (5.) We opted not to include t-statistics for the economy of space. Also, the inclusion of coefficient estimates and p-values effectively communicates the statistical significance of our results. This note also applies to Tables 3 and 4.

Table 4. Main empirical findings with the inclusion of the cultural identity (CULTURE_PCA) control variable.

Empirical Model:	Empirical Model 1	Empirical Model 2	Empirical Model 3	Empirical Model 4	Empirical Model 5	Empirical Model 6	Empirical Model 7
Period Examined:	Total	Before COVID-19	After COVID-19	After COVID-19	After COVID-19	After COVID-19	After COVID-19
Variable Symbol	Dependent Variable: D(NPLS)						
D(NPLS(-1))	−0.785255 ***	−0.019509	−1.704412 ***	−2.807484 ***	−1.125098 ***	−2.606772 ***	−2.542010 ***
D(UNEMP(-1))	−0.000233 ***	0.018100 ***	−0.135149 ***	0.054589	−0.055827	0.026299	0.048239
D(ROA(-1))	−0.003389 ***	0.005782	−1.482710 **	−1.177897 *	−0.529568	−1.178363	−1.486075 *
D(CPI(-1))	−0.000159 ***	0.001973	−0.020326	−0.031509 *	0.001022	−0.033672	−0.039164 *
D(CAP(-1))	0.304569 ***	−0.236408 ***	0.213311	1.441384 ***	0.716930	1.040842 *	1.862165 **
D(LOAN_DISBRS(-1))	-7.51×10^{-5}	0.009489	−0.169374	−0.352748 *	−0.047177	−0.206157	−0.286899 *
R_GDP_Q2Q(-1)	−0.000220 ***	0.016549	−0.024380	−0.041457 *	−0.030766	−0.031095	−0.037303

Table 4. *Cont.*

Empirical Model:	Empirical Model 1	Empirical Model 2	Empirical Model 3	Empirical Model 4	Empirical Model 5	Empirical Model 6	Empirical Model 7
Period Examined:	Total	Before COVID-19	After COVID-19	After COVID-19	After COVID-19	After COVID-19	After COVID-19
Variable Symbol	Dependent Variable: D(NPLS)						
COVID19_DUMMY	0.005049						
COVID19_VACCINATED			-2.13×10^{-9}		-0.099896 ***	0.025171	0.023815
GOVT_RESP_STR			0.000196 **				1365.652 ***
CONTNMN						-9.796766	
GOVT_ECON_SUP			0.000135				
D(ASSET_TO_GDP)				-46.53331			-13.21214 **
COVID19_DEATHS				-84.60416	11.64699 *		
CULTURE_PCA	-0.000780 ***	0.007468	-0.169073 ***	-0.105757 *	-0.412581 ***	-0.137801 *	-0.114973 *
C	-0.001125	-0.353802	-0.693554	-1.527742	-0.259571	-1.184283	-1.366887
Observations:	355	261	94	87	82	94	87
R-squared:	0.993921	0.723856	0.958924	0.976036	0.956702	0.945497	0.978265
F-statistic:	6.799609	8.493018	8.646293	4.329519	9.722032	6.425089	1.166894
Prob(F-stat):	0.000000	0.000000	0.000000	0.000476	0.000000	0.001973	0.001288

Note(s): (1.) Model 1 refers to the total sample period: 2015Q1–2021Q4, aiming to explore the COVID-19 impact on the change in NPLs of the European banks controlling for the cultural identity. Model 2 refers to the period before the COVID-19 pandemic: 2015Q1–2019Q4, aiming to examine the macro and bank-specific variables' effect on the change in NPLs controlling for the cultural identity. Models 3 to 7 refer to the period after the pandemic: 2020Q1–2021Q4, aiming to examine the effect of both COVID-19 and policy response variables on the change in NPLs, as well as the effect of the central bank and government policy support measures controlling for the cultural identity, respectively. The number of observations was adjusted in each model to account for the inclusion of the CULTURE_PCA variable. (2.) OLS methodology was employed for the regression model estimation. More specifically, Fixed Corrected Panel Effects estimations with country-fixed effects were utilized for all models because of the Hausman test. The table presents the values of the coefficients, while the significance of the *p*-value is presented with an asterisk: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. (3.) NPLS stands for aggregate non-performing loans to total gross loans; UNEMP represents the % of unemployment; ROA stands for return on assets; CPI stands for quarterly consumer price index; CAP represents the bank capital and reserves to total assets; LOAN_DISBRS represents the loan disbursements to customers; R_GDP_Q2Q stands for the quarterly percentage growth rate of the real GDP; COVID19_DUMMY stands for the COVID-19 pandemic existence; COVID19_VACCINATED represents the vaccinated population against the COVID-19 pandemic; GOVT_RESP_STR stands for government response stringency index; CONTNMN stands for government response containment index; GOVT_ECON_SUP represents the government response economic support index; ASSET_TO_GDP stands for the total assets/quarterly gross domestic product at market prices (a representative of quantitative easing measures); COVID19_DEATHS represents the COVID-19 deaths; finally, the variable CULTURE_PCA represents the synthetic cultural identity variable. (3.) The introduction of the variable CULTURE_PCA in Table 4 led to a reduction in the total number of observations, as evident from the models presented in Table 4. This decrease in observations was attributed to the presence of missing values for the newly included variable. It is important to note that the same pattern was observed in Table 5, as the inclusion of CULTURE_PCA also impacted the overall sample size.

Table 5. Robustness stepwise regression (forward) empirical findings with the inclusion of the cultural identity (CULTURE_PCA) control variable.

Empirical Model:	Empirical Model 1	Empirical Model 2	Empirical Model 3	Empirical Model 4	Empirical Model 5	Empirical Model 6	Empirical Model 7
Period Examined:	Total	Before COVID-19	After COVID-19	After COVID-19	After COVID-19	After COVID-19	After COVID-19
Variable Symbol	Dependent Variable: D(NPLS)						
D(NPLS(-1))	-0.181402 **	-0.019509	-0.301656 *	-0.278239 *	-0.498079 **	-0.271632 *	-0.270032 *
D(UNEMP(-1))	-0.005908 *	0.003338 *	-0.000312 *	-0.001151	0.013107	0.000785	0.002276
D(ROA(-1))	-0.298137 *	0.019812	-0.305789 *	-0.368425 *	-0.318869	-0.279971	-0.281101 *
D(CPI(-1))	-0.000642 *	0.002699	-0.001511	-0.000592 *	-0.000191	-0.000225	-0.000689 *
D(CAP(-1))	0.100329 *	-0.144452 *	0.221153	0.248323 *	0.041593	0.137912 *	0.167602 *
D(LOAN_DISBRS(-1))	-0.007724	0.006638	-0.002502	-0.010842 *	0.026314	0.009312	0.004685
R_GDP_Q2Q(-1)	-0.000593 **	0.011218	-0.003963	-0.003617 *	-0.002826	-0.002181	-0.002388
COVID19_DUMMY	0.003182						
COVID19_VACCINATED			-0.030673 ***		-0.032643 **	-0.030712 ***	-0.030607 ***
GOVT_RESP_STR			180.3136 *				198.7774 *
CONTNMN						-179.7631 *	
GOVT_ECON_SUP			5.78×10^{-5}	4.84×10^{-5}			
D(ASSET_TO_GDP)				-10.65967			-3.051043 *
COVID19_DEATHS					0.410834 *		
CULTURE_PCA	-0.020594 *	-0.015615	-0.002635 **	-0.005780 *	-0.013931 **	-0.005301 *	-0.011990 *
C	-0.025936	-0.042776	-0.131299	-0.008730	-0.165630	-0.063473	-0.034680

Table 5. *Cont.*

Empirical Model:	Empirical Model 1	Empirical Model 2	Empirical Model 3	Empirical Model 4	Empirical Model 5	Empirical Model 6	Empirical Model 7
Period Examined:	Total	Before COVID-19	After COVID-19	After COVID-19	After COVID-19	After COVID-19	After COVID-19
Variable Symbol	Dependent Variable: D(NPLS)						
Observations:	355	261	94	87	82	94	87
R-squared:	0.549349	0.587550	0.740867	0.705801	0.762225	0.731266	0.736555
F-statistic:	9.351394	4.843548	16.133090	13.471660	16.768050	16.745510	14.378710
Prob(F-stat):	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.001288

Note(s): (1.) Model 1 refers to the total sample period: 2015Q1–2021Q4. Model 2 refers to the pre-pandemic period: 2015Q1–2019Q4. Models 3 to 7 refer to the post-pandemic period: 2020Q1–2021Q4. (2.) OLS methodology was employed for the regression model estimation. More specifically, Fixed Corrected Panel Effects estimations with country-fixed effects were utilized for all models because of the Hausman test. The table presents the values of the coefficients, while the significance of the *p*-value is presented with an asterisk: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. (3.) NPLS stands for aggregate non-performing loans to total gross loans; UNEMP represents the % of unemployment; ROA stands for return on assets; CPI stands for quarterly consumer price index; CAP represents the bank capital and reserves to total assets; LOAN_DISBRS represents the loan disbursements to customers; R_GDP_Q2Q stands for the quarterly percentage growth rate of the real GDP; COVID19_DUMMY stands for the COVID-19 pandemic existence; COVID19_VACCINATED represents the vaccinated population against the COVID-19 pandemic; GOVT_RESP_STR stands for government response stringency index; CONTNMN stands for government response containment index; GOVT_ECON_SUP represents the government response economic support index; ASSET_TO_GDP stands for the total assets/quarterly gross domestic product at market prices (a representative of quantitative easing measures); COVID19_DEATHS represents the COVID-19 deaths; finally, the variable CULTURE_PCA represents the synthetic cultural identity variable. (3.) The introduction of the variable CULTURE_PCA in Table 4 led to a reduction in the total number of observations, as evident from the models presented in Table 4. This decrease in observations was attributed to the presence of missing values for the newly included variable.

Regarding the results related to the entire analysis period, we observed a statistically insignificant effect of the pandemic on the change in NPLs (variable COVID19_DUMMY), whereas bank profitability and unemployment rate had a statistically significant negative effect on the change in NPLs. The statistically insignificant impact of COVID-19 on the change in NPLs during the total period indicates that the strong capital accumulation of banks after the GFC increased bank profitability (ROA), rendering them resilient to the pandemic's effect.

Regarding the results related to the pre-COVID-19 period, our regression estimates showed that loan disbursements exerted a positive and statistically significant effect on the change in NPLs. This implies that banks should implement new risk auditing policies when granting loans. We also found that bank capitalization was significant and negatively affected the change in NPLs. This suggests that more capitalized banks were able to fare better during the crisis, which aligns with Demir and Danisman's work (2021). In line with Makri et al. (2014), we also find that the unemployment rate was statistically significant, positively affecting the change in NPLs, suggesting that higher unemployment rates were associated with an increase in the change in NPLs.

Coming to the results related to the post-COVID-19 period, we observed that both bank capitalization and strictness of lockdown measures were statistically significant factors that positively affected the change in NPLs. Our findings align with those of Yi et al. (2022) and Apergis (2022). Additionally, the QE measures assisted borrowers in meeting their regular loan repayment obligations despite the adverse macroeconomic conditions. We also found that government economic policies had a statistically insignificant effect on the change in NPLs. The unpredictability, severity, and scale of the pandemic posed challenges for tailored government economic policies to effectively address the situation. This result aligns with the conclusions of Dunbar (2022). Additionally, the effectiveness of these policies may have been overshadowed by the combined force of significant capital accumulation facilitated by the banks following the GFC crisis, along with the implementation of QE measures. Similar to the findings reported by Cowling et al. (2022), as COVID-19 vaccinations increased, more borrowers were able to generate income, enabling them to meet their loan obligations, eventually reducing the NPLs. Additionally, the negative sign of the coefficient of the UNEMP during the post-pandemic period suggests that other mitigating factors, such

as government support measures and improving economic conditions as vaccinations increase (Table 3, empirical model 3), outweighed the immediate impact of the increased unemployment rate.

The results of Table 4 present the regression estimates with the inclusion of the CUL-TURE_PCA factor.

Regarding the total period of analysis (Table 4, empirical model 1), contrary to Ari et al. (2021), who anticipated a surge in NPLs, we observed a statistically insignificant effect of COVID19_DUMMY on the change in NPLs, while national culture was a statistically significant factor posing a negative effect on the change in NPLs. This implies that cultural-driven economies support economic growth, enabling debtors to effectively cope with their debt obligations. Adding to the results of Table 3, regarding the total period of analysis, debtors have continued to effectively meet their loan obligations, despite the challenges posed by the pandemic, translating into an outcome where the effect of the pandemic on bank stability remained statistically insignificant. This finding is consistent with the work of Demir and Danisman (2021), who stressed the crucial role of strong capital buffers. We also noticed that incorporating the CULTURE_PCA variable into our analysis reduced the effect of COVID19_DUMMY on the change in NPLs. This finding underscores the importance of cultural-driven economies in maintaining financial stability.

Regarding the pre-COVID-19 period, we did not find any statistically significant effect of national culture on economic growth and the change in NPLs. In the post-COVID-19 period, we observed a positive effect of national culture on bank capital, bank profitability, and economic growth, implying that borrowers' commitment to national values resulted in economic expansion and, consequently, increased bank capital and bank profitability. This finding aligns with the conclusions of Gaganis et al. (2020), who found that cultural factors significantly influenced financial stability and economic performance. Better capitalized banks reporting high profitability ratios could easily absorb the negative spillover effects of COVID-19 and avert a new wave of NPLs. Additionally, we found that the implementation of QE measures (differentiated variable: ASSET_TO_GDP) in cultural-driven economies led to a significant NPL reduction.

The pandemic introduced unparalleled economic uncertainty (Yi et al. 2022). Also, the pandemic's impacts, as well as the economic support policies implemented to mitigate the pandemic's economic effect, were not uniform across European countries and cultures. This dynamic interplay caused individuals and businesses to reassess their financial decisions and, consequently, their attitudes toward NPL repayment activities. Additionally, while in the pre-COVID-19 period the cultural norms were overshadowed by economic factors and the regulatory environment, during the lockdown period of COVID-19, debt repayment was primarily influenced by the inherent cultural values.

5.3. Subsample Analysis

Tables A5–A10 of Appendix A depict the empirical results, distinguishing by European subregion, core, peripheral European countries, NPL type, and NPL economic sector. While the baseline estimations yielded encouraging results, the subsample analysis revealed the specific arrears being affected, ultimately experiencing an increase in their NPLs. Sections 5.3.1 and 5.3.2 summarize the key findings.[10]

5.3.1. The Entire Period

From Table A5 (MODELS 1–7), we found that COVID19_DUMMY did not exert a significant effect on NPLs. Consistent with the findings of Makri et al. (2014), the unemployment rate was significant and exerted a positive effect on the change in NPLs. The change in NPLs of the prior period was significant and exerted a positive effect on the change in NPLs of the current period. The net purchases at book value (PEPP_PURCHASES) variable was significant and exerted a negative effect on the change in NPLs. This finding is supported by Yi et al. (2022), who emphasized the positive impact of QE policies on

financial stability during the pandemic. Bank capital was significant and exerted a positive effect on the change in NPLs.

5.3.2. Pre-COVID-19 Period

From Panel A (Tables A6–A8) of Appendix A, we observed that South Europe was the most vulnerable to external macroeconomic forces. Similar to the findings of Demir and Danisman (2021), we also observed that core European economies with high profitability ratios were more resilient. The mortgage NPLs (NPL_RATIO_MORT), and the NPLs related to small and medium-sized enterprises (SMEs), non-financial corporations (NPL_RATIO_NFCs), households (NPL_RATIO_HHs), and commercial real estate (CRE), were found to be vulnerable to external macroeconomic shocks. Notably, NPL_RATIO_CRE, NPL_RATIO_SME, and NPL_RATIO_NFCs NPL portfolios are well capitalized, offering a buffer against potential macroeconomic turbulences. This observation is supported by Bitar and Tarazi (2022), who emphasized that higher capitalization ratios help banks absorb economic shocks more effectively.

5.3.3. Post-COVID-19 Period

From Table A5 (Model 7), Panel B (Tables A6–A8), as well as Tables A9 and A10 of Appendix A, we derived that the NPLs were negatively affected by the increase in COVID-19 deaths (Rizwan et al. 2020). The presence of QE measures enhanced the banks' capital, preventing a new wave of NPLs, consistent with the observations of Yi et al. (2022). Core economies fared better in comparison to peripheral economies due to a sounder financial system (Apergis 2022). The results of Table A8 indicate that the strictness of lockdown measures hindered the ability, in particular, of SMEs, to repay the loans, unlike larger firms. This was in line with the findings of Cowling et al. (2022). The findings presented in Table A9 suggest that sectors that were considered essential and continued their operations during lockdown periods were not as severely affected, whereas sectors involving physical work-related activities experienced an increase in their NPLs due to the strictness of lockdown measures. Those sectors were the following: "agriculture, forestry, and fishing" (NFCNPL_AGR), "education" (NFCNPL_EDU), "information and communication" (NFCNPL_INF), "manufacturing" (NFCNPL_MAN), "professional, scientific, and technical activities" (NFCNPL_PRF), "accommodation and food service activities" (NFCNPL_ACC), "administrative and support service activities" (NFCNPL_ADM), and "human health services and social work activities" (NFCNPL_HUM). These findings are in line with Sharif et al. (2020), who stated that the pandemic may be viewed as an economic crisis, implying the differential impacts of the pandemic on various economic sectors. Moreover, from Table A5 (Model 7), we also found that the banks' securitization strategy was statistically significant and reduced the NPLs.

Regarding the role of national culture, from Table A10 we deduced that cultural influences in central European economies exerted a positive impact on borrowers' willingness to fulfill their loan commitments, ultimately leading to a decrease in the NPLs. Borrowers from southern European economies, with strong cultural ties, were more likely to get vaccinated against COVID-19, which in turn revitalized the economy and resulted in a decrease in NPLs. Northern European economies benefited from a strong cultural identity, which contributed to economic prosperity, bank profitability, and reduced NPLs. Culture had a significant negative effect, particularly in SME NPL portfolios. There was a statistically significant relationship between national culture and borrowers' willingness to receive a COVID-19 vaccine in various NPL sectors, which implies increased borrower cash flows and a subsequent reduction in NPLs associated with these sectors. Notably, we deduced a significant negative effect, particularly in the NPLs of the "electricity, gas, steam, and air conditioning supply" sector (NFCNPL_ELE).

Based on these findings, we also deduced that in countries where tradition and benevolence are prevalent, policies that emphasize social responsibility and community welfare are likely to be more effective. Additionally, banks should incorporate cultural

assessments into their risk management frameworks, tailoring financial products and services to align with the cultural values of their customers to enhance borrower loyalty and reduce default rates. Community-based financial education programs that resonate with local cultural values can improve financial literacy and behaviors, thereby reducing non-performing loans (NPLs). Moreover, promotional strategies that emphasize the alignment of financial products and policies with cultural norms can increase the adoption of financial products and improve compliance with repayment obligations.

6. Robustness Tests

The results reported in the previous section were based on the application of the fixed-effects method for the entire sample period. First, regarding the variation in sample size between the dependent variable (684 observations) and the other variables in our empirical estimates, we conducted sensitivity analyses to assess the potential impact on our findings. This involved systematically testing our models with various subsamples and configurations. These sensitivity tests considered scenarios where we first included and then excluded certain quarters and countries to gauge the robustness of our results to changes in sample composition. The sensitivity analyses reaffirmed the stability of our key findings.

For robustness, we also proceeded by estimating alternative econometric models. Specifically, we first regressed each independent variable against the dependent variable. We then proceeded by successively including each independent variable, while regressing with the change in NPLs. Those models confirmed the subsample analysis, as reported in Section 5.3. They also provided additional insights and helped mitigate the concerns related to the smaller sample size in the post-COVID-19 period.[11] It was confirmed that banks' securitization strategy was statistically significant and positively associated with NPL reduction of all NPL portfolios, except HH NPLs. It was derived that bank risk indicators, such as risk capital, operational risk, and risk-weighted assets (RWA), were statistically significant and exerted a positive effect on NPLs. A rise in the NPLs due to the increased COVID-19 deaths in South Europe was averted due to the government's financial support. Additionally, a rise in HH and MORT NPLs was averted due to the government's financial aid, which enabled the debtors to continue meeting their loan repayment obligations. On the other hand, the other NPL types relied on strong capitalization, high profitability, and QE measures. Although the MORT NPL portfolio was covered with enhanced provisioning, HH NPLs exhibited increased risk exposure (results available upon request).

We also conducted the same empirical analysis by excluding an important economic center, the United Kingdom, from the period 2020Q1 to 2021Q4. The analysis revealed that the magnitude of the coefficients and the significant indicators slightly increased. This indicates that, even without the United Kingdom, the remaining European countries had enough financial strength to handle the negative economic spillovers of COVID-19.[12] The third robustness test was related to the dependent variable. More specifically, we conducted a series of empirical estimations using the NPL ratio from the EBA database as an alternative response variable. The derived results confirmed the findings obtained from the initial empirical models.[13]

Additionally, we conducted a series of robustness checks to validate our findings regarding the effect of national culture on NPLs. We calculated the average of the ten Schwartz national culture dimensions for each country and period (Schwartz 1994) and performed the same analysis. We also independently employed each Schwartz cultural dimension for both the pre- and post-COVID-19 periods. Our robustness checks confirmed the pre-COVID-19 period results of Table 4. Moreover, in the post-COVID-19 period, all national cultural values were found to negatively affect the change in NPLs, with tradition (IMPTRAD), benevolence (IPEQOPT), power (IPRULE), and security (IPSTRGV) showing the most significant negative effects. This implies that adherence to national traditions, rules, and feelings of safety and security were associated with lower NPL ratios. Banks operating in countries with these cultural values could tailor their financial products and

repayment plans to align with prudence and security, thereby reducing default risks (results not reported due to space limitations).

As an alternative methodology, we applied stepwise regression (forward). The results are reported in Table 5. Despite some differences in the magnitude of the coefficients, the results confirmed the findings of Table 4.

To address the issue of endogeneity in terms of policy responses and confirm the validity of the baseline estimations, we utilized the Arellano and Bond (1991) difference GMM method (Hansen 1982) for dynamic panels by employing the dependents' variable one lag period as an instrumental variable, since the current NPLs are also a byproduct of the NPLs of the prior period, making the lagged dependent variable an appropriate instrument to account for endogeneity. Furthermore, we employed the second lag of control variables as additional instrumental variables, enhancing the exogenous variation in our model and the causal relationship between the policy responses and the growth of NPLs. The empirical results of GMM were reported to be quite similar in terms of the magnitude and sign of the coefficients to the empirical results reported in Sections 5.2 and 5.3. This implies that the primary methodology employed was robust and effectively addressed endogeneity that may arise due to reverse causality and omitted variable bias. GMM results supported our baseline estimations and highlighted that peripheral economies were able to withstand the negative economic spillovers of COVID-19 due to the combination of capital accumulation and government economic support, while core economies were able to quickly recover from the pandemic due to sounder financial systems, enabling them to respond with a faster speed and a better solution to COVID-19. To validate the GMM model's results, we used the J-Test for over-identification restrictions, which was found to be valid. As an additional alternative methodology, we also applied Robust Least Squares (RLS), which yielded similar results to the prior section (the empirical estimates of GMM and RLS are not reported due to space limitations).[14] Furthermore, we included interaction terms to capture the interplay between 'bank capital'–'government economic support', 'bank capital'–'QE policy measures', 'securitization'–'government economic support', and 'cultural identity'–'GDP growth' and the growth of NPLs. The interaction terms analysis validated our findings that peripheral economies exhibited resilience against the pandemic's economic spillovers due to the synergy between capital accumulation and government economic support. Welch's *t*-tests and Kruskal–Wallis tests were conducted to compare the means of the subsamples analyzed, indicating significant differences between the means and, therefore, reinforcing our earlier findings related to the combined impact of securitization, wealth, and government economic support on the growth of NPLs.

Finally, as an alternative research approach, we also followed a difference-in-differences (DID) research approach to investigate the effects of the COVID-19 pandemic on the change in NPLs by comparing changes over time between two groups. In this approach, we created a treatment group representing the post-COVID-19 period and a control group representing the pre-COVID-19 period. Following this strategy, we derived that in the treatment group, the change in NPLs continued to decrease, compared to the control group, further strengthening the validity of our primary pre- and post-COVID-19 research approach. We also excluded outlier periods characterized by extreme values in key variables, such as NPLs, COVID-19 deaths, and government response indices. Specifically, quarters with Z-scores greater than 3 or less than -3 for these variables were removed from the analysis. After excluding these outliers, we found that our results remained consistent, suggesting that our findings were not driven by these extreme values. Lastly, we also performed a rolling window analysis with an eight-quarter window to observe the stability of our results over time. The rolling window analysis confirmed that the relationships between COVID-19 measures, economic support policies, and NPLs were stable across different sub-periods, further validating the robustness of our findings (robustness tests available upon request).

7. Conclusions and Future Research

This paper examined the effects of the COVID-19 pandemic on the European Union's NPLs. This research is the first to analyze this effect by European subregion, on country-level prosperity, distinguishing NPL type and NPL economic sector.

Our empirical results indicated that the extensive loan disbursements during the pre-pandemic period contributed to the rise in NPLs. This suggests that European banks should establish additional risk auditing policies. Despite the adverse economic spillovers of COVID-19, the accumulation of bank capital after the GFC, along with the government and central bank economic support provided, resulted in a substantial NPL reduction. Specifically, we found that peripheral economies were able to withstand the negative economic spillovers of COVID-19, primarily due to the combination of capital accumulation and government economic support. On the other hand, core economies were able to quickly recover due to their robust profitability ratios. The successive lockdowns particularly affected the NPL growth of SMEs, while larger firms performed better. In line with Dunbar (2022), households were able to continue meeting their loan repayment obligations due to the government's financial support. Additionally, in line with Cowling et al. (2022), physical work-related activities were severely affected by the successive lockdowns, resulting in higher NPLs, while vital sectors that continued their normal operations were not affected. In line with Cicchiello et al. (2022), while vaccinations increased, NPLs decreased, enabling a functional economy and leading to high loan repayment rates. Additionally, bank risk indicators increased dramatically during the pandemic, suggesting the need for the implementation of new and effective risk management practices. Finally, we also concluded that even during the pandemic, the brutal securitization strategy that banks pursued, along with the economic support policies, resulted in a substantial decrease in NPLs.

This study was also innovative by being the first to highlight the effect of cultural values on both borrowers' and lenders' behavior. More specifically, borrowers in culturally driven countries encourage innovation and competitiveness, ultimately boosting the economy. Despite the increased levels of economic uncertainty, we provided evidence that the rate of debt repayment increased in conjunction with cultural values, ultimately reducing the NPLs.

Policymakers and financial institutions can use these insights to mitigate the impact of future economic shocks by enhancing risk auditing policies, encouraging capital accumulation, implementing dynamic stress testing, and providing targeted support for SMEs and vulnerable sectors. By understanding and leveraging cultural factors, financial policies can be more effectively tailored to promote economic resilience and stability. Moreover, centralized support mechanisms, continuous monitoring and dynamic adaptation of economic policies, coordination between monetary and fiscal policies, as well as the implementation of advanced risk management practices, are essential in preparing for and responding to future crises. These measures, combined with robust securitization frameworks, can significantly reduce the risk of NPLs and maintain financial stability during economic downturns.

Future studies could examine the effect of COVID-19 on NPLs utilizing additional candidate predictors. They could also examine the relationship between environmental (E), social (S), and governmental (G) factors and NPLs. Moreover, they could extend the temporal coverage by including recent economic events, such as the geopolitical conflict between Russia and Ukraine. Finally, they could also examine the effect of the energy crisis on the NPLs or conduct a county-level analysis, considering the cultural identity.

Finally, this research provided robust results for both scientific and policymaking purposes. This research is also expected to pave the way for a new branch of literature related to the factors affecting the NPLs, ultimately leading to a revised strategy for resolving NPLs, not only for Europe but also on a global scale.

Author Contributions: J.H.P.: Conceptualization, Investigation, Writing—Original Draft, Validation, Visualization. D.K.: Conceptualization, Supervision, Writing—Review and Editing, Validation,

Visualization. G.A.S.: Conceptualization, Investigation. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Table A1. Country sample.

Country	Observation per Country	Cumulative Observation Count	Subregion Categorization	Core/Periphery Categorization
Denmark	28	28	Northern Europe	Intermediate group
Spain	28	56	Southern Europe	Extended Periphery
United Kingdom	28	84	Northern Europe	Intermediate group
France	28	112	Southern Europe	Hard-Core group
Italy	28	140	Southern Europe	Extended Periphery
Ireland	28	168	Northern Europe	Extended Periphery
Finland	28	196	Northern Europe	Extended Periphery
Portugal	28	224	Southern Europe	Extended Periphery
Sweden	28	252	Northern Europe	Intermediate group
Greece	28	280	Southern Europe	Extended Periphery
Austria	28	308	Central Europe	Hard-Core group
Belgium	28	336	Northern Europe	Hard-Core group
Germany	28	364	Central Europe	Hard-Core group
Netherlands	28	392	Central Europe	Hard-Core group
Bulgaria	28	420	Southern Europe	Extended Periphery
Croatia	28	448	Southern Europe	Extended Periphery
Czech Republic	28	476	Central Europe	Intermediate group
Estonia	28	504	Northern Europe	Intermediate group
Hungary	28	532	Central Europe	Extended Periphery
Latvia	28	560	Northern Europe	Intermediate group
Lithuania	28	588	Northern Europe	Intermediate group
Luxembourg	28	616	Central Europe	Intermediate group
Malta	28	644	Southern Europe	Extended Periphery
Poland	28	672	Central Europe	Intermediate group
Romania	28	700	Central Europe	Extended Periphery
Slovenia	28	728	Southern Europe	Intermediate group
Slovakia	28	756	Central Europe	Intermediate group
Cyprus	28	784	Southern Europe	Extended Periphery
Total	**784**	**784**	-	-

Notes: (1) This table presents the sample of countries that synthesize the data of our research, the observation distribution by country, their categorization per subregion, and core/peripheral economies. (2) The core/peripheral economies are distinguished based on the Phillips and Sul (2007) approach. The total number of country data points used in our research was 784 country observations.

Table A2. Data sources and description for the dependent variables.

Variable	Variable Role	Variable Group	Explanation	Source	Parameter Shown
NPLS	Main dependent variable	NPL Ratio	Aggregate non-performing loans to total gross loans	ECB	Percentage (%)
NPL_RATIO_HHS	Secondary Dependent variable	NPL Type	Aggregate non-performing loans to total gross loans—Households	EBA	Percentage (%)
NPL_RATIO_MORT	Secondary Dependent variable	NPL Type	Aggregate non-performing loans to total gross loans—Mortgages	EBA	Percentage (%)

Table A2. *Cont.*

Variable	Variable Role	Variable Group	Explanation	Source	Parameter Shown
NPL_RATIO_NFCS	Secondary Dependent variable	NPL Type	Aggregate non-performing loans to total gross loans—Non-financial corporations	EBA	Percentage (%)
NPL_RATIO_SME	Secondary Dependent variable	NPL Type	Aggregate non-performing loans to total gross loans—Small and medium-sized enterprises	EBA	Percentage (%)
NPL_RATIO_CRE	Secondary Dependent variable	NPL Type	Aggregate non-performing loans to total gross loans—Commercial real estate	EBA	Percentage (%)
NFCNPL_AGR	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—A: Agriculture, forestry, and fishing	EBA	Percentage (%)
NFCNPL_MIN	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—B: Mining and quarrying	EBA	Percentage (%)
NFCNPL_MAN	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—C: Manufacturing	EBA	Percentage (%)
NFCNPL_ELE	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—D: Electricity, gas, steam, and air conditioning supply	EBA	Percentage (%)
NFCNPL_WAT	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—E: Water supply	EBA	Percentage (%)
NFCNPL_CON	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—F: Construction	EBA	Percentage (%)
NFCNPL_WRT	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—G: Wholesale and retail trade	EBA	Percentage (%)
NFCNPL_TRA	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—H: Transport and storage	EBA	Percentage (%)
NFCNPL_ACC	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—I: Accommodation and food service activities	EBA	Percentage (%)
NFCNPL_INF	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—J: Information and communication	EBA	Percentage (%)

Table A2. *Cont.*

Variable	Variable Role	Variable Group	Explanation	Source	Parameter Shown
NFCNPL_FIN	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—K: Financial and insurance activities	EBA	Percentage (%)
NFCNPL_REA	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—L: Real estate activities	EBA	Percentage (%)
NFCNPL_PRF	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—M: Professional, scientific, and technical activities	EBA	Percentage (%)
NFCNPL_ADM	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—N: Administrative and support service activities	EBA	Percentage (%)
NFCNPL_PAD	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—O: Public administration and defense, compulsory social security	EBA	Percentage (%)
NFCNPL_EDU	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—P: Education	EBA	Percentage (%)
NFCNPL_HUM	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—Q: Human health services and social work activities	EBA	Percentage (%)
NFCNPL_ART	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—R: Arts, entertainment, and recreation	EBA	Percentage (%)
NFCNPL_OTH	Secondary Dependent variable	NPL Economic Sector	Aggregate non-performing loans to total gross loans—Non-financial corporations—S: Other services	EBA	Percentage (%)

Notes: (1) This table presents the data, their explanation, as well as the data sources of the dependent variables used in this research. (2) All variables are depicted at an aggregated country level, whereas before they were employed for empirical testing, they were all transformed to first or second differences because of unit root testing.

Table A3. Data sources and descriptions of the candidate predictors and the control variables.

Variable	Variable Role	Variable Group	Explanation	Source	Parameter Shown
UNEMP	Control variable	Macroeconomic Variables	Percentage (%) of unemployment	DataStream	Percentage (%)
CPI	Control variable	Macroeconomic Variables	Quarterly Consumer Price Index	DataStream	No.
R_GDP_Q2Q	Control variable	Macroeconomic Variables	Quarterly percentage growth rate of real GDP	IMF	Percentage (%)
GDP_MARKET	Control variable	Macroeconomic Variables	Quarterly gross domestic product at market prices	Eurostat	No.

Table A3. *Cont.*

Variable	Variable Role	Variable Group	Explanation	Source	Parameter Shown
NPLS (-1)	Control variable	Bank-specific Variables	Previous quarter aggregate non-performing loans to total gross loans	ECB	Percentage (%)
ROA	Control variable	Bank-specific Variables	Return on assets: profit or loss for the year/total assets	DataStream	Percentage (%)
CAP	Control variable	Bank-specific Variables	Bank capital and reserves to total assets	DataStream	Percentage (%)
LOAN_DISBRS	Control variable	Bank-specific Variables	Loan disbursements to customers	DataStream	Percentage (%)
FINANCIAL_ASSETS	Control variable	Bank-specific Variables	Total financial instruments on the asset side	EBA	No.
PROVISIONS	Control variable	Bank-specific Variables	Impairments (credit risk losses)/equity	EBA	Percentage (%)
RISK_CAPITAL	Control variable	Bank-specific Variables	Total risk exposure amount for position, foreign exchange, and commodities risks/total risk exposure amount	EBA	Percentage (%)
OPER_RISK	Control variable	Bank-specific Variables	Total risk exposure amount for OpePercentage (%) ns/total risk exposure amount	EBA	No.
LIABILITIES	Control variable	Bank-specific Variables	Total deposits other than from banks/total liabilities	EBA	No.
CASH_BALANCES	Control variable	Bank-specific Variables	Cash positions/total assets	EBA	Percentage (%)
FINANCIAL_ASSETS	Control variable	Bank-specific Variables	Total financial instruments on the asset side	EBA	No.
EQUITY	Control variable	Bank-specific Variables	Equity instruments/total assets	EBA	Percentage (%)
TOTAL_ASSETS	Control variable	Bank-specific Variables	Total assets	EBA	No.
RETAINED_EARNINGS	Control variable	Bank-specific Variables	Retained earnings/Tier 1 capital volume	EBA	Percentage (%)
DERIVATIVES	Control variable	Bank-specific Variables	Derivatives/total assets	EBA	Percentage (%)
CRED_DEPOSITS	Control variable	Bank-specific Variables	Deposits from credit institutions/total liabilities	EBA	Percentage (%)
TIER1_CAP	Control variable	Regulatory Variables	Additional Tier 1 capital	EBA	No.
COVER_Percentage (%)	Control variable	Regulatory Variables	Accumulated impairment, accumulated negative changes in fair value due to credit risk for non-performing loans and advances/total gross non-performing loans and advances	EBA	Percentage (%)
RWA_VOLUME	Control variable	Regulatory Variables	RWA volume	EBA	No.
OWN_FUNDS_TIER1	Control variable	Regulatory Variables	Tier 1 capital volume	EBA	No.
SECURITIZATION	Control variable	Regulatory Variables	Securitization positions/risk-weighted exposure amounts for credit, counterparty credit, and dilution risks and free deliveries	EBA	Percentage (%)
PEPP_PURCHASES	Candidate predictor	Quantitative Easing Variables	Net purchases at book value	ECB	No.
ASSET_TO_GDP	Candidate predictor	Quantitative Easing Variables	Total assets/quarterly gross domestic product at market prices	ECB	Percentage (%)

Table A3. *Cont.*

Variable	Variable Role	Variable Group	Explanation	Source	Parameter Shown
QE_ANNOUNCEMENT	Candidate predictor	Quantitative Easing Variables	Quantitative Easing (QE) Announcement: 1 Corresponding to Dates: 18 March 2020 and 4 June 2020.	(Hoang et al. 2021)	Binary (1/0)
EXP_ASSET_PURC	Candidate predictor	Quantitative Easing Variables	Expanded Asset Purchase Program (APP)	ECB	No.
BOND_PURC	Candidate predictor	Quantitative Easing Variables	Covered bonds purchases at book value (CBPP3)	ECB	No.
COVID19_DUMMY	Candidate predictor	COVID-19 Variables	COVID-19 pandemic existence	Author's Calculations	Binary (1/0)
COVID19_VACCINATED	Candidate predictor	COVID-19 Variables	COVID-19 vaccinated population	DataStream	No.
COVID19_DEATHS	Candidate predictor	COVID-19 Variables	COVID-19 deaths	DataStream	No.
CONTNMN	Candidate predictor	COVID-19 Government Response Variables	Government response containment index	DataStream	Index
GOVT_RESP_STR	Candidate predictor	COVID-19 Government Response Variables	Government response stringency index	DataStream	Index
GOVT_ECON_SUP	Candidate predictor	COVID-19 Government Response Variables	Government response economic support index	DataStream	Index

Notes: (1) This table presents the data, their explanation, as well as the data sources of the candidate predictors and the control variables employed. (2) All variables are depicted at the aggregated country level, whereas before they were employed for empirical testing, they were transformed to first or second differences because of unit root testing.

Table A4. Data sources and descriptions of the cultural dimensions.

Literature	Variable Symbol	Cultural Dimensions	Short Definition	ESS (European Social Survey) Question	Values/Answer Range from ESS (European Social Survey)	
	ipcrtiv	Self-direction	Independent thought and action	Important to think new ideas and be creative	Value	Category
	ipgdtim	Stimulation	Excitement, novelty, and challenge in life	Important to have a good time	1	Very much like me
	ipudrst	Hedonism	Pleasure or sensuous gratification for oneself	Important to understand different people	2	Like me
	ipshabt	Achievement	Personal success through demonstrating competence according to social standards	Important to show abilities and be admired	3	Somewhat like me
Schwartz National Culture Values (Schwartz 1994)	ipfrule	Power	Social status, prestige, control, or dominance	Important to do what is told and follow rules	4	A little like me
	ipstrgv	Security	Safety, harmony, and stability of society, of relationships, and of self	Important that government is strong and ensures safety	5	Not like me
	ipbhprp	Conformity	Restraint of actions, inclinations, and impulses likely to upset or harm others and violate social expectations or norms	Important to behave properly	6	Not like me at all
	imptrad	Tradition	Respect, commitment, and acceptance of the customs and ideas that one's culture or religion provides	Important to follow traditions and customs	7	Refusal *

Table A4. *Cont.*

Literature	Variable Symbol	Cultural Dimensions	Short Definition	ESS (European Social Survey) Question	Values/Answer Range from ESS (European Social Survey)	
Schwartz National Culture Values (Schwartz 1994)	ipeqopt	Benevolence	Preserving and enhancing the welfare of those with whom one is in frequent personal contact	Important that people are treated equally and have equal opportunities	8	Don't know *
	impenv	Universalism	Understanding, appreciation, tolerance, and protection for the welfare of all people and for nature	Important to care for nature and environment	9	No answer *
Author's Calculations	CULTURE_PCA	National Cultural Identity Variable	Percentage (%)	Cultural Identity	(*) Missing Value	

Notes: (1.) This table presents the data, their explanation, as well as the data sources of the variables employed. The variables depicted in this table are related to the Schwartz (1994) cultural dimensions, as derived from the European Social Survey (ESS). More specifically, the second column refers to the name of the cultural value, the third column provides a short description of the respective cultural dimension, the fourth column depicts the ESS question, from which the data for each variable were derived, the fifth column represents the name of the variable, as depicted in the ESS survey, and finally, the last column depicts the respective questions represented in the ESS survey for each cultural dimension. (2.) The asterisk * corresponds to missing values in the European Social survey (ESS). (3.) All variables are depicted at the aggregated country level. No unit root testing was implemented for those variables since those variables were not directly used in the empirical estimations. Instead, we proceeded by forming a new cultural dimension variable, by utilizing principal component analysis (PCA) methodology (=CULTURE_PCA). More specifically, the variable presented in the last row of the above table was calculated utilizing the PCA and was not derived from the ESS survey. Instead, this variable was the culmination of the Schwartz (1994) cultural dimensions.

Table A5. Regression results for total sample.

This Table Presents the Empirical Results Related with the Total Sample of Analysis (2015Q1–2021Q4) as Well as the Post-COVID-19 Period (2020Q1–2021Q4)

Regression Results—Total Sample and Post-COVID-19 Period

		Dependent Variable						
		Total Period: 2015Q1–2021Q4						Post-COVID-19 Period: 2020Q1–2021Q4
Total Period Analysis		MODEL (1)	MODEL (2)	MODEL (3)	MODEL (4)	MODEL (5)	MODEL (6)	MODEL (7)
Variable Group	Variable Symbol	D(NPL_RATIO_CRE)	D(NPL_RATIO_HHS)	D(NPL_RATIO_MORT)	D(NPL_RATIO_NFCS)	D(NPL_RATIO_SME)	D(NPLS)	D(NPLS)
Macroeconomic Variables	D(UNEMP(-1))	0.001179	0.000583	0.000565	0.000917	0.000928	0.002346 ***	0.006601
	D(CPI(-1))						−0.000506	−0.001593
	R_GDP_Q2Q(-1)	−0.000444	−0.000288	−0.000272	−0.000368	−0.000450	−0.000135	0.002276
	D(NPLS(-1))						0.690926 ***	−0.674424 ***
	D(NPL_RATIO_CRE(-1))	0.798120						
	D(NPL_RATIO_HHS(-1))		0.832194					
	D(NPL_RATIO_MORT(-1))			0.830631				
	D(NPL_RATIO_NFCS(-1))				0.791745			
	D(NPL_RATIO_SME(-1))					0.812455		
Bank-specific Variables	D(OPER_RISK,1)							
	RISK_CAPITAL						−0.000573	2.536535 *
	D(SECURITIZATION,1)						−0.000573	2.771233 *
	D(TIER1_CAP,2)							−2.804435 *
	D(RWA_VOLUME,1)							−1.855125 **
	D(TOTAL_ASSETS,1)	−0.000360	−0.000377	−0.000418	−0.000410	−0.000445	−0.000573	0.009934 *
	D(ROA(-1))						0.111898	−1.212907
	D(CAP(-1))						0.251789 ***	0.225163
	D(CAP,2)	0.001959	0.001511	0.001626	0.001581	0.001826	0.000688	
	D(LOAN_DISBRS(-1))						−0.008183	0.002465
Quantitative Easing Variables	PEPP_PURCHASES	-4.59×10^{-8}	-3.07×10^{-7}	-3.68×10^{-7}	-1.34×10^{-7}	-2.61×10^{-7}	-8.67×10^{-7} **	-7.45×10^{-2} *
	BOND_PURC	-5.00×10^{-5}	-2.77×10^{-5}	-2.52×10^{-5}	-4.01×10^{-5}	-4.57×10^{-5}	1.28×10^{-6}	
	EXP_ASSET_PURC	−0.001098	−0.000698	−0.000644	−0.000911	−0.001090	-8.64×10^{-5}	

Table A5. *Cont.*

This Table Presents the Empirical Results Related with the Total Sample of Analysis (2015Q1–2021Q4) as Well as the Post-COVID-19 Period (2020Q1–2021Q4)

Regression Results—Total Sample and Post-COVID-19 Period

				Dependent Variable				
			Total Period: 2015Q1–2021Q4					Post-COVID-19 Period: 2020Q1–2021Q4
Total Period Analysis		MODEL (1)	MODEL (2)	MODEL (3)	MODEL (4)	MODEL (5)	MODEL (6)	MODEL (7)
Variable Group	Variable Symbol	D(NPL_RATIO_CRE)	D(NPL_RATIO_HHS)	D(NPL_RATIO_MORT)	D(NPL_RATIO_NFCS)	D(NPL_RATIO_SME)	D(NPLS)	D(NPLS)
COVID-19 Variables	COVID19_DEATHS							−4.632856 *
	COVID19_DUMMY	−0.012205	−0.007001	−0.011405	−0.009049	−0.007541	0.041887	
Regression Main Statistics	R-squared	0.836537	0.841355	0.832740	0.823983	0.833629	0.987545	0.935943
	Adjusted R-squared	0.698222	0.707117	0.691212	0.675045	0.692853	0.977007	0.829180
	F-statistic	6.048053	6.267637	5.883921	5.532407	5.921688	9.370878	8.766596
	Prob(F-statistic)	0.001587	0.001335	0.001811	0.002426	0.001757	0.000000	0.000000
	Durbin–Watson stat	1.969687	1.972348	1.971576	1.970468	1.974364	1.930466	2.855695

Note(s): (1.) Table A5 presents the regression results related to both the total and post-COVID-19 periods. More specifically, Models 1 to 6 present the empirical results referring to the total period (2015Q1–2021Q4), while Model 7 presents the empirical results referring to the post-COVID-19 period (2020Q1–2021Q4). (2.) OLS methodology was employed for the regression model estimation. More specifically, Fixed Corrected Panel Effects estimations with country-fixed effects were utilized for all models because of the Hausman test. The table presents the values of the coefficients, while the significance of the p-value is presented with an asterisk: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. (3.) The (−1) denotes one period lag. This note also applies to the subsequent tables. (4.) The variable NPLs stands for the aggregate non-performing loans to total gross loans; NPL_RATIO_CRE represents the commercial real estate NPLs to total gross loans (aggregate); NPL_RATIO_HHS stands for household NPLs to total gross loans (aggregate); NPL_RATIO_MORT stands for mortgage NPLs to total gross loans (aggregate); NPL_RATIO_NFCS represents the non-financial corporations' NPLs to total gross loans (aggregate); NPL_RATIO_SME represents the small and medium-sized enterprises' NPLs to total gross loans (aggregate); CPI stands for quarterly consumer price index; R_GDP_Q2Q represents the quarterly percentage growth rate of real GDP; OPER_RISK stands for total risk exposure amount for operations/total risk exposure amount; RISK_CAPITAL denotes the total risk exposure amount for position, foreign exchange, and commodities risks/total risk exposure amount; SECURITIZATION represents the securitization positions/risk-weighted exposure amounts for credit, counterparty credit, and dilution risks and free deliveries; TIER1_CAP denotes the additional Tier 1 capital; RWA_VOLUME stands for RWA volume; TOTAL_ASSETS represents the total assets; ROA denotes the return on assets: profit or loss for the year/total assets; CAP represents the bank capital and reserves to total assets; LOAN_DISBRS stands for loan disbursements to customers; PEPP_PURCHASES represents the net purchases at book value; BOND_PURC stands for covered bonds purchases at book value (CBPP3); EXP_ASSET_PURC represents the Expanded Asset Purchase Program (APP); COVID19_DEATHS denotes the COVID-19 deaths; finally, COVID19_DUMMY stands for COVID-19 pandemic existence. (5.) Even though sample sizes are not included, the main statistics of the regression estimates imply that our empirical models demonstrated strong explanatory power. The high R-squared value and significant F-statistic reinforce the validity and reliability of the results. This note also applies to the subsequent tables. (6.) We opted not to include t-statistics since the inclusion of coefficient estimates and p-values effectively communicate the statistical significance of our results. This note also applies to the subsequent tables.

Table A6. Regression results for European subregions.

PANEL A. This Table Presents the Empirical Results for the European Subregions. The Period of Analysis Is the Pre-COVID-19 Period (2015Q1–2019Q4)

PANEL A. Regression Results—European Subregions—Pre-COVID-19 Period

		Dependent Variable					
		Pre-COVID-19: 2015Q1-2019Q4					
Subregional Analysis		Central Europe	Northern Europe	Southern Europe	Central Europe	Northern Europe	Southern Europe
Variable Group	Variable Symbol	MODEL (1)	MODEL (2)	MODEL (3)	MODEL (4)	MODEL (5)	MODEL (6)
		D(NPLS)	D(NPLS)	D(NPLS)	D(NPLS)	D(NPLS)	D(NPLS)
Macroeconomic Variables	D(UNEMP(-1))	-0.013766	0.006658	0.025692 *	-0.002803	0.007229	0.022256 **
	D(CPI(-1))	-0.005350	0.055737 *	-0.012663	-0.007320	0.014939	-0.012663
	R_GDP_Q2Q(-1)	0.033917	0.038799 ***	0.148603	-0.012171	0.022272	0.039282
	D(GDP_MARKET,1)	-1.76×10^{-5}	1.40×10^{-5}	2.95×10^{-6}			
Bank-specific Variables	D(NPLS(-1))	0.269035	-0.285082 **	-0.391149 **	0.269035	-0.158322	-0.427347
	D(ROA(-1))	0.065223	0.034237	-2.829566 **	0.065223	0.034237	0.773295
	D(CAP(-1))	-0.582087	-0.076589	-1.438326	-0.135500	-0.293470	-0.355399
	D(LOAN_DISBRS(-1))	-0.006321	0.001147	-0.009063	-0.038671 ***	0.016848	0.007029
	D(PROVISIONS,1)	-3.839253	-3.145415	-7.427375			
	RISK_CAPITAL	-2.266459	-7.837162	-4.265118			
	D(OPER_RISK,1)	0.432963	-4.806479	-4.247923			
	D(LIABILITIES,1)	-0.018438	0.001512	0.028989			
	D(CASH_BALANCES,1)	-4.410540	-3.112317	-5.878817			
	D(FINANCIAL_ASSETS,1)	0.014633	-0.001524	-0.032837			
	D(EQUITY)	-4.167760	-2.249403	2.292337 **			
	D(RETAINED_EARNINGS,1)	-0.645146	-0.529536	-1.352197 **			
	D(DERIVATIVES2)	-7.776273	-9.068875	6.795368 ***			
	D(CRED_DEPOSITS,1)	-2.575671	-8.317736	9.03572 *			
Regulatory Variables	D(TIER1_CAP2)	0.041501	-0.302660	0.080518			
	D(COVER_RATIO2)	0.661557	-2.656202	1.833429 **			
	D(RWA_VOLUME,1)	-0.000586	-4.85×10^{-5}	-0.000966			
	D(OWN_FUNDS_TIER1(-1),1)	0.008595	-0.000198	-0.00130			
	D(SECURITIZATION,1)	-7.580479	-8.324792	-5.782130			
Regression Main Statistics	R-squared	0.900685	0.669831	0.942711	0.351186	0.588546	0.835070
	Adjusted R-squared	-1.681505	0.273628	0.721740	-0.197811	0.358132	0.695514
	F-statistic	0.348807	1.690627	4.266221	0.639686	2.554299	5.983770
	Prob(F-statistic)	0.897632	0.091925	0.026995	0.767729	0.019794	0.001671
	Durbin-Watson stat	2.449610	1.750368	2.499906	2.220722	1.649272	1.861230

PANEL B. This Table Presents the Empirical Results for the European Subregions. The Period of Analysis Is the Post-COVID-19 Period (2020Q1–2021Q4).

PANEL B. Regression Results—European Subregions—Post-COVID-19 Period

		Dependent Variable		
		Post-COVID-19: 2020Q1-2021Q4		
Subregional Analysis		Central Europe	Northern Europe	Southern Europe
Variable Group	Variable Symbol	MODEL (1)	MODEL (2)	MODEL (3)
		D(NPLS)	D(NPLS)	D(NPLS)
Macroeconomic Variables	D(UNEMP(-1))	0.000194	0.001718	-0.095467
	D(CPI(-1))	0.001380	0.000358	0.002404
	R_GDP_Q2Q(-1)	0.001591	0.000165	0.001966
Bank-specific Variables	D(NPLS(-1))	0.114139	0.078584	-0.914587 ***
	D(ROA(-1))	0.178231	-0.128961	-0.724560
	D(CAP(-1))	0.006007	0.289453 ***	0.484247
	D(LOAN_DISBRS(-1))		0.009299	0.006653
COVID-19 Variables	COVID19_VACCINATED	-0.007391 ***	0.002064	-0.143976 **
Regression Main Statistics	R-squared	0.445240	0.396354	0.871000
	Adjusted R-squared	-0.035552	0.010021	0.731250
	F-statistic	0.926056	1.025940	6.232562
	Prob(F-statistic)	0.550839	0.464432	0.001604
	Durbin-Watson stat	2.642044	2.190621	1.409280

Note(s): (1.) Table A6 presents the regression results related to the central, as well as north European subregions. More specifically, PANEL A presents the empirical results referring to the pre-COVID-19 period, while PANEL B presents the empirical results referring to the post-COVID-19 period. (2.) OLS methodology was employed for the regression model estimation. More specifically, Fixed Corrected Panel Effects estimations with country-fixed effects were utilized for all models because of the Hausman test. The table presents the values of the coefficients, while the significance of the p-value is presented with an asterisk: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. (3.) The variable NPLs stands for the aggregate non-performing loans to total gross loans; UNEMP represents the % of unemployment; CPI represents the quarterly consumer price index; R_GDP_Q2Q stands for quarterly percentage growth rate of real GDP; GDP_MARKET stands for quarterly gross domestic product at market prices; ROA represents the return on assets: profit or loss for the year/total assets; CAP represents the bank capital and reserves to total assets; LOAN_DISBRS stands for loan disbursements to customers; PROVISIONS stands for impairments (credit risk losses)/equity; RISK_CAPITAL represents the total risk exposure amount for position, foreign exchange, and commodities risks/total risk exposure amount; OPER_RISK stands for total risk exposure amount for operations/total risk exposure amount; LIABILITIES denotes the total deposits other than from banks/total liabilities; CASH_BALANCES represents the cash positions/total assets; FINANCIAL_ASSETS denotes the total financial instruments on the asset side; EQUITY stands for equity instruments/total assets; RETAINED_EARNINGS represents the retained earnings/Tier 1 capital volume; DERIVATIVES denotes the derivatives/total assets; CRED_DEPOSITS represents the deposits from credit institutions/total liabilities; TIER1_CAP stands for additional Tier 1 capital; COVER_RATIO represents the accumulated impairment, accumulated negative changes in fair value due to credit risk for non-performing loans and advances/total gross non-performing loans and advances; RWA_VOLUME stands for RWA volume; OWN_FUNDS_TIER1 represents the Tier 1 capital volume; SECURITIZATION denotes the securitization positions/risk-weighted exposure amounts for credit, counterparty credit, and dilution risks and free deliveries; finally, COVID19_VACCINATED stands for COVID-19 vaccinated population.

Table A7. Regression results for European prosperity.

PANEL A. This table presents the empirical results for the Prosperity dimension. The period of analysis is the Pre-COVID-19 Period (2015Q1–2019Q4).

PANEL A Regression
Results—Prosperity—Pre-COVID-19 Period

Dependent Variable

Core—Periphery

Variable Group	Variable Symbol	Hard-Core Country Group		Intermediate Country Group		Extended Periphery Country Group	
		MODEL (1)	MODEL (2)	MODEL (3)	MODEL (4)	MODEL (5)	MODEL (6)
		D(NPLS)	D(NPLS)	D(NPLS)	D(NPLS)	D(NPLS)	D(NPLS)
Macroeconomic Variables	D(UNEMP(-1))	0.011557	-0.009357 ***	0.024228	-0.015295	-0.005513	-0.023370
	D(CPI(-1))	0.009861	0.001172	0.027906	0.013778	0.007535	0.000790
	R_GDP_Q2Q(-1)	-0.000169	3.81×10^{-5}	0.054228	0.007601	0.016925	0.038665
	D(GDP_MARKET,1)	-1.85×10^{-7}		1.74×10^{-6}		2.32×10^{-5}	
Bank-specific Variables	D(NPLS(-1))	0.440823	-0.561929 **	-0.142991 **	-0.288937 **	0.061585	-0.596344 ***
	D(ROA(-1))	-0.150535	-0.171537 **	0.155848	-0.051783	0.257572	-2.461562
	D(CAP(-1))	0.009861	0.150353	0.402339	-0.150081	-1.119386	-1.690086 **
	D(LOAN_DISBRS(-1))	-0.001410	-0.008798	-0.009894	0.000947	0.108287	0.039455
	D(PROVISIONS,1)	-2.167336		-3.487687		-3.303196	
	RISK_CAPITAL	-0.540978		-1.937135		-2.298918	
	D(OPER_RISK,1)	-0.778553		-4.675385		-1.184725	
	D(LIABILITIES,1)	-0.001180		-2.034571		-0.089222	
	D(CASH_BALANCES,1)	-6.340108 **		2.14×10^{-5}		-4.961550	
	D(FINANCIAL_ASSETS,1)	0.00218		-4.47762		0.064814	
	D(EQUITY)	-8.068984		3.96×10^{-5}		-1.091987	
	D(RETAINED_EARNINGS,1)	0.042142		-4.959733		0.770446	
	D(DERIVATIVES,2)	-3.186733		-0.158886		-3.608660	
	D(CRED_DEPOSITS,1)	-1.481367		-1.114752		-7.557611	
				-0.308019			
Regulatory Variables	D(TIER1_CAP,2)	0.112170		-0.226567		-0.454617	
	D(COVER_RATIO,2)	-2.815370		-2.034571		0.091373	
	D(RWA_VOLUME,1)	0.000232		0.000988		0.001412	
	D(OWN_FUNDS_TIER1(-1),1)	-0.000182		-0.001128		-0.005464	
	D(SECURITIZATION,1)	-6.982720		-2.651291		-4.707613	
Regression Main Statistics	R-squared	0.816686	0.618123	0.549040	0.638041	0.963809	0.759030
	Adjusted R-squared	0.109617	0.294996	0.007888	0.450821	0.022846	0.491285
	F-statistic	1.155030	1.912941	1.014577	3.407970	1.024279	2.834898
	Prob(F-statistic)	0.454674	0.132621	0.489720	0.002260	0.667778	0.066135
	Durbin-Watson stat	2.169879	1.812284	1.501292	2.290640	1.313388	1.706952

PANEL B. This Table Presents the Empirical Results for the Prosperity Dimension. The Period of Analysis Is the Post-COVID-19 Period (2020Q1–2021Q4).

PANEL B Regression
Results—Prosperity—Post-COVID-19 Period

Dependent Variable
Post-COVID-19: 2020Q1-2021Q4

Core—Periphery

Variable Group	Variable Symbol	Hard-Core Country Group	Intermediate Country Group	Extended Periphery Country Group
		MODEL (1)	MODEL (2)	MODEL (3)
		D(NPLS)	D(NPLS)	D(NPLS)
Macroeconomic Variables	D(UNEMP(-1))	-0.003286	-0.011902	-0.079501
	D(CPI(-1))	0.000860	0.001357	0.004810
	R_GDP_Q2Q(-1)	-0.000730	-0.000103	-0.010835
Bank-specific Variables	D(NPLS(-1))	-0.140869	-0.004967	-0.749463
	D(ROA(-1))	0.142126 **	0.090678	-0.487957
	D(CAP(-1))	-0.144226	0.090874	-1.531234
	D(LOAN_DISBRS(-1))	-0.01496	0.017691	-0.109999
Regulatory Variables	D(TIER1_CAP,2)	-4.418525	-3.864481 **	-8.000965
	D(COVER_RATIO,2)	-2.616419	0.708464	-3.686813
COVID-19 Variables	COVID19_VACCINATED	-0.09943	0.002023	-0.143110 **
Regression Main Statistics	R-squared	0.528754	0.382402	0.906250
	Adjusted R-squared	-0.028174	0.064245	0.765624
	F-statistic	0.949412	1.201930	6.44429
	Prob(F-statistic)	0.541448	0.315479	0.006610
	Durbin-Watson stat	2.335049	1.984339	1.199686

Note(s): (1). Table A7 presents the regression results related to the core, as well as the peripheral countries. More specifically, PANEL A presents the empirical results referring to the pre-COVID-19 period, while PANEL B presents the empirical results referring to the post-COVID-19 period. (2.) OLS methodology was employed for the regression model estimation. More specifically, Fixed Corrected Panel Effects estimations with country-fixed effects were utilized for all models because of the Hausman test. The table presents the values of the coefficients, while the significance of the p-value is presented with an asterisk: *** $p < 0.01$, ** $p < 0.05$. (3.) The variable NPLs represents the aggregate non-performing loans to total gross loans; UNEMP represents the % of unemployment; CPI stands for quarterly consumer price index; R_GDP_Q2Q stands for quarterly percentage growth rate of real GDP; GDP_MARKET represents the quarterly gross domestic product at market prices; ROA represents the return on assets: profit or loss for the year/total assets; CAP stands for bank capital and reserves to total assets; LOAN_DISBRS stands for loan disbursements to customers; PROVISIONS represents the impairments (credit risk losses)/equity; RISK_CAPITAL stands for total risk exposure amount for position, foreign exchange, and commodities risks/total risk exposure amount; OPER_RISK denotes the total risk exposure amount for operations/total risk exposure amount; LIABILITIES represents the total deposits other than from banks/total liabilities; CASH_BALANCES denotes the cash positions/total assets; FINANCIAL_ASSETS stands for total financial instruments on the asset side; EQUITY represents the equity instruments/total assets; RETAINED_EARNINGS denotes the retained earnings/Tier 1 capital volume; DERIVATIVES represents the derivatives/total assets; CRED_DEPOSITS stands for deposits from credit institutions/total liabilities; TIER1_CAP represents the additional Tier 1 capital; TIER1 denotes the Tier 1 capital volume; OWN_FUNDS_TIER1 represents the RWA volume; SECURITIZATION stands for securitization positions/risk-weighted exposure amounts for credit, counterparty credit, and dilution risks and free deliveries; finally, COVID19_VACCINATED stands for COVID-19 vaccinated population.

Table A8. Regression results for NPL type.

PANEL A. This Table Presents the Empirical Results per NPL Type. The Period of Analysis Is the Pre-COVID-19 Period (2015Q1–2019Q4).

PANEL A Regression Results—NPL Type—Pre-COVID-19 Period

Variable Group	Variable Symbol	MODEL (1) D(NPLS)	MODEL (2) D(NPL_RATIO_CRE)	MODEL (3) D(NPL_RATIO_HHS)	MODEL (4) D(NPL_RATIO_MORT)	MODEL (5) D(NPL_RATIO_NFCS)	MODEL (6) D(NPL_RATIO_SME)
				Dependent Variable			
Macroeconomic Variables	D(UNEMP(-1))	-0.01412 ***	0.004447 *	0.004028 *	0.003909 *	0.003590 *	0.004816 *
	D(CPI(-1))	-0.00439	-0.006894 *	-0.003656 *	-0.003369 *	-0.005314 *	-0.006599 *
	R_GDP_Q2Q(-1)	0.021946	0.001510	0.001451	0.001427	0.001649	0.002270
Bank-specific Variables	D(NPL_RATIO_CRE(-1))		-0.160619 ***				
	D(NPL_RATIO_HHS(-1))			-0.240182 *			
	D(NPL_RATIO_MORT(-1))				-0.243654 *		
	D(NPL_RATIO_NFCS(-1))					-0.167326 **	
	D(NPL_RATIO_SME(-1))						-0.185631
	D(ROA(-1))	-0.033750	-0.011243	0.000901	0.001561	-0.005135	-0.004125
	D(CAP(-1))	-0.388216 *	-0.038143 **	-0.016790 *	-0.014518 *	-0.026650 **	-0.029528 ***
	D(SECURITIZATION,1)		5.699070*	4.843643 *	4.933218	4.301169 *	5.470455 *
	D(LOAN_DISBRS(-1))	0.018008 **	0.000427	-0.000102	-0.000174	0.000174	-1.13×10^{-5}
Regression Main Statistics	R-squared	0.705563	0.645273	0.688260	0.686605	0.677843	0.720262
	Adjusted R-squared	0.596847	0.514297	0.573155	0.570890	0.558893	0.582164
	F-statistic	6.489998	4.926644	5.979449	5.933580	5.698537	5.215574
	Prob(F-statistic)	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
	Durbin-Watson stat	1.885753	2.582328	2.393623	2.391076	2.533094	2.639241

PANEL B. This Table Presents the Empirical Results referring to the Post-COVID-19 Period (2020Q1–2021Q4).

PANEL B Regression Results—NPL Type—Post-COVID-19 Period

Variable Group	Variable Symbol	MODEL (1) D(NPLS)	MODEL (2) D(NPL_RATIO_CRE)	MODEL (3) D(NPL_RATIO_HHS)	MODEL (4) D(NPL_RATIO_MORT)	MODEL (5) D(NPL_RATIO_NFCS)	MODEL (6) D(NPL_RATIO_SME)
				Dependent Variable			
Macroeconomic Variables	D(UNEMP(-1))	-0.045555 *	-0.000300	0.000116	0.000156	-0.000161	-9.26×10^{-5}
	D(CPI(-1))	-0.01392	-3.02×10^{-5}	5.55×10^{-7}	-5.91×10^{-6}	-1.70×10^{-5}	-5.81×10^{-6}
	R_GDP_Q2Q(-1)	-0.000562	8.21×10^{-7}	-2.30×10^{-5}	-1.58×10^{-5}	-1.11×10^{-5}	-9.82×10^{-6}
Bank-specific Variables	D(NPLS(-1))	-1011580 *					
	D(NPL_RATIO_CRE(-1))		-0.674931 *				
	D(NPL_RATIO_HHS(-1))			-1.739795 *			
	D(NPL_RATIO_MORT(-1))				-1.181618 *		
	D(NPL_RATIO_NFCS(-1))					-0.656276*	
	D(NPL_RATIO_SME(-1))						-0.767985 *
	D(ROA(-1))	-0.240106	-0.004929	-0.000432 ***	-0.008763 **	-0.004389	-0.005456
	D(CAP(-1))	0.465061 ***	-3.02×10^{-5}	0.001676	0.002876	0.004136 ***	0.004710
	D(LOAN_DISBRS(-1))	-0.020088	1.34×10^{-5}	5.55×10^{-5}	6.69×10^{-5}	5.17×10^{-5}	4.96×10^{-5}
	D(FINANCIAL_ASSETS,1)	-0.000826					
Quantitative Easing Variables	D(ASSET_TO_GDP)	-2.694894 **					
	D(QE_ANNOUNCEMENT)	-0.000822					
COVID-19 Variables	COVID19_VACCINATED		-0.000363 ***	-0.000272	-0.000277	-0.000234 ***	-0.000307 ***
COVID-19 Government Response Variables	GOVT_RESP_STR	3.76×10^{-7}	3.46×10^{-6} *	3.76×10^{-7}	8.09×10^{-8}	2.40×10^{-6} **	1.69×10^{-6} **
Regression Main Statistics	R-squared	0.748355	0.470007	0.557662	0.506621	0.631051	0.586299
	Adjusted R-squared	0.644502	0.288693	0.406336	0.337834	0.504832	0.444770
	F-statistic	7.205881	2.592234	3.685163	3.001533	4.999635	4.142598
	Prob(F-statistic)	0.000000	0.000701	0.000005	0.000107	0.000000	0.000001
	Durbin-Watson stat	2.097308	1.990526	0.867767	0.878659	2.153291	1.652758

Note(s): (1.) PANEL A presents the empirical results referring to the pre-COVID-19 period, while PANEL B presents the empirical results referring to the post-COVID-19 period. (2.) OLS methodology was employed for the regression model estimation. More specifically, Fixed Corrected Panel Effects estimations with country-fixed effects were utilized for all models because of the Hausman test. The table presents the values of the coefficients, while the significance of the *p*-value is presented with an asterisk: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. (3.) The variable NPLs represents the aggregate non-performing loans to total gross loans; UNEMP represents the % of unemployment; CPI stands for quarterly consumer price index; R_GDP_Q2Q stands for quarterly percentage growth rate of real GDP; NPL_RATIO_CRE represents the commercial real estate NPLs to total gross loans (aggregate); NPL_RATIO_HHS represents the household NPLs to total gross loans (aggregate); NPL_RATIO_MORT stands for mortgage NPLs to total gross loans (aggregate); NPL_RATIO_NFCS stands for non-financial corporations' NPLs to total gross loans (aggregate); NPL_RATIO_SME represents the small and medium-sized enterprises' NPLs to total gross loans (aggregate); ROA stands for return on assets: profit or loss for the year/total assets; CAP denotes the bank capital and reserves to total assets; SECURITIZATION represents the securitization positions/risk-weighted exposure amounts for credit, counterparty credit, and dilution risks and free deliveries; LOAN_DISBRS denotes the loan disbursements to customers; FINANCIAL_ASSETS stands for total financial instruments on the asset side; ASSET_TO_GDP represents the total assets/quarterly gross domestic product at market prices; QE_ANNOUNCEMENT denotes the Quantitative Easing (QE) Announcement: 1 Corresponding to dates: 18 March 2020 and 4 June 2020; COVID19_VACCINATED represents the COVID-19 vaccinated population; finally, the variable GOVT_RESP_STR stands for government response stringency index.

Table A9. Regression results for NPL sector.

PANEL A. This Table Presents the Empirical NPL Sectoral Results. The Period of Analysis Is the Post-COVID-19 Period (2020Q1–2021Q4). Empirical MODELS 1 to 10 Are Presented for Economy of Space.

PANEL_A Regression Results—NPL Sector—Post-COVID-19 Period

Dependent Variable — Post-COVID-19: 2020Q1–2021Q4

NPL Type Analysis		MODEL (1)	MODEL (2)	MODEL (3)	MODEL (4)	MODEL (5)	MODEL (6)	MODEL (7)	MODEL (8)	MODEL (9)	MODEL (10)
Variable Group	Variable Symbol	D/NFCNPL_AGR	D/NFCNPL_ART	D/NFCNPL_CON	D/NFCNPL_EDU	D/NFCNPL_ELE	D/NFCNPL_FIN	D/NFCNPL_HUM	D/NFCNPL_INF	D/NFCNPL_MAN	D/NFCNPL_MIN
Macroeconomic Variables	D(UNEMP(-1))	0.000567	−0.000175	−0.000175	−0.000147	-9.92×10^{-5}	0.000117	−0.000156	−0.000450 ***	−0.000145	0.000461
	D(CPI(-1))	−0.000154 **	-2.09×10^{-5}	-2.09×10^{-5}	-2.96×10^{-5}	1.69×10^{-5}	−0.000107	3.71×10^{-5}	-1.78×10^{-5}	-2.09×10^{-5}	-3.12×10^{-6}
	R_GDP_Q2Q(-1)	7.71×10^{-5}	-5.36×10^{-5}	-5.36×10^{-5}	-1.49×10^{-5}	-5.49×10^{-5}***	−0.000127	-8.60×10^{-5}***	3.21×10^{-5}	1.80×10^{-6}	-1.58×10^{-5}
Bank-specific Variables	D(NFCNPL_AGR(-1))	−0.507225 *									
	D(NFCNPL_ART(-1))		−0.336184 *								
	D(NFCNPL_CON(-1))			−0.250137							
	D(NFCNPL_EDU(-1))				−0.397752 *						
	D(NFCNPL_ELE(-1))					−0.347851 *					
	D(NFCNPL_FIN(-1))						−0.376004 *				
	D(NFCNPL_HUM(-1))							−0.422962 *			
	D(NFCNPL_INF(-1))								−0.464776 *		
	D(NFCNPL_MAN(-1))									−0.762066 *	
	D(NFCNPL_MIN(-1))										−0.168169
	D(ROA(-1))	−0.007102	−0.005355	−0.005535	−0.008331 ***	0.002382	−0.013318	−0.005936	−0.002028	−0.003229	−0.015687 ***
	D(CAP(-1))	0.010036	0.005510	0.005510	0.010548 **	3.86×10^{-5}	0.025841	0.005728	0.000372	0.005174 ***	−0.011644
	D(LOAN_DSBRS(-1))	0.000305	3.13×10^{-6}	3.13×10^{-6}	3.56×10^{-5}	-5.18×10^{-5}	−0.000143	0.000238	-9.97×10^{-5}	0.00027	0.000979 ***
COVID-19 Variables	COVID19_VACCINATED	−0.000394	−0.000201	−0.000201	-2.99×10^{-5}	4.71×10^{-5}	0.000528	0.000133	−0.000217	−0.000294 ***	0.000319
COVID-19 Government Response Variables	GOVT_RESP_STR	7.92×10^{-6} *	1.74×10^{-6}	1.74×10^{-6}	2.68×10^{-6} **	5.89×10^{-7}	3.19×10^{-6}	-2.13×10^{-6} ***	1.71×10^{-6} ***	2.10×10^{-6} **	-1.11×10^{-6}
Regression Main Statistics	R-squared	0.601971	0.305408	0.327182	0.382561	0.240557	0.229595	0.501099	0.711797	0.551537	0.325625
	Adjusted R-squared	0.471172	0.067785	0.097007	0.171332	−0.019252	−0.033965	0.330422	0.613201	0.398116	0.094917
	F-statistic	4.495360	1.285261	1.421452	1.81117	0.925899	0.871129	2.935950	7.219349	3.594917	1.411419
	Prob(F-statistic)	0.000000	0.199114	0.120823	0.024287	0.573029	0.644102	0.000144	0.000000	0.000000	0.125524
	Durbin-Watson stat	2.302737	2.230066	1.802728	2.387571	2.518205	2.556315	2.225116	2.508357	2.086867	1.992636

PANEL_B Regression Results—NPL Sector—Post-COVID-19 Period

PANEL B. This table presents the empirical NPL sectoral results. The period of analysis is the Post-COVID-19 Period (2020Q1–2021Q4). Empirical MODELS 11 to 19 are presented for economy of space.

Dependent Variable — Post-COVID-19: 2020Q1–2021Q4

NPL Type Analysis		MODEL (11)	MODEL (12)	MODEL (13)	MODEL (14)	MODEL (15)	MODEL (16)	MODEL (17)	MODEL (18)	MODEL (19)
Variable Group	Variable Symbol	D/NFCNPL_OTH	D/NFCNPL_PAD	D/NFCNPL_REA	D/NFCNPL_PRF	D/NFCNPL_WRT	D/NFCNPL_TRA	D/NFCNPL_WAT	D/NFCNPL_ACC	D/NFCNPL_ADM
Macroeconomic Variables	D(UNEMP(-1))	−0.000202	−0.000291	-9.75×10^{-5}	-5.82×10^{-5}	-6.88×10^{-5}	0.000174	1.26×10^{-5}	−0.001094 ***	2.49×10^{-5}
	D(CPI(-1))	-1.35×10^{-5}	−0.000144	1.06×10^{-6}	-4.38×10^{-5}	9.97×10^{-6}	2.83×10^{-5}	-1.40×10^{-5}	−0.000103 **	-5.34×10^{-5}
	R_GDP_Q2Q(-1)	8.21×10^{-5}	0.00279	3.26×10^{-5}	9.65×10^{-7}	-5.91×10^{-5}	-4.35×10^{-5}	-3.51×10^{-5}	8.33×10^{-5}	4.08×10^{-6}
Bank-specific Variables	D(NFCNPL_OTH(-1))	−0.627122 *								
	D(NFCNPL_PAD(-1))		−0.320885 *							
	D(NFCNPL_REA(-1))			−0.051564						
	D(NFCNPL_PRF(-1))				−0.423346 *					
	D(NFCNPL_WRT(-1))					−0.758765 *				
	D(NFCNPL_TRA(-1))						−0.383743 *			
	D(NFCNPL_WAT(-1))							−0.103849		
	D(NFCNPL_ACC(-1))								−0.339492 *	
	D(NFCNPL_ADM(-1))									−0.196234
	D(ROA(-1))	−0.004911	−0.019703	−0.003910	−0.003256	−0.02876	0.000996	−0.004275	0.006024	−0.006901
	D(CAP(-1))	−0.000148	0.020447	0.002607	0.00890	0.03934	0.02277	0.005790 **	0.002347	0.003552
	D(LOAN_DSBRS(-1))	−0.000509	4.50×10^{-5}	-1.44×10^{-5}	−0.000121	1.91×10^{-5}	0.000205	−0.000152	−0.000174	−0.000172

Table A9. *Cont.*

PANEL B. This table presents the empirical NPL sectoral results. The period of analysis is the Post-COVID-19 Period (2020Q1–2021Q4). Empirical MODELS 11 to 19 are presented for economy of space.

PANEL B Regression Results—NPL Sector—Post-COVID-19 Period					Dependent Variable				
					Post-COVID-19: 2020Q1–2021Q4				
NPL Type Analysis	MODEL (11)	MODEL (12)	MODEL (13)	MODEL (14)	MODEL (15)	MODEL (16)	MODEL (17)	MODEL (18)	MODEL (19)
Variable Symbol	D/NFCNPL_OTH	D/NFCNPL_PAD	D/NFCNPL_REA	D/NFCNPL_PRF	D/NFCNPL_WRT	D/NFCNPL_TRA	D/NFCNPL_WAT	D/NFCNPL_ACC	D/NFCNPL_ADM
COVID-19 Variables COVID19_VACCINATED	0.000158	-3.92×10^{-5}	−0.000147	−0.000212	-0.000575 *	-0.000246 ***	−0.000196	0.000230	-0.000998 *
COVID-19 Government Response Variables GOVT_RESP_STR	-1.07×10^{-6}	4.84×10^{-6}	1.18×10^{-6}	3.76×10^{-6} **	1.62×10^{-6}	6.89×10^{-7}	7.68×10^{-7}	6.87×10^{-6} *	2.79×10^{-6} ***
Regression Main Statistics R-squared	0.623453	0.405312	0.551938	0.409734	0.567890	0.437779	0.440247	0.451692	0.352797
Adjusted R-squared	0.494635	0.201866	0.398654	0.207801	0.420063	0.245440	0.195069	0.264113	0.131385
F-statistic	4.839774	1.992235	3.600751	2.029057	3.841585	2.276082	1.950726	2.408007	1.593397
Prob(F-statistic)	0.000000	0.010911	0.000007	0.009249	0.000003	0.003008	0.013133	0.001640	0.061077
Durbin–Watson stat	2.091972	2.941706	2.348241	2.452917	1.952255	2.298212	1.767624	2.202710	1.823173

Note(s): (1.) PANEL B presents the NPL sectoral empirical estimation results, referring to the post-COVID-19pPeriod. (2.) OLS methodology was employed for the regression model estimation. More specifically, Fixed Corrected Panel Effects estimations with country-fixed effects were utilized for all models because of the Hausman test. The table presents the values of the coefficients, while the significance of the *p*-value is presented with an asterisk: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. (3.) The variable NFCNPL_OTH represents the aggregate NPLs—non-financial corporations—S: Other services; NFCNPL_PAD represents the aggregate NPLs—non-financial corporations—O: Public administration and defense, compulsory social security; NFCNPL_REA stands for aggregate NPLs—non-financial corporations—L: Real estate activities; NFCNPL_PRF stands for aggregate NPLs—non-financial corporations—M: Professional, scientific, and technical activities; NFCNPL_WRT represents the aggregate NPLs—non-financial corporations—G: Wholesale and retail trade; NFCNPL_TRA represents the aggregate NPLs—non-financial corporations—H: Transport and storage; NFCNPL_WAT stands for aggregate NPLs—non-financial corporations—E: Water supply; NFCNPL_ACC stands for aggregate NPLs—non-financial corporations—I: Accommodation and food service activities; NFCNPL_ADM represents the aggregate NPLs—non-financial corporations—N: Administrative and support service activities; UNEMP stands for % of unemployment; CPI denotes the quarterly consumer price index; R_GDP_Q2Q represents the quarterly percentage growth rate of real GDP; ROA denotes the return on assets: profit or loss for the year/total assets; CAP stands for bank capital and reserves to total assets; LOAN_DISBRS represents the loan disbursements to customers; COVID19_VACCINATED denotes the COVID-19 vaccinated population; finally, GOVT_RESP_STR represents the government response stringency index.

Table A10. Regression estimates per European subregion, prosperity, NPL type, and NPL sector dimensions, with the inclusion of the CULTURE_PCA variable.

PANEL A. This Table Presents the Empirical Results per European Subregion, Prosperity, NPL Type, and NPL Sector Dimensions, with the Inclusion of the CULTURE_PCA Variable. The Period of Analysis Is the Post-COVID-19 Period (2020Q1–2021Q4).

PANEL A Regression results for dimensions with the inclusion of the CULTURE_PCA variable: Subregion/Prosperity/NPL Type/NPL Sector Dimensions—Post-COVID-19 Period

Dimension		Subsample Analysis: Central Europe	Subsample Analysis: Northern Europe	Subsample Analysis: Southern Europe	Subsample Analysis: Prosperity (Hard-Core / Intermediate / Extended Periphery)			Subsample Analysis: NPL Type					Subsample Analysis: NPL Sector				
	Period	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4	Post-COVID-19: 2020Q1–2021Q4
	Model	MODEL (1)	MODEL (2)	MODEL (3)	MODEL (4)	MODEL (5)	MODEL (6)	MODEL (7)	MODEL (8)	MODEL (9)	MODEL (10)	MODEL (11)	MODEL (12)	MODEL (13)	MODEL (14)	MODEL (15)	MODEL (16)
Variable Group	Variable Symbol	D(NPLS)	D(NPLS)	D(NPLS)	D(NPLS)	D(NPLS)	D(NPLS)	D(NPL_RATIO_CRE)	D(NPL_RATIO_HHS)	D(NPL_RATIO_MORT)	D(NPL_RATIO_NFCS)	D(NPL_RATIO_SME)	DNFC NPL_AGR	DNFC NPL_ART	DNFC NPL_CON	DNFC NPL_EDU	DNFC NPL_ELE
Macroeconomic Variables	D(UNEMP(-1))	-0.007931	0.011335	-0.111700 *	0.003481	0.006406	0.053175	0.001879	-6.89×10^{-5}	6.31×10^{-5}	-0.000329 **	-0.000326 **	0.000238	-6.76×10^{-5} *	-6.30×10^{-5}	-1.63×10^{-5} *	-0.000208
	D(CPI(-1))	0.001898	-0.007608	-0.019499 **	0.001383	-0.000461 ***	0.005369	-1.67×10^{-5} *	-2.19×10^{-6}	7.42×10^{-6}	-6.53×10^{-5} **	-1.45×10^{-5}	-9.36×10^{-6}	-7.78×10^{-6}	3.51×10^{-5}	4.44×10^{-6}	-4.57×10^{-5} **
	R_GDP_Q2Q(-1)	0.002009	-0.004334 **	-0.004334 **	0.000253	-0.000215 **	-0.034766 **	-5.12×10^{-5} *	-1.56×10^{-5}	-7.21×10^{-6}	1.88×10^{-5}	-0.000349 ***	-5.58×10^{-6}	-1.17×10^{-6}	2.63×10^{-5}	8.68×10^{-7}	2.32×10^{-5}
Bank-specific Variables	Dep. Variable one lag	0.464575	0.049432	-3.775014 ***	-0.865746 *	-0.246730	-1.815510	0.981349 ***	2.083129 ***	1.533147 ***	0.801145 ***	0.105791	-0.534687 *	-0.028861	0.130361	-1.112437 ***	0.284846 *
	D(ROA(-1))	0.119711	-0.415089 ***	0.302717	-0.100564	-0.035518	-0.106373	-0.009202 *	-0.002407 *	-0.000588 *	-0.001405 **	-0.004141 **	-0.001777	-0.000458	0.000849	-0.000192 *	-7.87×10^{-5}
	D(CAP(-1))	-0.129604 *	0.072743	-0.457852	-0.011465	-0.074462	-7.204265 ***	-0.008419 *	-0.001205 *	-0.001692 *	0.002382 *	0.000753	-0.005993 **	8.83×10^{-5}	-0.003073 *	-0.000146	-0.002629
	D(LOAN_DISBRS(-1))	-0.034574	0.032301 *	-0.323045 *	0.051394	0.028744	0.242544	0.000381	3.00×10^{-5}	-3.65×10^{-5}	7.86×10^{-5}	0.000546	0.000467	-1.99×10^{-5}	-1.82×10^{-5}	-1.03×10^{-5}	3.9×10^{-5}
COVID-19 Government Response Variables	GOVT_RESP_STR	6.68×10^{-5}	0.000300 *	-0.000541	3628751	3193318 **	4671116 ***	0.919157 *	2676209 *	0.168257	0.973022 *	1.674120 **	0.694395 *	-0.37435	-2701637 **	-0.133923 *	0.762508 *
COVID-19 Variables	COVID19_VACCINATED	-0.009501 *	0.01454	-0.021299 *	-0.007642 *	-0.000959	-0.114527 *	-0.000249 ***	-6.55×10^{-5}	-3.1×10^{-5}	-5.19×10^{-5} ***	-4.53×10^{-5} ***	2.98×10^{-5}	5.22×10^{-5} *	0.000158 *	-2.22×10^{-5}	-0.000138 **
Cultural Dimension Variables	CULTURE_PCA	0.037043 ***	-0.277838 *	-0.835484 *	-2104299 **	-0.013389 **	-2.946179 **	-6.49×10^{-6} *	-0.000118 *	-3.69×10^{-5} *	-7.52×10^{-5} *	-0.000767 ***	-0.000437 **	4.74×10^{-5}	-0.000303 *	2.97×10^{-6}	-0.000449 ***
Regression Main Statistics	R-squared	0.545894	0.810609	0.999153	0.960615	0.446338	0.969491	0.494017	0.897462	0.920042	0.569513	0.777499	0.671304	0.470087	0.517033	0.945442	0.198118
	Adjusted R-squared	0.405554	0.210872	0.992798	0.812922	-0.031825	0.870338	0.381139	0.881349	0.907824	0.394627	0.554999	0.342609	0.325187	0.320828	0.890885	0.009440
	F-statistic	1.641136	1.351608	1.572255	6.504124	0.933442	9.777706	4.233321	5.569747	7.352284	3.254491	3.494368	2.042329	1.871016	2.635166	1732.921	1.050030
	Prob(F-statistic)	0.051928	0.075763	0.006337	0.041659	0.056696	0.020277	0.000130	0.000000	0.000000	0.003229	0.012823	0.096970	0.090358	0.012770	0.000002	0.020250
	Durbin–Watson stat	2.271937	2.005223	2.235409	2.112149	1.950082	1.457893	1.948111	1.535112	1.535685	2.259323	2.021428	2.025776	2.415203	1.614852	1.863350	2.449369

Table A10. *Cont.*

PANEL B. This table presents the empirical results per NPL SECTOR dimension, with the inclusion of the CULTURE_PCA variable. The period of analysis is the Post-COVID-19 Period (2020Q1–2021Q4).

Dimension	Variable Group	Variable Symbol	Post-COVID-19: 2020Q1–2021Q4 MODEL (17) DNFC NPL_FIN	Post-COVID-19: 2020Q1–2021Q4 MODEL (18) DNFC NPL_HUM	Post-COVID-19 2020Q1–2021Q4 MODEL (19) DNFC NPL_INF	Post-COVID-19: 2020Q1–2021Q4 MODEL (20) DNFC NPL_MAN	Post-COVID-19: 2020Q1–2021Q4 MODEL (21) DNFC NPL_MIN	Post-COVID-19: 2020Q1–2021Q4 MODEL (22) DNFC NPL_OTH	Post-COVID-19: 2020Q1–2021Q4 MODEL (23) DNFC NPL_PAD	Post-COVID-19: 2020Q1–2021Q4 MODEL (24) DNFC NPL_REA	Post-COVID-19: 2020Q1–2021Q4 MODEL (25) DNFC NPL_PRF	Post-COVID-19: 2020Q1–2021Q4 MODEL (26) DNFC NPL_WRT	Post-COVID-19: 2020Q1–2021Q4 MODEL (27) DNFC NPL_TRA	Post-COVID-19: 2020Q1–2021Q4 MODEL (28) DNFC NPL_WAT	Post-COVID-19: 2020Q1–2021Q4 MODEL (29) DNFC NPL_ACC	Post-COVID-19: 2020Q1–2021Q4 MODEL (30) DNFC NPL_ADM
	Macroeconomic Variables	D(UNEMP(-1))	-0.000391 *	8.31×10^{-5}	0.000189	4.35×10^{-5}	1.80×10^{-5}	-0.000182	0.000107 **	-0.000717	2.82×10^{-5}	-0.000160	5.40×10^{-6}	-1.69×10^{-5}	-4.55×10^{-5}	0.000147
		D(CPI(-1))	1.53×10^{-6}	1.86×10^{-6}	-1.19×10^{-5}	2.36×10^{-5}	-3.10×10^{-5} ***	-4.81×10^{-5}	-3.47×10^{-6}	5.41×10^{-5}	1.65×10^{-5}	-2.67×10^{-5}	3.56×10^{-5}	-2.63×10^{-6}	-1.55×10^{-5}	-0.000343 **
		R_GDP_Q2Q(-1)	6.36×10^{-5}	-2.42×10^{-5} ***	5.05×10^{-5}	-5.52×10^{-5} *	-3.15×10^{-7}	-7.68×10^{-5}	-2.42×10^{-5} ***	0.000304	-2.16×10^{-5} *	-0.000238	6.26×10^{-5} **	-2.64×10^{-5}	1.05×10^{-5}	6.63×10^{-7}
	Bank-specific Variables	Dep. Variable one lag	0.456051 *	0.457979	0.334241	-0.328069 ***	-0.438023 ***	0.156705	0.859949 ***	1.079198 ***	-0.834081 ***	0.681852 *	-1.062127 ***	-0.549158 **	-0.379988	-0.487171
		D(ROA(-1))	-0.000366	0.001105 *	0.000567	0.001338	-0.000457 *	-0.004947 *	-4.57×10^{-5}	0.000216	0.011596	0.000774	0.000237	0.000200	0.000902	-0.009518 *
		D(CAP(-1))	-0.003463 *	0.000811	0.003319 ***	0.003320 *	-0.000639	-0.002034	-0.000316 *	0.002944	-0.001678	-0.003390	-0.004556 **	3.51×10^{-5} *	0.001617 **	-0.008479 *
		D(LOAN_DISBRS(-1))	-0.000125	4.56×10^{-5} *	1.32×10^{-5}	-0.000485 ***	-4.09×10^{-5}	0.000381	-7.39×10^{-5}	-0.002081 *	0.000104	0.000494	0.000719 *	-6.96×10^{-5}	-0.000389 **	0.001390
	COVID-19 Government Response Variables	GOVT_RESP_STR	0.187973 *	0.294227 *	0.677779 **	-1.258017	6.409342 ***	-0.726005	0.140850 **	-1.307522	0.791880	-1.124358	8.066302	0.060199	-1.656750	4.252682
	COVID-19 Variables	COVID19_VACCINATED	1.64×10^{-5}	-6.01×10^{-5} *	-2.82×10^{-5}	-3.35×10^{-5} *	-0.000124 ***	0.000181	8.28×10^{-6} *	2.24×10^{-5}	-7.76×10^{-5}	-8.56×10^{-5} *	-0.000631 *	5.28×10^{-5}	7.91×10^{-5}	-0.003825 **
	Cultural Dimension Variables	CULTURE_PCA	-0.000120 *	-0.000109 *	0.000103	-8.92×10^{-5} *	-0.000131	-1.88×10^{-5}	-3.01×10^{-5} *	-0.000687 *	-0.004236	-0.000264 *	-0.002122 *	2.44×10^{-6}	0.000107	-0.012083 *
	Regression Main Statistics	R-squared	0.497624	0.465839	0.334178	0.418653	0.628992	0.404087	0.559674	0.538699	0.364145	0.657442	0.809130	0.544230	0.720942	0.703607
		Adjusted R-squared	0.352909	0.348837	0.326687	0.330861	0.478271	0.393234	0.419347	0.477398	0.305829	0.314883	0.618259	0.488461	0.441884	0.407213
		F-statistic	1.624843	2.146698	1.235432	2.490588	4.173201	0.679097	1.271042	1.167781	1.409689	1.919211	4.239156	1.194091	2.583481	2.373894
		Prob(F-statistic)	0.129538	0.039114	0.091043	0.011063	0.000487	0.061679	0.029081	0.087839	0.008506	0.047040	0.005344	0.072296	0.043295	0.058736
		Durbin-Watson stat	1.467275	2.08264	2.282751	1.867849	2.051844	2.006813	1.321082	2.106355	1.933857	1.741933	1.870760	1.081447	1.835548	1.648639

Note(s): (1.) PANEL B presents the regression estimates related to the remaining NPL sectors, with the inclusion of the CULTURE_PCA variable for the post-COVID-19 period. (2.) OLS methodology was employed for the regression model estimation. More specifically, Fixed Corrected Panel Effects estimations with country-fixed effects were utilized for all models because of the Hausman test. The table presents the values of the coefficients, while the significance of the *p*-value is presented with an asterisk: *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$. (3.) The variable NFCNPL_FIN represents the aggregate NPLs—non-financial corporations—K: Financial and insurance activities; NFCNPL_HUM represents the aggregate NPLs—non-financial corporations—Q: Human health services and social work activities; NFCNPL_INF stands for aggregate NPLs—non-financial corporations—J: Information and communication; NFCNPL_MAN stands for aggregate NPLs—non-financial corporations—C: Manufacturing; NFCNPL_MIN represents the aggregate NPLs—non-financial corporations—B: Mining and quarrying; NFCNPL_OTH represents the aggregate NPLs—non-financial corporations—S: Other services; NFCNPL_PAD stands for aggregate NPLs—non-financial corporations—O: Public administration and defense, compulsory social security; NFCNPL_REA stands for aggregate NPLs—non-financial corporations—L: Real estate activities; NFCNPL_PRF represents the aggregate NPLs—non-financial corporations—M: Professional, scientific, and technical activities; NFCNPL_WRT stands for aggregate NPLs—non-financial corporations—G: Wholesale and retail trade; NFCNPL_TRA denotes the aggregate NPLs—non-financial corporations—H: Transport and storage; NFCNPL_WAT represents the aggregate NPLs—non-financial corporations—E: Water supply; NFCNPL_ACC denotes the aggregate NPLs—non-financial corporations—I: Accommodation and food service activities; NFCNPL_ADM stands for aggregate NPLs—non-financial corporations—N: Administrative and support service activities; UNEMP represents the % of unemployment; CPI denotes the quarterly consumer price index; R_GDP_Q2Q represents the quarterly percentage growth rate of real GDP; ROA stands for return on assets: profit or loss for the year/total assets; CAP represents the bank capital and reserves to total assets; LOAN_DISBRS stands for loan disbursements to customers; GOVT_RESP_STR represents the government response stringency index; COVID19_VACCINATED denotes the COVID-19 vaccinated population; finally, CULTURE_PCA stands for the national cultural identity variable.

Notes

1. The United Kingdom is included in our selected country dataset, even though it exited the EU in January 2020.
2. The missing data belong to the following countries/periods: Country group 1: Bulgaria, Croatia, Cyprus, Latvia, Slovakia/Period: 2015Q1–2018Q3, Country group 2: Italy/Period: 2015Q–2016Q3, Country group 3: Greece/Period: 2015Q1–2020Q3.
3. Levin–Lin–Chu, Im–Pesaran–Shin, ADF–Fisher Chi-square, and PP–Fisher Chi-square tests were employed to account for data stationarity. Unit root tables for level, first, and second differences are available upon request.
4. Second differences only applied on the variables: COVER_RATIO, DERIVATIVES, TIER1_CAP, and CAP.
5. All unit root test results are available upon request.
6. Durbin–Watson statistic results are depicted in respective tables of Appendix A.
7. For instance, the statistics of COVID-19 deaths (COVID19_DEATHS) as well as of the vaccinations against COVID-19 (COVID19_VACCINATED), respectively, are only available in the post-COVID-19 period and not in the pre-COVID-19 period.
8. Binary variable with values 1/0, where 1 denotes the existence of COVID-19 and 0 the non-existence of COVID-19.
9. The descriptive statistics related to the secondary dependent variables and the control variables employed in this study, as well as the correlation matrix, are not depicted due to space limitations, but are available upon request.
10. Tables depicting the regression results related with the NPL sector for the pre-pandemic period, as well as alternative econometric results generated for all the subsamples of the current research, are not included due to space limitations. All regression models are available upon request.
11. Table 4 serves as both a supplement and a robustness check for the primary results pertaining to the entire sample period. All other robustness models are available upon request.
12. Detailed robustness check results related to the exclusion of the United Kingdom are available upon request.
13. Detailed results of the NPL ratio dependent variable collected from the EBA database are available upon request.
14. Detailed results of robustness checks related to the alternative econometric methods used are available upon request.

References

Ahmed, Shakeel, M. Ejaz Majeed, Eleftherios Thalassinos, and Yannis Thalassinos. 2021. The Impact of Bank Specific and Macro-Economic Factors on Non-Performing Loans in the Banking Sector: Evidence from an Emerging Economy. *Journal of Risk and Financial Management* 14: 217. [CrossRef]
Alessi, Lucia, Brunella Bruno, Elena Carletti, Katja Neugebauer, and Isabella Wolfskeil. 2022. Cover your assets: Non-performing loans and coverage ratios in Europe. *Economic Policy* 36: 685–733. [CrossRef]
Alnabulsi, Khalil, Emira Kozarević, and Abdelaziz Hakimi. 2023a. Non-performing loans and net interest margin in the MENA region: Linear and non-linear analyses. *International Journal of Financial Studies* 11: 64. [CrossRef]
Alnabulsi, Khalil, Emira Kozarević, and Abdelaziz Hakimi. 2023b. Non-performing loans as a driver of banking distress: A systematic literature review. *Commodities* 2: 111–30. [CrossRef]
Apergis, Nicholas. 2022. Convergence in non-performing loans across EU banks: The role of COVID-19. *Cogent Economics & Finance* 10: 2024952. [CrossRef]
Arellano, Manuel, and Stephen Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Review of Economic Studies* 58: 277–97. [CrossRef]
Ari, Anil, Sophia Chen, and Lev Ratnovski. 2020. COVID-19 and Non-Performing Loans: Lessons from past Crises. *SSRN Electronic Journal*. [CrossRef]
Ari, Anil, Sophia Chen, and Lev Ratnovski. 2021. The dynamics of non-performing loans during banking crises: A new database with post-COVID-19 implications. *Journal of Banking and Finance* 133: 106140. [CrossRef]
Ashraf, Badar Nadeem, Changjun Zheng, and Sidra Arshad. 2016. Effects of national culture on bank risk-taking behavior. *Research in International Business and Finance* 37: 309–26. [CrossRef]
Banks, James, Heidi Karjalainen, and Carol Propper. 2020. Recessions and Health: The Long-Term Health Consequences of Responses to the Coronavirus. *Fiscal Studies* 41: 337–44. [CrossRef]
Bassani, Giovanni. 2021. Of Viruses, Economic Crises and Banks: The European Banking Union and the Response to COVID-19. *European Business Law Review* 32: 437–72. [CrossRef]
Bernanke, Ben S. 2009. *The Crisis and the Policy Response: A Speech at the Stamp Lecture*. London: London School of Economics, January 13. Speech 442, Board of Governors of the Federal Reserve System (U.S.).
Bettinger, Eric P. 2010. Instrumental Variables. In *International Encyclopedia of Education*, 3rd ed. Edited by P. Peterson, E. Baker and B. McGaw. Amsterdam: Elsevier, pp. 223–28. [CrossRef]
Bitar, Mohammad, and Amine Tarazi. 2022. Individualism, Formal Institutional Environment and Bank Capital Decisions. *Journal of Corporate Finance* 76: 102244. [CrossRef]
Boubakri, Narjess, Ali Mirzaei, and Anis Samet. 2017. National culture and bank performance: Evidence from the recent financial crisis. *Journal of Financial Stability* 29: 36–56. [CrossRef]

Boubakri, Narjess, Zheng Cao, Sadok El Ghoul, Omrane Guedhami, and Xinjie Li. 2023. National culture and bank liquidity creation. *Journal of Financial Stability* 64: 101086. [CrossRef]

Brunnermeier, Markus, and Arvind Krishnamurthy. 2020. The macroeconomics of corporate debt. *Review of Corporate Finance Studies* 9: 656–65. [CrossRef]

Ceylan, Rahmiye Figen, Burhan Ozkan, and Esra Mulazimogullari. 2020. Historical evidence for economic effects of COVID-19. *European Journal of Health Economics* 21: 817–623. [CrossRef]

Cicchiello, Antonella Francesca, Mariacristina Cotugno, Silvia Monferrà, and Stefano Perdichizzi. 2022. Credit spreads in the European green bond market: A daily analysis of the COVID-19 pandemic impact. *Journal of International Financial Management & Accounting* 33: 383–411. [CrossRef]

Colak, Gonul, and Oğuzhan Öztekin. 2021. The impact of COVID-19 pandemic on bank lending around the world. *Journal of Banking and Finance* 133: 106207. [CrossRef]

Couppey-Soubeyran, Jézabel, Erica Perego, and Fabien Tripier. 2020. 'European banks and the COVID-19 crash test'. *Europe Econpol* 4.

Cowling, Marc, Nick Wilson, Paul Nightingale, and Matej Kacer. 2022. Predicting future default on the COVID-19 bounce back loan scheme: The £46.5 billion question. *International Small Business Journal* 40: 650–66. [CrossRef]

Demir, Ender, and Gamze Ozturk Danisman. 2021. Banking sector reactions to COVID-19: The role of bank-specific factors and government policy responses. *Research in International Business and Finance* 58: 101508. [CrossRef]

Di Tommaso, Caterina, and Vincenzo Pacelli. 2022. Does nonperforming loan securitization affect credit default swap spreads? Evidence from European banks. *Journal of International Financial Management & Accounting* 33: 285–306. [CrossRef]

Duan, Yuejiao, Sadok El Ghoul, Omrane Guedhami, Haifeng Li, and Xinjie Li. 2021. Bank systemic risk around COVID-19: A cross-country analysis. *Journal of Banking and Finance* 133: 106299. [CrossRef] [PubMed]

Dunbar, Kwamie. 2022. Impact of the COVID-19 event on U.S. banks' financial soundness. *Research in International Business and Finance* 59: 101520. [CrossRef]

Durand, Pierre, and Gaëtan Le Quang. 2022. Banks to basics! Why banking regulation should focus on equity. *European Journal of Operational Research* 301: 349–72. [CrossRef]

EBA. 2019. EBA REPORT ON NPLs, www.eba.europa.eu. Available online: https://www.eba.europa.eu/sites/default/documents/files/document_library/Risk%20Analysis%20and%20Data/Risk%20Assessment%20Reports/2019/Final/%20EBA%20Report%20on%20NPLs-for%20publication_final.pdf (accessed on 6 October 2023).

Foglia, Matteo, Abdelhamid Addi, and Eliana Angelini. 2022. The Eurozone banking sector in the time of COVID-19: Measuring volatility connectedness. *Global Finance Journal* 51: 100677. [CrossRef]

Gaganis, Chrysovalantis, Iftekhar Hasan, and Fotios Pasiouras. 2020. National culture and housing credit. *Journal of Empirical Finance* 56: 19–41. [CrossRef]

Giannetti, Mariassunta, and Yishay Yafeh. 2012. Do cultural differences between contracting parties matter? Evidence from syndicated bank loans. *Management Science* 58: 365–83. [CrossRef]

Hansen, Lars Peter. 1982. Large sample properties of generalized method of moments estimators. *Econometrica* 50: 1029–54. [CrossRef]

Ho, Kung-Cheng, Hsiu-Yun Huang, Zheng Pan, and Yifan Gu. 2023. Modern pandemic crises and default risk: Worldwide evidence. *Journal of International Financial Management & Accounting* 34: 211–42. [CrossRef]

Hoang, Thi-Hong-Van, Elysé A. Segbotangni, and Amine Lahiani. 2021. ESG Performance and COVID-19 Pandemic: An Empirical Analysis of European Listed Firms. *SSRN Electronic Journal*. [CrossRef]

Hofstede, Geert. 2001. *Culture's Consequences: Comparing Values, Behaviors, Institutions and Organizations across Nations*. Newcastle upon Tyne: Sage.

Jameaba, Muyanja-Ssenyonga. 2020. Why the Italian Banking System that "Cruised Over" the Global Financial Crisis, Found Itself Roiled by the Sovereign Debt Crisis: Reviewing and Highlighting the Key Issues. *SSRN Electronic Journal*. [CrossRef]

Kindleberger, Charles P., and Robert Z. Aliber. 2011. *Manias, Panics, and Crashes: A History of Financial Crises*, 6th ed. London: Palgrave Macmillan.

Kostis, Pantelis C., Kyriaki I. Kafka, and Panagiotis E. Petrakis. 2018. Cultural change and innovation performance. *Journal of Business Research* 88: 306–13. [CrossRef]

Kozak, Sylwester. 2021. The Impact of COVID-19 on Bank Equity and Performance: The Case of Central Eastern South European Countries. *Sustainability* 13: 11036. [CrossRef]

Kryzanowski, Lawrence, Jinjing Liu, and Jie Zhang. 2022. Effect of COVID-19 on non-performing loans in China. *Finance Research Letters* 52: 103372. [CrossRef] [PubMed]

Laeven, Luc, and Fabian Valencia. 2021. Systemic Banking Crises: A New Database. *SSRN Electronic Journal*. [CrossRef]

Laeven, Luc, and Fabian Valencia. 2018. Systemic Banking Crises Revisited. *IMF Working Papers* 18. [CrossRef]

Laeven, Luc, and Fabian Valencia. 2020. Systemic Banking Crises Database II. *IMF Economic Review* 68: 307–61. [CrossRef]

Loang, Ooi Kok, Zamri Ahmad, and R. V. Naveenan. 2023. Non-performing loans, macroeconomic and bank-specific variables in Southeast Asia during COVID-19 pandemic. *The Singapore Economic Review* 68: 941–61. [CrossRef]

Makri, Vasiliki, Athanasios Tsagkanos, and Athanasios Bellas. 2014. Determinants of non-performing loans: The case of Eurozone. *Panoeconomicus* 61: 193–206. [CrossRef]

Mateev, Miroslav, Ahmad Sahyouni, and Turki Al Masaeid. 2022. Bank performance before and during the COVID-19 crisis: Does efficiency play a role? *Review of Managerial Science* 18: 29–82. [CrossRef]

Mayordomo, Sergio, Maria Rodriguez-Moreno, and Juan Ignacio Peña. 2014. Derivatives holdings and systemic risk in the U.S. banking sector. *Journal of Banking and Finance* 45: 84–104. [CrossRef]

Minsky, Hyman P. 1992. The Financial Instability Hypothesis. *SSRN Electronic Journal*. [CrossRef]

Naili, Maryem, and Younès Lahrichi. 2022. Banks' credit risk, systematic determinants and specific factors: Recent evidence from emerging markets. *Heliyon* 8: e08960. [CrossRef] [PubMed]

Ozili, Peterson K. 2019. Non-performing loans and financial development: New evidence. *Journal of Risk Finance* 20: 59–81. [CrossRef]

Ozili, Peterson K., and Erick Outa. 2017. Bank loan loss provisions research: A review. *Borsa Istanbul Review* 17: 144–63. [CrossRef]

Dursun-de Neef, H. Özlem, and Alexander Schandlbauer. 2021. COVID-19 and lending responses of European banks. *Journal of Banking and Finance* 133: 106236. [CrossRef] [PubMed]

Park, Cyn-Young, and Kwanho Shin. 2021. COVID-19, nonperforming loans, and cross-border bank lending. *Journal of Banking and Finance* 133: 106233. [CrossRef]

Petrakis, Panagiotis E., Pantelis C. Kostis, and Dionysis G. Valsamis. 2015. Innovation and competitiveness: Culture as a long-term strategic instrument during the European Great Recession. *Journal of Business Research* 68: 1436–38. [CrossRef]

Petrakis, Panagiotis, and Pantelis Kostis. 2013. Economic growth and cultural change. *The Journal of Socio-Economics* 47: 147–57. [CrossRef]

Phillips, Peter C. B., and Donggyu Sul. 2007. Transition modeling and econometric convergence tests. *Econometrica* 75: 1771–855. [CrossRef]

Rinaldi, Laura, and Alicia Sanchis-Arellano. 2021. Household Debt Sustainability: What Explains Household Non-Performing Loans? An Empirical Analysis. *SSRN Electronic Journal*. [CrossRef]

Rizwan, Muhammad Suhail, Ghufran Ahmad, and Dawood Ashraf. 2020. Systemic risk: The impact of COVID-19. *Finance Research Letters* 36: 101682. [CrossRef] [PubMed]

Roodman, David. 2009. How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal* 9: 86–136. [CrossRef]

Salazar, Yadira, Paloma Merello, and Ana Zorio-Grima. 2023. IFRS 9, Banking Risk and COVID-19: Evidence from Europe. *Finance Research Letters* 56: 104130. [CrossRef]

Schwartz, Shalom H. 1994. Beyond individualism/collectivism: New cultural dimensions of values. In *Individualism and Collectivism: Theory, Method, and Applications*. Edited by U. Kim, H. C. Triandis, C. Kagitcibasi, S.-C. Choi and G. Yoon. Thousand Oaks: Sage Publications, pp. 85–119.

Sharif, Arshian, Chaker Aloui, and Larisa Yarovaya. 2020. COVID-19 pandemic, oil prices, stock market, geopolitical risk and policy uncertainty nexus in the US economy: Fresh evidence from the wavelet-based approach. *International Review of Financial Analysis* 70: 101496. [CrossRef] [PubMed]

Shehzad, Khurram, Liu Xiaoxing, and Hayfa Kazouz. 2020. COVID-19's disasters are perilous than Global Financial Crisis: A rumor or fact? *Finance Research Letters* 36: 101669. [CrossRef] [PubMed]

Wellalage, Nirosha Hewa, Vijay Kumar, Arshad Iqbal Hunjra, and Mohammad Al-Faryan. 2022. Environmental performance and firm financing during COVID-19 outbreaks: Evidence from SMEs. *Finance Research Letters* 47: 102568. [CrossRef]

Xie, Xin, Nabeel Mirza, Muhammad Umar, and Xiang Ji. 2024. COVID-19 and market discipline: Evidence from the banking sector in emerging markets. *International Review of Economics & Finance* 89: 612–21. [CrossRef]

Yang, Jianlei, and Chunpeng Yang. 2021. Economic policy uncertainty, COVID-19 lockdown, and firm-level volatility: Evidence from China. *Pacific Basin Finance Journal* 68: 101597. [CrossRef]

Yi, Xingjian, Sheng Liu, and Zhouheng Wu. 2022. What drives credit expansion worldwide?—An empirical investigation with long-term cross-country panel data. *International Review of Economics & Finance* 80: 225–42. [CrossRef]

Zheng, Chen, and Junru Zhang. 2021. The impact of COVID-19 on the efficiency of microfinance institutions. *International Review of Economics & Finance* 71: 407–23. [CrossRef]

Žunić, Amila, Kemal Kozarić, and Emina Žunić Dželihodžić. 2021. Non-Performing Loan Determinants and Impact of COVID-19: Case of Bosnia and Herzegovina. *Journal of Central Banking Theory and Practice* 10: 5–22. [CrossRef]

Article

Financial Fragility and Public Social Spending: Unraveling the Endogenous Nexus

Dionysios Kyriakopoulos [1], John Yfantopoulos [2] and Theodoros Stamatopoulos [3,*]

[1] Department of Political Science and Public Administration, School of Economics and Political Sciences, National and Kapodistrian University of Athens, 106 78 Athens, Greece; dionysis.kyriakopoulos@gmail.com
[2] Athens MBA, School of Economics and Political Sciences, National and Kapodistrian University of Athens, 106 78 Athens, Greece; yfantopoulos@gmail.com
[3] Department of Accounting and Finance, School of Administrative, Economics and Social Sciences, University of West Attica, 122 41 Athens, Greece
[*] Correspondence: stamth@uniwa.gr; Tel.: +30-210-5381-789

Abstract: This article provides both stylized facts and estimations of the endogenous nexus of the financial fragility hypothesis (FFH) with public social spending (PSS) for a paradigmatic Eurozone member country. The sample period 1995–2022 includes three major economic crises, the global financial crisis 2007–2009, the European debt crisis 2010–2015 and the COVID-19 pandemic one in 2020–2022. Within the context of the financialization literature, this paper is founded, for the first time, as far as we know, on the "financial fragility hypothesis", combining the effects of both Minsky's "financial instability", as it has been extended for open economies, and the "Eurozone fragility one". Similar to the relevant literature, the findings show that the PSS is associated, in a long-term steady state (cointegration), with the financial fragility process, starting, firstly, from the hedge-financing structure with high profitability of firms, when PSS decreases; secondly, to hyper-speculative financing with risky options, supported by bank credit and openness, indebtedness or discretionary fiscal policy, when PSS rises; thirdly, to the hyper-speculative or even Ponzi financing structures with over-indebtedness (leverage) from the global capital market, inflated asset prices and internationalized fragility, when PSS also rises, and so on. Our conclusion validates Minsky's famous saying, "stability breeds instability", also in the architecturally incomplete Eurozone. Policy implications are straightforward and discussed.

Keywords: public social spending; financialization; Minsky's extended financial instability hypothesis; Eurozone's fragility hypothesis; animal spirits

JEL Classification: F36; F44; F62; G28; H12; H55; H63; I38

1. Introduction

Our neoclassical model of standard economics proves that it is not sufficient to simultaneous interpretation, on the one hand, the continuous transformation of economies and, on the other hand, the permanence of the market as a mode of production (Boyer 2022). The more resounding example is the global financial crisis of 2008 (GFC-2008). The fundamental three assumptions of this prevailing model, self-regulating and efficient markets, the rational behavior of actors and equilibrium, are obviously inadequate in explaining contemporary structural crises (Boyer and Saillard 2002). Today, we recognize the fact that the long-term dynamic development of the countries of East Asia was built on "proactive state policy" (Wade 1990). China's resilience and better performance than North America after the GFC-2008 have shown that there may be more effective alternative development models (Boyer et al. 2011; Alary and Lafaye de Micheaux 2015). Especially in the European Union (EU), we should take into account that given complementarities among institutional forms, such as transfer payments and social security systems, the productive system and

specialization imply more interdependence (not only competition), provided that a stable international regime prevails (Boyer 2022).

A crucial transformation of the contemporary Western economies (North America and Europe) concerns the institution of the "financial system", which, in the last forty years or so, has gradually prevailed in almost all countries. More strictly speaking, this is the "financialization" phenomenon that shaped the transformation of countries' growth models, e.g., the export-led, the welfare-led or even the consumption-led, to that of the finance-led growth regime (Boyer 2000). However, there is no unanimity yet for its definition. Mader et al. (2020) analytically explain a widely employed definition of the "financialization" offered by Epstein (2005), which is "*the increasing roles of financial motives, financial markets, financial actors and financial institutions, in the operation of the domestic and international economies*". Alternatively, scholarship on financialization (van der Zwan 2014; Besedovsky 2018; Mader et al. 2020) has identified the following four different aspects: (1) the emergence of a new regime of accumulation, (2) the dominance of shareholder value, (3) the financialization of everyday life, and (4) structured finance and cultural as well as calculative transformation of credit rating agencies (CRAs). The literature on financialization can be classified into two groups depending on whether the focus of the research is microeconomic or macroeconomic (Cibils and Allami 2013). The last one includes economists from the following three schools of thought; first, "Régulation" (Boyer and Saillard 2002), second, "Post-Keynesian" (Minsky 1975, 1977, 1978, 1982, 1983, 1986, 1992a, 1992b, 1995, 1996), and third, "Radical" (Krippner 2005; Epstein 2005; Lapavitsas 2009).

The aim of this paper is to examine if the "financialization" of a European Monetary Union's (EMU) member country (namely Greece[1]) can explain its public social spending (PSS) during the period 1995–2022. The famous welfare state, an element of the "identity" of the countries of the European Union (EU), is profoundly transformed[2] in more and more financialized economies. Thus, in seeking relevant evidence, we investigate the research question "Is the financial fragility hypothesis (FFH) compatible with long run or steady state public social spending (PSS) of an EMU's member-country (namely Greece), over the sample period 1995–2022?" The term FFH includes[3] Minsky's (1986) financial instability hypothesis (FIH) as it has been extended by Arestis and Glickman (2002) for open economies (eFIH) and the Eurozone fragility hypothesis (EZFH) (De Grauwe 2011, 2012, 2013; De Grauwe and Ji 2022), since Greece belongs to the core of the EU and that of the EMU.

To our knowledge, there is no other study to date that applies the FFH theory and our methodology. Moreover, the case of Greece as a small-open EMU economy is "paradigmatic" because (a) it is internationally deficient, specialized mainly in services such as shipping or tourism, its financialization has been proven (Kyriakopoulos et al. 2022), while this literature is compatible with the endogenous "FFH-PSS" nexus we study here; (b) it was the only one of the EMU member countries for which the political system (Greece–EU) allowed the liquidity crisis of 2010 to slide into a solvency crisis and finally a sovereign default in 2012; it is the same country that, until the GFC-2008, had approximately the same credit rating by the CRAs as the other EMU member countries, but none of them (which experienced the Eurozone-sovereign or banking crisis, namely Italy, Portugal, Spain and Ireland) except Greece signed three MoUs[4] with the Troika (IMF, European Commission and ECB[5]).

We focus on the theory of FFH as an appropriate stream of the financialization process transforming our economies based on the theoretical idea of *endogenous* financial instability or even crises, as well as because it seems that it satisfactorily interprets the PSS (Boyer 2013; De Grauwe 2013; Rossi 2013). The sample period 1995–2022 covers one and a half business cycles with three global or regional crises, the GFC-2008, the Eurozone debt crisis 2010–2015 and the economic crisis of COVID-19 pandemic (2020–2022). Eurozone fragility, always present over the sample period, is conceived on either issuing debt denominated in euro that member countries do not control or where no centralized fiscal budget exists, or even that the European Central Bank (ECB) cannot, by its statute, act as a lender of last resort for

any country of the zone. That is why we have introduced a "Hyper-speculative" financing structure in order to take into account this permanent source of regional instability (EZFH).

The stylized facts we analyzed based on many time series graphs showed that the sample period could be divided into four sub-periods identified by the FFH as (a) the "super-speculative financing structure" 1995–2002, (b) a mix of "hedge and hyper-speculative financing structure" 2002–2008, (c) the "hyper-speculative financing structure" 2008–2018 and (d) the "hyper-speculative" but with some indications towards the "hedge-financing structure" 2018–2022. Furthermore, within this context, we provide consistent interpretations for the Greek PSS based on empirical models [ARDL (p, q1, ..., qk)] estimated at long-term steady-state relationships (cointegrated) with the main factors of the FFH theory; under the limitation of Eurozone vulnerability, these could be allocated to the following stages: (a) clear profitability of the private sector; (b) its risky options supported by bank credits and discretionary fiscal policy; (c) openness to the global capital market and over-indebtedness, revealing "animal spirits" ... until the produced instability leads to a new structure of markets restoring stability, and so on.

We contribute the relevant theory by providing the new term of "hyper-speculative financing" and the FFH economic interpretations of the public social spending of an EMU member country. The most important regional lesson resulting from the conclusion of this paper seems to has already been taken from the European authorities; for instance, one can see the reversal of the European monetary (PEPP[6] by the European Central Bank (ECB) instead of OMT[7]) and fiscal (RRF[8] by the European Commission instead of austerity) policies during the economic crisis of the COVID-19 pandemic (2020–2022) versus the Eurozone debt crisis (2010–2015).

Our research can be generalized to the Eurozone member countries. This is the goal for our next paper. Furthermore, the generalization question of this paper is equivalent to asking whether the Eurozone continues to be fragile. As De Grauwe and Ji (2022) explain, the ECB's willingness to be the lender of last resort in member state bond markets is uncertain. President Draghi did so in 2012 with his famous "whatever it takes" and the OMT program, for which he did not have to put up a single euro. Will Lagarde or whichever next ECB president do so if need be? Since the problem Greece faced in 2010–2012 has not been institutionally solved, it could happen to any other member country of the EMU as long as bond holders, e.g., fund managers, fear (for any reason) that bonds will not be repaid at maturity, which could trigger a self-fulfilling liquidity crisis and even a default.

The structure of the paper is as follows. In the Section 2, we discuss the theoretical underpinnings of the financial fragility hypothesis, for which we document its explanation of public social spending or social security payments. In the Section 3, we describe the research design, modeling and the data analyzed in the paper. The Section 4 presents the empirical analysis and results with the relevant interpretations we offer, while the Section 5 concludes the paper.

2. Theoretical Foundation of the Financial Fragility Hypothesis

Public social spending (PSS) includes both demographic and economic categories of benefits. Demographic ones can be pensions or survivors' and family support benefits, while those linked to the economic cycle can be sickness, disability, unemployment and social exclusion benefits. EMU members like Greece, as aging societies, are expected to have increasing needs for demographic categories anyway.

Economically, however, it should be stressed that, as Darby and Melitz (2008) have shown[9], PSS for aspects like unemployment, age and health, and incapacity and sickness functions as an *automatic stabilizer*, i.e., their payments during recessions increase and conversely, in recovery, decrease. This evidence is very important because although PSS as an automatic stabilizer is proven to work in favor of a "smooth landing or take-off" of the economy, they do not seem to disprove the endogenous nature of cycles or even crises (as this paper also shows), as Minsky (1986) summarized in his famous phrase "stability

breeds instability". Thus, the automatic stabilizer argument enhances the respective of the introduction section that this paper's evidence could be generalized in the Eurozone.

Thus, as *transfer payments* are recorded in government expenditures, PSS follows the mechanism of income redistribution (social policy) within the effectiveness of fiscal policy. More specifically, the European System of Accounts (ESA 2010, § 4.83) defines the (PSS) "*Social contributions and benefits (D.6) as: social benefits are transfers to households, in cash or in kind, intended to relieve them from the financial burden of a number of risks or needs*[10] *[(a) <u>sickness</u>; (b) invalidity, disability; (c) occupational <u>accident</u> or <u>disease</u>; (d) old <u>age</u>; (e) <u>survivors</u>; (f) maternity; (g) <u>family</u>; (h) promotion of employment; (i) <u>unemployment</u>; (j) <u>housing</u>; (k) education; (l) general neediness], made through collectively organized <u>schemes</u>, or outside such schemes by government units and Nonprofit Institutions Serving Households (NPISHs); they include payments from general government to producers which individually benefit households and which are made in the context of social risks or needs*".

In order to develop our research question, we start with the broad theoretical body of "financialization", which has been recently excellently overviewed by Mader et al. (2020). We can distinguish the economists who work at a microeconomic level from those who are interested in the great picture or the macroeconomy. The latter include the Regulationist (Boyer and Saillard 2002; Boyer 2000, 2022), the post-Keynesian (Palley 2007) and the Radical (Krippner 2005; Epstein 2005; Lapavitsas 2009) schools of thought. Among post-Keynesians, the financial instability hypothesis (FIH) theory proposed by Minsky (1957, 1975, 1982, 1986) holds a dominant position in the literature because of its great interpretative capacity for business cycles and crises, like the GFC-2008 or the Eurozone's (2010–2018) one.

Minsky introduced the term "*money manager capitalism*" to describe not only a version of capitalism that is dominated by financial motives and activities but rather to define capitalism as financialized by default (Sotiropoulos and Hillig 2020; Christophers and Fine 2020). In this perspective, his FIH is founded on the idea that any (capitalist) economy *endogenously* establishes a financial structure which is susceptible to crises (Minsky 1983). Its economic performance is mainly determined by the way firms finance their fixed capital investments. Thus, in the first stage of the upward trend of a business cycle, often called the "recovery phase", economic actors (primarily the businesses) are able to finance from their operations both the interests and principal of their loans; Minsky called this financial structure "*hedge* finance". The reason seems to be that both lenders and borrowers do remember recent depression times, so they behave conservatively, while investment plans are mainly financed by internal sources (retained earnings) rather than through banks or other external sources. So, corporate "profitability" seems to be not only the starting point for the *hedge* financing stage but also its foundation, due to the fact that the previous period of high risk is still "alive".

Nevertheless, "A break in the boom occurs whenever... reversals in present-value relations take place. Often this occurs after the increase in demand financed by speculative finance has raised interest rates, wages of labor, and prices of material so that profit margins and thus the ability to validate the past has eroded" (Minsky 1986, p. 220). Arestis and Glickman (2002) underline here what is the most important, which is in contrast with the neo-classical model, though strongly compatible with financial (in origin) crises, i.e., the endogenous nature of the events that break the boom reached by the economy. Although Minsky (1977) accepts that the catalyst of a crisis could be an "external" event, he steadily denies the "exogenous shocks or accidents or even policy errors", writing that "our economy endogenously develops fragile or crisis-prone financial structures" (Minsky 1977, pp. 139–40). Thus, the memories fade due to the increased output (fueled by raised investments), which causes optimism to prevail in markets. This triggers entrepreneurs (see, Schumpeter's sense) to promote projects on innovative products, which inevitably come, this time, with more risky debts. Hence, some of these risky investments could fail (due to the mentioned reversal of present values caused by raised income, interest rate and so on), while if the failures last, then the respective borrowers could not be able to repay (at least) part of their loan. If the bank still believes in the project, it may refinance

the principal while it steadily receives interest. Minsky calls this second stage of the FIH "speculative" finance. External financing and growing debts characterize the speculative financing stage. The duration of this speculative financing stage could be prolonged by reckless leveraging, which actually could be translated as the exposure to risk of asset prices' collapse (Dow 2020). Thus, in Minsky's thought, financial leverage, increasing the ratio of debts in the capital structure, is really significant for the liquidity, the solvency and finally the bankruptcy of any agent.

When a "market" is witnessed of a number of insolvent borrowers, or if it happens that a policy is reversed by the authorities, then firms, sectors or even the economy as a whole could enter the final (third) stage of the Minskyan FIH, that of "*Ponzi*" finance. In the latter, the borrower can repay neither the capital nor the interests of their debts from operational cash flows; thus, they are bankrupted.

The direct consequence of the FIH is the famous Minskyan quote "*stability breeds instability*" over the long term, because of the *endogenous* devaluing of liquidity, the easy lending standards by the deregulated banking system, and the ensuing private debt—leverage—in order to support inflated asset values and the rising expensive capital stock (Holloway and Eloranta 2014; De Grauwe 2013).

In addition, innovation could also come from credit institutions, as happens in reality. This concerns their well-structured processes and their financial products: first, the worldwide *deregulation* process, starting in the early mid-70s from the US, which mainly gave birth to the *globalization* process (Sen 2020); second, the *liberation* of international financial transactions, notably free and easy capital flows (after the collapse of the Bretton Woods system in 1973); third, crucial financial innovation *instruments*, i.e., mainly *derivatives*, like forward, futures, options, and swaps originating from commodity markets or even collateralized debt obligations (CDO) or mortgage- (or asset-) backed securities (MBS or ABS) or credit default swaps (CDS) and the like; it is noteworthy that when financial assets are not backed by physical ones, then they constitute the essential components of an aspect of the financialization process (Blanchard et al. 2021). We refer in summary to the capabilities offered by all these processes and instruments as "*the drive towards financial innovation*". The latter is explained in detail by Arestis and Glickman (2002) who extended the Minskyan FIH (eFIH) in order to show the state of the *internationalized financial fragility*, whereby an economy could fall when, in addition, domestic agents borrow in foreign exchange. They distinguish three (3) potential scenarios: (i) a crisis that is domestic (d) in origin but impacts its external (e) situation (they give this the term "*d to e crisis*"); (ii) a crisis that is external (e) in origin but impacts its domestic (d) situation (they give this the term "*e to d crisis*"); and (iii) crisis-intensifying interactions between (i) and (ii).

The first (i) potential scenario, a "*d to e crisis*", as already mentioned, starts from a rising cost of domestic capital goods, while, as Minsky argues, the result will be present-value reversal and a decline in asset prices. However, the speculative financed units in the closed economy of the early Minskyan analysis are transformed now into the open one to "*superspeculative*" finance because of the huge outflows of international portfolio investments (both by non-residents and residents) who cannot afford the devaluation of their domestic assets. The latter causes the home currency to devalue too, or it could even trigger an exchange rate crisis. Thus, domestic firms in the open economy are now vulnerable due to the fact that even if they were hedge-financed in the beginning (matching asset and liability maturities), they have incurred debts in foreign exchange, while their cash flows are denominated in the home currency. The contagion effects could now be internationalized. As regards the openness of the country to global financial markets, this could cause the second (ii) scenario, a "*e to d crisis*", through the foreign exchange market or the form of its international integration, e.g., it could be a member of a monetary union (e.g., EMU). So, in the stand-alone case, the residents could accumulate debts denominated in foreign currencies; as long as the central bank can also increase the official reserves so as to be able to finance the relevant liabilities with no need for intervention in the foreign exchange market, then *hedge*-financing could be sustainable; however, when the endogenous process

drives up the debt-to-reserves ratio (especially short-term debt), creating doubts regarding the ability of monetary authorities to preserve the purchasing power of the home currency, then the market expectations for the exchange rate could become a source of uncertainty, destabilizing the domestic economy vs. the rest of the world; thus, possible fire-sales of the home currency could, in effect, downgrade the finance of domestic units, including that of the state, to the so called *super*-speculative (Arestis and Glickman 2002) or even the Ponzi financing final stage (equivalent to bankruptcy). The external finance of Minsky's speculative stage is now exacerbated by foreign exchange markets' effects, hence the term "super"-speculative.

Furthermore, in the case of the member countries of an incomplete monetary union like the Eurozone (or EMU), the *super*-speculative stage can become even worse. This is because they have delegated their monetary policy to the ECB, which, by its mandate, cannot act as a lender-of-last-resort, and hence, the members of the EMU cannot give a 100% guarantee to their bondholders that they will have the necessary liquidity to pay them out at maturity (Boyer 2013; De Grauwe 2013; Rossi 2013; De Grauwe and Ji 2022). Put differently, Greece, like every member country of the EMU, issues government bonds in the home currency, the euro, over which it has no control. Thus, as has historically been confirmed, after the GFC-2008 and the ensuing Eurozone sovereign debt crisis that erupted in 2010 (EZ-2010), the unified till then bonds' risk was priced differently since then by the relevant capital market for the South and West Euro-Area Periphery countries (SWEAP) than for the core ones (CORE). This special risk that a government of an EMU member state can run out of cash, which creates the potential for self-fulfilling liquidity crises that may force it to default (as it was the case for Greece in 2010–2012), cannot arise in standalone countries because the central bank of the latter can issue whatever amount in home currency to repay its government bondholders at maturity. Thus, in order to take into account this additional Eurozone-country risk to the super-speculative scheme of Arestis and Glickman (2002), we introduce the term "*hyper*-speculative" financing units, so as to complete the financial fragility hypothesis (FFH). The latter is theoretically expected to affect the EMU's government expenditures of the sample era, part of which are the (Greek) PSS we analyze in this paper.

It is worth noting that this aforementioned risk fragmentation of the EZ-2010 (coming from the diversity of the government bond yields, distinguished into bad and good clusters) and the ensuing burst of the Eurozone's fragility[11] did not happen during the economic crisis of the COVID-19 pandemic, and thus did not trigger a new sovereign crisis. The reason was probably the monetary (the unconditional PEPP of the ECB) and fiscal (the first "Eurobonds"—NGEU[12] program) policies applied by the European authorities (De Grauwe and Ji 2022).

In Figure 1, the main factors of the FFH affecting PSS are schematically allocated to the following three stages:

(a) Starting from "stability economic conditions", the domestic results mainly increase the profitability of the private sector. "Hedge financing structure" prevails. In general, PSS is expected to decrease as output raises [↓ E(PSS)].

(b) The entrance of new foreign direct investments (FDIs) or even short-run portfolios increase risky options for domestic small-medium enterprises (SMEs), which seek bank credits and discretionary fiscal policy resulting in financial leverage—indebtedness— while the inflated asset prices and exposure to foreign funds feed instability conditions of "hyper-speculative financing structure". In general, PSS is expected to increase as output is squeezed [↑ E(PSS)].

(c) The "animal spirits" of the financial downturn should be "paid" by the government through austerity policies. . . "Ponzi-finance" is exacerbated by the Eurozone's inherent fragility. In general, PSS is expected to increase as output is squeezed [↑ E(PSS)]. New markets' structures give birth to a new round from stability to instability, and so on.
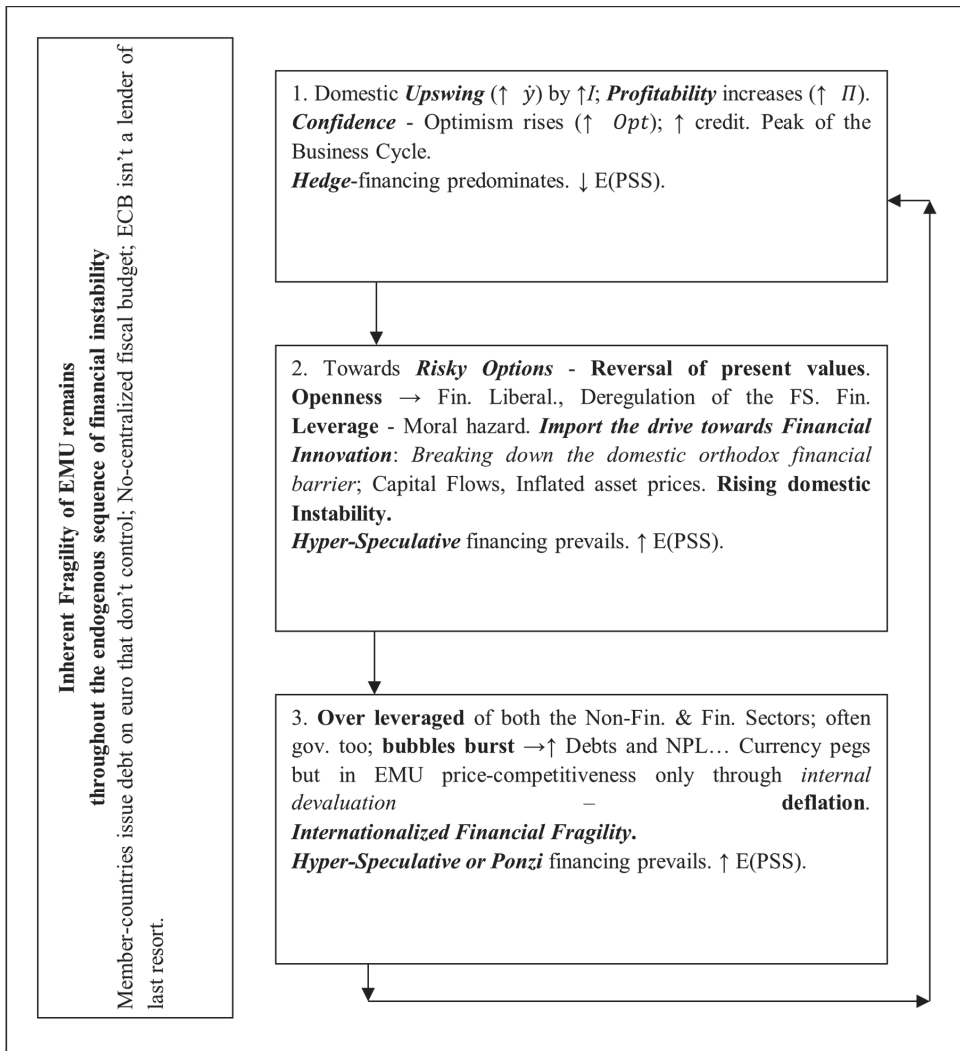
Figure 1. Financial fragility hypothesis (FFH): an *endogenous* process. Notes: $\dot{y}$ = real gross domestic product (GDP) growth; I = business (fixed) investments; E(PSS) = theoretically expected value of public social spending (PSS) or social security payments; Π = profitability; Opt = optimism; EMU = European Monetary Union or Eurozone (EZN) member countries; FS = financial system; ECB = European Central Bank; NPL = non-performing loans.

Based on the aforementioned theoretical analysis of FFH (Minsky 1986; Arestis and Glickman 2002; De Grauwe 2011; Mader et al. 2020), it is considered plausible to empirically investigate the following research question: "*Is the financial fragility hypothesis (FFH) compatible with long run or steady state public social spending (PSS) of an EMU's member-country (namely Greece), over the sample period 1995–2022?*"

Kyriakopoulos et al. (2022) provided theoretical support and empirical evidence for the financialization of the Greek economy based on Boyer's (2000) seminal paper. They confirmed the compatibility of a number of mechanisms that originated from the financial system (see Boyer 2000, Figure 3) and impacted the dependent variable of PSS. However, in this paper, we have proceeded a step further: we elaborate on the specific theoretical

underpinnings of the FFH theory in this section; we appropriately apply, in the following section, the ARDL (p, q) econometric method of Pesaran et al. (2001); and finally, we propose the variable selections tailored to explore the endogenous impacts of financial instability on PSS.

3. Research Design, Modeling and Data

The epistemological problem of the difference between research data and research purpose, is also present in this study, where our effort is to identify if the whole economy or a sector is in a specific Minsky financial stage. However, we can obtain a wider picture, as for instance, when during the Eurozone sovereign debt crisis in 2010–2012, the bond market sentiments transformed the Greek government's liquidity crisis to a solvency one and eventually forced it into bankruptcy; this obviously constitutes the Ponzi finance stage. In addition, we would be making an epistemological error if we tried to "prove" our *falsifiable* research question (Popper 2002; Kuhn 1962; Hepburn and Andersen 2021). We only try to test if the FFH could be considered compatible with the long term or steady state of an EMU's member country (Greece) PSS (cointegrated) over the sample period 1995–2022.

As explained earlier, the financial fragility of a small open Eurozone-member economy arises when endogenous factors predicted by the theory appear and cause either ups and downs of business cycles or, even worse, crises. Thus, we develop a double test for the relationship, "FFH, the cause—PSS, the effect-purpose"; first, we use descriptive statistics with many time series graphs in order to see if the stylized facts of the Greek economy's impacts on PSS tell the story of the FFH; second, in the econometric analysis, we identify autoregressive distributed lag [ARDL (p, q_1, . . ., q_k)] empirical models so as to test if there are long-run (cointegrated) relationships between the dependent of the PSS and a number of "independent" variables expressing the factors predicted by the FFH theory.

More specifically, the social contributions and benefits (D.6) paid (PSS) in Greece include[13] seven out of twelve risks or needs provided by the (ESA 2010), that is, (1) old age (pensions), (2) sickness, (3) survivors, (4) disability, (5) family, (6) unemployment, and since 2014, (7) social exclusion[14] (see Appendix A, Figure A13 evolution of the PSS composition).

We expect that the main factors of the FFH affecting PSS, as they have been explained in the previous section and presented in Figure 1, can be detected from estimations of our empirical ARDL (p, q) models. However, a limited effect of PSS functioning as an automatic stabilizer is expected, especially during the period of the Eurozone debt crisis (2010–2015), not only because of the Stability and Growth Pact (SGP of the EMU) predictions[15] still in place, but mainly because of the harsh austerity–deflation policies imposed on the country by the Troika (International Monetary Fund—IMF, European Commission—EC, and ECB) through the memoranda (MoU) signed by the government. They treated PSS as a luxury good with low-income elasticity.

So, first, we investigate whether the "profitability" of sectors or the economy as a whole is such that the constant expansion of the real output can be justified, i.e., whether we can detect the "stability conditions" expressing "hedge finance". If the data (primarily of the non-financial sector) record a sharp increase in credit and then loans with a parallel increase in asset prices, this could suggest a robust financial structure in that period (hedge finance). Conversely, if endogenous forces push asset prices down, while private sector borrowing increases, alongside anemic growth in real output (a widening output gap), then our open Eurozone-member economy slides into a "hyper-speculative financial" structure. Hence, this could reflect an advanced stage of the Minsky moment "*stability breeds instability*". Further, for Greece, which belongs to the EMU, the previously mentioned Eurozone fragility factor should be added. That cumulative effect of "free capital movements plus the membership of the incomplete EMU", we have called the "hyper-speculative" financial structure. Finally, the "Ponzi" finance stage can be easily ascertained from declared solvency crisis—defaulting.

We specify the long-term model in Equation (1).

$$scbtotgdp_t = a + X_{jt} + \varepsilon_t \tag{1}$$

where $scbtotgdp_t$ = Greek social contributions and benefits paid (the dependent variable); $X_{jt} = 1, \ldots, k$ explanatory variables; t = time in quarters; ε_t = the disturbance term.

By implementing the Pesaran et al. (2001) bounds testing approach for the cointegration, and by rewriting Equation (1) in an error-correction model (ECM) form, the short-run effects of the FFH factors on PSS are presented in the Equation (2) (Hajilee et al. 2021).

$$\Delta scbtotgdp_t = a + \sum_{k=1}^{nj} \beta_{t-k} \Delta X_{jt-\kappa} + \lambda_j X_{jt-1} + \xi_t \varepsilon_{t-1} + \mu_t \tag{2}$$

To demonstrate the long-term relationship, we need to control two criteria, namely the sign and significance level of the error-correction coefficient (ξ_t); a significant negative ξ_t is an indication of a long-term relationship or cointegration among the variables.

Equation (2) has been estimated in five (5) different models according to the FFH structure summarized in Figure 1, while estimations of the variables are presented in Table 3. The definitions and sources of the variables are reported in Table A1 of Appendix A. So, the vectors of independents include Model 1, focusing on the "profitability factor" X_{jt} = (*invgdp*, *realgdp10yy*, *discrpol*, *goss11gdp*, *goss12gdp*); Model 2a, focusing on the "risky options and financial development factors" X_{jt} = (*realgdp10yy*, *discrpol*, *credtnfgdp*[16], *credtfgdp*); Model 2b, focusing on the "Openness and indebtness -leverage" X_{jt} = (*realgdp10yy*, *discrpol*, *cagdp*, *s11debtgdp*, *pudbtgr*, *s1ltloanss2tot*); Model 2c, stressing the "Inflated asset prices and foreign funds exposure" X_{jt} = (*realgdp10yy*, *cagdp*, *s11debtgdp*, *s1ltloanss2tot*, *hpiyoy*, *asegspiyoy*, *gr10ygby*); and Model 3, focusing on "Animal spirits "paid" by the gov." X_{jt} = (*discrpol*, *m3outsgdp*, *nlbs13gdp*, *rgsnowb*).

An example of the analytical form of Model 1 (Table 3) as a special case of Equation (2) is presented in Equation (2a).

$$\begin{aligned} \Delta scbtotgdp_t = \quad & a + \sum_{k=1}^{n1} \beta_{t-k} \Delta invgdp_{t-\kappa} + \sum_{k=1}^{n2} \beta_{t-k} \Delta realgdp10yy_{t-\kappa} + \sum_{k=1}^{n3} \beta_{t-k} \Delta discrpol_{t-\kappa} \\ & + \sum_{k=1}^{n4} \beta_{t-k} \Delta goss11gdp_{t-\kappa} + \sum_{k=1}^{n5} \beta_{t-k} \Delta goss12gdp_{t-\kappa} + \lambda_1 invgdp_{t-1} + \lambda_2 realgdp10yy_{t-1} \\ & + \lambda_3 discrpol_{t-1} + \lambda_4 goss11gdp_{t-1} + \lambda_5 goss12gdp_{t-1} + \xi_t \varepsilon_{t-1} + \mu_t \end{aligned} \tag{2a}$$

As Kyriakopoulos et al. (2022) mentioned, the Pesaran et al. (2001) estimation approach used in this paper has three main advantages over other methods of cointegration: it obviates the unit root pretests to identify the degree of integration of the time series; it can be used with either I(0) or I(1) variables, but not I(2); a one-step simultaneous estimation on both long-term and short-term models is applied. The procedure involves three stages (Goel et al. 2008): first, searching the long-term (level) relationship among the variables applying the bound tests through the estimation of a conditional ECM; second, the lagged dependent-variable term and the one-period lag on regressors are tested for joint significance via an F-test, under the (null) H0 "variables have not relation in levels" and using the critical values of Pesaran et al. (2001), and a supplementary *t*-test is available for the significance of the lagged dependent variable, with critical values again provided by Pesaran et al. (2001); third, if from the previous tests a level relationship cannot be rejected, then the long-term or cointegrated one is estimated through the ARDL procedure, as proposed by Pesaran and Shin (1999).

In Table 3, only the estimated coefficients (λ_i and ξ_t) of the ECM are presented, since we are interested only in long-term cointegrated relationships.

As regards the "discretionary fiscal policy" variable (*discrpol*), this was based on the methodology of Fatás and Mihov (2003), and we have used it as a proxy of the unobserved government expenditures such as the "interest paid for the public debt" or "countercyclical

additional payments". It comes from the residuals of the ordinary least squares (OLS), estimated using Equation (3)

$$\Delta lgovcons_t = \Delta lgovcons_{t-1} + \Delta lgdp_t + inflr_t + inflr_t^2 + fuelspiyy_t + \varepsilon_t \qquad (3)$$

where $\Delta lgovcons_t$ is the growth rate (yoy) of the government consumption expenditures; $\Delta lgdp_t$ is the growth rate of the nominal gross domestic product (GDP); $inflr_t$ is the inflation rate as measured by the general consumer price index; $inflr_t^2$ is the squared inflation; $fuelspiyy_t$ is the growth rate of fuel price index (yoy).

All data used in the paper have been drawn either from the Refinitiv (LSEG) database, the Hellenic Statistical Authority[17] or the Bank of Greece[18].

4. Empirical Analysis and Discussion

We begin with a wider picture of the economy 1995–2023 based on the raw variables and stylized facts shown in the graphs provided in Appendix A. It should be pointed out that although PSS payments are transfer payments, they are nevertheless recorded in government expenditures and hence in the state budget; see fiscal policy.

So, in the first step of (descriptive statistical) empirical analysis, one can find out if the national accounts data, which we processed and present in Figures A1–A13 in Appendix A, are compatible with the FFH as summarized schematically in Figure 1. From the latter arise the turning points one looks for: (a) profitability feeding a climate of confidence–optimism and rising credits, which could express the Minsky "hedge financing" stage; (b) a sequence from the undertaking of risky business investment plans with extra leverage in foreign exchange currency and the reversal of present values with shrinking profit margins, to a rise in outstanding debts and pessimism and capital outflows with likely liquidity crisis, which could be translated as the "hyper-speculative financing" stage; (c) according to the finance-led growth regime adopted as part of the insertion of the country into the incomplete EMU (Kyriakopoulos et al. 2022), this choice could also lead to bankruptcy, or, in Minskyan terms, to a "Ponzi financing" structure.

We think that the accumulative information of the facts reported in Figures A1–A13 of Appendix A need not be explained. A few comments could be useful:

Given that we are interested in (endogenous) *financial* instability, the starting point should be the overall performance of the economy recorded by the history of the GDP (see Figure A2). The global financial crisis of 2008 (GFC-2008) and the Eurozone debt crisis 2010–2014 had a decisive structural impact on the economy, which is reflected in the large decline in employment in manufacturing ($\simeq$32%) over the 2008–2014 recession and its stagnation since then, while the respective approximately constant increase in the financial sector is reflected in a jump in their ratio of FI/MNF (see Figure A1). The profitabilities, measured by the gross operating surplus[19] to GDP, of the non-financial sector ($\downarrow$18%) and financial sector ($\uparrow$67%) follow opposite tendencies and translate the increase in the financialization of the economy as well as the transition from a hedge–financial structure to a hyper-speculative one (Figures A2 and A3).

The snowball effect shown in Figure A11 can express the change in the climate of confidence from euphoria during 1995–2008 to pessimism or animal spirits over the deep recession 2008–2014, stagnation 2014–2018 and some recovery since then. The most important structural damage in a long-term perspective is demonstrated in the de-investment of the economy (see Figures A6 and A8). In financial terms, the reversal in profit margins (2008–2014) can be confirmed by respective asset price falls (dwellings, the stocks in the Athens exchange and government bonds) (see Figures A4, A5 and A9). The hyper-speculative financing structure can once again be detected.

The international investment position (BP-IIP[20]) of the country is steadily negative and in fact continuously deteriorated over the 2002–2022 period (see Figure A12). The stylized facts of the international borrowing of the country are compatible with the FFH and are similar to those in the literature (Vigny 2022; Amoutzias 2019).

The facts could be interpreted as showing a de facto "Ponzi financing structure", as the productive private sector (S.11) was obliged to work with no credit since 2014 while outstanding debt accumulated (see Figures A6–A8). Regarding the overall performance of institutional sectors [(non-financial (S.11), financial (S.12), government (S.13) and Households and NPISH (S1.M), and the domestic economy (S.1)] can be approximated by the measurement of "net lending (+ surplus)/borrowing[21] (- deficit)"[22] to GDP (Figure A10).

Thus, in Minskyan terms and based on Table 1, it seems[23] that respective data are compatible with a "speculative" financial structure during the period before the introduction of the euro (1995–2002); then, the structure was "hyper-speculative" (2002–2012); after that, the "Ponzi-financing" structure dominated since the government defaulted and bank recapitalized in 2012 till 2015. From 2016, there was a return to a recovery trend with "hyper-speculative financing", because, despite the change in European policies (see PEPP and RRF at least), Eurozone fragility remains (De Grauwe and Ji 2022).

Table 1. Indications from our sample for the Minskyan financial structures (averages of periods).

	1995–2001	2002–2011	2012–2015	2016–2022
Gross Operating Surplus Non-Financ. Firms/GDP (Figure A2)	0.180	0.170	0.160	0.140
Gross Operating Surplus Financ. Firms/GDP (Figure A2)	0.020	0.020	0.010	0.030
Total Net Operating Income of Credit Institutions/Net Capital (Figure A3)		0.560	0.490	0.280
Net lending (+) of Non-Financial firms (Figure A10)	0.025	0.115	0.199	0.018
Net lending (+) of Financial firms (Figure A10)	1.215	0.350	1.696	0.283
Net borrowing (−) of General Government (Figure A10)	−0.498	−0.592	−0.486	−0.142
Athens Stock Exchange General Index * (Figure A5)	0.438	0.595	−0.733	−0.832
Price of 10-year government bond * (Figure A5)	0.559	0.242	−1.547	0.253
Urban areas' residential price index * (Figure A5)	−1.293	0.949	−0.223	−0.339
Credit to private non-financial sector/GDP (Figure A6)	0.050	0.196	0.012	−0.209
Outstanding Debt of Non-Financial Firms/GDP (Figure A8)	0.418	0.565	0.691	0.633
Credit to Households/GDP (Figure A6)	0.103	0.442	0.651	0.575
Outstanding Debt of Households/GDP (Figure A8)	0.123	0.408	0.652	0.574
Non-Performing Loans/Total Gross Loans of Banks (Figure A6)		5.403	27.459	32.780
Gross Fixed Capital Formation/GDP (Figure A8)	0.227	0.219	0.111	0.119
Current Account/GDP	−0.047	−0.102	−0.018	−0.044
International Investment Position # (Figure A12)		−184	−229	−274
Snowball Effect (Figure A11)	0.055	0.034	0.134	0.026

* Standardized values; # billion Euros; snowball effect = r − g; r = the weighted average of the interest rates of 10-year government bonds; g = nominal GDP growth rate. All calculations are based on the dataset; we constructed the reported Figures and graphs.

In the second step of the empirical analysis, we provide econometric estimations (Table 3) and interpretations compatible with the FFH. The social contributions and benefits paid as a ratio of GDP (*scbtotgdp*) are explicitly defined as the dependent variable of the estimated empirical models identified in ECM forms as presented by Equation (2) or (2a). In Table 2, the summary statistics of the variables used in the ARDL (p, q) estimations are shown. The availability of data has defined their time span as well, so that we do not have a balanced time series sample.

Table 2. Summary statistics of the variables used in ARDL (p, q) estimations.

Variable	Obs	Mean	Std. Dev.	Min	Max
scbtotgdp (the dependent var.)	97	0.4284	0.05	0.34	0.52
invgdp (model 1)	113	0.1803	0.06	0.08	0.29
realgdp10yy (model 1, 2)	114	0.0126	0.08	−0.43	0.22
discrpol (model 1, 2 & 3)	109	0.0000	0.04	−0.11	0.09
goss11gdp (model 1)	97	0.1601	0.02	0.10	0.20
goss12gdp (model 1)	97	0.0187	0.01	0.00	0.04
credtnfgdp (model 2)	112	0.5278	0.14	0.29	0.72
credthgdp (model 2)	112	0.4201	0.22	0.06	0.67
cagdp (model 2)	102	−0.0604	0.06	−0.19	0.12
s11debtgdp (model 2)	95	0.5879	0.10	0.37	0.73
pudbtgr (model 2)	87	0.0079	0.03	−0.20	0.10
s1ltloanss2tot (model 2)	102	0.4662	0.18	0.15	0.78
nprtfinflowsgdp (model 2)	85	−0.0214	0.13	−0.79	0.23
hpiyoy (model 2)	109	0.0350	0.08	−0.13	0.16
asegspiyoy (model 2)	108	0.0700	0.39	−0.66	1.67
gr10ygby (model 2)	104	0.0653	0.05	0.01	0.25
m3outsgdp (model 3)	108	3.7618	0.55	2.68	5.29
nlbs13gdp (model 3)	97	−0.0648	0.06	−0.31	0.06
rgsnowb (model 3)	103	0.0508	0.06	−0.07	0.26

Note: See Table A1 in Appendix A for the necessary details of all used variables; the variables are measured as ratios or growth rates (not %) or even, in one case, are defined as residuals (*discrpol*). The dependent variable in all models is the "social contributions and benefit paid as ratio to GDP" = *scbtotgdp*; for convenience we repeat: (private fixed investment)/gross domestic product (GDP) = *invgdp*; real GDP growth rate (annual basis) = *realgdpanngr*; discretionary policy (see Equation (3)) = *discrpol*; (gross operating surplus of the non-financial sector)/GDP = *goss11gdp*; (gross operating surplus of the financial sector)/GDP = *goss12gdp*; (credit to non-financial sector)/GDP = *credtnfgdp*; (credit to households and NPISH[24])/GDP = *credthgdp*; (current account balance)/GDP = *cagdp*; (outstanding debt of the non-financial sector)/GDP = *s11debtgdp*; growth rate of the public debt = *pudbtgr*; (long-term (LT) loans to the non-financial sector coming from the *external* sector)/(total LT loans) = *s1ltloanss2tot*; housing price index growth [year-over-year (yoy)] = *hpiyoy*; Athens exchange general share price index (yoy) = *asegspiyoy*; Greek (GR) government 10 years bond yield = *gr10ygby*; (M3[25] outstanding)/GDP = *m3outsgdp*; (general government net lending or borrowing)/GDP = *nlbs13gdp*; snowball effect (r-g)[26] = *rgsnowb*.

In the estimated models[27] presented in Table 3, it was not possible to distinguish sub-periods of the business cycle detected in the previous descriptive analysis, for econometric reasons, due to the insufficient data in the demanding process of ARDL (p, q) co-integration relations. We should start the evaluation from the "Diagnostic Statistics, Pesaran et al. (2001) bounds tests and Adjustment EC-Term". Both the *F*-test and *t*-test confirm the existence of a cointegration relationship for all the five models; *this is equivalent to saying that our research question cannot be rejected* (at 1% significance level) given the second condition of the negative sign and strong significance of the estimated coefficients of the ECM; the latter vary between −0.76 and −1.51, indicating fast and very fast adjustment of the system. That is, between 76% and 151% of the total difference from the short-run dynamics to the long-term equilibrium trend was occurring within the coming quarter of the sample period. Based on the remaining diagnostic tests, the estimated models cannot be rejected.

Table 3. ARDL estimations.

	Model 1	Model 2a	Model 2b	Model 2c	Model 3
	Profitability	Risky options + financial developm.	Openness and indebtedness—leverage	Inflated asset prices and foreign fund exposure	Animal spirits "paid" by the gov.
D.scbtotgdp	1999q4–2022q3	2007q4–2022q3	2001q1–2021q4	1999q4–2021q4	1999q4–2022q1

Table 3. *Cont.*

	Model 1	Model 2a	Model 2b	Model 2c	Model 3
Panel A. Long-term (cointegrated) estimations					
(Priv. Fix. Investm.)/GDP	−0.148 ***				
Real GDP growth r.	−0.224 ***	−0.306 ***	−0.248 ***	−0.250 ***	
Discret. Pol.	0.204 **	0.279 **	0.087 *		0.110 **
(Gross Operat. Surpl. Non-Fin. Sect.)/GDP	−0.418 **				
(Gross Operat. Surpl. Financ. Sect.)/GDP	−0.961 **				
(Credit to Non-Fin. Sect.)/GDP		0.272 ***			
(Credit to Households and NPISH)/GDP		0.075 ***			
(Current Account)/GDP			−0.180 **	−0.341 ***	
(Outstand. Debt Non-Fin. Sect.)/GDP			0.152 ***	0.176 ***	
Growth rate of thePublic debt			−0.093 **		
(LT loans Non-Fin. sect. from external sect.)/(Tot. LT loans)[28]			0.057 **	0.107 ***	
Housing price index growth (yoy)				0.062 *	
Athens Exch. gen. sh. pr. index (yoy)				0.019 ***	
GR gov. 10 y. Bond Yield				0.140 ***	
(M3 Outst.)/GDP					0.011 **
(Gen. Gov. net Lend. or Borr.)/GDP					−0.119 ***
Snowball eff. (r-g).					0.223 ***
Trend	0.0005 **		0.0009 ***	0.001 ***	0.001 ***
Constant	0.3640 ***	0.256 ***	0.172 ***	0.173 ***	0.129 ***
Panel B. Diagnostic Statistics					
Bounds test (H0: no level relationship)	reject at 1% level	reject at 1% level	reject at 1% level	reject at 1% level	reject at 1% level
Adjustment EC Term: L.scbtotgdp	−0.856 ***	−0.757 ***	−1.127 ***	−1.512 ***	−1.028 ***

Table 3. *Cont.*

	Model 1	Model 2a	Model 2b	Model 2c	Model 3
Ramsey RESET F-test (H0: Model has no omitted variables)	Prob > F = 0.2182	Prob > F = 0.5505	Prob > F = 0.4110	Prob > F = 0.0826	Prob > F = 0.1404
Breusch–Godfrey LM test for AR (H0: no serial correlation)	Prob > chi2 = 0.2852	Prob > chi2 = 0.3143	Prob > chi2 = 0.7718	Prob > chi2 = 0.2268	Prob > chi2 = 0.3592
Breusch–Pagan/Cook–Weisberg test (H0: Constant variance)	Prob > chi2 = 0.2574	Prob > chi2 = 0.7495	Prob > chi2 = 0.7341	Prob > chi2 = 0.9560	Prob > chi2 = 0.7774
Mean VIF	4.18	5.71	7.57	9.46	4.12
Observations	92	60	84	89	90
Adj R-squared	0.8099	0.8825	0.8177	0.8437	0.7959

Notes: Only long-term cointegration estimations are presented. *** stands for statistically significant at 1% level or lower; ** stands for statistically significant at 5% level or lower; * stands for statistically significant at 10% level or lower.

Model 1 (see Table 3) provides statistically significant estimations for the factors concerning "profitability", private investments, growth and discretionary fiscal policy (see also Figure 1, stage 1 seems to prevail here), which are proved to be in a long-term steady state with the dependent of the "*social contributions and benefits to the GDP*" (hereafter PSS). Thus, when the gross operating surplus of the non-financial sector and the financial one (as a proxy of the "profitability" of these main sectors of the economy) rise, the PSS decreases[29] because the increasing output, income and employment reduce the need for social support, mainly for old age, sickness and survivors[30] ($\uparrow goss11(12)gdp \rightarrow \downarrow scbtotgdp$). The same reasoning for the rising efficiency of the economy can also be used to justify the even more statistically significant (at less than 0.001 level) negative relationship of both private investments and the growth rate of the economy with the dependent of the PSS [$\uparrow(invgdp, realgdpyy) \rightarrow \downarrow scbtotgdp$]; see relevant Figures A1–A4, A10 and A13 in Appendix A. These findings are in line with Pierros (2020). The opposite statistically significant positive steady state relationship between "discretionary policy" and PSS can be justified by the government expenditure that does not concern either permanent behavior, nor the increase in economic activity, nor inflation or fuel subsidies[31], but rather and probably the electoral cycle or populist extraordinary benefits, etc. ($\uparrow discrpol \rightarrow \uparrow scbtotgdp$).

In Model 2, we focused on the explanatory variables concerning risky options—financial development, openness and indebtedness—leverage of the economy (see also Figure 1, stage 2 seems to prevail here). Credits can increase the corporate turnover, which is associated with employees' increased sickness or invalidity or occupational accident and disease, as well as greater requirements for supporting maternity and family, housing or education from the social security system; hence, there is a positive estimated coefficient for the relationship between "credits to the non-financial sector" and PSS (Model 2a) ($\uparrow credtnfgdp \rightarrow \uparrow scbtotgdp$). Also, the higher the credits to households, the more the family can be expected to grow, and that is why, for the majority of employees or workers (middle or lower incomes), more PSS is needed for housing, education, health and so on; this can be a reasonable interpretation of the positive relationship with the PSS estimated in Model 2a ($\uparrow credthgdp \rightarrow \uparrow scbtotgdp$); see relevant Figure A6 in Appendix A. These findings are also similar to those of Amoutzias (2019).

Rising current account balances or a shrinking deficit, and an implicit increase in the international competitiveness of domestic production, translates a respective increase in net private investments, while both factors cause GDP growth ($\uparrow cagdp \rightarrow \downarrow scbtotgdp$); the latter, as explained earlier, is negatively related to the automatic stabilizer PSS; see relevant

Figure A12 in Appendix A for the international investment position of the country (BP-IIP). Furthermore, the outstanding debt of the non-financial sector (S.11), which has been estimated to be positively related to PSS, can pass, in good times (see the 1995–2008 period here), the threshold of heavy financial leverage, which, in turn, refers to the transition from the Minskyan "hedge to hyper-speculative financing". To put it differently, the excessive growth of private debt in times of euphoria, with an increase in the cost of capital (and the lack of capital goods produced by domestic manufacturing[32]), due to its equally excessive demand, since it reverses present values, leads to corporate bankruptcies. Thus, it causes a decrease in production, income and employment, and therefore, there is a strong need to increase social support for the affected (i.e., raise of the PSS); hence, there is a positive and strongly significant estimated relation "outstanding debt of the non-financial sector—PSS" in Models 2b and 2c ($\uparrow s11debtgdp \rightarrow \uparrow scbtotgdp$); see relevant Figure A8 in Appendix A.

However, the growth rate of public debt is not a surprise that is estimated to negatively affect PSS because it focuses on the rate of increase and not on the level of this debt. So, because this factor is perceived as synonymous with the fear or even the panic of the capital markets, to whose rising they react with fire-sales of bonds or other assets to push up their yields, real interest rates and so on, provoking once again the reversal of present values, investment failures and bankruptcies. This process ends in a recession that demands governments to release (or at least not brake or restrain) automatic stabilizers like PSS in order to support the affected. Of course, this process in the Greek case was much worse because the Troika imposed deflation policies—austerity—although they found a liquidity crisis in the onset at 2010 and not a solvency crisis! The deflation process blocked the operation of PSS—automatic stabilizers ($\uparrow pudbtgr \rightarrow \downarrow scbtotgdp$)! These began to operate in a limited and gradual manner after 2012 and Draghi's "whatever it takes", the restructuring programme (PSI), and the change in the composition of public debt from bonds to mortgage loans from the Eurozone member countries through the ESM; see relevant Figure A9 in Appendix A. So, in practice, it seems that the monetary policy determines the sustainability of the public debt, as is predicted by post-Keynesians.

The variable concerning the long-term borrowing of the non-financial sector (S.11) from abroad as a ratio to the total of its corresponding loans (Models 2b, 2c) is also of paramount importance. A statistically significant positive (cointegration) relationship with PSS was estimated. Its interpretation is similar to that of the outstanding debt of the sector, but here, it is not enough to add the risk of the foreign currency. We should also add that of the Eurozone's fragility hypothesis (De Grauwe 2011). That is why we propose to expand the Arestis and Glickman (2002) term of "super-speculative" with that of "hyper-speculative" in order to also include the risk that the EMU financing units issue their debts on the euro that no country controls ($\uparrow s1ltloanss2tot \rightarrow \uparrow scbtotgdp$); see relevant Figure A12 in Appendix A.

The estimated long-term relationship between asset prices (housing, stock prices—general price index, as well as 10-year government bond yield) and PSS theoretically only has an expected positive sign for the last one, the yield of 10-year government bonds (Model 2c). When this yield rises, it translates the decrease in the price of traded bonds due to some kind of fear. Domestic interest rates are gradually influenced upwards, which reduces aggregate demand, output, income and employment. To the extent that public social spending (PSS) acts as an automatic stabilizer, it will increase as estimated in Model 2c [$\uparrow gr10ygby \rightarrow \uparrow scbtotgdp$]. As regards housing prices and prices of the Athens stock exchange, the estimated positive cointegrated relation could be explained like this: in times of crises (GFC-2008, EZ 2010–2015, COVID-19 2020–2022), the restructuring of the goods–services market due to bankruptcies or M&As[33] of firms (usually small and medium-sized enterprises[34]) generally increases the profit margins (determined independently of unemployment) of the survived, while labor wages decrease due to the large increase in unemployment caused by the recession[35]. Then, logically, automatic stabilizers such as PSS work in the opposite way, and therefore, there is a positive correlation between asset prices

(especially equity prices) and social spending on relief for the affected [↑ (*hpiyoy*, *asegspiyoy*) → ↑*scbtotgdp*]; see relevant Figures A4 and A5 in Appendix A.

Finally, in Model 3, we present estimations concerning indicative monetary and fiscal policy measures reacting to over-leverage and internationalized financial fragility. It should be noted that regardless of the investment grade given by the CRAs[36] for Greek government bonds (junk until 2023), the ECB bought them ("waiver") for reasons of financial stability of the Eurozone; the latter caused a credit boom from the financial sector of the core EMU countries towards Greece as a result of the increase in its private debt (especially due to a real estate boom; see Figures A1 and A6–A8 in the Appendix A) up to the GFC-2008 (Boyer 2013). Thus, the expansionary monetary policy (M3/GDP) of the ECB through Keynesian short-term processes drives lower interest rates and higher outputs, whereby income and employment are compatible with less need for PSS. However, as prices rise due to excess aggregate demand, in the long term, real money balances fall, interest rates increase with output, and incomes and employment decrease; hence, there is a rise in automatic stabilizers of PSS, such as employment promotion policies (↑*m3outsgdp* → ↑*scbtotgdp*). This evidence is similar to that in Rossi (2013). The snowball effect (r-g) has been estimated as a positive steady-state relationship with PSS; when it rises, the public debt rises too, while contractionary effects result in a lower output, income and employment, causing an increase in PSS (↑*rgsnowb* → ↑*scbtotgdp*). In the end, the increase in net government lending (+) (i.e., the primary fiscal budget surpluses, not only from the austerity policies implemented in crises but also due to the SGP of the EMU before them) ameliorates the mood in the markets, causing a decrease in interest rates and a respective increase in profit margins (especially if public property is sold as it was here—a condition of the Troika lenders) and so on as aforementioned; hence, lower amounts of PSS are needed, justifying the estimated negative relationship (↑*nlbs13gdp* → ↓*scbtotgdp*); see relevant Figures A10 and A11 in Appendix A.

The findings reported in Table 3 are in line with the financialization literature, particularly with the FFH, including the incomplete Eurozone architecture and asymmetric governance (among others, Vigny 2022; De Grauwe and Ji 2022; Pierros 2020; Amoutzias 2019; Boyer 2013; Rossi 2013).

5. Conclusions

In this paper, we have investigated the research question "Is the financial fragility hypothesis (FFH) compatible with the long-term or steady-state public social spending (PSS) of an EMU member country during the period 1995–2022?" Based on stylized facts and relevant econometric estimations, we have reported (for Greek data) that we could not disprove it. Methodologically, transfers of PSS are included (ESA 2010) in government consumption, so they have to be treated as part of the general framework of the fiscal policy. The latter require that we analyze the whole economic system, within the limitations of the Eurozone's fragility, and referring to the FFH, as endogenously breeding instability and activating the automatic stabilizers like PSS.

Within this context, we provided interpretations for PSS, which was estimated [with ARDL (p, q)] to be in long-term steady-state relationships with the main factors of the FFH schematically and comprehensively presented in Figure 1. In the estimated models, it is considered that it has been interpreted either in the first stage of the FFH concerning "profitability", with private investments, growth and discretionary fiscal policy according to a predominant hedge–financing structure while PSS decreases; or the second stage, concerning risky options—financial development, openness and indebtedness—that leverage the economy by a prevailing structure of hyper-speculative financing while PSS rises; or even the third stage of the FFH, concerning indicative monetary and fiscal policies reacting to over-leverage and internationalized financial fragility of the EMU by a prevailing system of hyper-speculative or Ponzi financing while PSS also rises.

We contributed the relevant theory by providing the new term of "hyper-speculative financing" and economic interpretations of the behavior of PSS. This became possible in the context of endogenous financial fragility (FFH, part of the financialization literature) of the

Greek economy, a member country of the EMU, which offers useful policy implications for academic research, investors and policy makers. The main domestic lessons are compatible with the FFH and are to do with ensuring hedge–financing structures, while the most important external ones are those we have learned from the reversal of EU policies during the COVID-19 pandemic crisis, both fiscal with the quasi-Eurobond–RRF and monetary with the ECB's PEPP, which were introduced almost unconditionally for member states; this reversal of the European policy helped to overcome that crisis without the panic of the Eurozone sovereign or banking crisis, and especially of Greece during the period of 2010–2015. Our findings are similar to the relevant literature.

The main lesson the paper offers, especially to investors and policy makers[37], is that economic policy is needed since we could not reject the hypothesis that the system affecting the PSS is endogenous, which seems equivalent to Minskyan's famous quote "stability breeds instability".

In other words, economic policy does matter. Our sample economy has been shown to endogenously produce vulnerability when competition is not ensured in all markets or when economic policy is only concerned with nominal values (like discretionary policies in the euphoria period 1995–2008 or austerity applied in the turbulent times of 2010–2015) and not with the convergence or complementarity of real production patterns in the EMU. This latter, i.e., the real convergence in the EU, has urgently required since the 2020 pandemic–economic crisis, as well as the transformations of contemporary capitalisms, especially in Asia (Boyer 2022) and recent geopolitical uncertainties such as wars in Europe and the Middle East. We should correct the incomplete Eurozone by creating a central budget to make fiscal policy effective at a European "united" level, like the monetary one by the ECB; this is justified by the paper's findings, which proved the endogenous nexus of financial fragility—public social spending in (Greece), a paradigmatic member country of the Eurozone.

This paper is limited by its target, focused on EMU member countries and especially on a small open economy at the border of the EU. The authors aim in the near future to replicate this research for countries at the periphery of the EMU, comparing the findings with those of its core.

Appendix A

Table A1. Definition, labels and sources of variables.

Definition	Label (Location)	Sources
Employment in Public Administration, Social Security, Education, and Health as a ratio of the total employment.	PASSEH-TOT (Figure A1)	Quarterly National Accounts, copyright OECD (The Organization for Economic Cooperation and Development) via Refinitiv (LSEG).
Manufacturing employment as a ratio of the total.	MNF-TOT (Figure A1)	Quarterly National Accounts, copyright OECD via Refinitiv. . . . Manufacturing . . .
Financial and Insurance Activities' employment as a ratio of the total.	FI-TOT (Figure A1)	Quarterly National Accounts, copyright OECD via Refinitiv. . . . Financial and Insurance Activities . . .
Real Estate Activities' employment as a ratio of the total.	RE-TOT (Figure A1)	Quarterly National Accounts, copyright OECD via Refinitiv. . . . Real Estate Activities . . .
Financial and Insurance Activities' employment as a ratio of the Manufacturing one.	FI-MFN (Figure A1)	Authors' calculations based on OECD data via Refinitiv.
Gross Operating Surplus (GOS) of the Non-Financial Firms (S.11).	*GOS-S11GDP* (Figure A2 & Model 1)	Hellenic Statistical Authority and Refinitiv/OECD data.
Gross Operating Surplus of the Financial Firms (S.12).	*GOS-S12GDP* (Figure A2 & Model 1)	Hellenic Statistical Authority and Refinitiv/OECD data via Refinitiv.
Ratio of the GOS of the financial to the non-financial sector.	S12/S11 (Figure A2)	Authors' calculations based on Hellenic Statistical Authority and OECD data via Refinitiv.
Gross domestic product, current prices, seasonally adjusted	GDPCUSA (Figure A2)	Hellenic Statistical Authority via Refinitiv.
Total Net Operating Income of Credit Institutions in Greece (=Net Interest Income + Net Fee and Commission Income + Dividend Income + Net Gains on Financial Transactions + Other Income) as Ratios of their Total Assets.	TNOI-TASS (Figure A3)	Authors' calculations based on Bank of Greece data.
Total Net Operating Income of Credit Institutions in Greece as Ratios of their Net Capital.	TNOI-NCAP (Figure A3)	Authors' calculations based on Bank of Greece data.
Index prices of dwellings (historical series), urban areas. 1997 = 100	DWELLPI (Figure A4)	Bank of Greece.
Greece, Capital Markets, Market Capitalization of Listed Domestic Companies (% of Gross Domestic Product)	MRKCAPGDP (Figure A4)	World Bank WDI via Refinitiv.

Table A1. *Cont.*

Definition	Label (Location)	Sources
Standardized form of the ASE general price index based on the share price indices of the Athens Exchange (ASE)	ASE-GI_sz (Figure A5)	Authors' calculations based on Bank of Greece data.
Standardized form of the price of the 10 years Greek government's bond based on the Financial Markets, Greek Government Securities.	PRICEGR10YB_sz (Figure A5)	Authors' calculations based on Bank of Greece data.
Standardized form of the index prices of dwellings based on the residential Property Indices.	DWEPRIND_sz (Figure A5)	Authors' calculations based on Bank of Greece data.
Credit to the Non-Financial Corporations as a ratio of the GDP.	*CDT-NONFGDP* (Figure A6 & Model 2)	Authors' calculations based on Bank of International Settlements via Refinitiv data.
Credit to the Households as a ratio of the GDP.	*CDT-HOUSGDP* (Figure A6 & Model 2)	Authors' calculations based on Bank of International Settlements via Refinitiv data.
Nonperforming loans to total gross loans. Financial Soundness Indices.	NPLTGL (Figure A6)	Authors' calculations based on International Monetary Fund (IMF) via Refinitiv data.
Gross fixed capital formation to GDP ratio.	*GFCF-GDP* (Figure A6 and A8 and Model 1)	Authors' calculations based on Hellenic Statistical Authority data.
Loans to Non-financial corporations as a ratio of total Banking Loans.	NFI-S.11/DL (Figure A7)	Authors' calculations based on Bank of Greece data.
Loans to Financial institutions as a ratio of total Banking Loans.	FI-S.12/DL (Figure A7)	Authors' calculations based on Bank of Greece data.
Loans to the General Government as a ratio of total Banking Loans.	GG-S.13/DL (Figure A7)	Authors' calculations based on Bank of Greece data.
Loans to the Households & NPISH ratio of total Banking Loans.	Oth-S.1M/DL (Figure A7)	Authors' calculations based on Bank of Greece data.
Households and NPISH Debt Outstanding to GDP.	HDEBTOUTSGDP (Figure A8)	Authors' calculations based on European Central Bank via Refinitiv data.
Nonfinancial Corporations Debt Outstanding to GDP.	NFIDEBTOUTSGDP (Figure A8)	Authors' calculations based on European Central Bank via Refinitiv data.
Ratio of the Greek general government bonds held by domestic banks to total issued.	BGRbanks-GRGGB (Figure A9)	Authors' calculations based on Hellenic Statistical Authority and Bank of Greece data.
Ratio of the Greek general government bonds held by domestic banks to their total assets.	BGRbanks-TASSB (Figure A9)	Authors' calculations based on Hellenic Statistical Authority and Bank of Greece data.
Ratio of Net lending (+)/net borrowing (−) (B.9) of the total economy (S.1) to GDP (B.1g)	NLNDBRR/GDP-S.1 (Figure A10)	Authors' calculations based on Hellenic Statistical Authority data.
Ratio of Net lending (+)/net borrowing (−) (B.9) of the nonfinancial corporations (S.11) to GDP (B.1g)	NLNDBRR/GVA-S.11 (Figure A10)	Authors' calculations based on Hellenic Statistical Authority data.
Ratio of Net lending (+)/net borrowing (−) (B.9) of the financial sector (S.12) to GDP (B.1g)	NLNDBRR/GVA-S.12 (Figure A10)	Authors' calculations based on Hellenic Statistical Authority data.

Table A1. *Cont.*

Definition	Label (Location)	Sources
Ratio of Net lending (+)/net borrowing (−) (B.9) of the general government (S.13) to GDP (B.1g)	NLNDBRR/GVA-S.13 (Figure A10)	Authors' calculations based on Hellenic Statistical Authority data.
Ratio of Net lending (+)/net borrowing (−) (B.9) of the households and NPISH (S.1M) to GDP (B.1g)	NLNDBRR/GVA-S.1M (Figure A10)	Authors' calculations based on Hellenic Statistical Authority data.
Spread between the 10 years government bond yield (r) and the growth rate of nominal GDP (g).	*R-G_SNOWB* (Figure A11 & Model 3)	Authors' calculations based on Main Economic Indicators, copyright OECD via Refinitiv data.
Interest rate margin (average interest rates on loans—deposits) of domestic banks.	interstRmrg (Figure A11)	Authors' calculations based on Bank of Greece data.
Greece, Portfolio investments in domestic stocks from non-residents (liabilities)	INFL-STOCKS (Figure A12)	Authors' calculations based on Bank of Greece data.
Greece, Portfolio investments in domestic bonds and treasury bills from non-residents (liabilities)	INFL-BONDSTB (Figure A12)	Authors' calculations based on Bank of Greece data.
Greece, Portfolio investments in foreign stocks from residents (requirements)	OUTFL-STOCKS (Figure A12)	Authors' calculations based on Bank of Greece data.
Greece, Portfolio investments in foreign bonds and treasury bills from residents (requirements)	OUTFL-BONDSTB (Figure A12)	Authors' calculations based on Bank of Greece data.
Greece, Balance of Payments, Balance of financial transactions, portfolio investments.	BPFT-PRTFINV (Figure A12)	Authors' calculations based on Bank of Greece data.
Greece, Balance of Payments, International investment position.	BP-IIP (Figure A12)	Authors' calculations based on Bank of Greece data.
Greece, Gross Domestic Product, Market Prices, Annualized Rate, Constant Prices, AR, 2010 Prices	*realgdp10yy* (model 1, 2)	Quarterly National Accounts, copyright OECD via Refinitiv.
Discretionary policy of the expenditures of the general government (Fatás and Mihov 2003).	*discrpol* (model 1, 2 & 3)	Authors' calculations based on OECD via Refinitiv and Hellenic Statistical Authority data.
Balance of current account as a ratio of GDP.	*cagdp* (model 2)	OECD Economic Outlook.
Greece, Sector Accounts, Other, Nonfinancial Corporations, Debt Outstanding to Gross Domestic Product.	*s11debtgdp* (model 2)	ECB (the European Central Bank) via Refinitiv.
The growth rate of the Greek, Public Debt, General Government, Long-Term, Total, Current Prices, not seas. adj., Euro.	*pudbtgr* (model 2)	World Bank QPSD via Refinitiv.
The ratio of Long Term (LT) Loans of residents (S1) from non-residents (S2) to Total LT loans of residents (S2/F42).	s1ltloanss2tot (model 2)	Authors' calculations based on Bank of Greece data.
Ratio of the net portfolio investments inflows to GDP. Balance of financial transactions, Balance of payments (BPM6).	*nprtfinflowsgdp* (model 2)	Authors' calculations based on Bank of Greece data.
Greece, Prices of Dwellings, Urban Areas (nsa., 1997 = 100).	*hpiyoy* (model 2)	Oxford Economics via Refinitiv.

Table A1. *Cont.*

Definition	Label (Location)	Sources
Athens stock exchange general stock price index (year-over-year).	*asegspiyoy* (model 2)	Athens Exchange.
Greece, Long-Term Government Bond Yields, 10-Year, Main (Including Benchmark), Yield 10-Year Government Bonds.	*gr10ygby* (model 2)	OECD Main Economic Indicators via Refinitiv.
The ratio for Greece, Money Supply M3 Outstanding Amounts, Mill. Euro to GDP.	*m3outsgdp* (model 3)	Authors' calculations based on Bank of Greece data. OECD via Refinitiv for the GDP.
The ratio (for Greece) of net lending (+) or borrowing (−) of the general government (S13) to GDP. Quarterly non-financial accounts of institutional sectors.	*nlbs13gdp* (model 3)	Authors' calculations based on Hellenic Statistical Authority data.
Social Contributions and Benefits Paid (ratio to GDP).	*scbtotgdp* (the dependent var. in all models)	Greece, Total Transactions (ESA 2010), Social Contributions and Benefits: Paid, Current Prices, Euro. Refinitiv/Datastream/Eurostat.
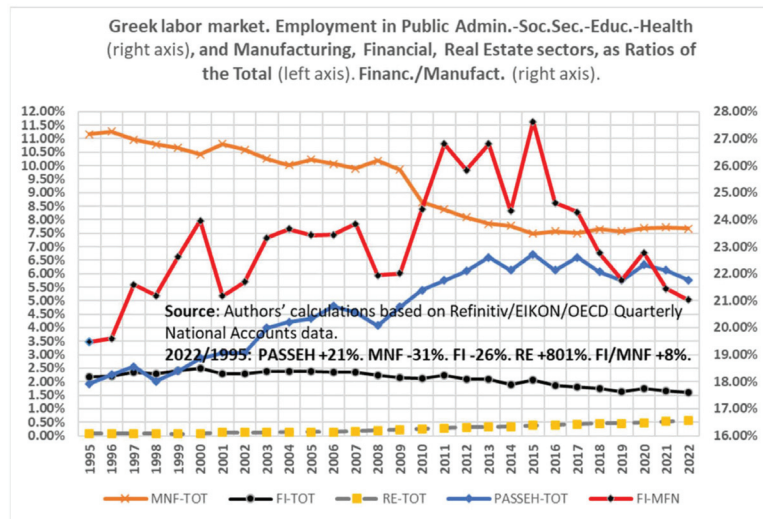
Figure A1. Huge changes in the labor market especially since the GFC-2008. PASSEH-TOT = employment in public administration, social security, education, and health as a ratio of the total employment; MFN (FI) [RE]-TOT = manufacturing (financial and insurance activities) [real estate activities] employment as a ratio of the total; FI-MFN = financial and Insurance Activities' employment as a ratio of the manufacturing one.
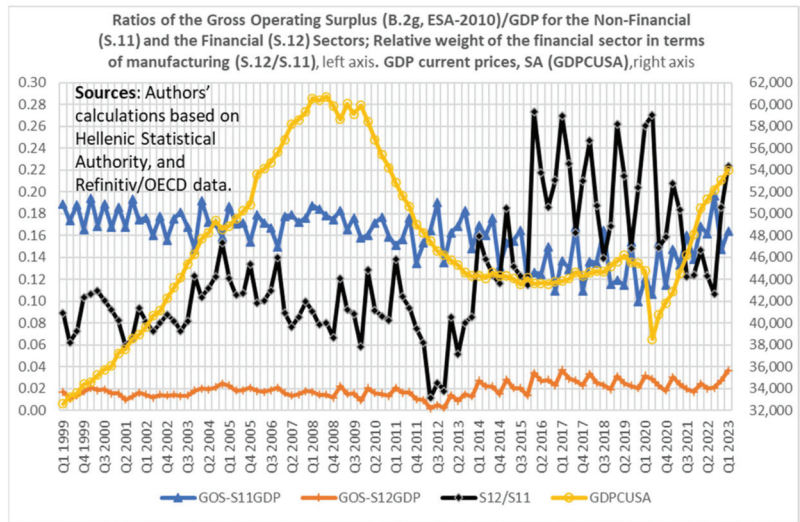
Figure A2. Profitability of the financial sector and manufacturing, as well as current GDP (sa). GOS-S11GDP = gross operating surplus (GOS) of non-financial firms (S.11) to GDP ratio; GOS-S12GDP = gross operating surplus (GOS) of financial firms (S.12) to GDP ratio; S12/S11 = ratio of the GOS of the financial to the non-financial sector; GDPCUSA = gross domestic product, current prices, seasonally adjusted.
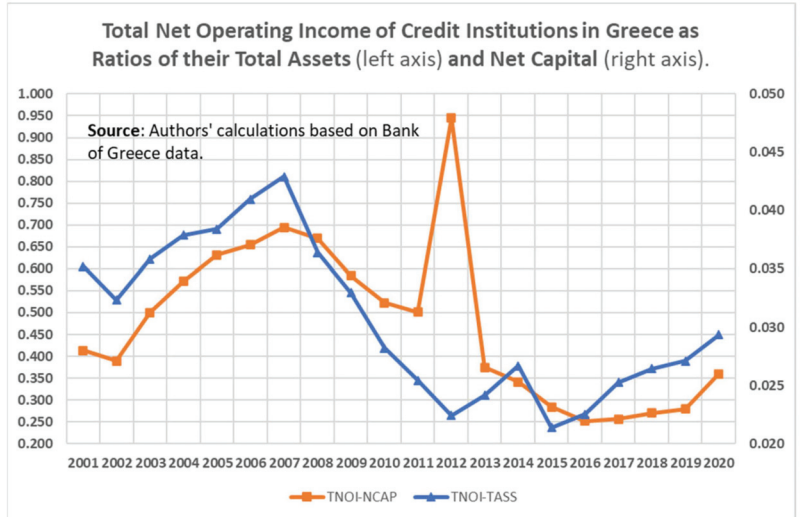
Figure A3. Profitability of credit Institutions. TNOI-TASS = total net operating income of credit institutions in Greece (=net interest income + net fee and commission income + dividend income + net gains on financial transactions + other income) as ratio of their total assets; TNOI-NCAP = Total net operating income of credit institutions in Greece as ratios of their net capital.
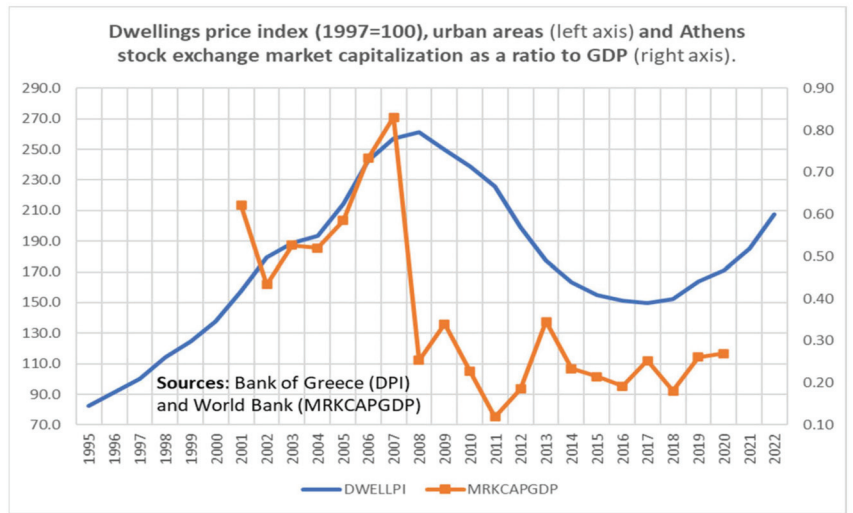
Figure A4. Asset prices and relative values. DWELLPI = index prices of dwellings (historical series), urban areas, 1997 = 100; MRKCAPGDP = Greece, capital markets, market capitalization of listed domestic companies (% of gross domestic product).
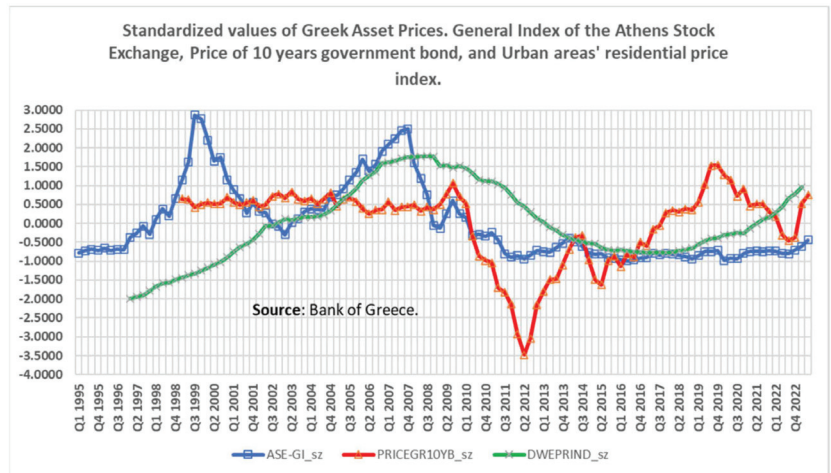
Figure A5. Volatility of asset prices. ASE-GI_sz = standardized form of the ASE general price index based on the share price indices of the Athens Exchange (ASE); PRICEGR10YB_sz = standardized form of the price of 10-year Greek government bonds based on the financial markets, Greek government securities; DWEPRIND_sz = standardized form of the index prices of dwellings based on the residential property indices.
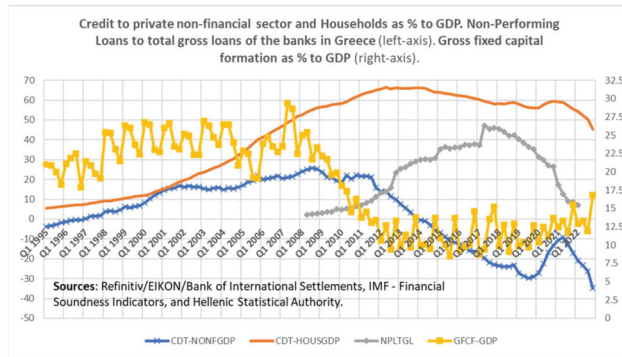
Figure A6. Credits, NPL and private investments to GDP. CDT-NONFGDP = credit to non-financial corporations as a ratio of the GDP; CDT-HOUSGDP = credit to households as a ratio of the GDP; NPLTGL = nonperforming loans to total gross loans, financial soundness indices; GFCF-GDP = gross fixed capital formation (private fixed investments) to GDP ratio.
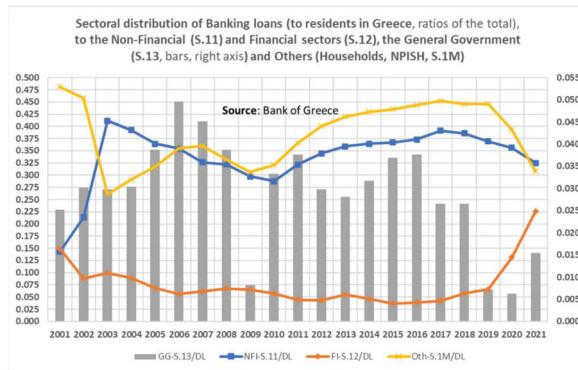
Figure A7. Distribution of banking loans. NFI-S.11/DL = loans to non-financial corporations as a ratio of total banking loans; FI-S.12/DL = loans to financial institutions as a ratio of total banking loans; GG-S.13/DL = loans to the general government as a ratio of total banking loans; Oth-S.1M/DL = loans to households and NPISH as a ratio of total banking loans.
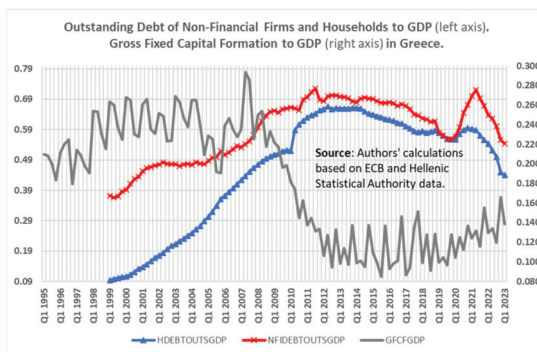
Figure A8. Debts and investments. HDEBTOUTSGDP = households and NPISH debt outstanding to GDP; NFIDEBTOUTSGDP = nonfinancial corporations debt outstanding to GDP; GFCFGDP = gross fixed capital formation (private fixed investments) to GDP ratio.
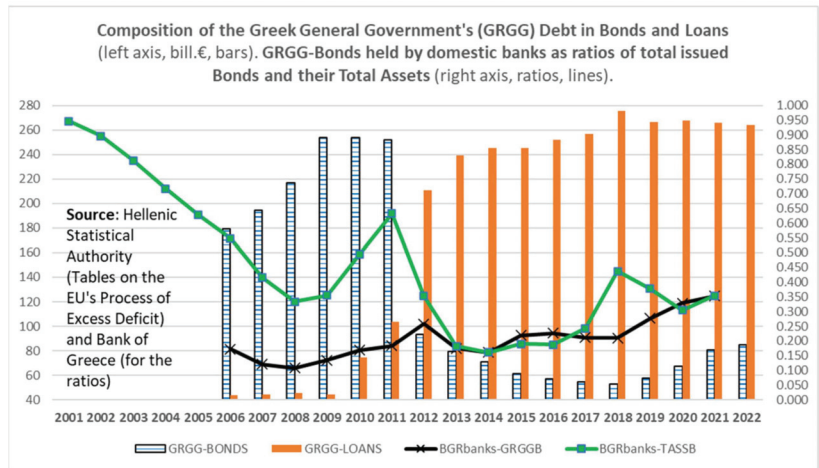
Figure A9. Government's bonds and loans, as well as domestic banks' holding. BGRbank-GRGGB (TASSB) = ratio of the Greek general government bonds held by domestic banks to total issued (to their total assets); GRGG-BONDS = Greek bonds held by domestic banks; GRGG-LOANS = Greek loans of the general government's debt (end of periods, current prices, mill. €).
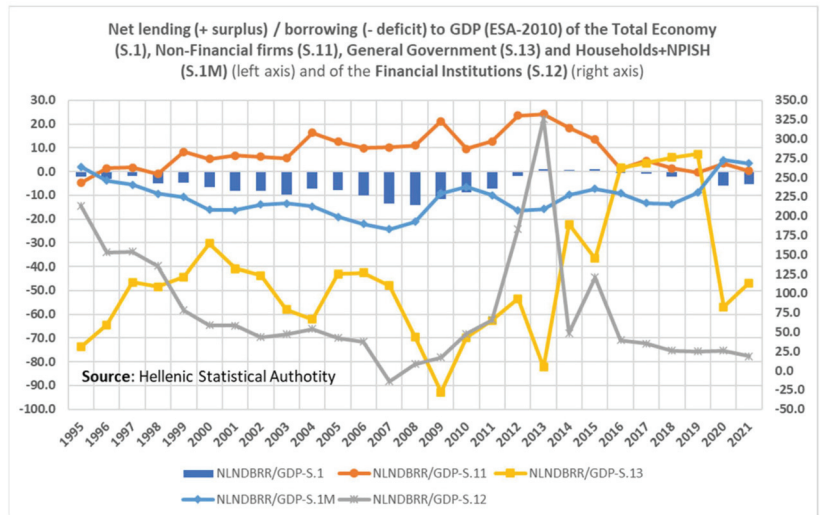
Figure A10. Overall performance of institutional sectors. Where: NLNDBRR/GDP-S.1 = ratio of net lending (+)/net borrowing (-) (B.9) of the total economy (S.1) to GDP (B.1g); NLNDBRR/GDP-S.11(S.12) [S.13] {S.1M} = ratio of net lending (+)/net borrowing (-) (B.9) of non-financial corporations (financial institutions) [general government] {households and NPISH} to GDP (B.1g).
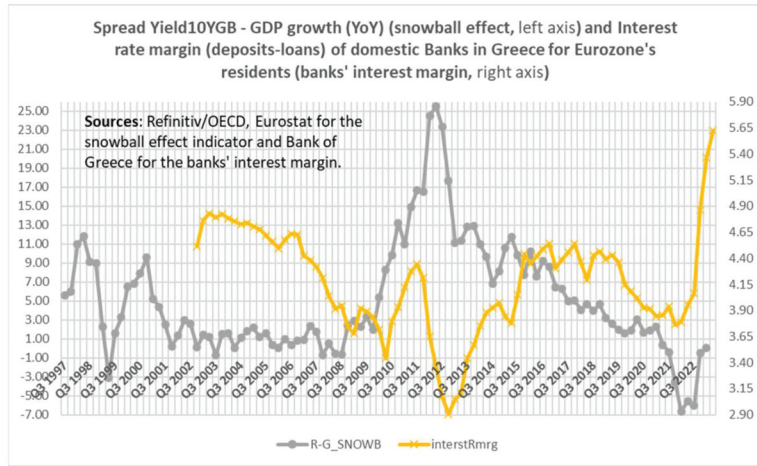
Figure A11. Snowball effect and interest rate margin of domestic banks. R-G_SNOWB = spread between the 10 years government bond yield (r) and the growth rate of nominal GDP (g); InterstRmrg = interest rate margin (average interest rates on loans—deposits) of domestic banks.
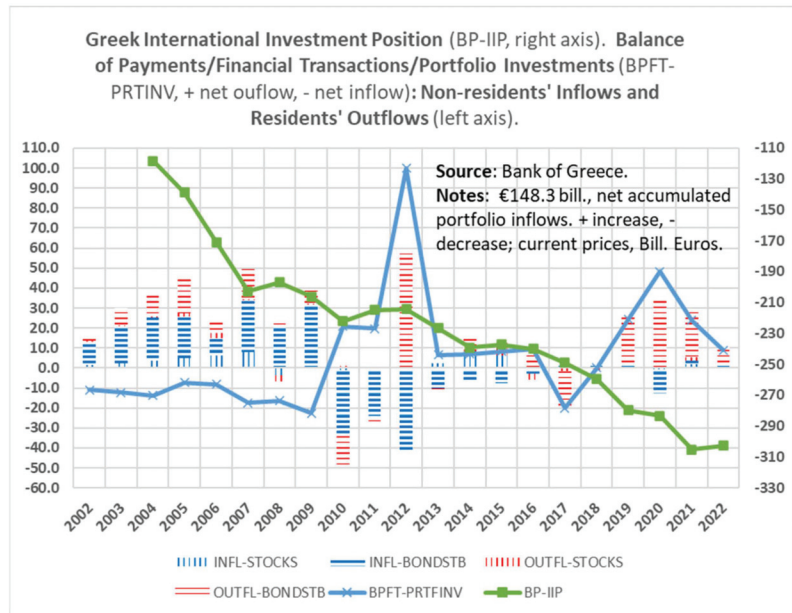
Figure A12. Obvious domestic structural deficiencies and double financial fragility. INFL-STOCKS = Greece, portfolio investments in domestic stocks from non-residents (liabilities); INFL-BONDSTB = Greece, portfolio investments in domestic bonds and treasury bills from non-residents (liabilities); OUTFL-STOCKS = Greece, portfolio investments in foreign stocks from residents (requirements); OUTFL-BONDSTB = Greece, portfolio investments in foreign bonds and treasury bills from residents (requirements); BPFT-PRTFINV = Greece, balance of payments, balance of financial transactions, portfolio investments; BP-IIP = Greece, balance of payments, international investment position.
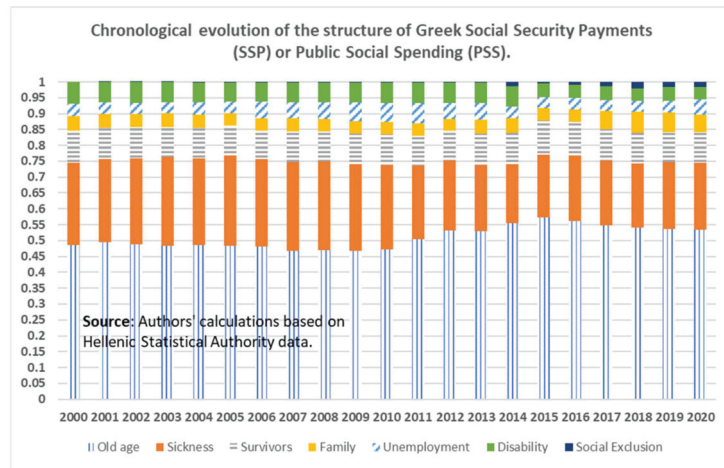
Figure A13. The anelastic composition of the PSS.

Notes

[1] Strictly speaking, Greece, along with Italy, Spain, Portugal and Ireland, is considered to belong to the South-West Euro Area Periphery (SWEAP) as opposed to the core member countries of the Eurozone (Aizenman et al. 2013).

[2] European System of Integrated Social Protection Statistics (ESSPROS): a nominal convergence of Greece towards EMU is reported in aggregated benefits and grouped schemes—in % of GDP; these total social transfers were, on average (stdev), 19.3% (1.4%) for Greece during 1995–2008, and for the Euro area (19 countries), 25.1% (0.4%) during 2000–2008; for Greece, 25.9% (1.0%) and for the Euro area (19 countries), 27.9% (0.3%) during 2009–2019. https://ec.europa.eu/eurostat/databrowser/view/spr_exp_gdp_ _custom_9184255/default/line?lang=en accessed on 17 January 2024.

[3] So, FFH ≡ eFIH + EZFH.

[4] Memorandum of understanding.

[5] International Monetary Fund (IMF), European Commission (EC) and European Central Bank (ECB).

[6] Pandemic emergency purchasing program of the European Central Bank (ECB).

[7] Outright monetary transactions of the ECB.

[8] Recovery and resilience fund.

[9] In their sample of 21 OECD countries during the period 1982–2003, Luxembourg and Greece are the missing EU members.

[10] Only the underlined categories of benefits are offered in Greece.

[11] Resulting from wrong economic policies adopted by the European authorities, such as the OMT (outright monetary transactions) program announced by the ECB but under the condition of austerity rules imposed by the ESM (European Stability Mechanism). This was the case of the SWEAP countries.

[12] Next-generation European Union plan.

[13] Based on the Hellenic Statistical Authority's (HAS's) data and communications (https://www.statistics.gr/en/statistics/-/ publication/SHE24/- accessed on 17 January 2024). The hierarchical presentation of these categories of benefits is based on the realized average payments in 2020, which is referred by the HSA as the last year of available annual data 2000–2020.

[14] The latest available annual HSA data (Hellenic Statistical Authority) show that, on average (standard deviation), the share of pensions (age plus survivors), health and disability–incapacity in public social spending was 91% (2.1%) during the period 2000–2020; the rest was either unemployment or family benefits, with 4% (0.9%) each.

[15] The balanced fiscal budget should, in the recession of the crisis period, be in surplus so as to find ways of repaying the sovereign debt. Harsh austerity–deflation policy. . .

[16] Proxy of the "Financial development" (Choi et al. 2017).

[17] https://www.statistics.gr/.

[18] https://www.bankofgreece.gr/en/homepage.

[19] *"The first account in the sequence is the production account, which records the output and inputs of the production process, leaving value added as the balancing item. The value added is taken forward to the next account which is the generation of income account. Here the compensation of employees in the production process is recorded, as well as taxes due to government because of the production, so that the operating surplus (or mixed income from the self-employed of the households' sector) can be derived as the balancing item for each sector."*

(ESA 2010, p. 53). That is, the gross operating surplus (GOS) used here as a measure of the profitability of sectors does not include any financial expenses or receipts (interest or principal of a debt).

[20] Balance of payments—international investment position.

[21] The term 'net lending/net borrowing' is a sort of terminological shortcut. When the variable is positive (meaning that it shows a financing capacity), it should be called net lending (+); when it is negative (meaning that it shows a borrowing need), it should be called net borrowing (−). (ESA 2010, p. 466).

[22] The net lending (+) or borrowing (−) of the total economy is the sum of the net lending or borrowing of the institutional sectors. It represents the net resources that the total economy makes available to the rest of the world (if it is positive) or receives from the rest of the world (if it is negative). The net lending (+) or borrowing (−) of the total economy is equal but of opposite sign to the net borrowing (−) or lending (+) of the rest of the world. (ESA 2010, p. 306).

[23] It is not possible from macroeconomic data to "prove" whether a firm's or household's loan is being repaid "regularly" (hedge financing structure), or in a Eurozone country, or if only interest is being paid by firms (hyper-speculative) or not serviced at all (Ponzi). However, when the outstanding debts of the non-financial sector increase, while at the same time, it is excluded from bank credit, NPLs proliferate or, finally, the government defaults or the international investment position of the country continuously deteriorates, the evidence is sufficiently strong to characterize the sub-periods of the sample according to the Minsky classification.

[24] Non-profit institution serving households (see, ESA 2010).

[25] Money supply measure.

[26] r = weighted average of the interest rate of the government's outstanding debt; g = real gross domestic product (GDP) growth rate.

[27] STATA/SE version 17 software has been used, provided by the University of West Attica academic license.

[28] When we alternatively used the variable "*nprtfinflowsgdp*" (also drawn from the Bank of Greece dataset and balance of financial account), the estimated coefficient also statistically significant at 5% level this time had a negative sign, −0.058.

[29] We emphasize that this is performed in a long-term equilibrium, or the series are cointegrated.

[30] Which (old age + sickness + survivors) aggregate to an average of almost 85% of the total PSS (see, Figure A13 in Appendix A).

[31] That have taken into account in the determinants of Equation (3).

[32] This characterizes almost permanently the lack of international competitiveness and the respective trade balance deficit of the country.

[33] Mergers and acquisitions.

[34] The majority of more than 80% of the Greek production sector.

[35] It should be stressed that the austerity–deflationary policies that were implemented (which reduced real Greek GDP by 25% and increased unemployment to 27% in the period 2008–2015) not only reduced aggregate demand but also aggregate supply due to the restructuring of sectors and markets brought about by the successive crises.

[36] Credit rating agencies.

[37] It is a matter of special paper(s) to reconsider democratic decision making (including economic policy) in Europe... Greece marked human history and civilization, both with "polis («πόλις»)-democracy" and with "critical thinking" who lived in ancient Greek «πόλεις» or Eastern Romanian Empire or even under Ottoman Empire and occupation, at least until 1832. So, it worked historically. Few words can be reported: the first one, (πόλις) democracy, means another way of collectiveness institutionally and functionally homologous to "truth" (i.e., that which does not wear out does not die), while the criterion in the second (critical thought), i.e., how can one distinguish right–wrong, truth–false, valid knowledge–illusion, is the "common experience" of all citizens, and not principles of utilitarian efficiency that we economists use, insisting only on the "society of needs" and not, as people worldwide demand, also the "society of human relations". (Yannaras 2011).

References

Aizenman, Joshua, Michael Hutchison, and Yothin Jinjarak. 2013. What is the Risk of European Sovereign Debt Defaults? Fiscal Space, CDS Spread and Market Pricing of Risk. *Journal of International Money and Finance* 34: 37–59. [CrossRef]

Alary, Pierre, and Elsa Lafaye de Micheaux. 2015. *Capitalismes Asiatiques et Puissance Chinoise: Diversité et Recomposition des Trajectoires Nationales*, 1st ed. Paris: Sciences Po. Les Presses. ISBN 2724616200.

Amoutzias, Chrysovalantis. 2019. An investigation of the effects of income inequality on financial fragility: Evidence from Organization for Economic Co-operation and Development countries. *International Journal of Finance & Economics* 24: 241–59.

Arestis, Philip, and Murray Glickman. 2002. Financial crisis in Southeast Asia: Dispelling illusion the Minskyan way. *Cambridge Journal of Economics* 26: 237–60. [CrossRef]

Besedovsky, Natalia. 2018. Financialization as calculative practice: The rise of structured finance and the cultural and calculative transformation of credit rating agencies. *Socio-Economic Review* 16: 61–84. [CrossRef]

Blanchard, Olivier, Alessia Amighini, and Francesco Giavazzi. 2021. *Macroeconomics: A European Perspective*, 4th ed. London: Pearson. Available online: https://www.pearson.com/en-gb/subject-catalog/p/macroeconomics-a-european-perspective/P200000003 747?view=educator (accessed on 12 December 2023).

Boyer, Robert. 2000. Is a Finance-led Growth Regime a Viable Alternative to Fordism? A Preliminary Analysis. *Economy and Society* 29: 111–45. [CrossRef]

Boyer, Robert. 2013. Origins and ways out of the euro crisis: Supranational institution building in the era of global finance. *Contributions to Political Economy* 32: 97–126. [CrossRef]

Boyer, Robert. 2022. The Transformations of Contemporary Capitalisms: Seven Lessons. In *Rethinking Asian Capitalism: The Achievements and Challenges of Vietnam Under Doi Moi*. Edited by Thi Anh-Dao Tran. London: Palgrave Macmillan, pp. 25–58.

Boyer, Robert, and Yves Saillard, eds. 2002. *Régulation Theory: The State of the Art*. London: Routledge.

Boyer, Robert, Hiroyasu Uemura, and Akinori Isogai, eds. 2011. *Diversity and Transformations of Asian Capitalisms*. London: Routledge. [CrossRef]

Choi, Sangyup, Davide Furceri, and João Tovar Jalles. 2017. *Fiscal Stabilization and Growth: Evidence from Industry-Level Data for Advanced and Developing Economies*. IMF Working Paper No. 2017/198. Washington: International Monetary Fund.

Christophers, Brett, and Ben Fine. 2020. The value of financialization and the financialization of value. In *The Routledge International Handbook of Financialization*. Edited by Philip Mader, Daniel Mertens and Natascha van der Zwan. New York: Taylor & Francis, pp. 19–30. Available online: https://bookshelf.vitalsource.com/books/9781351390361 (accessed on 15 June 2023).

Cibils, Alan, and Cecilia Allami. 2013. Financialisation vs. Development Finance: The Case of the Post-Crisis Argentine Banking System. *Revue de la Régulation* 13. [CrossRef]

Darby, Julia, and Jacques Melitz. 2008. Social spending and automatic stabilizers in the OECD. *Economic Policy* 23: 715–56. [CrossRef]

De Grauwe, Paul. 2011. *The European Central Bank: Lender of Last Resort in the Government Bond Markets?* Working Paper No. 3569. Munich: Center for Economic Studies, Ludwig-Maximilians-Universität, and Ifo Institute (CESifo). Available online: https://www.cesifo.org/en/publications/2011/working-paper/european-central-bank-lender-last-resort-government-bond-markets (accessed on 15 June 2023).

De Grauwe, Paul. 2012. The governance of a fragile Eurozone. *The Australian Economic Review* 45: 255–68. [CrossRef]

De Grauwe, Paul. 2013. *Design Failures in the Eurozone: Can They Be Fixed?* LSE Europe in Question Discussion Paper Series No. 57; Available online: https://ec.europa.eu/economy_finance/publications/economic_paper/2013/ecp491_en.htm (accessed on 15 June 2023).

De Grauwe, Paul, and Yuemei Ji. 2022. The fragility of the Eurozone: Has it disappeared? *Journal of International Money and Finance* 120: 102546. [CrossRef]

Dow, Sheila. 2020. Financialization, money and the state. In *The Routledge International Handbook of Financialization*. Edited by Philip Mader, Daniel Mertens and Natascha van der Zwan. New York: Taylor & Francis, pp. 56–67. Available online: https://bookshelf.vitalsource.com/books/9781351390361 (accessed on 15 June 2023).

Epstein, Gerald A. 2005. *Financialization and the World Economy*. Northampton: Edward Elgar. ISBN 978 1 84376 874 6.

European System of Accounts. 2010. Available online: https://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/ks-02-13-269 (accessed on 8 May 2024).

Fatás, Antonio, and Ilian Mihov. 2003. The case for restricting fiscal policy discretion. *The Quarterly Journal of Economics* 118: 1419–47. [CrossRef]

Goel, Rajeev K., James E. Payne, and Rati Ram. 2008. R&D expenditures and U.S. economic growth: A disaggregated approach. *Journal of Policy Modeling* 30: 237–50. [CrossRef]

Hajilee, Massomeh, Donna Y. Stringer, and Linda A. Hayes. 2021. On the link between the shadow economy and stock market progress: An asymmetry analysis. *The Quarterly Review of Economics and Finance* 80: 303–16. [CrossRef]

Hepburn, Brian, and Hanne Andersen. 2021. Scientific Method. In *The Stanford Encyclopedia of Philosophy*, Summer 2021 ed. Edited by Edward N. Zalta. Available online: https://plato.stanford.edu/archives/sum2021/entries/scientific-method/ (accessed on 1 May 2024).

Holloway, Matthew, and Jari Eloranta. 2014. "Stability breeds instability?" A Minskian analysis of the Asian crisis of the Tigers in the 1990s. *Investigaciones de Historia Económica—Economic History Research* 10: 115–26. [CrossRef]

Krippner, Greta R. 2005. The financialization of the American economy. *Socio-Economic Review* 3: 173–208. [CrossRef]

Kuhn, Thomas S. 1962. *The Structure of Scientific Revolutions*. Chicago: University of Chicago Press.

Kyriakopoulos, Dionysios, John Yfantopoulos, and Theodoros V. Stamatopoulos. 2022. Social Security Payments and Financialization: Lessons from the Greek Case. *Journal of Risk and Financial Management* 5: 615. [CrossRef]

Lapavitsas, Costas. 2009. Financialised Capitalism: Crisis and Financial Expropriation. *Historical Materialism* 17: 114–48. [CrossRef]

Mader, Philip, Daniel Mertens, and Natascha van der Zwan. 2020. *The Routledge International Handbook of Financialization*. New York: Taylor & Francis. Available online: https://bookshelf.vitalsource.com/books/9781351390361 (accessed on 15 June 2023).

Minsky, Hyman P. 1957. Central banking and money market changes. *Quarterly Journal of Economics* 71: 171–87. [CrossRef]

Minsky, Hyman P. 1975. *John Maynard Keynes*. New York: Columbia University Press.

Minsky, Hyman P. 1977. A theory of systemic fragility. In *Financial Crises: Institutions and Markets in a Fragile Environment*. Edited by Edward I. Altman and Arnold W. Sametz. New York:: John Wiley & Sons, pp. 138–52. Available online: https://digitalcommons.bard.edu/hm_archive/231/ (accessed on 10 May 2023).

Minsky, Hyman P. 1978. *The Financial Instability Hypothesis: A Restatement*. Annandale-on-Hudson: Levy Economics Institute of Bard College. Available online: https://digitalcommons.bard.edu/hm_archive/180/ (accessed on 10 May 2023).

Minsky, Hyman P. 1982. *Can "It" Happen Again? Essays in Instability and Finance*. Reprint by Routledge 11 April 2016. Armonk: M. E. Sharpe. ISBN 9781138641952. Available online: https://www.routledge.com/Can-It-Happen-Again-Essays-on-Instability-and-Finance/Minsky/p/book/9781138641952 (accessed on 10 May 2023).

Minsky, Hyman P. 1983. The financial instability hypothesis: An interpretation of Keynes and an alternative to "standard" theory. In *John Maynard Keynes: Critical Assessments*. Edited by John Cunningham Wood. London: Macmillan, pp. 282–92.

Minsky, Hyman P. 1986. *Stabilizing an Unstable Economy*. Annandale-on-Hudson: Levy Economics Institute of Bard College. Available online: https://digitalcommons.bard.edu/hm_archive/144/ (accessed on 10 May 2023).

Minsky, Hyman P. 1992a. *The Financial Instability Hypothesis*. Working Paper No. 74. Annandale-on-Hudson: Levy Economics Institute of Bard College.

Minsky, Hyman P. 1992b. Hyman P. Minsky. In *A Bibliographical Dictionary of Dissenting Economists*. Edited by Philip Arestis and Malcolm C. Sawer. Aldershot: Edward Elgar. Available online: https://www.e-elgar.com/shop/gbp/a-biographical-dictionary-of-dissenting-economists-second-edition-9781858985602.html (accessed on 10 May 2023).

Minsky, Hyman P. 1995. Financial factors in the economics of capitalism. *Journal of Financial Services Research* 9: 197–208. [CrossRef]

Minsky, Hyman P. 1996. Uncertainty and the Institutional Structure of Capitalist Economies. *Journal of Economic Issues* 30: 357–68. [CrossRef]

Palley, Thomas I. 2007. *Financialisation: What It Is and Why It Matters*. Working Paper No. 525. Annandale-on-Hudson: The Levy Economics Institute of Bard College.

Pesaran, M. Hashem, and Yongcheol Shin. 1999. An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis. In *Econometrics and Economic Theory in the 20th Century: The Ragnar Frisch Centennial Symposium*. Edited by Steinar Strøm. Cambridge: Cambridge University Press, pp. 371–413. [CrossRef]

Pesaran, M. Hashem, Yongcheol Shin, and Richard J. Smith. 2001. Bounds testing approaches to the analysis of level relationships. *Journal of Applied Economics* 16: 289–326. [CrossRef]

Pierros, Christos. 2020. Income distribution, structural competitiveness and financial fragility of the Greek economy. *International Review of Applied Economics* 34: 50–74. [CrossRef]

Popper, Karl. 2002. *The Logic of Scientific Discovery*. New York: Routledge. ISBN 9780415278447.

Rossi, Sergio. 2013. Financialisation and monetary union in Europe: The monetary–structural causes of the euro-area crisis. *Cambridge Journal of Regions, Economy and Society* 6: 381–400. [CrossRef]

Sen, Sunanda. 2020. Financialization, speculation and instability. In *The Routledge International Handbook of Financialization*. Edited by Philip Mader, Daniel Mertens and Natascha van der Zwan. New York: Taylor & Francis, pp. 448–57. Available online: https://bookshelf.vitalsource.com/books/9781351390361 (accessed on 15 June 2023).

Sotiropoulos, Dimitris P., and Ariane Hillig. 2020. Financialization in heterodox economics. In *The Routledge International Handbook of Financialization*. Edited by Philip Mader, Daniel Mertens and Natascha van der Zwan. New York: Taylor & Francis, pp. 125–35. Available online: https://bookshelf.vitalsource.com/books/9781351390361 (accessed on 15 June 2023).

van der Zwan, Natascha. 2014. Making sense of financialization. *Socio-Economic Review* 1: 99–129. [CrossRef]

Vigny, Léo. 2022. The Greek Sovereign Crisis: A Post-Keynesian Synthesis. *International Journal of Political Economy* 51: 151–69. [CrossRef]

Wade, Robert. 1990. *Governing the Market: Economic Theory and the Role of Government in East Asian Industrialization*. Princeton: Princeton University Press. [CrossRef]

Yannaras, Cristos. 2011. *Six Philosophical Paintings*. Athens: IKAROS Publishers. ISBN 13 9789609527248.

Article

Advanced Statistical Analysis of the Predicted Volatility Levels in Crypto Markets

Vadim Azhmyakov [1,*,†], Ilya Shirokov [1,†] and Luz Adriana Guzman Trujillo [2,†]

1   1ex Corporation, Dubai P.O. Box 9305, United Arab Emirates; iyushirokov@1ex.com
2   LARIS, Université d'Angers, 49000 Angers, France; luzadriguz@gmail.com
*   Correspondence: vadim.azhmiakov@1ex.com
†   These authors contributed equally to this work.

Abstract: Our paper deals with an advanced statistical tool for the volatility prediction problem in financial (crypto) markets. First, we consider the conventional GARCH-based volatility models. Next, we extend the corresponding GARCH-based forecasting and calculate a specific probability associated with the predicted volatility levels. As the probability evaluation is based on a stochastic model, we develop an advanced data-driven estimation of this probability. The novel statistical estimation we propose uses real market data. The obtained analytical results for the statistical probability of the levels are also discussed in the framework of the integrated volatility concept. The possible application of the established probability estimation approach to the volatility clustering problem is also mentioned. Our paper includes a concrete implementation of the proposed volatility prediction tool and considers a novel trading and volatility estimation module for crypto markets recently developed by the 1ex Trading Board group in collaboration with GoldenGate Venture. We also briefly discuss the possible application of a model combined with the data-driven volatility prediction methodology to financial risk management.

Keywords: technical analysis; formal volatility models; volatility prediction; statistical probability of levels; trading algorithms

1. Introduction and Motivation

In financial engineering, volatility is usually defined as the dispersion of a return series and is computed by taking the (sample) standard deviation (see, e.g., Brooks 2015; Greene 2011; Poon and Granger 2003). Volatility is the most important parameter in the pricing of crypto derivatives, and the trading volume has drastically increased in recent years. To price an option, one needs to know (or estimate) the volatility of the underlying asset from the real-time instant until the option expires. Note that the probability distributions of financial returns are characterized by high volatility persistence and thick tails (see, e.g., Andersen et al. 2001; Cont 2001). In the realm of modern crypto markets, the manifestation of volatility constitutes a very important indicator of the inherent fluctuations in the main market characteristics (Danielsson et al. 2018; Poon and Granger 2003; Schwert 1990). It gives crucial information to crypto traders and constitutes a pivotal element of many effective cryptocurrency trading strategies. Volatility information is very important for assessing risk and pricing derivative products, as well as for developing trading strategies.

The volatility prediction problem is an important topic in the technical analysis of financial markets (see Bollerslev 1986; Bollerslev and Wooldridge 1992; Haas and Peter 2024; Poon and Granger 2003; Shah et al. 2018; Verhoeven et al. 2002; Wang et al. 2020, and the references therein). It is common knowledge that various time-series models are widely used for handling the data of highly volatile financial markets. For example, the relatively simple generalized autoregressive conditional heteroskedasticity (GARCH) models can effectively be applied to the volatility forecasting problem (see Bollerslev 1986;

Bollerslev and Wooldridge 1992; Francq and Zakoian 2010; Huang et al. 2008; Sen et al. 2021; Taylor 1986; Wang et al. 2020). On the other hand, the majority of the existing time-series techniques for volatility prediction involve the celebrated quasi-maximum likelihood (QML) estimation approach (Birge and Louveaux 2011; Bollerslev and Wooldridge 1992; Franses and Ghijsels 1999; Gallager 2013; Poznyak 2009). One can use this conventional estimation methodology for necessary parameter identification in a selected GARCH(p,q) model. We refer to Azhmyakov et al. (2021), Lewis (1986), and Poznyak (2009) for the general identification theory.

However, the conventional QML estimation approach mentioned above was initially designed under the classic normality assumption (Franses and Ghijsels 1999; Poon and Granger 2003; Poznyak 2009; Taylor 1986). Thus, it is inefficient if the volatility proxies are non-Gaussian (see Fan et al. 2014). Otherwise, there are several specific features of financial market volatility that are well-documented. These "stylized facts" include the fat-tailed distributions of asset returns, volatility clustering, asymmetry, and mean reversion. It is well known that the QML method is non-robust in the presence of outliers, even with fat-tailed and skewed distributions. Therefore, the classic (Gaussian) QML method for the GARCH parameter identification problem needs to be improved and extended with some additional analytical tools.

Although the development of mathematical models for crypto markets and new trading algorithms has been a major topic of research, forecasting financial market volatility is more difficult. Surprisingly, although high volatility can pose a considerable menace to shareholders, it can also be a source of significant financial returns. Even when stock markets oscillate, fall, or skyrocket, there is always a possibility to profit if market volatility is exploited.

In this paper, we extend the existing GARCH volatility prediction technique using an additional useful tool. We introduce predicted volatility levels and calculate the probability that financial market volatility will not fall below these (predicted) levels. This formal mathematical technique involves the well-known stochastic volatility model (see, e.g., Poon and Granger 2003). Additionally, we use real market data and develop a constructive lower estimation for the probability evaluation mentioned above. This data-driven estimation of the probability of volatility levels integrates the model parameters and real market data. The advanced statistical analysis we propose can be considered for integrated volatility frameworks. It can also be useful in the context of the volatility clustering problem (Lux and Marchesi 2000; Nikolova et al. 2020). Our paper includes a short presentation of a practically oriented volatility analysis module recently developed by the 1ex Trading Board group in collaboration with GoldenGate Venture. This novel tool includes the volatility prediction methodology discussed in this paper and some related trading algorithms.

Let us also note that robust volatility forecasting plays an important role in financial risk management (Kahneman and Tversky 2013; Schwert 1990; Ziemba and Vickson 1975). Various hedge funds, banks, financial groups, and trading houses use the well-known value-at-risk (VaR) indicators. Modern VaR estimators essentially use some volatility predictors. Moreover, a credible volatility prediction scheme can also be applied to optimize the design of novel profitable trading algorithms for crypto markets (Azhmyakov et al. 2022, 2023; Barmish and Primbs 2016; Baumann 2017; Formentin et al. 2018). It is well known that the highly fluctuating crypto exchange prices and the corresponding very frequent changes in the main market indicators make accurate price forecasting nearly impossible. In this situation, consistent volatility prediction can essentially improve concrete trading strategies.

The remainder of our paper is organized as follows: Section 2 contains the formal volatility prediction problem formulation in the framework of the general GARCH model. We also examine concrete GARCH abstractions, discuss some useful mathematical and financial facts, and introduce the concept of the predicted volatility levels. Section 3 includes a critical consideration of the conventional QLM technique for model (parameter) identification in a general GARCH model. We focus our attention on the conceptual difficulties of this widely used methodology in the case of non-Gaussian stochastic errors.

The criticism in this section helps in understanding the necessity of some additional and novel predictive tools. Section 4 is devoted to the development of a novel probabilistic tool for volatility prediction. We use the well-known stochastic volatility model for this purpose. The application of an advanced mathematical technique makes it possible to calculate the characteristic probability associated with the predicted volatility levels. Concretely, we evaluate the exact probability of financial market volatility not falling below a prescribed (predicted) level. In Section 5, we perform a statistical analysis of financial market data and derive a novel, lower estimation of the probability associated with predicted volatility. This data-driven version of the probability estimation is a formal consequence of the stochastic volatility model studied in the previous section. Section 6 contains a short description of the practically oriented volatility estimation module "AI NEWS", recently developed by the 1ex Trading Board group in collaboration with GoldenGate Venture. Section 7 summarizes our paper.

2. Volatility Prediction in Financial Markets Using GARCH Models

Consider a series $\{p_s\}_{s=0,1,...,T}$ of prices of an asset at $(T+1)$ time points and introduce the corresponding logarithmic return (log return):

$$r_s := \ln(p_s/p_{s-1}),$$
$$s = 1, ..., T.$$

Using the obtained data set, $\{r_s\}_{s=1,...,t}$, we now define the (sample) volatility, σ_t^2, for the given time period:

$$\sigma_t^2 = \frac{1}{t-1} \sum_{s=1}^{t} (r_s - \bar{r}_t)^2. \tag{1}$$

Here,

$$\bar{r}_t := \frac{1}{t} \sum_{s=1}^{t} r_s$$

is a sample mean return. Evidently, (1) constitutes a method of moment unbiased estimation of the second moment (i.e., variance of return) for the observable series of returns, $\{r_s\}_{s=1,...,t}$. Here, we do not assume the covariance stationarity of $\{r_s\}_{s=1,...,t}$. The complete time period $s = 0, 1, ..., T$ can be interpreted as a full time frame associated with a complete series of historical prices.

Note that there are a number of theoretical and practical advantages to using log returns in finance (see, e.g., Brooks 2015; Greene 2011). In financial engineering, volatility is often defined as the square root of (1) (the sample standard deviation); however, the square root of σ_t^2 in (1) constitutes a biased estimation of the corresponding standard deviation (see, e.g., Poznyak 2009; Taylor 1986 for details).

As mentioned in the introduction, there are various methods for estimating the volatility σ_{t+1}^2 in (1). Next, we focus our attention on a simple but effective volatility forecasting procedure that uses a relatively simple GARCH model (see Francq and Zakoian 2010; Greene 2011, and the references therein). Recall that the generic GARCH(p,q) abstraction has the following formal expression:

$$r_t = \mu_t + \epsilon_t, \ t \in \mathbb{N}$$
$$\epsilon_t = \tilde{\sigma}_t z_t \tag{2}$$
$$\tilde{\sigma}_t^2 = \omega_t + \Sigma_{i=1}^{q} \alpha_t^i \epsilon_{t-i}^2 + \Sigma_{j=1}^{p} \beta_t^j \tilde{\sigma}_{t-j}^2.$$

Here,

$$\mu_t := E[r_t | \mathcal{F}_{t-1}]$$

is a conditional mean and $\mathcal{F}_{t-1}$ is a sigma-algebra generated by the data that are available up to the time instant $t-1$. Moreover, the possibly non-normal random variables, z_t, $t \in \mathbb{N}$, are assumed to be independent and identically distributed with

$$E[z_t|\mathcal{F}_{t-1}] = 0,$$
$$E[z_t^2|\mathcal{F}_{t-1}] = 1.$$

We also assume that the GARCH(p,q) coefficients

$$(\alpha_t^1, ..., \alpha_t^q),$$
$$(\beta_t^1, ..., \beta_t^p)$$

in (2) are known for every $t \in \mathbb{N}$. The deterministic value, ω_t, in model (2) is sometimes called an intercept. Note that we consider a general non-stationary GARCH model here. Evidently, the above non-stationary abstractions constitute an adequate modeling framework for modern, highly volatile crypto markets.

As the statistical properties of the sample mean, $\bar{r}_t$, in (1) make it a very inaccurate approximation of a true mean, taking the necessary deviations around zero instead of $\bar{r}_t$, as in Formulae (1), increases the accuracy of the volatility prediction. Therefore, we next assume that $\mu_t \equiv 0$ for all $t \in \mathbb{N}$ in (2).

The conditional variance $\tilde{\sigma}_t^2$ in the GARCH(p,q) model (2) constitutes a specific model-based volatility estimation. The main idea of the proposed GARCH model is that the conditional variance of returns has an autoregressive structure and is positively correlated to its own recent past. Note that this model also generates the volatility clustering effect.

In the case of a GARCH(1,1), we obtain

$$\tilde{\sigma}_t^2 = \omega_t + \alpha_t^1 \epsilon_{t-1}^2 + \beta_t^1 \tilde{\sigma}_{t-1}^2. \tag{3}$$

From (3), next, we derive the model-based unconditional variance estimation, $\tilde{\sigma}_{t+1}^2$, of the return r_{t+1}:

$$\tilde{\sigma}_{t+1}^2 = \omega_{t+1} + (\alpha_{t+1}^1 + \beta_{t+1}^1)\tilde{\sigma}_t^2, \ t \in \mathbb{N}. \tag{4}$$

The resulting volatility prediction expression (4) has a recursive nature. In the stationary case, namely, for

$$\alpha_t^1 \equiv \alpha > 0, \ \beta_t^1 \equiv \beta > 0,$$
$$\omega_t \equiv \omega > 0$$
$$1 - \alpha - \beta > 0$$

we evidently have the explicit time-invariant volatility prediction:

$$\tilde{\sigma}_{t+1}^2 = \tilde{\sigma}^2 = \frac{\omega}{1 - \alpha - \beta}. \tag{5}$$

The so-called volatility persistence is given here by $(\alpha + \beta)$.

Let us also present the resulting formulae for the estimated volatility, $\tilde{\sigma}_{t+1}^2$, of return associated with a stationary GARCH(2,2) predictive model:

$$\tilde{\sigma}_t^2 = \omega + \alpha^1 \epsilon_{t-1}^2 + \beta^1 \tilde{\sigma}_{t-1}^2 + \alpha^2 \epsilon_{t-2}^2 + \beta^2 \tilde{\sigma}_{t-2}^2. \tag{6}$$

The corresponding volatility prediction in that case can be expressed as follows:

$$\tilde{\sigma}_{t+1}^2 = \tilde{\sigma}^2 = \frac{\omega}{1 - (\alpha^1 + \beta^1) - (\alpha^2 + \beta^2)}. \tag{7}$$

We refer to Francq and Zakoian (2010) for the necessary mathematical formalism. Recall that the covariance-stationarity condition for the general GARCH(p,q) process (2) has the generic form

$$\Sigma_{i=1}^q \alpha_t^i + \Sigma_{j=1}^p \beta_t^j < 0.$$

Next, using a "predicted volatility level", we denote the value $\tilde{\sigma}_{t+1}^2$, calculated using one of the above Formulae (4), (5), or (7). In parallel with the common volatility definition, we also consider the well-known "integrated volatility" concept over the period t to $t + 1$:

$$\Phi_{t+1} := \int_0^1 \sigma_{t+\tau}^2 d\tau. \tag{8}$$

The integrated volatility concept (8) is of central importance in the pricing of crypto derivatives (see, e.g., Lux and Marchesi 2000 for details). Let us also refer to Poon and Granger (2003) for some existing integrated volatility estimators. Evidently, the GARCH-based predicted volatility levels mentioned above naturally imply the corresponding levels of integrated volatility in (8). For example, the combination of the simple trapezoidal rule and (4) implies the following estimation of Φ_{t+1}:

$$\tilde{\Phi}_{t+1} \approx \frac{1}{2}(\omega_{t+1} + (1 + \alpha_{t+1}^1 + \beta_{t+1}^1)\tilde{\sigma}_t^2) \tag{9}$$

Note that the integrated volatility, Φ_{t+1}, can also be estimated using the return samplings for a time interval of sufficiently frequent returns (see, e.g., Andersen et al. 2001):

$$\lim_{n\to\infty} P[|\Phi_{t+1} - \Sigma_{s=1}^n r_{n,(t+s)/n}^2| \le \delta] = 0 \ \forall \delta > 0. \tag{10}$$

Here, n is the sampling frequency, $r_{n,(t+s)/n}$ denotes a compound return, and $P[\cdot]$ is the probability associated with the exchange prices (exchange rates) under consideration. The basic relation (10) also involves a useful concept from modern financial engineering; namely, the so-called "realized volatility" (see Haas and Peter 2024; Poon and Granger 2003, and the references therein). Similar to the forecasting technique for the predicted volatility $\tilde{\sigma}_{t+1}^2$ discussed above, one can also estimate some of the additional important statistical characteristics of return. For example, the GARCH models presented in this section provide a consistent analytic basis for an adequate estimation of the corresponding kurtosis coefficients (see, e.g., Kim and White 2004 for details).

3. On the Critical Analysis of the QML Method for Parameter Identification

The general GARCH-based volatility model (2), as well as the concrete predictive relations (4), (5), (7), and (9), are derived under the assumption of the known model parameters (coefficients)

$$\{(\omega_t, \ \alpha_t^i), \ i = 1, ..., q\}$$

and

$$\{\beta_t^j, \ j = 1, ..., p\}.$$

However, practical application of these approaches involves a necessary identification procedure for defining the GARCH parameters mentioned above. Moreover, one also needs to estimate (identify) the standardized errors, z_t, in model (2).

The quasi-maximum likelihood (QML) method is widely used for the identification of the GARCH(p,q) model (2). We refer to Birge and Louveaux (2011); Bollerslev and Wooldridge (1992); Franses and Ghijsels (1999); Gallager (2013), and Poznyak (2009) for some mathematical details and concrete applications of the QML techniques. Under the standard normality assumptions, this method provides consistent and asymptotically normal estimations in the case of strictly stationary GARCH processes. Recall that the conventional QML involves maximizing the Gaussian log-likelihood, and the resulting

solution constitutes an adequate estimation of a parameter vector under the normality assumptions.

However, it is common knowledge that financial time series (for example, the crypto exchange rates) have the characteristics of being leptokurtotic and fat-tailed with skewness. Moreover, these series usually involve the so-called volatility clustering effect. The "stylized facts" about the financial market volatility mentioned above also include asymmetry and mean reversion. Note that these properties of the volatility dynamics are now well documented (see, e.g., Cont 2001; Poon and Granger 2003).

The non-regular behavior of the modern financial markets and the corresponding stylized facts about volatility make it impossible to consider the classical Gaussian assumption for the stochastic errors in GARCH(p,q) models. Note that the generic normal probability distributions do not involve outliers and are incompatible with the fat-tailed and skewness effects mentioned above. On the other side, it is well known that the QML method is non-robust in the presence of data outliers generated by fat-tailed and skewed distributions. It is remarkable that in some professional publications, and also in practical trading manuals, inconsistent normality assumptions are still followed. As a result, this simplified Gaussian-based modeling approach involves a deficient description and faulty forecasting of the real (crypto) market dynamics.

The above problem of an adequate modeling framework for the stochastic errors in the general GARCH(p,q) model (2) is crucially important for the resulting model-based volatility prediction. The basic QML estimation is inefficient if the volatility proxies are non-Gaussian (see, e.g., Poon and Granger 2003 and the references therein). As a consequence, one will obtain a possible inconsequential estimation, $\hat{\theta}_t$, for the parameters of the basic GARCH volatility model under consideration.

In this situation, one can consider some concrete fat-tailed and skewed probability distribution in order to examine and simulate the more realistic case studies of the modeled volatility dynamics. For example, one can use the "contaminated" and skewed normal distribution, skewed Student distribution, skewed generalized error distribution, and many others. These non-regular probability distributions generate various types of realistic additive and innovative outliers for the time-series-based modeling of financial time series. We refer to Azhmyakov et al. (2021), Fan et al. (2014), and Huber and Ronchetti (2005) for the corresponding research and simulation results.

Let us note that the fat-tailed and skewed distributions constitute an adequate modeling framework in the case of cryptocurrency time series (see Cont 2001). On the other hand, we usually have no information about a concrete real probability distribution associated with these specific financial series. The same is also true with respect to the series of returns. That means that the concrete non-Gaussian (fat-tailed and skewed) probability distributions of the stochastic errors in the GARCH model (2) are generally unknown.

The above fact constitutes the main motivation for the necessary methodological extension of the existing model-based techniques for volatility prediction and for developing some additional data-driven statistical tools. Next, we propose a novel statistical metric that can be used as an auxiliary analytic tool for the GARCH-based prediction of the volatility levels. This metric involves real market data and constitutes a quantitative method for seeing how well the model-based volatility prediction would have performed. The novel methodology we introduce in the next sections can also be implemented as a part of the common backtesting procedures for the design and verification of new algorithmic trading strategies.

4. Exact Probability Calculation for the Predicted Volatility Levels

This section presents a useful result that can be applied to the formal probabilistic analysis of the predicted volatility levels. By taking into consideration the conceptual difficulties of the OML method discussed in Section 3, we propose an auxiliary statistical-based predictive metric.

Consider the well-known stochastic volatility model (see, e.g., Poon and Granger 2003 and the references therein):

$$d\sigma_t^2 = (\mu_v - \beta\sigma_t^2)dt + \sigma_v\sigma_t^2 dW_v(t), \; t \in \mathbb{R}_+, \tag{11}$$

where μ_v is an average, β is the speed of the volatility process, and σ_v is called "volatility of volatility". Using $W_v(t)$, we denote a Wiener process with

$$W_v(0) = 0.$$

The above stochastic volatility model is usually considered in combination with the price dynamics:

$$dp_t = \mu_p p_t dt + \sigma p_t dW_p(t), \; t \in \mathbb{R}_+, \tag{12}$$

Similar to (11), we are dealing with an associated Wiener process here:

$$W_p(t), \; W_p(0) = 0.$$

Using μ_p in (12), we denote the mean. Many useful mathematical models of financial markets include the generic abstraction (12). Let us mention the classic Samuelson pricing model and the celebrated Black–Scholes theory (see Samuelson 1965; Black and Scholes 1973).

Note that the price and the volatility models (11) and (12) constitute an interconnected system of equations. This natural interconnection can be described by a specific correlation between $W_v(t)$ and $W_p(t)$. The fundamental system (11) and (12) can also be used for modeling the crypto markets. This model also generates some stylized facts about the volatility mentioned in the previous sections; namely, the Black–Scholes volatility smile and volatility clustering. We refer to Black and Scholes (1973) for further technical details.

We now consider the GARCH(1,1) and GARCH(2,2) models from Section 2 and the corresponding predicted volatility levels (4), (5), and (7). Let v_t^* be a required volatility level that is associated with the corresponding GARCH-based predictions. Roughly speaking, we have here the non-stationary level v_t^* such that

$$v_t^* < \tilde{\sigma}_t^2, \; \forall t \in \mathbb{R}_+$$

for (4), and the stationary level

$$v^* < \tilde{\sigma}^2$$

in the case of (5) and (7). The required probability associated with the predicted volatility level, v_t^*, can now be defined as follows:

$$P[\sigma_t^2 \in (v_t^*, +\infty) \, \big| \, \sigma_0^2 > v_t^*], \; \forall t \in [0, T]. \tag{13}$$

Where $T \in \mathbb{N}$. Recall that $P[\cdot]$ denotes the probability measure associated with the exchange prices, p_t (see Section 2). Note that the above diffusion Markov processes—namely, processes (11) and (12)—are assumed to be defined in the same probability space.

The proposed definition (13) expresses a probability that the volatility does not fall below a specific level (a level of v_t^*) for $t \in [0, T]$. This constitutes a kind of "consistency" for the model-based volatility prediction concept determined by GARCH(1,1) or GARCH(2,2). Note that in (13), we are dealing with a conditional probability and assuming that the initial volatility, σ_0^2, is higher than v_t^*. We now calculate the required probability determined by (13) using only the stochastic volatility abstraction discussed above. Note that this calculation does not involve the GARCH(p,q) volatility model.

In parallel with the characteristic probability (13), we introduce the formal complement

$$\rho(t, \sigma_t^2) := 1 - P[\sigma_t^2 \in (v_t^*, +\infty) \, \big| \, \sigma_0^2 > v_t^*], \; \forall t \in [0, T], \tag{14}$$

where the stochastic dynamics of σ_t^2 are given by (11). This complement expresses the probability of the "complementary" event

$$\sigma_t^2 \leq v_t^*$$

for $t < T$. The complementary probability given by (14) can now be evaluated. From the abstract result of Pontryagin et al. (1933), it follows that function $\rho(t, \sigma_t^2)$ in (14) satisfies the boundary value problem

$$
\begin{aligned}
&\frac{\partial \rho}{\partial t} = (\mu_v - \beta \sigma_t^2)\frac{\partial \rho}{\partial x} + \frac{1}{2}\sigma_v^2 \sigma_t^4 \frac{\partial^2 \rho}{\partial x^2}, \ \ t > 0, \ \sigma_t^2 > v_t^*, \\
&\rho(t, v_t^*) = 1, \ \ t \geq 0, \\
&\rho(0, \sigma_t^2) = 0, \ \ \sigma_t^2 > v_t^*,
\end{aligned}
\tag{15}
$$

where σ_t^2 is a solution of (11). From Pontryagin et al. (1933), we also deduce that the auxiliary function

$$\xi(t, \sigma_t^2) := (1 - \rho(t, \sigma_t^2)) \exp \frac{1}{\sigma_v^2 \sigma_t^4}(\mu_v - \beta \sigma_t^2)(\sigma_t^2 - v_t^*) + \frac{1}{2\sigma_v^2 \sigma_t^4}(\mu_v - \beta \sigma_t^2)^2 t$$

satisfies the boundary value problem for the conventional heat equation:

$$
\begin{aligned}
&\frac{\partial \xi}{\partial t} = \frac{1}{2}\sigma_v^2 \sigma_t^4 \frac{\partial^2 \xi}{\partial \sigma_t^4}, \ \ t > 0, \ \sigma_t^2 > v_t^*, \\
&\xi(t, v_t^*) = 0, \ \ t \geq 0, \\
&\xi(t, \sigma_t^2 - v_t^*) = \exp \frac{1}{\sigma_v^2 \sigma_t^4}(\sigma_t^2 - v_t^*), \ \ \sigma_t^2 > v_t^*.
\end{aligned}
\tag{16}
$$

The solution of the boundary value problem (16) can be written as follows (see, e.g., Kevorkian 2000):

$$
\begin{aligned}
\xi(t, \sigma_t^2) = \exp \Big\{ \frac{1}{2\sigma_v^2 \sigma_t^4}(\mu_v - \beta \sigma_t^2)^2 t \Big\} \times \\
\Big[\psi\Big(\frac{(\mu_v - \beta \sigma_t^2)t + (\sigma_t^2 - v_t^*)}{\sigma_v \sigma_t^2 \sqrt{t}}\Big) \times \exp \Big\{ \frac{1}{\sigma_v^2 \sigma_t^4}(\mu_v - \beta \sigma_t^2)(\sigma_t^2 - v_t^*) \Big\} - \\
\psi\Big(\frac{(\mu_v - \beta \sigma_t^2)t - (\sigma_t^2 - v_t^*)}{\sigma_v \sigma_t^2 \sqrt{t}}\Big) \times \exp \Big\{ -\frac{1}{\sigma_v^2 \sigma_t^4}(\mu_v - \beta \sigma_t^2)(\sigma_t^2 - v_t^*) \Big\} \Big].
\end{aligned}
\tag{17}
$$

Here, we use the following notation for the auxiliary function:

$$\psi(\zeta) := \frac{1}{2\pi}\int_{-\infty}^{\zeta} \exp \Big\{ -\frac{u^2}{2} \Big\} du,$$

where

$$\zeta := \mu_v - \beta \sigma_t^2.$$

We now consider the definition (17) of the auxiliary function $\xi(\cdot, \cdot)$ and obtain an explicit expression for the original function $\rho(\cdot, \cdot)$ in problem (15). This expression implies the corresponding formal result for the desired probability in (13):

$$
\begin{aligned}
P[\sigma_t^2 \in (v_t^*, +\infty) \mid \sigma_0^2 > v_t^*] = 1 - \rho(t, \sigma_t^2) = \\
1 + \psi\Big(\frac{(\mu_v - \beta \sigma_t^2)t - (\sigma_t^2 - v_t^*)}{\sigma_v \sigma_t^2 \sqrt{t}}\Big) \times exp\Big\{ -\frac{2}{\sigma_v^2 \sigma_t^4}(\mu_v - \beta \sigma_t^2)(\sigma_t^2 - v_t^*) \Big\} - \\
\psi\Big(\frac{(\mu_v - \beta \sigma_t^2)t + (\sigma_t^2 - v_t^*)}{\sigma_v \sigma_t^2 \sqrt{t}}\Big),
\end{aligned}
\tag{18}
$$

where $t \in [0, T]$.

A direct verification shows that the obtained complementary to (18)—namely, the probability $\rho(\cdot, \cdot)$—satisfies the boundary value problem (15). Moreover, for the function $\rho(\cdot, \cdot)$ and for the desired probability in (18), we can verify the natural condition

$$0 \le P[\sigma_t^2 \in (v_t^*, +\infty) \mid \sigma_0^2 > v_t^*] \le 1,$$

for $t \in [0, T]$.

We now conclude that the probability of the GARCH-based predicted volatility level, v_t^*, is explicitly given by the resulting relation (18). It expresses the probability of the event

$$\sigma_t^2 \in (v_t^*, +\infty)$$

for $[0, T]$, assuming that the initial volatility (for $t = 0$) satisfies the inequality condition

$$\sigma_0^2 > v_t^*.$$

Finally, let us note that the exact probability calculus developed in this section is based on the generic stochastic modeling approach (11) for volatility dynamics.

5. Statistics of the Predicted Volatility Levels

The probabilistic analysis of the predicted volatility levels performed in the previous section is based on an abstract mathematical model; namely, on the stochastic equation (11). Next, we consider this obtained theoretical technique and use it for an applied, data-driven statistical analysis of the GARCH-based estimations of volatility levels.

Consider the resulting Formulae (18) from Section 4 and put $t = T$. Our aim is to derive a lower estimation of the probability expressed in (18). We examine it for the constant predicted volatility levels, $\tilde{\sigma}^2$, in (5) (GARCH(1,1) model) and (7) (GARCH(2,2) model). For a constant volatility level, $v^* = \text{const}$, with

$$v^* < \tilde{\sigma},$$

we obtain

$$
\begin{aligned}
&P[\sigma_t^2 \in (v^*, +\infty) \mid \sigma_0^2 > v^*] \ge \\
&1 + \psi\left(\frac{(\mu_v - \beta\tilde{\sigma}^2)T - (\tilde{\sigma}^2 - v^*)}{\sigma_v\tilde{\sigma}^2\sqrt{T}}\right) \times exp\left\{-\frac{2}{\sigma_v^2\tilde{\sigma}^4}(\mu_v - \beta\tilde{\sigma}^2)(\tilde{\sigma}^2 - v^*)\right\} - \\
&\psi\left(\frac{(\mu_v - \beta\tilde{\sigma}^2)T + (\tilde{\sigma}^2 - v^*)}{\sigma_v\tilde{\sigma}^2\sqrt{T}}\right).
\end{aligned}
\tag{19}
$$

Here, $t \in [0, T]$.

We now examine the limit value of the probability expression in (19) for $T \to \infty$. We next interpret the resulting P_∞ as a probability that in the "foreseeable future" the volatility does not fall below the prescribed (constant) level, v^*. As

$$\lim_{\zeta \to \infty} \psi(\zeta) = 1,$$

we obtain

$$P_\infty \ge exp\left\{-\frac{2}{\sigma_v^2\tilde{\sigma}^4}(\mu_v - \beta\tilde{\sigma}^2)(\tilde{\sigma}^2 - v^*)\right\}. \tag{20}$$

Coming back to a real financial (crypto) market data, we introduce the number $M \in \mathbb{N}$, defined as follows:

$$M := \sum_{t \in \mathbb{N}} \mathbf{1}(\sigma_t^2 \ge v^*), \tag{21}$$

where σ_t^2 is the real market volatility. Function $\mathbf{1}(\cdot)$ in (21) is a generic indicator function,

$$\mathbf{1}(y) = 1 \text{ if } y \text{ is true;}$$
$$\mathbf{1}(y) = 0 \text{ if } y \text{ is false;}$$

for the discrete time $t \in \mathbb{N}$. Roughly speaking, the number $M \in \mathbb{N}$ indicates how many times the real market volatility σ_t^2 is higher as a given constant level, v^*. Using (20) and the basic properties of the exponential function, we deduce our final lower estimation of the limiting probability, P_∞:

$$P_\infty \geq M^{-2(\mu_v - \beta \tilde{\sigma}^2)(\tilde{\sigma}^2 - v^*)/\sigma_v^2 \tilde{\sigma}^4}. \tag{22}$$

Recall that the value $\tilde{\sigma}^2$ in (22) is determined by the stationary GARCH(1,1) predictive model (Formulae (5)) or by the stationary GARCH(2,2) predictive model (Formulae (7)).

The obtained final probability estimation (22) constitutes a data-driven (statistical) estimation of the probability that the volatility does not fall below a prescribed level v^*. The number M (determined above) describes the real behavior of the market volatility, σ_t^2. This number can be obtained from a concrete historical market data set. The same data set can also be used for the identification of the necessary parameters μ_v, β, and σ_v of the stochastic volatility model (11).

Consider now a simple example. In the simplified case,

$$\mu_v = 0,$$
$$\beta = 1,$$
$$\sigma_v = 1,$$

we obtain the following illustrative version of the general estimation (22):

$$P_\infty \geq M^{2(\tilde{\sigma}^2 - v^*)/\tilde{\sigma}^2}. \tag{23}$$

Note that there is no loss of generality in example (23), due to the parametric scalability of the obtained lower estimation (22).

The historical market data set used above for the evaluation of estimation (22) can also be applied to the so-called in-sample forecasting technique. This in-sample method can now be combined with the complementary out-of-sample testing. This approach is methodologically similar to the main idea of the celebrated Monte Carlo method (see Azhmyakov et al. 2023; Hammel and Paul 2002; Poznyak 2009; Rubinstein 1981, and the references therein).

The exact probability evaluation (18) from Section 4, as well as the corresponding data-driven statistical estimations (22) and (23), can also be performed in the context of the integrated volatility (8). One can use the exact value Φ_{t+1} of the integrated volatility in (8) or the simple approximate Formulae (9) for $\tilde{\Phi}_{t+1}$ and define the exact probability or the corresponding statistical estimation similar to (18) or (22), respectively. As many modern financial risk indicators, for example, the well-known value-at-risk (VaR) indicators, use volatility predictors, the probabilistic analysis of the predicted volatility levels presented in Sections 4 and 5 can also be applied to modern risk management.

6. Some Practical Implementations

We now discuss a concrete practical implementation of the statistical analysis for the volatility forecasting methodology developed in our paper. Concretely, we present the volatility prediction module for crypto markets recently developed by the 1ex Trading Board group (https://1ex.com/ (accessed on 11 February 2024)) in collaboration with GoldenGate Venture. This novel analytic tool, called "AI NEWS", has an interconnected structure that includes a generic GARCH(2,2) model and the associated statistical probability estimation (22) developed in the previous section.

The conceptual block diagram of the 1ex Trading Board volatility forecasting algorithm for the AI NEWS module mentioned above is given in Figure 1. Note that the three grouped sub-modules in Figure 1 represent a statistical block which complements the conventional GARCH-based volatility analysis and constitutes the main contribution of our paper.
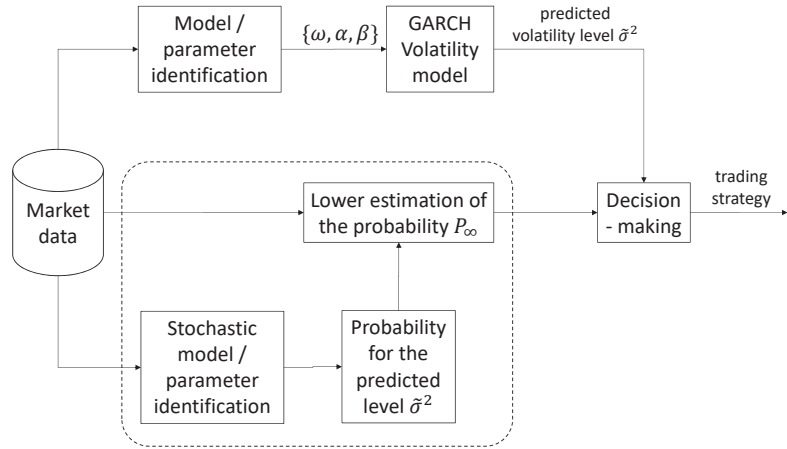
Figure 1. The conceptual block diagram of the 1ex Trading Board volatility prediction module.

This block diagram represents the necessary data and information flows, as well as the system inputs, outputs, and the operational sub-modules. The given sub-modules are necessary for the calculation of the predicted volatility levels and for the evaluation of the statistical estimation (22). The output of the presented block diagram constitutes a final trading decision. The current market data, as well as the involved volatility prediction and statistical probability of the levels, can be visualized by the designed module.

We now use the AI NEWS methodology mentioned above and depict (Figure 2) the market dynamics, the volatility prediction profile, and the corresponding statistical probability level for the concrete crypto trading pair BTC/USDT (date: 10 April 2024).

Figure 2. 1ex Trading Board volatility prediction module: BTC/USDT.

As one can see, the estimated statistical probability associated with the selected volatility level in that example is equal to 0.98.

Let us also apply the developed predictive statistical analysis to an alternative example; namely, to the trading pair SOL/USDT (date: 09/04/2024). The corresponding market dynamics, volatility prediction profile, and the calculated statistical probability level for this pair are shown in Figure 3.
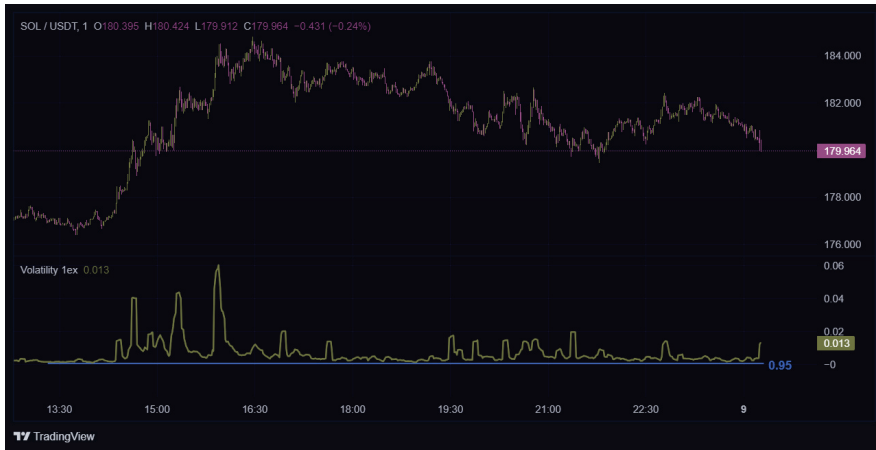
Figure 3. 1ex Trading Board volatility prediction module: SOL/USDT.

The statistical analysis of the predicted volatility levels studied in this section has the potential to be applied to risk management. The 1ex volatility tool mentioned above helps investors and portfolio managers to fix certain levels of risk which they can bear. An adequate forecast of the market volatility of asset prices over the investment holding period constitutes very useful initial information for assessing the investment risk.

Finally, note that many modern profitable cryptocurrency trading algorithms involve adequate volatility prediction schemes. Let us refer here to the class of feedback-type trading algorithms and also to the widely used family of pullback and drawdown trading strategies (see Azhmyakov et al. 2022, 2023; Barmish and Primbs 2016; Baumann 2017). The novel methodology of the predicted volatility levels developed in this paper can also be used in the celebrated Avellaneda–Stoikov market-making strategy.

7. Concluding Remarks

In this paper, we developed a complementary probabilistic tool for the conventional GARCH-based volatility predictor. The classic GARCH abstractions naturally generate some stationary or non-stationary predicted volatility levels. As mentioned in the Introduction, the real probability distributions of financial returns involve high volatility persistence, fat tails, and some additional effects. On the other hand, the widely used QLM parameter estimation methodology is closely related to the (non-realistic) normality assumption and the robust versions of this technique involve some restrictive technical assumptions.

The deficiency of the QLM method mentioned above, as well as the general methodological difficulties of the classic parameter identification approaches, have motivated the development of some additional (complementary) volatility forecasting techniques. From a formal point of view, a GARCH-based volatility predictor constitutes a model-based approach. The main idea of the approach proposed in this paper consists of using a novel data-driven volatility prediction metric that has an auxiliary character for the basic GARCH predictor. In this study, we developed a new additional predictive metric by applying some advanced probabilistic and statistical tools. Roughly speaking, for a model-based predicted volatility level we propose calculating a statistical probability that the financial market volatility does not fall below this level. We call it the "probability of predicted volatility levels".

As the exact result of computation of the above probability is given by a sophisticated (theoretical) expression, we next extended the proposed approach and calculated a constructive lower estimate of the "probability of predicted volatility levels". The obtained lower estimate involved the necessary real market data and constituted an implementable robust version of the prediction metric mentioned above. In fact, we finally proposed a combined prediction approach that contains the model-based and the data-driven elements. Moreover, this approach is compatible with some techniques of the celebrated Monte Carlo methodology, in view of the given (historical) data set.

The developed data-driven volatility prediction approach can be used as an auxiliary tool in many analytical concepts of modern financial engineering; for example, it can be studied in the context of general time-series forecasting. Robust and credible volatility prediction is also a part of many modern cryptocurrency trading algorithms. Moreover, the proposed estimation technique can be applied to the advanced characterization of the integrated volatility and to the important problem of volatility clustering in crypto markets. A robust volatility prediction technique plays a crucial role in modern financial risk management.

Let us also note that the proposed formal calculation of the statistical probability associated with the predicted volatility levels can be implemented with models other than the concrete GARCH-based models studied in this paper, as the developed technique is fully compatible with some alternative volatility forecasting approaches. Moreover, the proposed probabilistic and statistical techniques have a general analytic nature and do not depend on some specific financial data or market conditions.

The probabilistic analysis studied in our paper constitutes an initial theoretical development. In our study, we concentrated on some rigorous mathematical details of the proposed prediction schemes. However, these analytical techniques were implemented in a concrete module for crypto market volatility forecasting developed by the 1ex Trading Board group in collaboration with GoldenGate Venture. We discussed this practical tool briefly in our paper. The paper does not compare the proposed statistical approach to the existing volatility prediction methods. Due to a very large number of modern theoretical and applied results related to predicted volatility, we consider a necessary comparative analysis as a future work. Such a detailed comparative analysis constitutes an important and self-contained topic of a future paper.

Finally, note that the auxiliary statistical analysis of the predicted volatility levels developed in this study can also be considered in the context of novel forecasting approaches based on modern machine learning methodologies Jansen (2020); Shah et al. (2018). It seems possible to extend the proposed concepts of volatility levels and the corresponding statistical analysis developed in our paper to volatility forecasting schemes involving deep learning approaches.

Author Contributions: Conceptualization, V.A.; Methodology, L.A.G.T.; Software, I.S.; Validation, I.S.; Formal analysis, I.S.; Investigation, L.A.G.T.; Data curation, L.A.G.T.; Writing—review & editing, V.A. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Conflicts of Interest: Authors Vadim Azhmyakov and Ilya Shirokov are employed by the company 1ex Corporation. The remaining author declares that the research was conducted in the absence of any commercial or financial relationships that could be constructed as a potential conflict of interest.

References

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens. 2001. The distribution of realized stockr return volatility. *Journal of Financial Economics* 61: 43–76. [CrossRef]

Azhmyakov, Vadim, Ilya Shirokov, and Luz Adriana Guzman Trujillo. 2022. Application of a switched PIDD control strategy to the model-free algorithmic trading. *IFAC PapersOnline* 55: 145–50. [CrossRef]

Azhmyakov, Vadim, Ilya Shirokov, Yu Dernov, and Luz Adriana Guzman Trujillo. 2023. On a data-driven optimization approach to the PID-based algorithmic trading. *Journal of Risk and Financial Management* 16: 1–18. [CrossRef]

Azhmyakov, Vadim, Jose Pereira Arango, Moises Bonilla, Raymundo Juarez del Torro, and Stefan Pickl. 2021. Robust state estimations in controlled ARMA processes with the non-Gaussian noises: Applications to the delayed dynamics. *IFAC PapersOnline* 54: 334–39. [CrossRef]

Barmish, B. Ross, and James A. Primbs. 2016. On a new paradigm for stock trading via a model-free feedback controller. *IEEE Transactions on Automatic Control* 61: 662–76. [CrossRef]

Baumann, Michael H. 2017. On stock trading via feedback control when underlying stock returns are discontinuous. *IEEE Transactions on Automatic Control* 62: 2987–92. [CrossRef]

Birge, John R., and François Louveaux. 2011. *Introduction to Stochastic Programming*. New York: Springer.

Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–59. [CrossRef]

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]

Bollerslev, Tim, and Jeffrey M. Wooldridge. 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11: 143–72. [CrossRef]

Brooks, Chris. 2015. *Introductory Econometrics for Finance*. Glasgow: Cambridge University Press.

Cont, Rama. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1: 223–36. [CrossRef]

Danielsson, Jon, Marcela Valenzuela, and Ilknur Zer. 2018. Learning from history: Volatility and financial crises. *Review of Finance* 21: 2774–805.

Fan, Jianqing, Lei Qi, and Dacheng Xiu. 2014. Quasi-maximum likelihood estimation of GARCH models with heavy-tailed likelihoods. *Journal of Business and Economic Statistics* 32: 178–91. [CrossRef]

Formentin, Simone, Fabio Previdi, Gabriele Maroni, and Claudio Cantaro. 2018. Stock trading via feedback control: An extremum seeking approach. Paper presented at the Mediterranean Conference on Control and Automation, Zadar, Croatia, June 19–22 ; pp. 523–528.

Francq, Christian, and Jean-Michel Zakoian. 2010. *GARCH Models*, Wiltshire: Wiley.

Franses, Philip, and Hendrik Ghijsels. 1999. Additive outliers, GARCH and forecasting volatility. *International Journal of Forecasting* 15: 1–9. [CrossRef]

Gallager, Robert G. 2013. *Stochastic Processes*. New York: Cambridge University Press.

Greene, William H. 2011. *Econometric Analysis*. London: Pearson.

Haas, Martin G., and Franziska J. Peter. 2024. Implementing intraday model-free implied volatility for individual equities to analyze the return–volatility relationship. *Journal of Risk and Financial Management* 17: 2–19. [CrossRef]

Hammel, C., and Wolfgang B. Paul. 2002. Monte Carlo simulations of a trader-based market model. *Physica A* 313: 640–50. [CrossRef]

Huang, Da, Hansheng Wang, and Qiwei Yao. 2008. Estimating GARCH models: When to use what? *Econometric Journal* 11: 27–38. [CrossRef]

Huber, Peter J., and Elvezio M. Ronchetti. 2005. *Robust Statistics*. New York: Wiley.

Jansen, Stefan. 2020. *Machine Learning for Algorithmic Trading*. Birmingham: Packt.

Kahneman, Daniel, and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In *Handbook of the Fundamentals of Financial Decision Making*. Singapore: World Scientific Publishing.

Kevorkian, Jirair. 2000. *Partial Differential Equations: Analytical Solution Techniques*. Texts in Applied Mathematics; New York: Springer.

Kim, Tae-Hwan, and Halbert White. 2004. On more robust estimation of skewness and kurtosis. *Finance Research Letters* 1: 56–73. [CrossRef]

Lewis, Frank L. 1986. *Optimal Estimation*. New York: Wiley.

Lux, Thomas, and Michele Marchesi. 2000. Volatility clustering in financial markets: A microsimulation of interacting agents. International *Journal of Theoretical and Applied Finance* 3: 675–702. [CrossRef]

Nikolova, Venelina, Juan E. Trinidad Segovia, Manuel Fernández-Martínez, and Miguel A. Sánchez-Granero. 2020. A novel methodology to calculate the probability of volatility clusters in financial series: An application to cryptocurrency markets. *Mathematics* 8: 1216. [CrossRef]

Pontryagin, L. S., A. A. Andronov, and A. A. Vitt. 1933. On statistical analysis of dinamical systems. *Zhurnal Eksperimental'noi i Teoreticheskoi Fiziki* 3: 165–80.

Poon, Ser-Huang, and Clive W.J. Granger. 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature* 41: 478–539. [CrossRef]

Poznyak, Alexander. 2009. *Advanced Mathematical Tools for Automatic Control Engineers: Stochastic Tools*. New York: Elsevier.

Rubinstein, Reuven Y. 1981. *Simulation and the Monte Carlo Method*. New York: John Wiley Inc.

Samuelson, Paul A. 1965. Rational theory of warrant pricing. *Industrial Management Review* 6: 13–31.

Schwert, G. William. 1990. Stock volatility and crash of 87. *Review of Financial Studies* 3: 77–102. [CrossRef]

Sen, Jaydip, Sidra Mehtab, and Abhishek Dutta. 2021. Volatility modeling of stocks from selected sectors of the Indian economy using GARCH. Paper presented at the Asian Conference of Innovation in Technology, Pune, India, August 27–29; pp. 1–9.

Shah, Dev, Wesley Campbell, and Farhana H. Zulkernin. 2018. A comparative study of LSTM and DNN for stock market forecasting. Paper presented at the IEEE International Conference on Big Data, Seattle, WA, USA, December 10–13; pp. 4148–55.

Taylor, Stephen. 1986. *Modeling Financial Time Series*. Chichester: Wiley.

Verhoeven, Peter B., Berndt Pilgram, Michael McAleer, and Alistair Mees. 2002. Non-linear modelling and forecasting of S & P 500 volatility. *Mathematics and Computers in Simulation* 59: 233–41.

Wang, Lu, Feng Ma, Jing Liu, and Lin Yang. 2020. Forecasting stock price volatility: New evidence from the GARCH-MIDAS model. *International Journal of Forecasting* 36: 684–94. [CrossRef]

Ziemba, William T., and Raymond G. Vickson. 1975. *Stochastic Optimization Models in Finance*. New York: Academic Press.

Article

Penalized Bayesian Approach-Based Variable Selection for Economic Forecasting

Antonio Pacifico [1,*] and Daniela Pilone [2]

1   Department of Economics and Law, University of Macerata, Piazza Strambi 1, 62100 Macerata, Italy
2   Department of Economics and Finance, LUISS Guido Carli University, Viale Romania 32, 00198 Rome, Italy;
     daniela.pilone@studenti.luiss.it
*   Correspondence: antonio.pacifico@unimc.it or apacifico@luiss.it

Abstract: This paper proposes a penalized Bayesian computational algorithm as an improvement to the LASSO approach for economic forecasting in multivariate time series. Methodologically, a weighted variable selection procedure is involved in handling high-dimensional and highly correlated data, reduce the dimensionality of the model and parameter space, and then select a promising subset of predictors affecting the outcomes. It is weighted because of two auxiliary penalty terms involved in prior specifications and posterior distributions. The empirical example addresses the issue of pandemic disease prediction and the effects on economic development. It builds on a large set of European and non-European regions to also investigate cross-unit heterogeneity and interdependency. According to the estimation results, density forecasts are conducted to highlight how the promising subset of covariates would help to predict potential contagion due to pandemic diseases. Policy issues are also discussed.

Keywords: time-varying parameters; penalized approaches; machine learning techniques; Bayesian inference; forecasting; disease prediction

1. Introduction

In economics, machine learning (ML) techniques are a useful strategy for evaluating data mining because of gaining knowledge from the prior research and discovering hidden patterns in data. Generally speaking, an ML model splits the dataset in two parts: a training sample and a test sample. In the former, data information is uploaded within the system to make an inference on a set of predictors affecting the outcomes of interest. Then, the estimation results are in turn used in the test sample to check their degree of being unbiased and robustness by means of diagnostic tests.

This study focuses on supervised ML methods since they classify and group factors through labeled datasets predicting outcomes accurately. In this way, Bayesian inference can be addressed by assigning informative conjugate priors to every predictor affecting the outcomes. However, the data compression involved in these algorithms does not have any reference to the outcomes, and then they are unable to deal with some open related questions in variable selection problems, such as model uncertainty when a single model is selected a priori to be the true one (Madigan and Raftery 1994; Raftery et al. 1995; Breiman and Spector 1992), overfitting when multiple models are selected, providing a somewhat better fit to the data than simpler ones (Madigan et al. 1995; Raftery et al. 1997; Pacifico 2020), and structural model uncertainty when one or more functional forms of misspecification matter (Gelfand and Dey 1994; Pacifico 2020).

Overall, three main sparse penalized approaches are generally used to entail variable selection in large samples and highly correlated data: bridge regression (Wenjiang 1998), the smoothly clipped absolute deviation (SCAD; Fan and Li 2001), and the least absolute shrinkage and selection operator (LASSO, Tibshirani 1996). The first two approaches imply

nonconvex penalties, satisfying the oracle property for unbiased nonconvex penalized estimators. The oracle property refers to the statement that a given local minimum of the penalized sum of the squared residuals is asymptotically equivalent to the oracle estimator, which is in turn the ideal estimator obtained only with signal variables without penalization. However, it is computationally quite difficult to verify if reasonable local minima are asymptotically the oracle estimator. The LASSO technique is an alternative approach generally used for simultaneous estimation and variable selection by minimizing the residual sum of squares. Indeed, it is able to jointly choose the subset of covariates better for evaluating a model and yielding continuous variable selection, improving the prediction accuracy due to the bias variance trade-off. This research study builds on th latter in the context of time-varying parameters and large samples.

Consider the following high-dimensional time series model:

$$y_t = x_t' \beta + u_t \,, \tag{1}$$

where the stacked $i = 1, 2, \ldots, n$ denotes the units, $t = 1, 2, \ldots, T$ denotes time, y_{it} is an $n \cdot 1$ vector of the outcomes, $x_t = (x_{1t}, x_{2t}, \ldots, x_{nt})'$ is an $n \cdot 1$ vector of the endogenous covariates affecting the outcomes, β is an $n \cdot 1$ parameter vector, and u_t is an $n \cdot 1$ vector of the error terms. A high-dimensional time series model would matter when n is sufficiently larger than T.

According to Equation (1), the LASSO regression model can be obtained by maximizing the penalized likelihood:

$$L_m(\beta) = L(\beta) - \lambda \sum_{j=1}^{m} |\beta_j| \,, \tag{2}$$

where $L(\beta) = L(\beta|y)$ is the likelihood function, $j = 1, 2, \ldots, m$ denotes the variables, and λ controls the impact of the penalization term defined by the L_1 norm of the regression coefficients (Vidaurre et al. 2013; Wu and Wu 2016; Zhang and Zhang 2014). The form of penalization in Equation (2) is used for the variable selection by forcing some of the entries of the estimated β to be exactly zero (known as the sparse approach). Even if the LASSO is widely adopted in many fields of economic and medical data repositories thanks to its computational accessibility and sparsity,[1] it tends to suffer from several drawbacks, leading to it being inconsistent for model selection when data are high dimensional and highly correlated. This study addresses three of them: (1) no oracle properties because of a bias issue; (2) high false-positive selection rates and biases toward zero for large coefficients (Lu et al. 2012; Uematsu and Yamagata 2023; Adamek et al. 2023; Wong et al. 2020); and (3) bad and poor performance when the predictors are highly correlated.

This paper aims to overtake each of the aforementioned drawbacks when predicting pandemic diseases and their effects on productivity growth in a dynamic set-up. The contribution consists of proposing a simple Bayesian computational algorithm as improvements to the LASSO approach when handling high-dimensional and highly correlated data in multivariate time series. It takes the name of weighted LASSO Bayes (WLB) and focuses on machine learning penalized approaches for ruling out the predictors which are non-statistically significant and relevant to predicting the outcomes. In the WLB approach, variable selection acts as a strong case of Occam's razor; when a model receives less support from the data than any of their simpler submodels, it will be excluded and no longer considered. Thus, the model solution containing possible biased estimators will automatically be discarded, being far worse at predicting the data than other model solutions (drawback (1)). In this way, some variable selection problems such as overfitting and misspecification (or structural uncertainty) can be also dealt with (drawback (2)). To handle high-dimensional and highly correlated data, two penalty terms are added to prior specifications when computing the posterior inclusion probabilities (PIPs) to rule out from the variable selection potential nonsignificant estimators (drawbacks (2) and (3)).

Once the subset of model solutions (or a combination of predictors) better fitting the data is obtained, the best final subset will correspond to the one with the highest log

weighted likelihood ratio (lWLR). Here, the priors are the weighting functions and refer to the conjugate informative priors (CIPs) defined for every predictor. Finally, density forecasts can be constructed and performed to highlight how the final promising subset of covariates would help to predict the outcomes of interest.

The underlying logic is similar to analysis of Pacifico (2020), who developed a robust open Bayesian (ROB) procedure for improving Bayesian model averaging and Bayesian variable selection in high-dimensional linear regression and time series models. Similarities hold for acting as a strong form of Occam's razor to find the exact solution, involving a small set of models over which a model average can be computed, and for using CIPs for each predictor within the system. Nevertheless, the proposed approach differs from the ROB procedure in its computational strategy. This latter consists of three steps to find the final best subset of predictors affecting the outcomes, while the WLB algorithm finds the best model solution to be evaluated in a unique step. That feature is possible since the WLB procedure is able to rule out the covariates which are not statistically significant.

The WLB approach builds on the LASSO procedure by assigning additional weights in the form of penalty terms to maximize the likelihood-based analysis of state space models. The weights correspond to a threshold used in the shrinking procedure and a forgetting factor for the modeling and disentangling coefficient and volatility changes. The WLB approach also builds on further related studies proposing improvements to the traditional LASSO to deal with multicollinearity and high-dimensional problems in sparse modeling and variable selection (Mohammad et al. 2021; Jang and Anderson-Cook 2016; Ismail et al. 2023; Yang and Wen 2018).

The empirical example focuses on the predictive analysis of pandemic diseases among a large set of European and non-European regions, including either developed or developing countries. A high-dimensional set of data describing macroeconomic financial variables, socioeconomic and healthcare statistics, and demographic and environment indicators is addressed when making inferences. Some exogenous factors for dealing with region-specific characteristics are also added before computing the lWLR factor to deal with cross-unit heterogeneity and interdependency. The time period runs from 1990 to 2022. Density forecasts are then performed for the years 2023 and 2024 to study possible policy-relevant strategies and predict pandemic diseases or face potential contagion in the global economy.

The remainder of this paper is organized as follows. Section 2 discusses the Bayesian inference for modeling high-dimensional multivariate time series. Section 3 displays prior specification strategies and posterior distributions for conducting density forecasts. Section 4 describes the data and the empirical example addressed in this study. The final section contains some concluding remarks.

2. High-Dimensional Time Series and LASSO Bayes Inference

According to Equation (1), the predictive analysis of pandemic diseases is addressed by evaluating the following vector autoregressive (VAR) model:

$$y_t = \sum_{c=1}^{\bar{c}} \left(\Delta_c y_{t-c} + \Gamma_c x_{t-c} \right) + u_t \, , \tag{3}$$

where $c = 1, 2, \ldots, \bar{c}$ denotes lags, $t = 1, 2, \ldots, T$ denotes time, $y_{k,t}$ is a $K \cdot 1$ vector of the dependent variables (or the outcomes of interest) with $k = 1, 2, \ldots, K$, $y_{k,t-c}$ is a $K \cdot 1$ vector of the control variables referring to the lagged outcomes to capture persistence, $x_{s,t-c}$ is an $S \cdot 1$ vector of the endogenous (directly) observed factors with $s = 1, 2, \ldots, S$, Δ_c and Γ_c are the $K \cdot K$ and $S \cdot S$ matrices, respectively, satisfying appropriate stationarity, and $u_t \sim i.i.d.N(0, \sigma_u^2)$ is a $K \cdot 1$ vector of the error terms.

The equivalent equation-by-equation representation is

$$
\begin{aligned}
y_{k,t} \;=\;& \sum_{c=1}^{\bar{c}} \Big[\Delta_{k,1,c}, \Delta_{k,2,c}, \ldots, \Delta_{k,K,c} \Big] y_{t-c} + \sum_{c=1}^{\bar{c}} \Big[\Gamma_{s,1,c}, \Gamma_{s,2,c}, \ldots, \Gamma_{s,S,c} \Big] x_{t-c} + u_{k,t} = \\
=\;& \Big[y'_{t-1}, y'_{t-2}, \ldots, y'_{t-\bar{c}} \Big] \delta_k + \Big[x'_{t-1}, x'_{t-2}, \ldots, x'_{t-\bar{c}} \Big] \gamma_s + u_{k,t} \, .
\end{aligned}
\tag{4}
$$

Here, some considerations are in order. (1) The variables $k \in (1, 2, \ldots, K)$ and $s \in (1, 2, \ldots, S)$ can differ. (2) The homoskedasticity and weak exogeneity assumptions hold. More precisely, it is assumed that $E(u_t|y_{t-1}, \ldots, y_{t-\bar{c}}, x_{t-1}, \ldots, x_{t-\bar{c}}) = 0$. (3) Let weak exogeneity hold, let $\xi_t = (y'_t, x'_t)'$, and let there exist some constants $\bar{q} > q > 2$ and $d \geq max\left\{ 1, \left(\frac{\bar{q}}{q-1} \right) / (\bar{q} - 2) \right\}$. Then, the process $\xi_{j,t}$ is L_{2q} near-epoch-dependent (NED) with a size $-d$ and with positive bounded NED constants, where $j = 1, 2, \ldots, m$ denotes all predictors according to variables k and s.

The importance of this last assumption is twofold. First, it ensures that the error terms are contemporaneously uncorrelated with every predictor, and the process has finite and constant unconditional moments. Second, the NED framework allows for extending the methodology to other general forms or mixing processes such as linear processes, GARCH models, and nonlinear processes (see the following for further discussion: Wong et al. 2020; Wu and Wu 2016; Masini et al. 2022; Medeiros and Mendes 2016).

The computational approach takes the name of weighted LASSO Bayes (WLB) and aims to define conditional sets of regression parameters and coefficients to estimate the high-dimensional equation-by-equation VAR model in Equation (4).[2] We defined the process ξ_t and let $\theta = (\delta_k, \gamma_s)$. Equation (4) can be rewritten in simultaneous equation form:

$$
y_t = \xi_t \theta_t + u_t \, ,
\tag{5}
$$

where $y_t \in \mathbb{R}^n$ and $\theta = \{\theta_j\}$ is an auxiliary variable denoting the set of candidate predictors. Throughout this paper, an additional auxiliary variable $\chi = \{\chi_j\}$, where $\chi = (\chi_1, \chi_2, \ldots, \chi_m)'$, is defined as containing all possible 2^m model solutions, where $\chi_j = 0$ if θ_t is small (absence of the jth covariate in the model) and $\chi_j = 1$ if θ_t is sufficiently large (presence of the jth covariate in the model).

Let the full model be $M_F := \{1, 2, \ldots, m\}$, let $M_l = \{M_1, M_2, \ldots, M_p\}$ be the submodel class set for any subset of predictors obtained from the variable selection with $M_l \subseteq M_F$, and let $\theta_{M_l} := \{\theta_{l,M_l}\}_{l \in M_l} \in \mathbb{R}^{|M_l|}$ be the vector of regression coefficients better at fitting and predicting the data. The posterior inclusion probabilities are defined as follows:

$$
\sum_{l=1}^{p} \pi\Big(M_l|y_t, \chi\Big) = \frac{\sum_{l=1}^{p} \pi\Big(y_t|M_l, \chi\Big) \cdot \pi\Big(M_l\Big)}{\pi\Big(y_t|\chi\Big)} = \frac{\sum_{l=1}^{p} \pi\Big(y_t|M_l, \chi\Big) \cdot \pi\Big(M_l\Big)}{\sum_{j=1}^{\mathcal{J}} \pi\Big(y_t|M_j, \chi\Big) \cdot \pi\Big(M_j\Big)} \, ,
\tag{6}
$$

where $\mathcal{J}$ stands for the natural parameter space, $M_l \ll M_j$, and $l = 1, 2, \ldots, p$ with $l \ll j$.

A threshold τ is added to rule out from the variable selection potential nonsignificant estimators and then jointly deal with large sample sizes and selective inference. The latter refers to the problem of addressing issues when statistical hypotheses cannot be specified before data collection but are defined during the data analysis process. The final subset of predictors is achieved under the following condition:

$$
\mathcal{M}^{|M_l|} = \Big\{ M_l : M_l \subset M_j, \mathcal{M}^{|M_l|} \subset \mathcal{J}, \pi(M_l|y_t, \chi) > \tau \Big\} \, ,
\tag{7}
$$

where $\mathcal{M}^{|M_l|}$ is the submodel space based on the natural parameter space $\mathcal{J}$ and $\tau < 0.005$ (according to a two-sided alternative hypothesis).

Finally, let the possible (multi)collinearity problems matter in linear models because of highly correlated data. Once the final subset of model solutions is obtained, the lWRL factor of each M_l against M_j is computed, with the priors being the weighting functions. The highest lWLR will denote the final best subset of predictors to be chosen.

To deal with some variable selection problems such as model misspecification and overfitting, an additional penalty term is used to compute the posterior distributions. It corresponds to a forgetting (or decay) factor κ that varies in the range of [0.9–1.0] and controls the process of reducing past data at a constant rate over a period of time. Let the parameters be time-varying and the priors be defined before the data analysis process. The regression coefficients' dynamics might change over time because of the time components (trend and seasonal components) or multiple change points (structural breaks). In a linear context, just as in the proposed WLB approach, the usefulness of using a forgetting factor is in excluding the second case dealing with volatility changes. Thus, in a time of constant volatility ($\kappa \cong 0.9$), the model's prior choice for each M_l will be $= 1$, requiring a nonzero estimate for θ or that χ_j should be included in the model. Conversely, in the case of extremely large volatility changes ($\kappa \cong 1.0$), the estimated coefficient will be discarded and not accounted for anymore.

The underlying logic of using a weighted vector in the penalty term finds analogies with the adaptive LASSO method of Zou (2006), who proposed an improvement to the traditional LASSO procedure for handling high-dimensional data or highly correlated data and for satisfying oracle properties. However, in the WLB algorithm, the hyperparameter κ is built to weigh more according to the model size. Conversely, in the work of Zou (2006), the weighted vector was chosen to minimize cross-validation or generalized cross-validation errors, implying that a large enough weighted vector will lead the coefficients to become exactly equal to zero (ruled out from the shrinking procedure).

The log weighted likelihood ratio is computed as follows:

$$lWLR_{l,j} = log\left\{ \frac{\pi(M_l|y_t,\chi)}{\pi(M_j|y_t,\chi)} \right\}. \tag{8}$$

The model solution with the highest lWLR factor will correspond to the final best submodel solution M_l. The scale of evidence for interpreting the lWLR factor in Equation (8) is defined according to Kass and Raftery (1995):

$$\begin{cases} 0 < lWLR_{l,j} \leq 2 & \text{no evidence for submodel } M_l \\ 2 < lWLR_{l,j} \leq 6 & \text{moderate evidence for submodel } M_l \\ 6 < lWLR_{l,j} \leq 10 & \text{strong evidence for submodel } M_l \\ lWLR_{l,j} > 10 & \text{very strong evidence for submodel } M_l. \end{cases} \tag{9}$$

3. Prior Assumptions and Posterior Distributions

The variable selection procedure involved in the WLB algorithm entails estimating χ_j through θ_t with weights equal to κ. Thus, the posterior model probability, denoting the probability that a variable is *in* the model, corresponds to the mean value of the indicator χ_j. Let the indicator χ be unknown. The true value will be obtained by modeling the variable selection via a set of mixture CIPs:

$$\pi(\theta, \sigma^2, \chi) = \pi(\theta|\sigma^2, \chi) \cdot \pi(\sigma^2|\chi) \cdot \pi(\chi). \tag{10}$$

However, the auxiliary indicator χ depends on the realization of the θ values that are time-varying. To avoid this problem, the auxiliary parameter of θ is further modeled and assumed to follow a random walk process:

$$\theta_t = \theta_{t-1} + \epsilon_t \quad \text{with} \quad \epsilon_t \sim N(0, \Omega), \tag{11}$$

where $\Omega = diag(\omega_1, \omega_2, \ldots, \omega_m)$ is an $m \cdot m$ diagonal matrix and $\omega = (\omega_1, \omega_2, \ldots, \omega_m)'$ is an $m \cdot 1$ vector.

The matrix Ω can be a full covariance matrix, allowing for cross-correlation in the state vector θ_t. Even if that assumption would be counterproductive by increasing the model uncertainty, mainly in high dimensions, it would be necessary when predicting

pandemic diseases among different units (e.g., regions and countries). The error terms u_t and ϵ_t are assumed to be independent of one another to simplify the inference in a likelihood-based analysis of state space models. The unknown parameters to be estimated are then $(\theta_{1:T}, \sigma^2, \omega, \chi)$. According to these specifications, the CIPs in Equation (10) become

$$
\begin{aligned}
\pi(\theta_{1:T}, \sigma^2, \omega, \chi) &= \pi(\theta_{1:T}|\omega) \cdot \pi(\sigma^2) \cdot \pi(\omega) \cdot \pi(\chi) = \\
&= \prod_{t=1}^{T} \pi(\theta_t|\theta_{t-1}, \omega) \cdot \pi(\sigma^2) \cdot \prod_{t=1}^{T} \pi(\omega_j) \cdot \pi(\chi) ,
\end{aligned}
\tag{12}
$$

where

$$
\pi(\theta_t|y_t) = N\left(\bar{\theta}_{t-1|t-1}, \bar{R}_{t-1|t-1}\right) ,
\tag{13}
$$

$$
\pi(\sigma^2|y_t, \theta_t) = IG\left(\frac{\bar{\vartheta}}{2}, \frac{\bar{\rho}}{2}\right) ,
\tag{14}
$$

$$
\pi(\omega_0|\mathcal{F}_{t-1}) = IG\left(\frac{\alpha_0}{2}, \frac{\nu_0}{2}\right) .
\tag{15}
$$

Here, N(·) and IG(·) stand for the normal and (conjugate) inverse gamma distributions, respectively, and $\mathcal{F}_{t-1}$ refers to the information given up to time $t-1$. The latter is useful for dealing with potential coefficient changes due to persistent shocks (homoskedastic errors).

All hyperparameters are known and collected in a vector $\varrho = (\bar{\vartheta}, \bar{\rho}, \alpha_0, \nu_0)$. They are treated as fixed and obtained either from the data to tune the prior to the specific applications (such as $\bar{\rho}, \alpha_0$) or selected a priori to produce relatively loose priors (such as $\bar{\vartheta}, \nu_0$). Let θ_t be time-varying and defined as a random walk in Equation (11). Then, $\theta_{j,t}$ should be constructed to allow it to adapt to a new state in cases with larger ϵ_t values due to an unexpected shock at time t. Thus, the conjugate distribution of $\theta_{j,t}$, given χ, has the form

$$
\pi(\theta_j|\chi) = (1 - \chi_j) \cdot T\left(\varphi, 0, \varpi_\chi \cdot \mu_{0,\chi_j}\right) + \chi_j \cdot T\left(\varphi, 0, \varpi_\chi \cdot \mu_{1,\chi_j}\right) ,
\tag{16}
$$

where $T\left(\varphi, 0, \varpi_\chi \cdot \mu_{\phi,\chi_j}\right)$ with $\phi = (0, 1)$ is the T Student distribution with φ degrees of freedom and a scale parameter $\varpi_\chi \cdot \mu_{\phi,\chi_j}$.

The last parameter to be defined is the auxiliary indicator χ through the realization of θ_t. Let the framework be hierarchical. Then, the marginal prior $\pi(\chi)$ in Equation (12) contains the relevant information for the variable selection. More precisely, based on the data Y, $\pi(\chi)$ updates the probabilities on each of the 2^m possible values of χ. By identifying every χ with a submodel via $\chi_j = 1$ if and only if χ_j is included (presence of the jth covariate in the model), the χ values with higher probability would identify the promising submodels better fitting the data. Thus, according to Equation (16), the χ_j values might be treated as independent and evaluated with a marginal distribution:

$$
\pi(\chi_j) = w_{|\chi|} \cdot \binom{j}{|\chi|}^{-1} ,
\tag{17}
$$

where $w_{|\chi|}$ denotes the model's prior choice for the PIPs according to the model size $|\chi|$. This ensures assigning more weight to the parsimonious models by setting $w_{|\chi|}$ to be large for smaller $|\chi|$ values. More precisely, when the sample size is high dimensional, and the regression parameters are allowed to vary over time, the covariates would tend to be highly correlated. Then, let the model solutions fit the data similarly because of the conjugate priors. Simpler models with fewer parameters would be favored over more complex models with more parameters (overfitting).

Finally, let θ_t evolve over time according to Equation (11), and suppose that the data run from $(t = 0)$ to $(t = T)$. In order to obtain a training sample $(t - 1, 0)$, the Kalman

filter algorithm is used to generate the features of the θ_j values over time. Equation (13) is then rewritten as

$$\pi(\theta_t|\theta_{t-1}, y_t) = N\left(\bar{\theta}_{t|t}, \bar{R}_{t|t}\right), \tag{18}$$

where $\bar{\theta}_{t|t}$ and $\bar{R}_{t|t}$ denote the conditional distribution of θ_t and its variance-covariance matrix at time t given the information over the sample $(t-1, 0)$.

The conditional posterior distribution of $(\theta_1, \theta_2, \ldots, \theta_t|y_t)$ is computed through the forward recursions for the posterior means $(\bar{\theta}_{t|t+1})$ and covariance matrix $\bar{R}_{t|t+1}$:

$$\pi(\theta_t|\theta_{t-1}, y_t) = N\left(\bar{\theta}_{t|t+1}, \bar{R}_{t|t+1}\right), \tag{19}$$

where

$$\bar{\theta}_{t|t+1} = \bar{\theta}_{t|t} + \sum_{t=1}^{T} w_{|\chi|} \cdot \sqrt{\bar{R}_{t|t} \cdot \sigma^2}, \tag{20}$$

$$\bar{R}_{t|t+1} = \left[I_{nm} - \left(\bar{R}_{t|t} \cdot \bar{R}_{t-1|t-1}^{-1}\right)\right] \cdot \left(\bar{R}_{t|t}\right), \tag{21}$$

with

$$\bar{\theta}_{t|t} = \bar{\theta}_{t-1|t-1} + \sum_{t=1}^{T} (1-\kappa) \cdot \sqrt{\bar{R}_{t-1|t-1} \cdot \sigma^2}. \tag{22}$$

Here, $\bar{R}_{t|t}$ and $\bar{R}_{t-1|t-1}$ refer to the variance-covariance matrices of the conditional distributions of $\bar{\theta}_{t|t}$ at time t and $\bar{\theta}_{t-1|t-1}$ at time $t-1$, respectively, κ denotes the forgetting factor involved in the shrinking procedure, $\bar{\theta}_{t-1|t-1} \cong 0.01$, and $w_{|\chi|}$ denotes the PIPs obtained by the sum of the PMPs in Equation (6).

The computation of the penalty term κ aims to discard the estimated coefficients θ_j from the variable selection in case of extremely high volatility. More precisely, if volatility changes matter (temporarily larger ϵ_t), then the full covariance matrix Ω increases (larger ω), setting up the forgetting factor κ to be close to one. By construction, the second term in Equation (22) will be zero, automatically discarding $\theta_{j,t}$ from the shrinking procedure. Indeed, the conditional distributions at times t and $t-1$ would match ($\bar{\theta}_{t|t} = \bar{\theta}_{t-1|t-1} \cong 0.01$), and the PIPs in Equation (20) would decrease (lower $w_{|\chi|}$) because of the larger model size $|\chi|$ in accordance with Equation (17). Consequently, this implies that χ will require an estimate of zero or that χ should be excluded from the model.

Given Equation (19), the other posterior distributions are defined as follows:

$$\pi(\sigma^2|y_t) = IG\left(\frac{\hat{\vartheta}}{2}, \frac{\hat{\rho}}{2}\right), \tag{23}$$

$$\pi(\omega|y_t) = IG\left(\frac{\bar{\alpha}}{2}, \frac{\bar{\nu}}{2}\right). \tag{24}$$

Here, some considerations are in order. In Equation (23), $\hat{\vartheta} = \vartheta_0 \cdot \bar{\vartheta}$, $\hat{\rho} = \rho_0 \cdot \bar{\rho}$, $\vartheta_0 \cong 0.10$, and $\rho_0 \cong 1.0$ are hyperparameters collected in ϱ, $\bar{\vartheta} = 1 - \kappa$, and $\bar{\rho} = \bar{\nu}$. This means that in the case of volatility changes ($\kappa \cong 1.0$), the only relevant estimate will be the scale parameter $\hat{\rho}$ controlling the height of the distribution's peak.[3] Much higher volatility (higher σ^2) will be associated with a larger model size (high serial correlations among errors in the data) and then lower $w_{|\chi|}$, implying exclusion of θ_t from the variable selection.

In Equation (24), $\bar{\alpha} = \alpha_0 \cdot \kappa$, $\bar{\nu} = \nu_0 \cdot \bar{\kappa}$, $\alpha_0 \cong 0.01$, and $\nu_0 \cong 1.0$ denote the arbitrary degree of freedom and the arbitrary scale parameter, respectively, and $\bar{\kappa} = \kappa \cdot exp(0.5 \cdot b)$, where b is a nominal variable equaling one if volatility changes matter and is zero otherwise. In this way, at a time of constant volatility ($\kappa \cong 0.90$), $\bar{\nu}$ will be close to the forgetting factor. Conversely, in cases with extremely high volatility changes ($\kappa \cong 1.0$), $\bar{\nu}$ will assume higher values.

4. Empirical Example

The WLB algorithm was constructed and run on 277 regions for 23 countries, including European developed and developing economies, and non-European countries. The estimation sample was expressed in years spanning the period of 1990–2022 ($T = 33$). All data came from the Eurostat database.[4]

The panel set contained 98 directly observed variables, accounting for potential predictors affecting the outcomes. They were split into three groups: (1) 40 macroeconomic financial indicators, investigating the role of economic conditions in cases of pandemic contagion such as economic status, economic development, competitiveness, and imbalances; (2) 35 socioeconomic and healthcare factors, highlighting potential causes for health factors such as being overweight and tobacco and alcohol consumption, as well as health expenditures, hospital employment, and healthcare statistics; and (3) 23 demographic and environment indicators, understanding how the spreading of pandemic diseases is affected by, for example, urbanity, population, pollution, and internet use. The variable of interest refers to the real growth rate of the gross domestic product (GDP) at current market prices for every region (productivity hereafter).

In Table 1, the best subset predictors better fitting the data and then predicting the outcomes are displayed. They corresponded to 30 factors, where 10 of them referred to macroeconomic financial variables, 12 predictors denoted socioeconomic and healthcare indicators, and 8 factors accounted for demographic and environment statistics. In order to investigate how these predictors affected the depedent variable, the conditional posterior sign (CPS) indicator was evaluated, taking values of one or zero if a covariate in $\theta_{l,t}$ had a positive or negative effect on the outcomes, respectively. Let the CPS be close to one or zero for every predictor. Variable selection problems such as model uncertainty, overfitting, and model misspecification are dealt with.

Here, some interesting economic policy issues are in order. First, when studying pandemic and health diseases among regions, macroeconomic financial linkages should be accounted for. Second, a geographical statement, generally ruled out from disease prediction analyses, needs to also be addressed. Third, macroeconomic financial indicators tend to be relevant as much as socioeconomic and health factors, highlighting that economic conditions and development issues are important drivers affecting the spread and transmission of diseases.

The usefulness of the WLB procedure involves performing variable selection around the PIP for every predictor within the system in order to rule out whether not they are relevant for forecasting the variable of interest (PIPs $> \tau$).

To investigate how cross-unit interdependency and heterogeneity would matter when predicting pandemic diseases, an $n \cdot 1$ vector of strictly exogenous factors $d_i = (d_1, d_2, \ldots, d_n)$ accounting for region-specific and geographical characteristics was also added. They were included ex post the shrinking procedure but before computing the lWLR in Equation (8). More precisely, three dummy variables were used to improve pandemic disease prediction: d_{1t}, accounting for regional disparity (equaling one if the region belonged to a developed countries and being zero otherwise); d_{2t}, denoting the geographical position (equaling one if it is a northwestern or northeastern region and zero if it is a central, southwestern, or southeastern region); and d_{3t}, referring to the initial economic condition of every region to absorb potential convergence (or catch-up) effects (equaling one if the productivity is higher than the average value of the country and being zero otherwise). According to d_{2t}, it was constructed using as a midpoint the country of Italy. To evaluate their usefulness for predicting pandemic diseases, an F-test statistic was carried out to verify their joint significance. When letting the p values be rather close to zero, all three time-invariant factors were included within the system. These results confirm that pandemic diseases and their possible contagion would be also affected by the surrounding environment.

The results highlight some important findings. (1) When studying pandemic diseases and their effects on economic development, accurate variable selection needs to account for different sets of indicators, even if not strictly related to health conditions. (2) In the context of time-varying parameters and high-dimensional data, variable selection problems

have to be dealt with. (3) The lWLR containing the best final submodel solution was set to 15.89, highlighting quite strong evidence for M_l. (4) Another important issue to be addressed is the (unobserved) cross-unit specific characteristics. More precisely, health care systems differ among developed countries, even more so when the analysis is extended to developing economies. Divergence exponentially increases at the regional level. Thus, any policy strategy, government statement, and investments have to be strictly specific to the economic status of that region. Without the exact attention to the true need of a region, any type of support would be useless, leading to opposite results with respect to those expected because of the increase in disparity, heterogeneity, and divergence in the economy (Pulawska 2021; Gungoraydinoglu et al. 2021; Abrhám and Vošta 2022).

Table 1. Best subset predictors: WLB procedure.

Idx.	Predictor	Unit	CPS
MACROECONOMIC FINANCIAL INDICATORS			
1	unit labor cost	% values	0.984
2	consumer price index	% GDP	0.003
3	financial transactions	% GDP	0.976
4	employment by age (15–64)	% Tot. Pop.	0.981
5	labor force, age 15–64	logarithm	0.876
6	unemployment rate by age (15–74)	% Tot. Pop.	0.001
7	risk of poverty by age (15–74)	% Tot. Pop.	0.000
8	weighted income per capita	% logarithm	0.969
9	wage and salaried workers	% Tot. Emp.	0.944
10	gross fixed capital formation	% GDP	0.935
SOCIOECONOMIC AND HEALTHCARE FACTORS			
11	overweight	std. rates per 100 people	0.001
12	consumption of tobacco	% adults (15+)	0.001
13	consumption of alcohol	std. rates per 100 people	0.003
14	current health expenditure	% GDP	0.968
15	*R&D* expenditure	% GDP	0.952
16	fertility rate	% Tot. Pop.	0.981
17	capital health expenditure	% GDP	0.974
18	death rate	per 1000 people	0.001
19	secondary school enrollment	% Tot. Pop.	0.937
20	social participation	% Tot. Pop.	0.852
21	tertiary educational attainment	% Tot. Pop. (25–64)	0.983
22	household price index	% GDP	0.975
DEMOGRAPHIC AND ENVIRONMENT FACTORS			
23	rural population	% Tot. Pop.	0.868
24	urban population	% Tot. Pop.	0.837
25	population growth	% Tot. Pop.	0.974
26	total population	logarithm	0.873
27	energy use	% GDP	0.831
28	total CO_2 emission	% Tot. Pop.	0.004
29	human capital	logarithm	0.847
30	internet use	% GDP	0.782
REGION-SPECIFIC CHARACTERISTICS			
-	d_{1t}	1 = northwestern, northeastern Europe; 0 = southwestern, southeastern Europe	
-	d_{2t}	1 = developed; 0 = developing	
-	d_{3t}	1 = high economic status; 0 = low economic status	
-	Real Growth Rate of GDP	Percentage Change	

The Table is split as follows. The first column denotes the predictor number; the second column displays the predictors; the third column refers to the measurement unit; and the last column displays the CPSs. The last row refers to the outcome of interest. The contraction Tot. Pop. stands for 'total population', and Tot. Emp. stands for 'total employed people'. All data refer to Eurostat database.

To complete the analysis, three density forecasts on the outcome of interest were performed (Figure 1). The first case (panel A) evaluated only the macroeconomic financial indicators, the second forecast (panel B) accounted for socioeconomic and healthcare statistics as well, and the third and final case (panel C) was run on all subsets of predictors, including demographic and environment indicators and exogenous factors on the regional disparity and characteristics.

The density forecasts were performed by running 100,000 iterations per each random start, spanning the period from 1990 to 2022. The *h* steps ahead forecast period refers to the years 2023 and 2024 ($h = 2$) for replicating the productivity dynamics in the current year and studying them in the next future year (2024). The associated computational costs were minimized, ensuring consistent posterior estimates and dimension reduction.[5] The yellow and red lines denote the 95% confidence bands, and the blue and purple lines denote the conditional and unconditional projections of the outcomes of interest for each time period $T + h$, respectively.
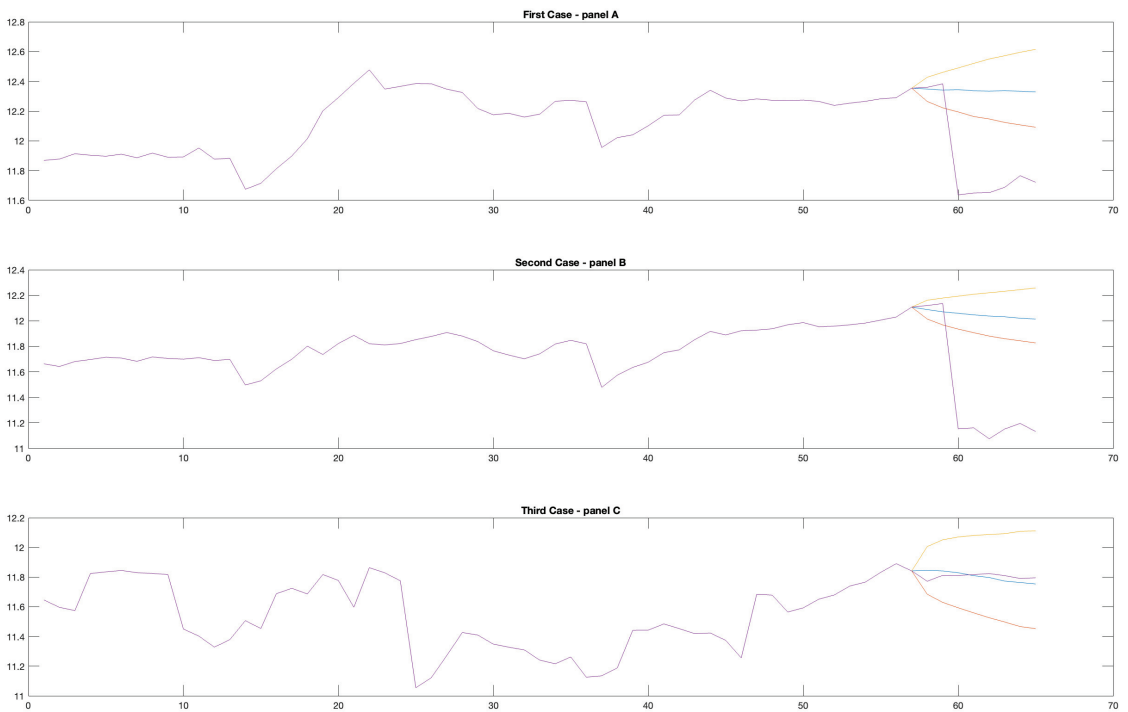
Figure 1. The plot for conditional and unconditional projections for the outcome of interest, spanning the period from 1990 to 2022. The forecast horizon refers to the years 2023 and 2024 ($h = 2$). The *Y* and *X* axes represent the conditional projections and sampling distribution in years, respectively. Concerning the latter, the reference period is 1990 (0); 1995 (10); 2000 (20); 2005 (30); 2010 (40); 2015 (50); 2022 (60); and 2024 (<70).

According to Figure 1, the forecasts in panel A matter more than the other two cases (panels B and C). This was an expected result when predicting the productivity growth while assuming an unexpected shock in the real and financial dimensions. Indeed, the density forecasts were higher, displaying larger productivity growth over time. In addition, when focusing on the conditional forecasts for the years 2023 and 2024 (blue line), the productivity tended to maintain a trend similar to the previous years. However, when

focusing on the unconditional projections (purple line), they strongly diverged because of the presence of endogeneity issues and misspecified dynamics.

In the second case (panel B), where socioeconomic and healthcare factors were also accounted for, the productivity dynamics changed. The density forecasts showed lower magnitudes over time, and the conditional prediction for the years 2023 and 2024 displayed a slightly pronounced decrease, reaching levels lower than the previous years (before the 2020 pandemic). Moreover, the confidence interval containing the conditional forecasting (yellow and red lines) were smaller and tighter than those in panel A. However, in this case, the unconditional projections also totally diverged, even at lower magnitudes. This suggests that the endogeneity and misspecification problems matter again.

Then, the third case included geographical and regional characteristics (panel C). In this specific case, four important results were achieved. First, the productivity dynamics showed strong seasonal components, mainly in accordance with some triggering events such as the 2008 Great Recession, post-crisis recovery programs, and the 2020 pandemic. Second, the upper and lower bound confidence intervals included either conditional or unconditional projections, minimizing the effects of potential endogeneity issues and misspecified dynamics. Third, the same confidence interval tended to be much larger and thicker than the ones observed in panels A and B because of significant cross-region heterogeneity. Fourth, the density forecasts for the years 2023 and 2024 were lower and showed a persistent decrease in the coming years (at least when focusing on the short- and medium-term periods).

To better address the question of if it is possible to predict pandemic diseases, the generalized Theil's entropy index was employed, and it displayed in Figure 2. It was computed by drawing the outcomes of interest for every region, weighted by the proportion of the population with respect to the total (blue line) and their conditional projections (red line), obtained through the forward recursions in Equation (19). The time period used ran from 2004 to 2024. The conditional projections were quite close to the observed weighted outcomes, highlighting the consistency and accuracy of the WLB algorithm in fitting the data. Focusing on the year 2019, a sudden fall in productivity was observed until 2020. Then, a totally opposite trend was achieved up to a recorded higher productivity level in 2021. However, in the year 2021, the productivty dynamics changed again by showing not only a significant decrease but also a downward trend with respect to the past. According to these findings, it is unlikely to predict 'ex ante' pandemic diseases, but it would be possible to control the contagion and then significantly face their aftereffects on the economy. For instance, when unexpected shocks significantly matter and affect productivity dynamics, more attention should be payed not only to the strictly related macroeconomic financial indicators but also the socioeconomic demographic factors that are similarly relevant nowadays.

During the 2020 pandemic, the global outbreak of public health emergencies merely focused on the health sector and disease-related costs. However, a more partial and comprehensive approach would be essential to evaluate the overall economic development impacts of the global pandemic. Indeed, widespread disease has led to economy-wide shocks to both the supply and consumption sides. In addition, the consistent decrease in consumption also negatively affected the global economy, which in turn caused spillover effects for China's regional economy through globalized international trade. Relevant government departments should have payed more attention to structural reforms not only for improving the public health system but also building a high-quality mode of economic growth and restructuring global value chains. All these features were present during the 2020 pandemic and still are today but without appropriate diversification purposes dealing with the compelling need of a country or specific region. Maybe, according to the productivity dynamics in Figure 2, a contagion lasting more than 2 years would have been addressed better through more specific and substantial measures and with lower development times.

From a global perspective, the 2020 pandemic has highlighted the need to jointly address health issues with the well-being and lives of citizens, prosperity and stability of societies and economies, and sustainable development. Health challenges are quickly evolving and rapidly changing the geopolitical environment due to the impact of three additional planetary crises: climate change, biodiversity, and pollution. At the same time, new opportunities linked to areas like research or digitalization have arisen. Thus, a global health strategy is needed to provide a new coherent, effective, and focused policy worldwide. The Global Gateway[6] represents the close strategy of the European Health Union, which protects the well-being of Europeans and the resilience of their health systems. The main aim of this strategy is that the European Union (EU) should deepen its interest in higher attainable standards of health based on fundamentally specific values such as solidarity, equity, and respect for human rights. Infectious diseases represent a heavy burden on many countries, and high infant and maternal mortality rates are matters to be accounted for. This highlights the need to address global health security programs to better prevent and be more resilient to future pandemics. The first two essential EU priorities are investing in the well-being of people and reaching universal health coverage with much stronger health systems. However, these priorities are rather different from 2010, and other related important drivers of ill health should be addressed in an integrated manner, such as climate change, environmental degradation, and humanitarian crises aggravated by the recent and current Russia-Ukraine war. Thus, it is essential to define a wide number of policies focusing on a global health agenda. Overall, global governance should require a new specific focus to keep strong and constant collaboration with the World Health Organization. Further cooperation should be built through the G7, G20, and other global and regional partners. The EU's policies should ensure coherent actions with them to avoid the existing gaps in global governance. To support these strategies' objectives, extremely strong cooperation with the private sector, civil society, and other stakeholders is needed.
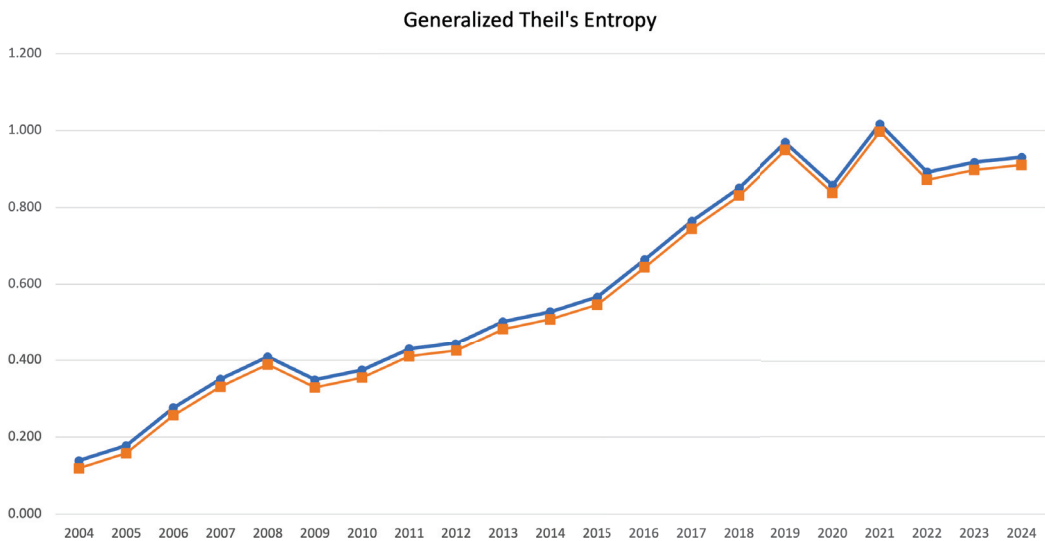
Figure 2. Plot of the generalized Theil's entropy index from 2004 to 2024. This corresponds to the outcomes of interest weighted by the proportion of the population with respect to the total (blue line) and their conditional projections (red line) obtained through the forward recursions in Equation (19).

Sensitivity Analysis and Heterogeneity Issues

In the current subsection, a counterfactual assessment is addressed to check the sensitivity and robustness analysis, and dynamic panel data with the generalized method of moments (DPDP-GMM) are assessed to investigate endogeneity issues in depth.

Concerning the former, the WLB procedure was performed by using a reduced time series on the same sample of 277 regions for 23 countries and 98 endogenous predictors over the period of 2000–2022. This time period was chosen for two reasons: a sufficiently large number of observations was still ensured ($T = 23$), and the last two main global crises (the 2008 financial crisis and the 2020 pandemic) followed to be dealt with. Moreover, different priors were defined for the hyperparameters, allowing for different bounds. In Equation (23), ϑ_0 and ρ_0 were set to 0.90 and 0.1, respectively. In Equation (24), α_0 and ν_0 were set to 1.0 and 0.1, respectively. Thus, the posterior $\hat{\rho} = \rho_0 \cdot \bar{\nu} \cong 0.10 \cdot (0.10 \cdot \bar{\kappa})$ would be equal to 0.009 in the case of constant volatility and 0.016 in the case of volatility changes. In both cases, the posterior $\hat{\rho}$ would approximately converge at the same value. The posterior $\hat{\vartheta}$ will be always close to zero if the volatilities matter and slightly larger otherwise ($\cong 0.10$). Finally, the posterior $\bar{\alpha} = \alpha_0 \cdot \kappa$ will assume larger values, corresponding exactly to the decay factor in the case of volatility ($\cong 1.0$) and no time shift ($\cong 0.90$). According to these model specifications, the final subset of predictors is unchanged with the variable selection focusing on 'inclusion' probabilities (PMPs). Moreover, the CPSs tend to decrease for the most predictors because of not properly disentangling volatilty from coefficient changes when computing $\hat{\rho}$, the relevant scale parameter for σ^2. However, by construction, the WLB procedure would always ensure sufficient accuracy. Indeed, when letting volatility changes be treated as permanent shifts in the case of dynamically changing variables, an unexpected shock would be absorbed in a certain M_l during the shrinking procedure and then included in the final submodel solution. The estimation results are displayed in Table 2.

Regarding endogeneity issues and further sensitivity analysis, four DPD-GMM models were evaluated according to the the estimation outputs in Table 1.[7] Model 1 includes all the predictors with CPSs strictly close to one or zero over the time period of 1990–2022 (full panel set). Model 2 includes the same previous number of predictors but over the subsample of 2000–2022 (reduced panel set). Model 3 refers to the only the predictors with CPSs strictly close to one or zero dealing with only two groups (socioeconomic healthcare and demographic environment factors) over the time period of 1990–2022 (full panel subset). Model 4 refers to the same previous number of predictors over the subsample of 2000–2022 (reduced panel subset). Here, some considerations are in order. (1) When addressing endogeneity issues (i.e., the error terms are serially correlated with potential covariates violating one of the assumptions of regression models (independent and identically distributed error terms)), dynamic panel data is a useful approach for modeling unobserved heterogeneity by using correct instruments for the endogenous variables from lower to higher orders (Arellano and Bond 1991; Blundell and Bond 1998). (2) In dynamic panel models, when series show strong linear dependencies and dominance of cross-sectional variability (just as in this case), a GMM system is an efficient method for modeling these instruments for effective treatment of endogeneity biases concerning the variables in the estimation. (3) Heteroskedasticity problems are dealt with using robust standard errors across all estimations. (4) Stationarity is checked using a Fisher-type test (Choi 2001), performing well for unbalanced panel datasets (just as in this case) and assuming independently distributed normal error terms for all units (i) and time (t).

In Table 3, two main diagnostic tests to check the validity of the instruments and efficiency of the estimates in the DPD-GMM models are displayed: Sargan's test for over-identification (Q_S), highlighting the performance and usefulness of the dynamic panel in dealing with endogenity issues and functional forms of misspecification, and the Arellano–Bond test (Q_{AB}) for the first- and second- order serial correlation of residuals (Arellano and Bond 1991). Two main findings can be addressed. First, the results in Models 1 and 2 were quite close, highlighting the performance of the WLB procedure in selecting the promising subset of predictors to be estimated and the usefulness of DPD-GMM models for

use in potential econometric approaches. Second, when considering a constrained panel set, endogeneity biases and serial correlations were not efficiently minimized because of omitting potential non-health-related indicators and region-specific characteristics when studying pandemic diseases and their effects on economic development. These results are in line with the density forecasts performed in Figure 1.

Table 2. Counterfactual assessment: sensitivity analysis.

Idx.	Predictor	Unit	CPS
	MACROECONOMIC FINANCIAL INDICATORS		
1	unit labor cost	% values	0.947
2	consumer price index	% GDP	0.011
3	financial transactions	% GDP	0.951
4	employment by age (15–64)	% Tot. Pop.	0.966
5	labor force, age 15–64	logarithm	0.839
6	unemployment rate by age (15–74)	% Tot. Pop.	0.009
7	risk of poverty by age (15–74)	% Tot. Pop.	0.033
8	weighted income per capita	% logarithm	0.947
9	wage and salaried workers	% Tot. Emp.	0.928
10	gross fixed capital formation	% GDP	0.911
	SOCIOECONOMIC AND HEALTHCARE FACTORS		
11	overweight	std. rates per 100 people	0.026
12	consumption of tobacco	% adults (15+)	0.013
13	consumption of alcohol	std. rates per 100 people	0.021
14	current health expenditure	% GDP	0.934
15	*R&D* expenditure	% GDP	0.917
16	fertility rate	% Tot. Pop.	0.885
17	capital health expenditure	% GDP	0.893
18	death rate	per 1000 people	0.027
19	secondary school enrollment	% Tot. Pop.	0.884
20	social participation	% Tot. Pop.	0.817
21	tertiary educational attainment	% Tot. Pop. (25–64)	0.934
22	household price index	% GDP	0.881
	DEMOGRAPHIC AND ENVIRONMENT FACTORS		
23	rural population	% Tot. Pop.	0.794
24	urban population	% Tot. Pop.	0.814
25	population growth	% Tot. Pop.	0.962
26	total population	logarithm	0.857
27	energy use	% GDP	0.791
28	total CO_2 emission	% Tot. Pop.	0.028
29	human capital	logarithm	0.815
30	internet use	% GDP	0.776
	REGION–SPECIFIC CHARACTERISTICS		
-	d_{1t}	1 = northwestern, northeastern Europe; 0 = southwestern, southeastern Europe	
-	d_{2t}	1 = developed; 0 = developing	
-	d_{3t}	1 = high economic status; 0 = low economic status	
-	Real Growth Rate of GDP	Percentage Change	

The Table is split as follows: the first column denotes the predictor number; the second column displays the predictors; the third column refers to the measurement unit; and the last column displays the CPSs. The last row refers to the outcome of interest. The contraction Tot. Pop. stands for 'total population', and Tot. Emp. stands for 'total employed people'. All data refer to Eurostat database.

Table 3. Diagnostic test in dynamic panel regressions with GMM system.

Test Statistic	Model 1	Model 2	Model 3	Model 4
	ROBUSTNESS			
Q_S (*p*-values)	236.04 (0.00)	217.03 (0.00)	102.01 (0.03)	97.28 (0.04)
Q_{AB} (*p*-values)	2.95 (0.40)	2.73 (0.32)	1.87 (0.07)	1.66 (0.09)
	OBSERVATIONS			
Year dummies	Yes	Yes	Yes	Yes
Region dummies	Yes	Yes	No	No
Regions	277	277	277	277
Time series (*T*)	33	23	33	23

The table addresses two main test statistics with related *p* values (in parentheses) to check validity of the instruments and efficiency of the estimates in the dynamic panel regressions. They are Sargan's test for over-identification (Q_S) and the Arellano–Bond serial correlation test (Q_A). Information on the data is also displayed.

5. Concluding Remarks

This study improves the existing literature on predicting pandemic diseases and their effects on economic development across a large set of European and non-European regions. A penalized Bayesian approach-based variable selection procedure is involved to reduce the dimensionality of the model and parameter space and thus select a promising subset of predictors affecting the outcomes of interest.

An empirical example on 277 regions for 23 countries described the estimating procedure and forecasting performance, covering the period of 1990–2022. The forecast horizon referred to the years 2023 and 2024 in order to replicate the productivity dynamics in the current year and study them in the future.

According to the estimation results, conditional and unconditional density forecasts on the productivity dynamics were conducted to highlight how the promising subset of covariates would help predict potential pandemic diseases. Here, three different cases were addressed, accounting for different subsets of predictors. The aim was to highlight how the presence of endogeneity issues and model misspecification affected the estimates when dealing with time-varying parameters and high dimensionality.

From a policy perspective, the global outbreak of public health emergencies has merely focused on the health sector and disease-related costs. Thus, relevant government departments should have paid more attention to structural reforms not only for improving the public health system but also building a high-quality mode of economic growth and restructuring the global value chains. In this way, through prudent long-term policies and more specific and substantial measures, the contagion would have been better addressed with lower development times and higher efficiency in terms of market opportunities, innovation, and consumption features.

The empirical results reported herein should be considered in light of some limitations. Firstly, even if the variables were time-varying, potential volatility changes were treated as permanent shifts and then replaced by coefficient changes. Future improvements might consider, for example, time-varying log volatilities and model them through MCMC implementations (e.g., Metropolis–Hastings and expectation–maximization algorithms). Then, an in-depth analysis might reveal the existence of further regional patterns, evaluated by involving in the variable selection procedure hierarchical fuzzy Bayesian clustering algorithms. Finally, this study did not investigate the causal relationship between the outcome of interest and the subset of predictors, which was beyond the scope of this paper. Related works might consider and discuss in depth the direct and indirect causal links between them.

Author Contributions: Conceptualization, D.P.; methodology, A.P.; software, A.P.; validation, A.P.; formal analysis, A.P.; investigation, A.P.; resources, A.P. and D.P.; data curation, A.P. and D.P.; writing—original draft preparation, A.P.; writing—review and editing, A.P.; supervision, A.P.; project administration, A.P. All authors have read and agreed to the published version of the manuscript.

Notes

1 The penalties are not differentiable at zero.
2 The analysis was conducted using the codes developed in Pacifico (2020) and adjusting them to include the LASSO procedure (`glmnet()` function as starting point) and penalized terms (`if{ }` and `else{}` functions). The statistical econometric software used was RStudio version 2023.12.0.
3 The higher the value of $\hat{\rho}$, the greater the spread will be (lower peak).
4 Source: Eurostat, October 2023.
5 The analysis was conducted using Matlab software, and all density forecasts were performed while waiting for less than a minute.
6 Source: Global Gateway, February 2024.
7 The analysis was conducted by using and modeling the `pdynmc` R-Studio documentation, which fits a linear dynamic panel data model based on moment conditions with the generalized method of moments.

References

Abrhám, James, and Milan Vošta. 2022. Impact of the COVID-19 pandemic on eu convergence. *Journal of Risk and Financial Management* 15: 384. [CrossRef]

Adamek, Robert, Stephan Smeekes, and Ines Wilms. 2023. Lasso inference for high-dimensional time series. *Journal of Econometrics* 235: 1114–43. [CrossRef]

Arellano, Manuel, and Stephen Bond. 1991. Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The Review of Economic Studies* 58: 277–97. [CrossRef]

Blundell, Richard, and Stephen Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87: 115–43. [CrossRef]

Breiman, Leo, and Philip Spector. 1992. Submodel selection and evaluation in regression. the x-random case. *International Statistical Review* 60: 291–319. [CrossRef]

Choi, In. 2001. Unit root tests for panel data. *Journal of International Money and Finance* 20: 249–72. [CrossRef]

Fan, Jianqing, and Runze Li. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* 96: 1348–60. [CrossRef]

Gelfand, Alan E., and Dipak K. Dey. 1994. Bayesian model choice: Asymptotics and exact calculations. *Journal of the Royal Statistical Society: Series B* 56: 501–14.

Gungoraydinoglu, Ali, Ilke Öztekin, and Özde Öztekin. 2021. The impact of COVID-19 and its policy responses on local economy and health conditions. *Journal of Risk and Financial Management* 14: 233. [CrossRef]

Ismail, Shah, Hina Naz, Sajid Ali, Amani Almohaimeed, and Showkat A. Lone. 2023. A new quantile-based approach for lasso estimation. *Mathematics* 11: 1452. [CrossRef]

Jang, Dae-Heung, and Christine M. Anderson-Cook. 2016. Influence plots for lasso. *Journal of Risk and Financial Management* 33: 1317–26. [CrossRef]

Kass, Robert E., and Adrian E. Raftery. 1995. Bayes factors. *Journal of American Statistical Association* 90: 773–95. [CrossRef]

Lu, Wenbin, Yair Goldberg, and J. P. Fine. 2012. On the robustness of the adaptive lasso to model misspecification. *Biometrika* 99: 717–31. [CrossRef]

Madigan, David, and Adrian E. Raftery. 1994. Model selection and accounting for model uncertainty in graphical models using occam's window. *Journal of American Statistical Association* 89: 1535–46. [CrossRef]

Madigan, David, Jeremy York, and Denis Allard. 1995. Bayesian graphical models for discrete data. *International Statistical Review* 63: 215–32. [CrossRef]

Masini, Ricardo P., Marcelo C. Medeiros, and Eduardo F. Mendes. 2022. Regularized estimation of high-dimensional vector autoregressions with weakly dependent innovations. *Journal of Time Series Analysis* 43: 532–57. [CrossRef]

Medeiros, Marcelo C., and Eduardo F. Mendes. 2016. l1-regularization of high-dimensional time-series models with non-gaussian and heteroskedastic errors. *Journal of Econometrics* 191: 255–71. [CrossRef]

Mohammad, Arashi, Asar Yasin, and Yüzbaşi Bahadır. 2021. Slasso: A scaled lasso for multicollinear situations. *Journal of Statistical Computation and Simulation* 91: 3170–83. [CrossRef]

Pacifico, Antonio. 2020. Robust open bayesian analysis: Overfitting, model uncertainty, and endogeneity issues in multiple regression models. *Econometric Reviews* 40: 148–76. [CrossRef]

Pulawska, Karolina. 2021. Financial stability of european insurance companies during the COVID-19 pandemic. *Journal of Risk and Financial Management* 14: 266. [CrossRef]

Raftery, Adrian E., David Madigan, and Chris T. Volinsky. 1995. Accounting for model uncertainty in survival analysis improves predictive performance. *Bayesian Statistics* 6: 323–49.

Raftery, Adrian E., David Madigan, and Jennifer A. Hoeting. 1997. Bayesian model averaging for linear regression models. *Journal of American Statistical Association* 92: 179–91. [CrossRef]

Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B* 58: 267–88. [CrossRef]

Uematsu, Yoshimasa, and Takashi Yamagata. 2023. Inference in sparsity-induced weak factor models. *Journal of Business and Economic Statistics* 41: 126–39. [CrossRef]

Vidaurre, Diego, Concha Bielza, and Pedro Larrañaga. 2013. A survey of l1 regression. *International Statistical Review* 81: 361–87. [CrossRef]

Wenjiang, J. Fu. 1998. Penalized regressions: The bridge versus the lasso. *Journal of Computational and Graphical Statistics* 7: 397–416.

Wong, Kam C., Zifan Li, and Ambuj Tewari. 2020. Lasso guarantees for beta-mixing heavy-tailed time series. *Annals of Statistics* 48: 1124–42. [CrossRef]

Wu, Wei B., and Ying N. Wu. 2016. Performance bounds for parameter estimates of high-dimensional linear models with correlated errors. *Electronic Journal of Statistics* 10: 352–79. [CrossRef]

Yang, Xiaoxing, and Wushao Wen. 2018. Ridge and lasso regression models for cross-version defect prediction. *IEEE Transactions on Reliability* 67: 885–96. [CrossRef]

Zhang, Cun-Hui, and Stephanie S. Zhang. 2014. Confidence intervals for low dimensional parameters in high dimensional linear models. *Journal of the Royal Statistical Society: Series B* 76: 217–42. [CrossRef]

Zou, Hui. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* 101: 1418–29. [CrossRef]

Article
Almost Perfect Shadow Prices

Eberhard Mayerhofer

Department of Mathematics & Statistics, University of Limerick, V94 T9PX Limerick, Ireland;
eberhard.mayerhofer@ul.ie

Abstract: Shadow prices simplify the derivation of optimal trading strategies in markets with transaction costs by transferring optimization into a more tractable, frictionless market. This paper establishes that a naïve shadow price ansatz for maximizing long-term returns, given average volatility yields a strategy that is, for small bid–ask spreads, asymptotically optimal at the third order. Considering the second-order impact of transaction costs, such a strategy is essentially optimal. However, for risk aversion different from one, we devise alternative strategies that outperform the shadow market at the fourth order. Finally, it is shown that the risk-neutral objective rules out the existence of shadow prices.

Keywords: transaction costs; portfolio choice; shadow prices; reflected diffusions

MSC: 91G10; 91G80

1. Introduction

With a little help of my friends.[1]

—The Beatles

Shadow prices (or *consistent price systems*) are a tool for characterizing the absence of arbitrage opportunities in markets with proportional transaction costs (see, for example, Czichowsky and Schachermayer (2016); Guasoni et al. (2010); Kabanov et al. (2002)), or for deriving optimal strategies for various objectives (see, for example, Guasoni and Muhle-Karbe (2013); Gerhold et al. (2013, 2014); Herdegen et al. (2023); Kallsen and Muhle-Karbe (2010)). This paper investigates the applicability of shadow prices to the optimization of long-term returns given average volatility.

Strategies that are optimal in frictionless markets[2] such as the delta-hedging of European-type options, or constant proportion strategies, lead to immediate bankruptcy under proportional costs.[3] To ensure solvency, trading frequency needs to be modulated to finite variation, trading as little as necessary to stay close to the target exposures. This paper relates to the objectives of long-run investors (Gerhold et al. (2014); Guasoni and Mayerhofer (2019, 2023); Taksar et al. (1988))[4], who consider it optimal to keep the fraction of wealth π invested in the risky asset within an interval around a target exposure by engaging only in trading whenever this fraction hits the boundaries of the interval.[5] For example, for constant investment opportunities and a sufficiently small relative bid–ask spread ε, the trading boundaries $\pi_- < \pi_+$ of an investor with risk-aversion γ are approximately

$$\pi_\pm = \pi_* \pm \left(\frac{3}{4\gamma}\pi_*^2(\pi_* - 1)^2\right)^{1/3}\varepsilon^{1/3}, \tag{1}$$

where $\pi_* = \frac{\mu}{\gamma\sigma^2}$ is the well-known Merton fraction, μ being the annualized average return of the risky asset, and σ its volatility.

Absurdly, such finite variation strategies are mathematically more challenging than the infinite variation strategies typically encountered in frictionless markets. Shadow prices

allow to transfer optimization into a more tractable, frictionless (but fictitious) market. A shadow price is a frictionless asset that evolves in the bid–ask spread of the risky asset and for which the optimal strategy buys (respectively sells) whenever its price agrees with the ask (respectively bid) of this risky asset. For objectives which are monotone functions of wealth, such as power utility, the strategy in the shadow market is also optimal in the original market because by trading in the shadow market, the investor is generally better off. Furthermore, shadow markets provide an elegant, intuitive derivation of optimal trading policies for different objectives and market models. It therefore may come as surprise that Guasoni and Mayerhofer (2019, 2023) use the more traditional Hamilton–Jacobi–Bellman equations both for the heuristic derivation of the candidate optimal control limit policy (with asymptotics (1)) *and* the verification of optimality. This is even more surprising, as the respective objectives lend themselves to very tractable candidate shadow prices and trading strategies (see Section 3.1 below). However, the local mean variance criterion is, in general, not monotone in wealth.[6] Therefore, the verification of optimality fails, leaving open the question of whether trading strategies derived in the shadow market are also optimal in the original market with transaction costs.

Guasoni and Mayerhofer (2019) show that, in the presence of transaction costs, maximizing returns is well posed, even without controlling for volatility—transaction costs act as a penalty in the objective. As a consequence, the efficient frontier is not a straight line as in the classical Merton problem but reaches a maximum for finite volatility, after which taking on even further risk may result in negative alpha. However, in frictionless markets, such an objective gives the incentive to seek arbitrary leverage, unless the asset has zero expected excess return. Thus, shadow prices are destined to fail as an optimization tool.

Nevertheless, a candidate shadow price can be found for a risk-averse investor. A construction similar to Gerhold et al. (2013) yields trading policies of the form (1), and thus, they are indistinguishable at the first order from the optimal one. Moreover, at the second order, they are distinguished by a mere change of sign in the second-order coefficient. Even more surprisingly, the equivalent safe rate of the shadow price trading strategy agrees at the third order with the maximum. In view of the second-order impact of transaction costs, it is essentially optimal. However, we devise trading policies that strictly outperform the shadow price trading strategy at fourth order.

Program of Paper

The paper is structured as follows: Section 2 presents the market model, encompassing a risky Black–Scholes asset with transaction costs, the mean–variance objective, and a recap of the optimal trading policies established in Guasoni and Mayerhofer (2019). Section 2.1 introduces control limit policies, evaluating their long-run performance along with small-transaction cost asymptotics (Lemma 3). In Section 3.1, a naïve ansatz for a shadow price is proposed, and asymptotic expansions of the trading boundaries are provided. Theorem 2 demonstrates their third-order asymptotic optimality and Theorem 3 establishes their strict sub-optimality. Section 3.3 provides a rigorous proof that for maximizing expected returns without controlling for volatility, no shadow price exists (Theorem 4). The final Section 4 summarizes our findings and points out directions for future research. The appendix computes a high-order approximation of the candidate shadow price to support the proof of Theorem 3.

2. Materials and Methods

The market model is comprised of two assets: a safe safe asset that is continuously compounded at a constant rate of $r \geq 0$ and a risky asset S purchased at its ask price S_t and satisfying the dynamics

$$\frac{dS_t}{S_t} = (\mu + r)dt + \sigma dB_t, \quad S_0, \sigma, \mu > 0, \tag{2}$$

where B is a standard Brownian motion. The risky asset's bid (selling) price is $(1-\varepsilon)S_t$, which implies a constant relative bid–ask spread of $\varepsilon > 0$, or, equivalently, constant proportional transaction costs.

Let w be the wealth associated with a self-financing trading strategy[7]. The mean–variance trade-off is captured by maximizing the equivalent safe rate,

$$\text{ESR} := \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T \frac{dw_t}{w_t} - \frac{\gamma}{2}\left\langle\int_0^{\cdot} \frac{dw_t}{w_t}\right\rangle_T\right]. \tag{3}$$

With π, the proportion of wealth invested in the risky asset, and with φ_t, the number of shares $\varphi_t = \varphi_t^{\uparrow} - \varphi_t^{\downarrow}$ being the difference of purchases $\varphi_t^{\uparrow}$ minus sales $\varphi_t^{\downarrow}$, one can rewrite the objective[8] as follows

$$\text{ESR} := r + \limsup_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T \left(\mu\pi_t - \frac{\gamma\sigma^2}{2}\pi_t^2\right)dt - \varepsilon\int_0^T \pi_t\frac{d\varphi_t^{\downarrow}}{\varphi_t}\right]. \tag{4}$$

In the absence of transaction costs ($\varepsilon = 0$), the objective is maximized by the constant proportion portfolio $\pi_* := \frac{\mu}{\gamma\sigma^2}$ dating back to Markowitz and Merton. The risk-neutral objective $\gamma = 0$ reduces to the average annualized return over a long horizon, which is well posed for transaction costs (Guasoni and Mayerhofer 2019, Theorem 3.2) but meaningless in the traditional framework with zero bid–ask spread, where a strategy can be arbitrarily levered. The case $\gamma = 1$ reduces to logarithmic utility, which is solved by the Taksar et al. (1988) for the unlevered case $\frac{\mu}{\gamma\sigma^2} < 1$.

An optimal strategy maximizing the equivalent safe rate exists. The following is a shortened version of (Guasoni and Mayerhofer 2019, Theorem 3.1), characterizing optimality.

Theorem 1. *Let $\frac{\mu}{\gamma\sigma^2} \neq 1$.*

1. *For any $\gamma > 0$, there exists $\varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$, there is a unique solution (W, ζ_-, ζ_+), with $\zeta_- < \zeta_+$ for the free boundary problem*

$$\frac{1}{2}\sigma^2\zeta^2 W''(\zeta) + (\sigma^2 + \mu)\zeta W'(\zeta) + \mu W(\zeta) - \frac{1}{(1+\zeta)^2}\left(\mu - \gamma\sigma^2\frac{\zeta}{1+\zeta}\right) = 0, \tag{5}$$

$$W(\zeta_-) = 0 \tag{6}$$

$$W'(\zeta_-) = 0, \tag{7}$$

$$W(\zeta_+) = \frac{\varepsilon}{(1+\zeta_+)(1+(1-\varepsilon)\zeta_+)}, \tag{8}$$

$$W'(\zeta_+) = \frac{\varepsilon(\varepsilon - 2(1-\varepsilon)\zeta_+ - 2)}{(1+\zeta_+)^2(1+(1-\varepsilon)\zeta_+)^2} \tag{9}$$

2. *The trading strategy that buys at $\pi_- := \zeta_-/(1+\zeta_-)$ and sells at $\pi_+ := \zeta_+/(1+\zeta_+)$ as little as possible to keep the risky weight π_t within the interval $[\pi_-, \pi_+]$ is optimal.*
3. *The maximum performance is*

$$\max_{\varphi\in\Phi} \lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\int_0^T\left(\mu\pi_t - \frac{\gamma\sigma^2}{2}\pi_t^2\right)dt - \varepsilon\int_0^T\pi_t\frac{d\varphi_t^{\downarrow}}{\varphi_t}\right] = \mu\pi_- - \frac{\gamma\sigma^2}{2}\pi_-^2, \tag{10}$$

where Φ is the set of admissible strategies in Definition 1 below, $\varphi_t = \pi_t w_t/S_t$ is the number of shares held at time t, and $\varphi_t^{\downarrow}$ is the cumulative number of shares sold up to time t.

4. *The trading boundaries π_- and π_+ have the asymptotic expansions*

$$\pi_{\pm} = \pi_* \pm \left(\frac{3}{4\gamma}\pi_*^2(\pi_*-1)^2\right)^{1/3}\varepsilon^{1/3} - \frac{(1-\gamma)\pi_*}{\gamma}\left(\frac{\gamma\pi_*(\pi_*-1)}{6}\right)^{1/3}\varepsilon^{2/3} + O(\varepsilon). \tag{11}$$

5. *The equivalent safe rate (ESR) has the expansion*

$$\text{ESR} = r + \frac{\gamma\sigma^2}{2}\pi_*^2 - \frac{\gamma\sigma^2}{2}\left(\frac{3}{4\gamma}\pi_*^2(\pi_* - 1)^2\right)^{2/3}\varepsilon^{2/3} + O(\varepsilon). \tag{12}$$

2.1. Admissible Strategies and Their Long-Run Performance

In view of transaction costs, only finite-variation trading strategies are consistent with solvency. This is illustrated by the following example.

Example 1. *Consider the dynamic hedging part of $1/\varepsilon$ variance swaps[9] on the asset S with maturity $T = 2$, that requires to hold*

$$\varphi_t = \frac{1}{\varepsilon S_t}$$

units of the underlying at each time $t \geq 0$. Trading discretely, along a mesh of size Δ, one needs to sell at $t + \Delta$ if and only if $S_{t+\Delta} > S_t$, which incurs a cost of

$$\varepsilon \times 1/\varepsilon \times S_{t+\Delta}(1/S_t - 1/S_{t+\Delta}) = (S_{t+\Delta}/S_t - 1).$$

Let $x_+ = \max(0, x)$ and $N = T/\Delta$, then the total transaction cost amounts to

$$C_N = \sum_{i=0}^{N-1}(S_{(i+1)\Delta}/S_{i\Delta} - 1)_+.$$

Note that this sum counts all positive simple returns of the asset, which can be approximated by logarithmic returns. Thus, as $N \to \infty$, $C_N \to C$, the semivariation of a Brownian motion B with drift,

$$C = \lim_{\Delta t \to 0}\sum_{i=0}^{T/\Delta t - 1}(B_{(i+1)\Delta t} - B_{i\Delta t})_+ = \infty,$$

almost surely. This shows that, under proportional transaction costs, such a dynamic trading strategy results in immediate bankruptcy.

Denote by X_t and Y_t the wealth in the safe and risky positions, respectively, and by $(\varphi_t^\uparrow)_{t\geq 0}$ and $(\varphi_t^\downarrow)_{t\geq 0}$, the cumulative number of shares bought and sold, respectively. The self-financing condition prescribes that (X, Y) satisfies the dynamics

$$dX_t = rX_t dt - S_t d\varphi_t^\uparrow + (1-\varepsilon)S_t d\varphi_t^\downarrow, \quad dY_t = S_t d\varphi_t^\uparrow - S_t d\varphi_t^\downarrow + \varphi_t dS_t. \tag{13}$$

A strategy is admissible if it is non-anticipative and solvent, up to a small increase in the spread.

Definition 1. *Let $x > 0$ (the initial capital) and let $(\varphi_t^\uparrow)_{t\geq 0}$ and $(\varphi_t^\downarrow)_{t\geq 0}$ be continuous, increasing processes, adapted to the augmented natural filtration of B. Then, $(x, \varphi_t = \varphi_t^\uparrow - \varphi_t^\downarrow)$ is an admissible trading strategy if the following apply:*

1. *Its liquidation value is strictly positive at all times: there exists $\varepsilon' > \varepsilon$ such that the discounted asset $\widetilde{S}_t := e^{-rt}S_t$ satisfies*

$$x - \int_0^t \widetilde{S}_s d\varphi_s + \widetilde{S}_t\varphi_t - \varepsilon'\int_0^t \widetilde{S}_s d\varphi_s^\downarrow - \varepsilon'\varphi_t^+\widetilde{S}_t > 0 \qquad a.s. \text{ for all } t \geq 0. \tag{14}$$

2. *The following integrability conditions hold:*

$$\mathbb{E}\left[\int_0^t |\pi_u|^2 du\right] < \infty, \quad \mathbb{E}\left[\int_0^t \pi_u \frac{d\|\varphi_u\|}{\varphi_u}\right] < \infty \quad \text{for all } t \geq 0, \tag{15}$$

where $\|\varphi_t\|$ denotes the total variation of φ on $[0, t]$.

The family of admissible trading strategies is denoted by Φ.

The following lemma describes the dynamics of the wealth process w_t, the risky weight π_t, and the risky-safe ratio ζ_t.

Lemma 1. *For any admissible trading strategy φ:*

$$\frac{d\zeta_t}{\zeta_t} = \mu dt + \sigma dB_t + (1 + \zeta_t)\frac{d\varphi_t^{\uparrow}}{\varphi_t} - (1 + (1 - \varepsilon)\zeta_t)\frac{d\varphi_t^{\downarrow}}{\varphi_t}, \tag{16}$$

$$\frac{dw_t}{w_t} = rdt + \pi_t(\mu dt + \sigma dB_t - \varepsilon\frac{d\varphi_t^{\downarrow}}{\varphi_t}), \tag{17}$$

$$\frac{d\pi_t}{\pi_t} = (1 - \pi_t)(\mu dt + \sigma dB_t) - \pi_t(1 - \pi_t)\sigma^2 dt + \frac{d\varphi_t^{\uparrow}}{\varphi_t} - (1 - \varepsilon\pi_t)\frac{d\varphi_t^{\downarrow}}{\varphi_t}. \tag{18}$$

For any such strategy, the functional

$$F_T(\varphi) := \frac{1}{T}\mathbb{E}\left[\int_0^T \frac{dw_t}{w_t} - \frac{\gamma}{2}\left\langle \int_0^T \frac{dw_t}{w_t}\right\rangle_T\right] \tag{19}$$

can be rewritten as

$$F_T(\varphi) = r + \frac{1}{T}\mathbb{E}\left[\int_0^T \left(\mu\pi_t - \frac{\gamma\sigma^2}{2}\pi_t^2\right)dt - \varepsilon\int_0^T \pi_t\frac{d\varphi_t^{\downarrow}}{\varphi_t}\right]. \tag{20}$$

Proof. See (Guasoni and Mayerhofer 2019, Lemma A.2). □

Lemma 2. *Let $\eta_- < \eta_+$ be such that either $\eta_+ < -1/(1 - \varepsilon)$ or $\eta_- > 0$. Then, there exists an admissible trading strategy $\hat{\varphi}$ such that the risky-safe ratio η_t satisfies SDE (16). Moreover, $(\eta_t, \hat{\varphi}_t^{\uparrow}, \hat{\varphi}_t^{\downarrow})$ is a reflected diffusion on the interval $[\eta_-, \eta_+]$. In particular, η_t has stationary density equal to*

$$\nu(\eta) := \frac{\frac{2\mu}{\sigma^2} - 1}{\eta_+^{\frac{2\mu}{\sigma^2} - 1} - \eta_-^{\frac{2\mu}{\sigma^2} - 1}}\eta^{\frac{2\mu}{\sigma^2} - 2}, \quad \eta \in [\eta_-, \eta_+], \tag{21}$$

when $\eta_- > 0$, and otherwise equals

$$\nu(\eta) := \frac{\frac{2\mu}{\sigma^2} - 1}{|\eta_-|^{\frac{2\mu}{\sigma^2} - 1} - |\eta_+|^{\frac{2\mu}{\sigma^2} - 1}}|\eta|^{\frac{2\mu}{\sigma^2} - 2}, \quad \eta \in [\eta_-, \eta_+]. \tag{22}$$

Proof. See (Guasoni and Mayerhofer 2019, Lemma B.5). □

Definition 2. *For the rest of the paper, the strategy in Lemma 2 is called "control limit policy for $\eta_{\pm}$", an adaption of the name of similar policies in Taksar et al. (1988), where "limit" actually relates to the boundaries of the interval $[\eta_-, \eta_+]$. Note that the strategy in Theorem 1 (2) is exactly of this kind: it entails no trading as long as $\zeta_t \in (\zeta_-, \zeta_+)$ and trades as little as necessary at $\zeta_{\pm}$ to keep the risky-safe ratio in the interval $[\zeta_-, \zeta_+]$. Alternative strategies, such as trading into the middle of the no-trade region, incur significantly larger transaction costs.[10]*

The following computes the statistics contributing to the ESR of any trading strategy as in Lemma 2 (not just the optimal one) in terms of the risky-safe ratio.

Lemma 3. *Consider a control limit policy for $\eta_\pm$. Long-run mean $\hat{m}$, long-run standard deviation $\hat{\sigma}$ and average transaction costs ATC are given by the almost sure limits,*

$$\hat{m} = \lim_{T\to\infty} \frac{1}{T} \int_0^T \frac{dw_t}{w_t} dt = r + \mu \int_{\eta_-}^{\eta_+} \left(\frac{\zeta}{1+\zeta} \right) \nu(d\eta), \tag{23}$$

$$\hat{\sigma}^2 = \lim_{T\to\infty} \frac{1}{T} \left\langle \int_0^{\cdot} \frac{dw_t}{w_t} \right\rangle_T = \sigma^2 \int_{\eta_-}^{\eta_+} \left(\frac{\zeta}{1+\zeta} \right)^2 \nu(d\eta), \tag{24}$$

$$\text{ATC} = \varepsilon \lim_{T\to\infty} \frac{1}{T} \int_0^T \pi_t \frac{d\varphi_t^{\downarrow}}{\varphi_t} = \frac{\sigma^2(2\mu/\sigma^2 - 1)}{2} \left(\frac{\frac{\varepsilon\zeta_+}{(1+\zeta_+)(1+(1-\varepsilon)\zeta_+)}}{1 - \left(\frac{\zeta_-}{\zeta_+} \right)^{2\mu/\sigma^2 - 1}} \right), \tag{25}$$

where ν is the stationary density of Lemma 2.

Proof. All the formulae use the ergodic theorem and thus can be obtained with the methods of Guasoni and Mayerhofer (2019). In particular, identity (25) holds in the view of (Gerhold et al. 2014, Lemma C.1). $\square$

Using the analytic expressions of (23)–(25) with MATHEMATICA, we obtain explicit asymptotics, precise at the third order in $\varepsilon^{1/3}$:

Lemma 4. *For the optimal strategy of Theorem 1, the statistics of Lemma 3 satisfy the following asymptotics:*

$$\hat{m} = r + \frac{\mu^2}{\gamma\sigma^2} - \frac{\mu(2\pi_* - 1)}{\gamma} \left(\frac{\gamma\pi_*(\pi_* - 1)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon^{4/3}), \tag{26}$$

$$\hat{\sigma}^2 = \frac{\mu^2}{\gamma^2\sigma^2} - \frac{\sigma^2\pi_*(7\pi_* - 3)}{2\gamma} \left(\frac{\gamma\pi_*(\pi_* - 1)}{6} \right)^{1/3} \varepsilon^{2/3} + O(\varepsilon^{4/3}), \tag{27}$$

$$\text{ATC} = \frac{3\sigma^2}{\gamma} \left(\frac{\gamma\pi_*(\pi_* - 1)}{6} \right)^{4/3} \varepsilon^{2/3} - \frac{\mu(\gamma - 1)}{2\gamma} \pi_*(\pi_* - 1)\varepsilon + O(\varepsilon^{4/3}). \tag{28}$$

The maximum equivalent safe rate satisfies

$$\text{ESR} = r + \frac{\gamma\sigma^2}{2}\pi_*^2 - \frac{\gamma\sigma^2}{2} \left(\frac{3}{4\gamma} \pi_*^2 (\pi_* - 1)^2 \right)^{2/3} \varepsilon^{2/3} + \frac{\mu(\gamma - 1)}{2\gamma} \pi_*(\pi_* - 1)\varepsilon + O(\varepsilon^{4/3}). \tag{29}$$

Remark 1.

1. *All asymptotics of Lemma 4 improve those of (Guasoni and Mayerhofer 2019, Theorem 3.1) in precision by one order. Note that (26) is a corrected version of (Guasoni and Mayerhofer 2019, Theorem 3.1, eq. (3.8)), where the bracket $(5\pi_* - 3)$ is given, instead of the correct term $(2\pi_* - 1)$ in (26).*
2. *One can run a consistency check that compares the asymptotics (29) of the maximum ESR (computed, by developing $r + \hat{m} - \frac{\gamma}{2}\hat{\sigma}^2 - \text{ATC}$ into a formal power series in $\varepsilon^{1/3}$) with the asymptotic expansion of the shorter formula $r + \mu\pi_- - \frac{\gamma\sigma^2}{2}\pi_-^2$ from Theorems 1 and 3.*

3. Results

3.1. Asymptotically Optimal Shadow Policies

In this section, a shadow price for the mean–variance objective (3) is constructed, and asymptotic formulas for the implied strategy that is optimal in the shadow market, are derived. The exposition is motivated by the shadow price construction for log-utility investors, cf. (Gerhold et al. 2013, Chapter 3), and see also Guasoni and Muhle-Karbe (2013); Gerhold et al. (2014). Assume the following functional form of the shadow price $\widetilde{S}_t$,

$$\widetilde{S}_t = g(\pi_t)S_t, \tag{30}$$

where g satisfies the boundary conditions

$$g(\pi_-) = 1, \quad g(\pi_+) = (1 - \varepsilon), \tag{31}$$

reflecting that an optimal strategy (such as of Theorem 1) is a control limit policy for $\pi_\pm$, which buys (respectively sells) the frictionless asset $\widetilde{S}$ precisely when its price equals the ask price S, and sells precisely when it equals the bid price $(1 - \varepsilon)S$.

If $\widetilde{S}$ satisfies (30) with twice differentiable g, and if g satisfies (34), then Itô's formula yields the dynamics of instantaneous returns[11]

$$\frac{d\widetilde{S}_t}{\widetilde{S}_t} = rdt + d\widetilde{\mu}_t + \widetilde{\sigma}_t dB_t,$$

with

$$d\widetilde{\mu}_t = \mu dt + \frac{g'(\pi_t)\big(\pi_t(1 - \pi_t)\mu dt + \pi_t(1 - \pi_t)^2\sigma^2\big)}{g(\pi_t)} \tag{32}$$

$$+ \frac{\frac{1}{2}g''(\pi_t)\pi_t^2(1 - \pi_t)^2\sigma^2 dt + g'(\pi_t)\left(\pi_t\frac{d\varphi_t^{\uparrow}}{\varphi_t} - (1 - \varepsilon\pi_t)\frac{d\varphi_t^{\downarrow}}{\varphi_t}\right)}{g(\pi_t)}$$

and diffusion coefficient

$$\widetilde{\sigma}_t = (\sigma g(\pi_t) + g'(\pi_t)\pi_t(1 - \pi_t)\sigma)/g(\pi_t). \tag{33}$$

The smooth pasting condition

$$g'(\pi_-) = g'(\pi_+) = 0 \tag{34}$$

is imposed such that the instantaneous drift of the shadow price becomes absolutely continuous (the condition removes the local time terms $\frac{d\varphi_t^{\uparrow}}{\varphi_t}$ and $\frac{d\varphi_t^{\downarrow}}{\varphi_t}$), and thus $d\widetilde{\mu}_t = \widetilde{\mu}_t dt$, with

$$\widetilde{\mu}_t = \mu + \frac{g'(\pi_t)(\pi_t(1 - \pi_t)\mu + \pi_t(1 - \pi_t)^2\sigma^2) + \frac{1}{2}g''(\pi_t)\pi_t^2(1 - \pi_t)^2\sigma^2}{g(\pi_t)}. \tag{35}$$

The fraction of wealth $\widetilde{\pi}$ invested in the risky asset, evaluated at the shadow price, satisfies

$$\widetilde{\pi}_t = \frac{Y_t g(\pi_t)}{X_t + Y_t g(\pi_t)} = \frac{\pi_t g(\pi_t)}{(1 - \pi_t) + \pi_t g(\pi_t)}. \tag{36}$$

The mean–variance optimality in the shadow market holds when the proportion of wealth in the shadow market's risky asset $\widetilde{S}$ equals the Merton fraction, that is,

$$\widetilde{\pi}_t = \frac{\widetilde{\mu}_t}{\gamma\widetilde{\sigma}_t^2}.$$

Equating this solution with (36), and using (35), (33) entails that g satisfies the ODE

$$\frac{1}{2}g''(\pi)\pi^2(1 - \pi)^2\sigma^2 = \frac{\gamma\pi\sigma^2(g + g'(\pi)\pi(1 - \pi))^2}{1 - \pi + \pi g(\pi)} - \mu g(\pi) \tag{37}$$

$$- g'(\pi)(\pi(1 - \pi)\mu + \pi(1 - \pi)^2\sigma^2).$$

Define Ψ implicitly as

$$g(\pi) = \frac{\Psi(Y/X)}{Y/X} =: \frac{\Psi(\zeta)}{\zeta},$$

and set

$$\zeta_\pm = \frac{\pi_\pm}{1 - \pi_\pm}. \tag{38}$$

Then, (Ψ, ζ_-, ζ_+) satisfy the problem

$$\Psi''(\zeta) = \frac{2\gamma \Psi'^2(\zeta)}{(1 + \Psi(\zeta))} - \frac{2\mu}{\sigma^2}\frac{\Psi'(\zeta)}{\zeta}, \tag{39}$$

$$\Psi'(\zeta_-) = \Psi(\zeta_-)/\zeta_- = 1, \tag{40}$$

$$\Psi'(\zeta_+) = \Psi(\zeta_+)/\zeta_+ = (1 - \varepsilon). \tag{41}$$

This is a free boundary problem, because both Ψ and the trading boundaries $\zeta_\pm$ for the control limit policy are unknown.

Using the explicit solution Ψ of the corresponding initial value problem (39) and (40), and respecting terminal conditions (41), one obtains a non-linear system of equations for (ζ_-, ζ_+). For small ε, this very system allows a unique solution with asymptotic expansion[12]

$$\widetilde{\zeta}_\pm = \frac{\pi_*}{1 - \pi_*} \pm \left(\frac{3}{4\gamma}\right)^{1/3}\left(\frac{(\pi_*)^2}{1 - \pi_*}\right)^{2/3}\varepsilon^{1/3} - \frac{(1 + 2\gamma)\pi_*}{2\gamma(1 - \pi_*)^2}\left(\frac{\gamma\pi_*(\pi_* - 1)}{6}\right)^{1/3}\varepsilon^{2/3} + O(\varepsilon). \tag{42}$$

In comparison, the optimal strategy of Theorem 1 is a control limit policy whose limits $\zeta_\pm$, in terms of the risky-safe ratio, have the expansion[13]

$$\zeta_\pm = \frac{\pi_*}{1 - \pi_*} \pm \left(\frac{3}{4\gamma}\right)^{1/3}\left(\frac{\pi_*}{(\pi_* - 1)^2}\right)^{2/3}\varepsilon^{1/3} - \frac{(5 - 2\gamma)\pi_*}{2\gamma(\pi_* - 1)^2}\left(\frac{\gamma\pi_*(\pi_* - 1)}{6}\right)^{1/3}\varepsilon^{2/3} + O(\varepsilon). \tag{43}$$

Note the factor $(1 + 2\gamma)$ in the $\varepsilon^{2/3}$ term in (42), which differs from the factor $(5 - 2\gamma)$ in (43). Accordingly, the associated trading boundaries have an asymptotic expansion,

$$\widetilde{\pi}_\pm = \pi_* \pm \left(\frac{3}{4\gamma}(\pi_*)^2(1 - \pi_*)^2\right)^{1/3}\varepsilon^{1/3} + \frac{(1 - \gamma)\pi_*}{\gamma}\left(\frac{\gamma\pi_*(\pi_* - 1)}{6}\right)^{1/3}\varepsilon^{2/3} + O(\varepsilon). \tag{44}$$

The expansion (42), respectively, (44), agrees with the above expansions (43), respectively, (11), up to the first order (they agree in constant and in $\varepsilon^{1/3}$ terms). But they disagree in a quite subtle way for any $\gamma \neq 1$ at the second order: the absolute values, but not the signs of the second-order term of $\pi_\pm$ (see (44)) and $\widetilde{\pi}_\pm$ (see (11)), are identical.

The following establishes asymptotic optimality of the third order of the strategy obtained from the shadow market (the proof exclusively uses MATHEMATICA and higher-order expansions of (44)).

Theorem 2. *The asymptotic expansion of the equivalent safe rate and average transaction costs of the control limit policy for $\widetilde{\pi}_\pm$ are of the exact same form as (29), respectively, (25). Thus, the strategy is asymptotically optimal at the third order. The long-run mean and variance defined by Lemma 3 satisfy the following asymptotics:*

$$\widetilde{m} = r + \frac{\mu^2}{\gamma\sigma^2} - \frac{\mu(2\gamma\pi_* - 1)}{\gamma}\left(\frac{\gamma\pi_*(\pi_* - 1)}{6}\right)^{1/3}\varepsilon^{2/3} + O(\varepsilon^{4/3}), \tag{45}$$

$$\widetilde{\sigma}^2 = \frac{\mu^2}{\gamma^2\sigma^2} - \frac{\sigma^2\pi_*(\pi_*(8\gamma + 1) - 3)}{2\gamma}\left(\frac{\gamma\pi_*(\pi_* - 1)}{6}\right)^{1/3}\varepsilon^{2/3} + O(\varepsilon^{4/3}). \tag{46}$$

Remark 2. *Note that, almost miraculously,*

$$\hat{m} - \frac{\gamma}{2}\hat{\sigma}^2 = \widetilde{m} - \frac{\gamma}{2}\widetilde{\sigma}^2 = O(\varepsilon^{4/3})$$

because the average transaction costs as well as the equivalent safe rate agree for both strategies up to the third order, and the mean and variance's third-order terms vanish (compare (45) and (46) with the mean and variance of the optimal strategy in Lemma 4).

3.2. Outperforming the Shadow Market

In Theorem 2, it is shown that $\widetilde{S}$ is an asymptotic shadow price, as the strategy that is optimal in the frictionless market is even optimal at the third order in the original market. For the proof of this statement, it is crucial to have precise asymptotic expansions of the trading boundaries $\widetilde{\pi}_{\pm}$.

The objective of this section is to prove that this strategy is not optimal. To this end, it would be useful to have higher (fourth and fifth) order terms in the expansion of the optimal trading boundaries $\zeta_{\pm}$ and thus, the maximum performance (29). However, the free boundary problem of (A1)–(A5) associated with the optimal solution of Theorem 1 is notoriously difficult to deal with, even with MATHEMATICA, while the free boundary problem (A9) and (A10) arising from the shadow price ansatz is much more tractable. Therefore, instead of developing the maximum performance to even higher precision, a strategy is found that merely outperforms the shadow market.

Theorem 3. *Suppose $\gamma \notin \{0, 1\}$. For any $\theta \in \mathbb{R}$, the family of control limit policies for*

$$\widetilde{\pi}_{\pm}^{\theta} := \widetilde{\pi}_{\pm} + (\theta - 1) \times \frac{(\gamma - 1)(\pi_*)^2 (1 - \pi_*)}{6} \left(\frac{6}{\gamma \pi_* (1 - \pi_*)} \right)^{2/3} \varepsilon^{2/3} \tag{47}$$

has an equivalent safe rate

$$\begin{aligned} \text{ESR} = r &+ \frac{\gamma \sigma^2}{2} \pi_*^2 - \frac{\gamma \sigma^2}{2} \left(\frac{3}{4\gamma} \pi_*^2 (\pi_* - 1)^2 \right)^{2/3} \varepsilon^{2/3} + \frac{\mu(\gamma - 1)}{2\gamma} \pi_* (\pi_* - 1) \varepsilon \\ &- \frac{\sigma^2 \times k(\theta)}{20\gamma} \left(\frac{\gamma \pi_* (1 - \pi_*)}{6} \right)^{2/3} \varepsilon^{4/3} + O(\varepsilon^{5/3}), \end{aligned} \tag{48}$$

where

$$k(\theta) := -9 + 2\pi_* \Big(9 + \pi_* \big(3 + 12\gamma(\gamma - 2) + (10\theta + 5\theta^2)(\gamma - 1)^2 \big) \Big)$$

and, thus, is asymptotically optimal at the third order. For sufficiently small ε, the best performance at the fourth order is achieved for $\theta = -1$, strictly outperforming the shadow performance ($\theta = 1$).

Proof. Using the method in Appendix A, derive asymptotic expansions of c and s (whence $\widetilde{\zeta}_{\pm}$ up to the sixth order), satisfying the free boundary problem (A1)–(A5) at the same order. Modifying the second-order term by including a factor θ as in (47), one arrives at (48). The fourth-order coefficient $k(\theta)$ is a polynomial of the second order in θ, with a global minimum at $\theta = -1$. The comparison sign and magnitude of this factor are straightforward and reveal that $\theta = -1$ outperforms any other control limit policy for $\theta \neq -1$. □

Remark 3. *Note that for $\theta = -1$, the control limit policy for (47) is, up to order two, equal to the optimal strategy (11) of Theorem 1. This does not mean that it is optimal at order four or beyond, as higher-order coefficients of $\widetilde{\pi}_{\pm}^{\theta}$ may not agree with those of the optimal boundaries $\pi_{\pm}$.*

3.3. The Limits of Shadow Prices

Recall that a shadow price $\widetilde{S}$ is a frictionless process evolving in the bid–ask spread

$$(1 - \varepsilon) S_t \leq \widetilde{S}_t \leq S_t, \quad t \geq 0 \tag{49}$$

such that the optimal strategy φ is also optimal in the original market, and buys (respectively sells) precisely when $\widetilde{S}_t = S_t$ (respectively $\widetilde{S}_t = (1 - \varepsilon) S_t$).

To start with, the dynamics of the risky-safe ratio, wealth and proportion of wealth in the shadow market, for any finite variation strategy, is stated.

Lemma 5. *Suppose the shadow price satisfies the dynamics*

$$\frac{d\widetilde{S}_t}{\widetilde{S}_t} = (r + \widetilde{\mu}_t)dt + \widetilde{\sigma}_t dB_t. \tag{50}$$

For any finite variation trading strategy φ,

$$\frac{d\widetilde{\zeta}_t}{\widetilde{\zeta}_t} = \widetilde{\mu}_t dt + \widetilde{\sigma}_t dB_t + (1 + \widetilde{\zeta}_t)\frac{d\varphi_t^{\uparrow}}{\varphi_t} - (1 + \widetilde{\zeta}_t)\frac{d\varphi_t^{\downarrow}}{\varphi_t}, \tag{51}$$

$$\frac{d\widetilde{w}_t}{\widetilde{w}_t} = rdt + \widetilde{\pi}_t(\widetilde{\mu}_t dt + \widetilde{\sigma}_t dB_t), \tag{52}$$

$$\frac{d\widetilde{\pi}_t}{\widetilde{\pi}_t} = (1 - \widetilde{\pi}_t)(\widetilde{\mu}_t dt + \widetilde{\sigma}_t dB_t) - \widetilde{\pi}_t(1 - \widetilde{\pi}_t)\widetilde{\sigma}_t^2 dt + \frac{d\varphi_t^{\uparrow}}{\varphi_t} - \frac{d\varphi_t^{\downarrow}}{\varphi_t}. \tag{53}$$

Proof. A similar proof as that of Lemma 1 applies. □

Lemma 6. *If a shadow price exists, then for the optimal strategy, the cash positions in the original and shadow markets agree ($\widetilde{X} = X$), and the fraction of wealth invested in the shadow price satisfies*

$$\pi_t \leq \widetilde{\pi}_t \leq \frac{(1 - \varepsilon)\pi_t}{1 - \varepsilon\pi_t}.$$

In particular, if the optimal strategy satisfies $\pi_t \in [\pi_-, \pi_+]$, then

$$\pi_- \leq \widetilde{\pi}_t \leq \frac{(1 - \varepsilon)\pi_+}{1 - \varepsilon\pi_+} \leq (1 - \varepsilon)\pi_+(1 + \varepsilon\pi_+). \tag{54}$$

Proof. The optimal strategy trades the risky asset at the same price in both markets; therefore, the cash positions agree.

The lower bound is proved by observing that for $a, b > 0$, the function

$$\frac{a\xi}{-b + a\xi}$$

is strictly decreasing for any $\xi > b/a$ (which corresponds to positive wealth), and since $\widetilde{S} \leq S$,

$$\widetilde{\pi}_t = \frac{\varphi_t \widetilde{S}_t}{X_t + \varphi_t \widetilde{S}_t} \geq \frac{\varphi_t S_t}{X_t + \varphi_t S_t} = \frac{\varphi_t S_t}{w_t} = \pi_t.$$

Similarly, the upper bound follows from

$$\widetilde{\pi}_t = \frac{\varphi_t \widetilde{S}_t}{X_t + \varphi_t \widetilde{S}_t} \leq \frac{(1 - \varepsilon)\varphi_t S_t}{X_t + \varphi_t(1 - \varepsilon)S_t} = \frac{(1 - \varepsilon)\pi_t}{1 - \varepsilon\pi_t}.$$

The constant bounds in terms of the trading boundaries $\pi_{\pm}$ are an obvious conclusion. The last inequality in (54) follows from the summation formula of the geometric series, knowing that solvency implies $\varepsilon\pi_+ < 1$. □

For proportional transaction costs, maximizing the expected excess returns

$$\lim_{T \to \infty} \frac{1}{T}\mathbb{E}\left[\int_0^T \frac{dw_t}{w_t}\right]$$

over all admissible strategies $\varphi \in \Phi$ is well posed. By (Guasoni and Mayerhofer 2023, Theorem 3.2), for sufficiently small ε, there exists $0 < \pi_- < \pi_+ < \infty$ such that the trading strategy $\hat{\varphi}$ that buys at π_- and sells at π_+ to keep the risky weight π_t within the interval

$[\pi_-, \pi_+]$ is optimal. The maximum expected return of this optimal strategy is given by the almost sure limit

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \frac{dw_t}{w_t} = r + \mu \pi_- \tag{55}$$

and the trading boundaries have the series expansions

$$\pi_- = (1 - \kappa) \kappa^{1/2} \left(\frac{\mu}{\sigma^2} \right)^{1/2} \varepsilon^{-1/2} + 1 + O(\varepsilon^{1/2}), \tag{56}$$

$$\pi_+ = \kappa^{1/2} \left(\frac{\mu}{\sigma^2} \right)^{1/2} \varepsilon^{-1/2} + 1 + O(\varepsilon^{1/2}), \tag{57}$$

where $\kappa \approx 0.5828$ is the unique solution of

$$\frac{3}{2} \xi + \log(1 - \xi) = 0, \quad \xi \in (0, 1).$$

The remainder of this section is dedicated to showing that a shadow market does not exist.

For technical reasons, it is assumed in this section that any shadow price satisfies the following.

Assumption 1. *A shadow price $\widetilde{S}$ is a continuous process satisfying the dynamics* (50) *with drift and diffusion coefficients being ergodic in the sense that, almost surely, for some $\bar{\mu}, \bar{\sigma}^2 \in \mathbb{R}$,*

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T \widetilde{\mu}_t dt = \bar{\mu}, \quad \lim_{T \to \infty} \frac{1}{T} \int_0^T \widetilde{\sigma}_t^2 dt = \bar{\sigma}^2. \tag{58}$$

Remark 4. *The fairly general Assumption 1 is natural in that it applies to all known constructions of shadow prices in continuous-time models. In fact, typically, the ratio $\frac{\widetilde{S}_t}{S_t}$ is equal to $g(\pi_t)$, where g is a real analytic function and π_t is a stationary process: the optimal proportion of wealth in the risky asset, evolving within an interval $[\pi_-, \pi_+]$, where one buys (respectively sells) precisely at the trading boundary π_i (respectively π_+) and satisfying $g(\pi_-) = 1$ and $g(\pi_+) = 1 - \varepsilon$, reflecting the very definition of shadow price, agreeing with the ask (respectively bid) whenever shares are purchased (respectively sold). As these functions in the literature are all analytic, one can use Itô's formula to derive the dynamics* (50) *in a more explicit form.[14] There exist continuous functions h and H such that*

$$\widetilde{\mu}_t = h(\pi_t), \quad \widetilde{\sigma}_t^2 = H(\pi_t)$$

By the ergodic theorem (Borodin and Salminen 2002, II.35 and II.36), one obtains the finite limits in (58).

Lemma 7. *If $\mu > \sigma^2/2$, then $\bar{\mu} > \bar{\sigma}^2/2$. In particular, $\bar{\mu} \neq 0$.*

Proof. Write the fraction $\widetilde{S}_t/S_t$ as explicit solutions, where the accrual factor e^{rt} factors out. As $\mu > \sigma^2/2$, by the law of iterated logarithms, $e^{-rt} S_t$ almost surely tends to ∞ as $t \to \infty$. If $\bar{\mu} \leq \bar{\sigma}^2/2$, then for sufficiently large t, $\widetilde{S}_t$ behaves like a geometric Brownian motion with drift $\bar{\mu}$ and volatility $\bar{\sigma}$, whence by a similar argument, $\liminf_{t \to \infty} \widetilde{S}_t = 0$, almost surely. Thus, $\liminf_{t \to \infty} \widetilde{S}_t/S_t = 0 < (1 - \varepsilon)$, a contradiction to (49). □

Theorem 4. *If $\mu > \sigma^2/2$, then a shadow price satisfying the dynamics of* (50) *and Assumption 1 does not exist.*

Proof. Assume, for a contradiction, there exists a shadow price $\widetilde{S}$. By Lemma 5, the shadow wealth $\widetilde{w}_t = \varphi_t \widetilde{S}_t + \widetilde{X}_t$ satisfies the SDE (52). Furthermore, by Lemma 6

$$\int_0^t \widetilde{\pi}_t^2 \widetilde{\sigma}_u^2 du \leq (1 - \varepsilon)^2 (\pi_+^2 (1 + \varepsilon \pi_+)^2 \int_0^t \widetilde{\sigma}_u^2 du < \infty$$

and thus, the integral of the Brownian term is a martingale by Assumption 1. Thus, the strategy φ with associated wealth $\widetilde{w}$ achieves its optimum at

$$\lambda := \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\int_0^T \frac{d\widetilde{w}_t}{\widetilde{w}_t}\right] = r + \lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\int_0^T \widetilde{\mu}_t \widetilde{\pi}_t dt\right].$$

Note that, by Assumption 1,

$$\bar{\mu} = \lim_{T\to\infty}\frac{1}{T}\int_0^T \widetilde{\mu}_t dt = \lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\int_0^T \widetilde{\mu}_t dt\right]$$

and by Lemma 7, $\bar{\mu} > 0$. Furthermore, by Lemma 6, there exist $0 < L < U < \infty$ such that

$$L\bar{\mu} < \lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\int_0^T \widetilde{\mu}_t \widetilde{\pi}_t dt\right] < U\bar{\mu}. \tag{59}$$

Any alternative strategy $\varphi^\star$, whose proportion of wealth in the shadow price satisfies

$$\widetilde{\pi}^* \geq U \tag{60}$$

outperforms φ because

$$\lambda^* \geq r + U\bar{\mu} > \lambda.$$

Trading strategies that keep the exposure in the shadow asset constant to U exist, but they are of infinite variation. To obtain finite variation strategies satisfying (60), recall that by Lemma 5, the fraction of wealth in the shadow asset $\widetilde{w}$ associated with a finite variation strategy satisfies (53). One can modify this strategy by allowing bulk trades: let φ^* be the finite variations strategy that does refrain from trading whenever $\widetilde{\pi}^\star \in (U, 2U)$ but buys (respectively sells) the shadow asset in bulk whenever $\widetilde{\pi}^\star$ hits U (respectively $2U$) so as to reset $\widetilde{\pi}^\star$ to the midpoint $3U/2$. Such a strategy can be constructed pathwise and satisfies

$$\frac{d\widetilde{\pi}^\star_{t-}}{\widetilde{\pi}^\star_{t-}} = (1 - \widetilde{\pi}^\star_{t-})(\widetilde{\mu}_t dt + \widetilde{\sigma}_t dB_t) - \widetilde{\pi}^\star_{t-}(1 - \widetilde{\pi}^\star_{t-})\widetilde{\sigma}_t^2 dt + \frac{\Delta\varphi_t^{\star,\uparrow}}{\varphi_t^\star} - \frac{\Delta\varphi_t^{\star,\downarrow}}{\varphi_t^\star}.$$

The existence of such a strategy contradicts the optimality of φ, and thus a shadow price does not exist. □

Remark 5. *The finite variation strategy in the end of the proof cannot be replaced by a (standard) reflected diffusion with two reflecting boundaries because for the existence of strong solutions to the associated SDE on convex domains (Tanaka 1979, Theorem 4.1), one would need $\widetilde{\mu}_t$ and $\widetilde{\sigma}_t$ to be regular enough functions of $\widetilde{\pi}_t^\star$, an assumption that is too strong in this context. Also, it is unknown whether such a strategy is solvent in the original market with transaction costs.*

4. Discussion

Optimizing portfolios in continuous-time markets with proportional costs presents mathematically challenging problems. Strategies that are optimal in frictionless markets must be adjusted to prevent immediate bankruptcy as exemplified by the dynamic hedging component of a variance swap (see Example 1). The strategies considered in this paper are stationary[15] and, thus, ergodic theorems are used to determine their long-run performance. To gain insights into trading frequency, transaction costs, and long-run performance, we derive asymptotic expansions of the trading boundaries for small bid–ask spread.

The paper explores the (candidate) shadow prices for local mean–variance investors, with a threefold contribution.

First, we discover that the optimal strategy in the (candidate) shadow market differs from the optimal one in the original market but only in the second-order terms of the asymptotic expansion of the trading boundaries.[16] Theorem 2 demonstrates that, for risk

aversion $\gamma > 0$, the equivalent safe rate of the shadow market strategy agrees at the third order with the maximum. As transaction costs are of the second order, we conclude that the performance of the shadow market strategy is essentially optimal. It is worth noting that the same is true[17] for a long-run power–utility investor (cf. Gerhold et al. (2014)), as their trading boundaries also agree at the first order with (1). Second, Theorem 3 establishes that for $\gamma \neq 1$, the (potential) shadow market strategy $\tilde{\pi}$ is not optimal, as it can be outperformed. The alternative strategy is not necessarily optimal, even though it agrees up to the second order with the optimal one. In summary, the (candidate) shadow price is an asymptotic shadow price. Third, Theorem 4 demonstrates that for risk-neutral investors ($\gamma = 0$), no such shadow market exists.

The findings of this paper prompt the following research problems. First, we conjecture that a minor modification of the objective will render the shadow price candidate of Section 3.1 optimal in the original market with transaction costs. Motivated by Martin (2012, 2016), we propose to replace the equivalent safe rate in (3) by an infinite horizon, the local mean–variance utility function[18]

$$\text{ESR} := \mathbb{E}\left[\int_0^\infty \delta e^{-\delta t}\frac{dw_t}{w_t} - \frac{\gamma}{2}\int_0^\infty \delta e^{-\delta t}\left\langle \frac{dw_t}{w_t}\right\rangle_t\right] \tag{61}$$

for some discount rate $\delta > 0$. In the absence of transaction costs ($\varepsilon = 0$), the maximum equivalent safe rate agrees with that of the old objective (3). More importantly, this objective leads to the exact same shadow market construction as in Section 3.1. The question remains if our shadow market policy maximizes also (61) in the original market, surpassing its third-order optimality (Theorem 2).

Second, the mathematical treatment of optimization problems involving transaction costs is always uniquely tailored to a specific objective. This results in free boundary problems that vary significantly, encompassing scenarios from Riccati Differential Equations Gerhold et al. (2012) and linear equations (Guasoni and Mayerhofer 2019, Theorem 3.3) to the nonlinear problem (37) addressed in this paper, and even singular problems for zero risk-aversion (Guasoni and Mayerhofer 2019, Theorem 3.2). The question persists: can a unified approach be devised that accommodates a diverse range of objectives? To explore this possibility, one might aim for conformity to a common format—a second-order free boundary problem stated as follows:

$$F(g, g', g'') = 0, \tag{62}$$

$$g(\pi_-) = 1, \quad g(\pi_+) = 1 - \varepsilon, \tag{63}$$

$$g'(\pi_-) = 0, \quad g'(\pi_+) = 0. \tag{64}$$

This problem involves an unknown scalar function $g = g(\pi)$ that must satisfy a second-order nonlinear ODE (62), along with buy and sell boundaries π_- and π_+, respectively. The latter boundaries must adhere to zeroth-order boundary conditions (63) and first-order conditions (64).[19] In practical trading applications, a second-order approximation of the trading boundaries would suffice. Such an approximation might be achieved through a general polynomial ansatz for an approximation of (62).

Third, most of the literature[20] regarding the existence of an optimal strategy and its asymptotic expansion depends on the assumption of a "sufficiently small" bid–ask spread ε, without providing a minimum ε_0, for which these statements hold. Are they applicable to actual bid–ask spreads observed in markets (for liquid assets ranging in the basis points)? Addressing this question involves either demonstrating optimality for all $\varepsilon \in (0, 1)$ or identifying counterexamples where optimality breaks down for larger transaction costs, along with determining the explicit lower bound ε_0 at which control limit policies remain optimal. Such a lower bound would be contingent on model parameters γ, μ/σ^2, and risk aversion. Most likely, it will depend on the chosen objective.

Appendix A. The Free Boundary Problem for the Shadow Price Candidate

Let us introduce two new parameters $c = 1/\zeta_-$ and $s = \zeta_+/\zeta_-$. By defining the new function ϕ implicitly via

$$\Psi(\zeta) := \phi(c\zeta)/c,$$

the free boundary problem (39)–(41) produces a similar one for $\phi(z)$,

$$\phi''(z) = \frac{2\gamma\phi'^2(z)}{(c + \phi(z))} - 2\gamma\pi^*\frac{\phi'(z)}{z}, \tag{A1}$$

$$\phi(1) = 1, \tag{A2}$$

$$\phi'(1) = 1, \tag{A3}$$

$$\phi(s) = (1 - \varepsilon)s, \tag{A4}$$

$$\phi'(s) = (1 - \varepsilon). \tag{A5}$$

Remark A1. *Since for small transaction costs, trading strategies will be control limit policies on sufficiently small intervals, only the following cases need to be distinguished:*

- $\zeta_- < \zeta_+ < -1$ *(levered case): Then, $c < 0$, and therefore $z > 0$, so $s < 1$. Conversely, $s < 1$ implies $\zeta_- < -1$.*
- $0 < \zeta_- < \zeta_+$ *(unlevered case): Then, $c > 0$, and therefore the argument $z < 0$, so $s > 1$. Conversely, $s > 1$ implies $\zeta_- > 0$.*

For the sake of brevity, let us only consider the levered case, that is, $\zeta_- < \zeta_+ < -1$ and $\phi(\zeta) < -1$. Since also $c < 0$, one obtains $\phi(z) > -c$ for all z. Also, $c + 1 > 0$. Dividing (A1) by ϕ' and integrating once, one thus obtains

$$\log(\phi'(z)) = 2\gamma\log(c + \phi(z)) - 2\gamma\pi^*\log z - 2\gamma\log(c + 1),$$

where the initial condition (A3) is respected. Taking antilogarithms, one thus obtains

$$\frac{\phi'(z)}{(c + \phi(z))^{2\gamma}} = \frac{z^{-2\gamma\pi^*}}{(c + 1)^{2\gamma}}. \tag{A6}$$

Exclude in the following the singular cases $\gamma \neq 1/2$ and $\mu/\sigma^2 \neq 1/2$ (those cases can be dealt individually, leading to simpler solutions of the ODE (A6).). Integrating once again, one obtains

$$\frac{(c + \phi(z))^{1-2\gamma}}{1 - 2\gamma} = \frac{(z^{1-2\gamma\pi^*} - 1)}{(1 - 2\gamma\pi^*)(c + 1)^{2\gamma}} + \frac{(c + 1)^{1-2\gamma}}{1 - 2\gamma}, \tag{A7}$$

where the initial condition (A2) is respected. Thus,

$$\phi(z) = -c + \left(\frac{\frac{(1-2\gamma)}{(1-2\gamma\pi^*)}(z^{1-2\gamma\pi^*} - 1)}{(c + 1)^{2\gamma}} + (c + 1)^{1-2\gamma} \right)^{\frac{1}{1-2\gamma}}. \tag{A8}$$

Until this stage, the terminal boundary conditions (A4) and (A5) have not been involved. Those allow to reformulate the free boundary problem in terms of a system of non-linear equations for s and c.

Lemma A1. *Let $\gamma \neq 1/2$ and $\mu/\sigma^2 \neq 1/2$. ϕ, c, s is a solution to the free boundary problem* (A1)–(A5) *if and only if s and c satisfy the following system of non-linear equations:*

$$\left(\frac{c + (1-\varepsilon)s}{c+1}\right)^{1-2\gamma} - 1 = \frac{1-2\gamma}{1-2\gamma\pi^*}\frac{s^{1-2\gamma\pi^*} - 1}{c+1}, \tag{A9}$$

$$(1-\varepsilon)^{\frac{1}{2\gamma}}s^{\pi^*} = \frac{c + (1-\varepsilon)s}{c+1}. \tag{A10}$$

Proof. The initial value problem (A1)–(A3), parameterized in c, has the explicit solution (A8). What remains is to involve the boundary conditions (A4) and (A5). Starting from (A7) and using (A4) yields

$$\frac{(c + (1-\varepsilon)s)^{1-2\gamma}}{1-2\gamma} = \frac{(s^{1-2\gamma\pi^*} - 1)}{(1-2\gamma\pi^*)(c+1)^{2\gamma}} + \frac{(c+1)^{1-2\gamma}}{1-2\gamma}, \tag{A11}$$

from which (A9) follows. Using (A4)–(A6), one obtains

$$\frac{(1-\varepsilon)}{(c+(1-\varepsilon)s)^{2\gamma}} = \frac{s^{-2\gamma\pi^*}}{(c+1)^{2\gamma}}.$$

Taking the 2γths root, one obtains (A10). The proof of the converse implication is similar. $\square$

Appendix B. Asymptotics of the Free Boundaries

Recall that $\pi^* = \frac{\mu}{\gamma\sigma^2}$ and note that

$$\zeta_- = \frac{1}{c}, \quad \zeta_+ = \frac{s}{c} \tag{A12}$$

and the associated trading boundaries $\pi_\pm$ satisfy

$$\pi_\pm := \frac{\zeta_\pm}{1+\zeta_\pm}. \tag{A13}$$

We introduce the abbreviations

$$\bar{c} := \frac{1-\pi^*}{\pi^*}, \quad \Delta := \left(\frac{6}{\gamma\pi^*(1-\pi^*)}\right)^{1/3}\varepsilon^{1/3}.$$

Proposition A1. *For sufficiently small $\varepsilon > 0$, the free boundary problem* (A1)–(A5) *has a unique solution $(h(\zeta), c, s)$. Moreover, the following asymptotics hold as $\varepsilon \to 0$:*

$$c = \bar{c} + \frac{1-\pi^*}{2\pi^*}\Delta + \frac{(1-\pi^*)(3 - \pi^*(2\gamma+1))}{12\pi^*}\Delta^2 \tag{A14}$$
$$- \frac{(\pi^*-1)\left((4\gamma^2 + 22\gamma + 1)(\pi^*)^2 - 24(2\gamma+1)\pi^* + 36\right)}{360\pi^*}\Delta^3 + O(\varepsilon^{4/3}),$$

$$s = 1 + \Delta + \frac{\Delta^2}{2} + \frac{1}{180}\left(\left(4\gamma^2 - 8\gamma + 1\right)(\pi^*)^2 + 3(4\gamma - 3)\pi^* + 36\right)\Delta^3 \tag{A15}$$
$$+ \frac{(8\gamma^2 - 26\gamma + 2)(\pi^*)^2 + 2(17\gamma - 9)\pi^* + 27}{360}\Delta^4 + O(\varepsilon^{5/3}).$$

Proof. The proof is inspired by (Gerhold et al. 2013, Proposition 6.1), where a similar result is developed for log utility from consumption and for unlevered strategies.[21] Having already solved the initial value problem (A1)–(A3), parameterized in c, which has the

explicit solution (A8), it remains to involve the boundary conditions (A4) and (A5). A naïve approach would be to define for sufficiently small δ the map $F := (F_1, F_2)^\top$, where

$$F_1(\delta, c, s) = \left(\frac{c + (1 - \delta^3)s}{c + 1}\right)^{1 - 2\gamma} - 1 - \frac{1 - 2\gamma}{1 - 2\gamma\pi^*} \frac{s^{1 - 2\gamma\pi^*} - 1}{c + 1}, \tag{A16}$$

$$F_2(\delta, c, s) = (1 - \delta^3)^{\frac{1}{2\gamma}} s^{\pi^*} - \frac{c + (1 - \delta^3)s}{c + 1}, \tag{A17}$$

and to show, by means of the implicit function theorem, that F has a unique zero $(s(\delta), c(\delta))$ at $(c = \bar{c}, s = 1)$, which is analytic in δ. Note, however, that the implicit function theorem cannot be applied in this case: even though $F(\delta_0 = 0, c_0 = \bar{c}, s_0 = s) = 0$, the Jacobian vanishes at the critical point $(0, \bar{c}, 1)$.

Consider the levered case only, as the other case can be proved quite similarly. In this case, $s < 1$. Having a look at Equation (A6), one sees that for $z < 1$, z is sufficiently close to $z = 1$, $\phi'(z) > 0$, and since $\phi(1) = 1$, this implies $\phi(z) < 1$. Since $\phi = \phi(z, c)$ is an analytic function in (z, c) near $c = \bar{c}$ and $z = 1$, it satisfies an expansion of the form

$$\phi(z, c) = 1 + (z - 1) + \sum_{i \geq 2} \sum_{j \geq 0} a_{ij}(z - 1)^i (c - \bar{c})^j$$

with coefficients a_{ij}, which can be calculated recursively. Furthermore, $a_{0j} = a_{1j} = \delta_{0j}$ for $j \geq 0$ due to the initial conditions (A2) and (A3). One now solves for c, s, invoking the terminal conditions (A4) and (A5). The latter imply that

$$\varepsilon s = s - \phi(s, c), \quad \text{and} \quad \phi(s, c) - s\phi'(z = s, c) = 0.$$

Dividing by $s - 1$, reflecting that the solution $s = 1$ is not interesting, a Taylor expansion yields

$$\frac{\phi(s, c) - s\phi'(s, c)}{s - 1} = \sum_{i \geq 0} \sum_{j \geq 0} b_{ij}(s - 1)^i (c - \bar{c})^j \tag{A18}$$

for certain coefficients b_{ij}. By using (A1)–(A3) and L'Hospital's rule, one obtains

$$b_{0,0} = \lim_{z \to 1, c \to \bar{c}} \frac{\phi(z, c) - z\phi'(z, c)}{z - 1} = -\phi''(1, \bar{c}) = 0,$$

and, further by a twofold application of L'Hospital's rule,

$$b_{1,0} = \lim_{z \to 1, c \to \bar{c}} \frac{\phi(z, c) - z\phi'(z, c)}{(z - 1)^2} = -\lim_{z \to 1, c \to \bar{c}} \phi^{(3)}(z, c) = 2\gamma\pi^*(1 - \pi^*) \neq 0.$$

Hence, the implicit function theorem is applicable and yields $s(c) = H(c)$ as a function of c such that

$$H(\bar{c}) = 1, \quad H'(\bar{c}) = \frac{2\pi^*}{1 - \pi^*}.$$

Inserting this function into (A10), one obtains the problem

$$g(c, \delta) := (1 - \delta^3)^{\frac{1}{2\gamma}} H^{\pi^*}(c) - \frac{c + (1 - \varepsilon)H(c)}{c + 1} = 0.$$

Since $g(c = \bar{c}, \delta = 0) = 0$ and $\partial_c g(c, \delta) = \frac{1}{\pi^*} \neq 0$, one can apply the implicit function theorem which asserts that for sufficiently small δ, a unique and analytic solution $c = c(\delta)$ exists to $g(c, \delta) = 0$ and $c(0) = \bar{c}$. Therefore, $c(\delta), s(\delta) = H(c(\delta))$ is the unique solution of our problem for small δ.

Finally, one derives the asymptotic formulas (A14) and (A15): let (ϕ, c, s) be the unique solution of (A1)–(A5). Due to Lemma A1, s and c satisfy the system (A9) and (A10). Substitute $c = c(s)$ from (A10) into (A9), and replace ε by δ^3 in all equations. Then, one

plugs into the modified Equation (A9) a power series ansatz for s, namely, $s = 1 + s_1\delta + \ldots s_6\delta^6$. Developing both sides as power series in δ and comparing coefficients leads to (A15). This result is then plugged into (A10), yielding, quite similarly, (A14). $\square$

Remark A2. *Using the formulae* (A12), *the asymptotics* (42) *for the trading boundaries* $\widetilde{\zeta}_\pm$ *in terms of the risky-safe ratio follow from the asymptotics of Proposition A1. The asymptotics* (44) *for* $\widetilde{\pi}_\pm$ *then follows from the relationship* (A13).

Notes

[1] Some tedious computations in this paper where performed by MATHEMATICA. For motivating this research topic and providing feedback, I am indebted to Professor Paolo Guasoni.

[2] More generally, the term *market frictions* encompasses, for example, price impact, short-selling constraints, and margin requirements (see Guasoni and Muhle-Karbe (2013), Guasoni and Weber (2020) and Guasoni et al. (2023) and the references therein).

[3] Example 1 below shows this failure for a variance swap hedge.

[4] These references deal with particularly tractable, long-run problems of local or global utility maximization; however, the first papers in this field, starting with Magill and Constantinides (1976), where optimal investment and consumption problems on an infinite horizon, which exhibit similar strategies and asymptotics. For an overview of this research field, see (Guasoni and Mayerhofer 2019, Chapter 1) and Guasoni and Muhle-Karbe (2013).

[5] For these strategies, the name "control limit policy" from Taksar et al. (1988) is adopted, see Definition 2 below.

[6] When $\gamma = 1$, the local-mean variance objective agrees with logarithmic utility, for which monotonicity holds and the shadow market strategy is the optimal one cf. Gerhold et al. (2013).

[7] For the dynamics of the wealth process, see Lemma 1 below.

[8] This follows from the respective finite-horizon objective (20), expressed in terms of π_t and φ_t, see Lemma 1.

[9] It is well-known that a variance swap with maturity T on a continuous semimartingale S can be perfectly hedged by holding $2/(TS_t)$ units of the underlying at time $t \le T$ (the dynamic hedging term), and a static portfolio of European puts and calls with expiry T, Bossu et al. (2005).

[10] By ergodicity, the strategy that makes bulk trades into the middle of the optimal no-trade region incurs average transaction costs of higher order, namely proportional to $\varepsilon^{1/3}$. (Compare the ATC (28) which is of second order.)

[11] The product rule gives
$$\frac{d\widetilde{S}_t}{\widetilde{S}_t} = \frac{dS_t}{S_t} + \frac{dg}{g} + \frac{d\langle S_t, g\rangle}{S_t g} =: (\widetilde{\mu}_t + r)dt + \widetilde{\sigma}_t dB_t,$$
from which the particular form of drift and diffusion coefficients (32), (33) can be computed.

[12] For the details leading to this and other asymptotics, see Appendix A, Proposition A1 and Remark A2.

[13] This expression is readily obtained from (11) by expanding (38) into formal power series in $\varepsilon^{1/3}$.

[14] The general form of drift and diffusion coefficients follows from the typical smooth pasting conditions $g'(\pi_\pm) = 0$, along the same arguments as in Section 3.1 that turn (32) into (35), by removing local-time terms.

[15] More precisely, certain portfolio statistics, such as π_t or ζ_t, exhibit stationarity.

[16] That this second order discrepancy is not essential, can be seen also by a numerical robustness check, with trades at daily frequency and with a finite time horizon of, say five years. Numerical examples are already elaborated for a similar objective in great detail in (Guasoni and Mayerhofer 2023, Section 6 (Figures 4 and 5)).

[17] This assertion can be proven using the same method as in Theorem 2.

[18] Note that we use portfolio returns, as opposed to changes of wealth in Martin (2012, 2016). Besides, Martin's work cares about asymptotic optimality at lowest order, similar to Kallsen and Muhle-Karbe (2017).

[19] Such a general representation bears the advantage that the stochastic process $\widetilde{S}_t := g(\pi_t)S_t$ could be interpreted as a (candidate) shadow price.

[20] (Taksar et al. 1988, Theorem 6.16) appears to be an exception, which does not refer to te smallness of transaction costs.

[21] Similar methods to derive asymptotic expansions in small transaction costs are found in the papers Gerhold et al. (2012, 2014); Guasoni and Mayerhofer (2019, 2023).

References

Borodin, Andrei N., and Paavo Salminen. 2002. *Handbook of Brownian Motion: Facts and Formulae*. Berlin and Heidelberg: Springer.

Bossu, Sebastien, Eva Strasser, and Regis Guichard. 2005. Just what you need to know about variance swaps. *JPMorgan Equity Derivatives Report* 4: 1–29.

Czichowsky, Christoph, and Walter Schachermayer. 2016. Duality theory for portfolio optimisation under transaction costs. *Annals of Applied Probability* 26: 1888–941. [CrossRef]

Gerhold, Stefan, Johannes Muhle-Karbe, and Walter Schachermayer. 2012. Asymptotics and duality for the Davis and Norman problem. *Stochastics An International Journal of Probability and Stochastic Processes* 84: 625–41. [CrossRef]

Gerhold, Stefan, Johannes Muhle-Karbe, and Walter Schachermayer. 2013. The dual optimizer for the growth-optimal portfolio under transaction costs. *Finance and Stochastics* 17: 325–54. [CrossRef]

Gerhold, Stefan, Paolo Guasoni, Johannes Muhle-Karbe, and Walter Schachermayer. 2014. Transaction costs, trading volume, and the liquidity premium. *Finance and Stochastics* 18: 1–37. [CrossRef]

Guasoni, Paolo, and Eberhard Mayerhofer. 2019. The limits of leverage. *Mathematical Finance* 29: 249–84. [CrossRef]

Guasoni, Paolo, and Eberhard Mayerhofer. 2023. Leveraged funds: Robust replication and performance evaluation. *Quantitative Finance* 23: 1155–76. [CrossRef]

Guasoni, Paolo, and Johannes Muhle-Karbe. 2013. Portfolio choice with transaction costs: A user's guide. In *Paris-Princeton Lectures on Mathematical Finance 2013*. Cham: Springer, vol. 2081, pp. 169–201.

Guasoni, Paolo, and Marko Hans Weber. 2020. Nonlinear price impact and portfolio choice. *Mathematical Finance* 30: 341–76. [CrossRef]

Guasoni, Paolo, Eberhard Mayerhofer, and Mingchuan Zhao. 2023. Options Portfolio Selection with Position Limits. *Michael J. Brennan Irish Finance Working Paper Series Research Paper* 22: 1–33. [CrossRef]

Guasoni, Paolo, Miklós Rásonyi, and Walter Schachermayer. 2010. The fundamental theorem of asset pricing for continuous processes under small transaction costs. *Annals of Finance* 6: 157–91. [CrossRef]

Herdegen, Martin, David Hobson, and Alex S. L. Tse. 2023. *Optimal Investment and Consumption with Epstein-Zin Stochastic Differential Utility and Proportional Transaction Costs*. Private communication.

Kabanov, Yuri, Miklós Rásonyi, and Christophe Stricker. 2002. No-arbitrage criteria for financial markets with efficient friction. *Finance and Stochastics* 6: 371–82. [CrossRef]

Kallsen, Jan, and Johannes Muhle-Karbe. 2010. On using shadow prices in portfolio optimization with transaction costs. *Annals of Applied Probability* 20: 1341–58. [CrossRef]

Kallsen, Jan, and Johannes Muhle-Karbe. 2017. The general structure of optimal investment and consumption with small transaction costs. *Mathematical Finance* 27: 695–703. [CrossRef]

Magill, Michael J. P., and George M. Constantinides. 1976. Portfolio selection with transactions costs. *Journal of Economic Theory* 13: 245–63. [CrossRef]

Martin, Richard J. 2012. Optimal multifactor trading under proportional transaction costs. *arXiv* arXiv:1204.6488.

Martin, Richard J. 2016. Universal trading under proportional transaction costs. *arXiv* arXiv:1603.06558.

Taksar, Michael, Michael J. Klass, and David Assaf. 1988. A diffusion model for optimal portfolio selection in the presence of brokerage fees. *Mathematics of Operations Research* 13: 277–94. [CrossRef]

Tanaka, Hiroshi. 1979. Stochastic differential equations with reflecting boundary condition in convex regions. *Hiroshima Mathematical Journal* 9: 163–77. [CrossRef]

Article

Implementing Intraday Model-Free Implied Volatility for Individual Equities to Analyze the Return–Volatility Relationship

Martin G. Haas and Franziska J. Peter *

Department of Corporate Management and Economics, Zeppelin University Am Seemooser Horn 20, 88045 Friedrichshafen, Germany; haasmartin@live.de
* Correspondence: franziska.peter@zu.de

Abstract: We implement the VIX methodology on intraday data of a large set of individual equity options. We thereby consider approaches based on monthly option contracts, weekly option contracts, and a cubic spline interpolation approach. Relying on 1 min, 10 min, and 60 min model-free implied volatility measures, we empirically examine the individual equity return–volatility relationship on the intraday level using quantile regressions. The results confirm a negative contemporaneous link between stock returns and volatility, which is more pronounced in the tails of the distributions. Our findings hint at behavioral biases causing the asymmetric return–volatility link rather than the leverage and volatility-feedback effects.

1. Introduction

Volatility is a key input variable for assessing risk, pricing derivative products, and developing trading strategies. As such, it has been and still is a central topic of research in finance. While, in early times, volatility was most commonly calculated as realized volatility (RV), derived from observed prices and their returns, the advent of exchange-traded options enabled the derivation of so-called implied volatility (IV). An early example is the derivation of IV by Latané and Rendleman (1976), who reversed the process of option pricing to retrieve the volatility that is implied by traders and used the famous Black and Scholes (1973) (B&S) formula for option pricing. The development of further option pricing models, e.g., assuming stochastic volatility and interest rates or jumps in the underlying process, has led to a corresponding development of these model-based IV methods. Whereas the merit of this IV was initially challenged—based on forecasting performance—it is nowadays considered the superior method for this purpose (see, e.g., the seminal review of Poon and Granger 2003). Despite IV's superiority, its quality is restricted by the assumptions and parameters of the option pricing model—unless you dispense the idea of parametric models. The methodology developed in Britten-Jones and Neuberger (2000), Demeterfi et al. (1999), and Carr and Madan (1997) uses the fact that arbitrary payoffs can be replicated using prices from a portfolio of options—a strategy called "option spanning", which was developed much earlier by Breeden and Litzenberger (1978); Green and Jarrow (1987); Nachman (1988). As it turns out, a portfolio of out-of-the-money put and call options can be used to replicate the risk-neutral expected variance of the underlying stock, with the only assumptions that its price process is continuous and there is a constant risk-free rate. The perhaps most prominent application of this method is the 2003 VIX, a volatility index for the S&P 500, developed and published by the Chicago Board Options Exchange (CBOE). The VIX is designed to produce 30-day expected volatility of the U.S. stock market, calculated from mid-quote prices of the S&P 500 Index (SPX) call and put options. The VIX is defined as implied volatility for a fixed 30-day maturity, quoted in annualized terms. It was originally founded in 1993 by Robert Whaley as a

model-based IV index on the S&P 100 and was intended to provide a basis for the trading of volatility derivatives, which can be used for hedging portfolios against the risk of changes in volatility (see Whaley 1993). Famously called the "Investor Fear Gauge", it "is set by investors and expresses their concerns about future stock market volatility" (Whaley 2000, p. 12), implying that risk-averse investors "fear" increases in volatility. As a matter of fact, the VIX has a significant negative relationship with the underlying S&P 500 index, which supports this interpretation, although it is found to be "time-varying, asymmetric and influenced by VIX-computation errors" (Gonzalez-Perez 2015, p. 3). Since its founding, the VIX has been intensively studied and has sparked a branch of research on the development of further model-free IV methods. Empirical research on the VIX covers its distributional properties, its use as a proxy overall market risk factor, the development and pricing of volatility derivatives, as well as respective trading strategies and its informational content with respect to forecasting volatility. However, research on VIX has been mostly conducted for market indices, like the S&P 100 and 500 and Nasdaq 100, and other national stock market indices, like the German DAX. The CBOE already publishes VIX-like indices for various individual equities, with examples including large-cap stocks like Amazon, Google, Microsoft, Apple, and Goldman Sachs, beyond which research on individual equities is scarce. On the individual level, Taylor et al. (2010) analyzed the forecasting performance for the 149 S&P 100 constituents. Dennis et al. (2006) assessed the leverage effect for a portfolio of 50 individual stocks and the S&P 100. Regarding frequencies, however, research is almost exclusively conducted on a daily frequency and, if it covers intraday frequencies, employs data on larger indices and is concerned with the development of model-free IV methods. In her review, Gonzalez-Perez (2015) found only one paper that used VIX on an intraday basis in order to assess the financial leverage effect, which was by Ishida et al. (2011), and, more recently, another paper by Badshah et al. (2016) on intraday leverage using quantile regressions. On a more general note, Andersen et al. (2021) recently conducted a descriptive study of high-frequency trade and quote option data and emphasized the usefulness of high-frequency option data.

We, therefore, calculate model-free implied volatility (MFIV) for individual equities on an intraday level by using the full universe of options written on individual firm U.S. equity at a 1-min frequency during the period from 1 February 2017 to 30 June 2017. We offer descriptive insights into the data requirements, the calculation process, and the MFIV properties.[1] Subsequently, to showcase the usability of such measures, we apply them to analyze the return–volatility relationship for individual equities on an intraday basis using quantile regressions. We are able to implement reliable MFIV measures for a sample of 138 individual equity options, whose available option contracts fulfill the quality requirements put forward by Jiang and Tian (2005, 2007) at a 1 min frequency. Concerning the return–volatility relationship, we confirm an asymmetric relationship between contemporaneous returns and volatility, which is more pronounced within the tail of the distribution corresponding to previous index-level-based findings for lower frequencies, which have previously been explained to be potentially caused by behavioral biases (compare Hibbert et al. (2008), Talukdar et al. (2017), Daigler et al. (2014), and Badshah et al. (2016)). The remainder of the paper is structured as follows: Section 2 outlines the MFIV (VIX) methodology; Section 3.1 describes the data, while Section 3.2 outlines the implementation of MFIV on individual equity and offers descriptive insights. Section 4 analyzes the return–volatility relationship, and we conclude in Section 5.

2. The VIX Methodology

Since 2003, the VIX has been calculated from S&P 500 index options, and the model-free[2] approach described by Demeterfi et al. (1999) and Britten-Jones and Neuberger (2000). Based on an option set of at least 100 strike prices, the VIX replicates a variance swap on the S&P 500. In its current whitepaper on the VIX, the CBOE (2023) implemented the variance replication strategy and calculated implied volatility based on the following formula (for a detailed derivation, see Appendix A):

$$\sigma^2 := \frac{2}{T}\left(\sum_i \frac{\Delta K_i}{K_i^2} Q(K_i) e^{rT}\right) - \frac{1}{T}\left(\frac{F_0}{K_0} - 1\right)^2$$

where r denotes the risk-free rate, K denotes the strike price, and $Q(K_i)$ gives out-of-the money option prices. Note that the CBOE measures the time T in years and uses calendar days instead of trading days. The integral underlying this formula (see Appendix A) is replicated numerically with a weighted sum using the weights $\Delta K_i := (K_{i+1} - K_i)/2$. At the end of the strike-price range, the weight is calculated as the midpoint between the two last (first) strike prices. F_0 represents a theoretical forward price, calculated as

$$F_0 := K_i + e^{rT}(C(K_i) - P(K_i))$$

where the difference between call (C) and put (P) prices, $(C(K_i) - P(K_i))$, is smallest. Consequently, K_0 represents the strike price immediately at or below F_0. The CBOE applies a selection theme to options, which uses call options for strikes higher than K_0 and put options for calls lower than K_0. Ordering the options by their strike prices and ascending (descending) from K_0, all calls (puts) after two consecutive zero-bid quotes are left out of the calculation.

The CBOE VIX is calculated with a maturity of 30 days. For this purpose, a "near-term" and a "next-term" contract are designated, for which the MFIV is calculated. The annualized 30-day MFIV, quoted as a percentage, is then derived via linear interpolation of the MFIV from both contracts:

$$\sigma_{VIX} = 100\sqrt{\left(\omega T_1 \sigma_1^2 + (1 - \omega) T_2 \sigma_2^2\right) \frac{525,600}{43,200}}$$

with the weights $\omega = \frac{T_2 - 30}{T_2 - T_1}$, where the subscripts 1 and 2 correspond to the near- and next-term contracts, respectively.

The replication strategy underlying the above formula relies on the availability of a continuum of strike prices. Real-world applications, however, are commonly faced with a limited set of discrete strikes. Jiang and Tian (2007) showed that an insufficient range of strikes leads to a downward bias in the calculated MFIV. The discreteness of strikes also introduces errors if strikes are too widely dispersed. As Jiang and Tian (2005) showed, reliable MFIV estimates can be obtained if the truncation point of each tail is 3.5 standard deviations (SDs)[3] from the at-the-money forward price F_0. They further show that the "discretization error", which is induced by the spacing between adjacent strike prices, is negligible below strike-price increments of 0.35 SDs.

Furthermore, prior to the introduction of weekly SPX options in 2014 (2003 VIX method), the near-term contract was defined as having at least 7 days to expiration and the next-term contract as the one with the consecutive expiration date. This 7-day minimum was chosen to minimize pricing anomalies that might occur close to expiration. With the introduction of the weekly SPX options (2014 VIX method), the selected contracts were those with expiration dates greater than 23 and smaller than 37 days. Most CBOE equity option contracts expire monthly on the 3rd Friday. For a subset of individual equity, such as Apple, Amazon, or Goldman Sachs, contracts with weekly expirations are quoted. This means that, at best, a constant maturity MFIV of seven days (or multiples of seven days) on a weekly frequency can be calculated. In order to be able to offer higher frequencies and other maturities, the CBOE uses intrapolation and extrapolation methods, using the MFIV of two contracts with neighboring expiration dates.

The CBOE selection rule for the monthly option contracts only allows a minimum of 7 days to expiration, upon which the selection rolls over to the next set of contracts. In certain cases, the interpolation can become an extrapolation. Jiang and Tian (2007) described this as a potential source of error. However, weekly options, which allow for a more precise interpolation, are not quoted for all stocks, motivating the use of an alternative interpolation approach. For our implementation on individual equities, we therefore intrapolate the

30-day MFIV using the information from a larger set of options. This should not only reduce the size of jumps but also use the information content of the term structure.

3. Implementing Model-Free Implied Volatility for Individual Equities

3.1. Data and Data Processing

We obtain 1 min option data from the CBOE data store, which contains NBBO market quotes for all equity option contracts traded on the CBOE.[4] The sample covers the time period from 1 February 2017 to 30 June 2017 on a 1 min frequency, covering 105 trading days. The classic monthly contracts expire on the third Friday of each month and make up the dataset for the 2003 VIX methodology. The 2014 VIX method uses these monthly contracts and the weekly contracts that expire each Friday. For our cubic spline method, we select all monthly contracts that have an expiration within 365 days.

The dataset contains 268 stocks, for which both monthly and weekly option contracts are quoted. We structure our sample into common market capitalization categories defined as "Mega" (>USD 200B), "Large" (>USD 10B), "Medium" (>USD 2B), and "Small" (>USD 300M) in order to judge the influence of size, as in Dennis et al. (2006). If, for a given minute, there are no available traded strike prices, MFIV cannot be calculated. Furthermore, we drop all stocks that exhibit more than 10 such missing values a day, which restricts our sample to 178 stocks.

In order to prepare the MFIV calculation as per the CBOE (2023) whitepaper, option prices $Q(K)$ for each strike K are derived as the midpoint of the bid-ask spread. The prices P for the respective underlyings are calculated in the same manner. The risk-free rate data are based on "constant maturity treasury" or CMT rates, which can be obtained from the U.S. Department of Treasury official resource center. Using a cubic spline interpolation and the daily CMT quotes, the risk-free rate r is then calculated for each time to maturity up to a one-minute-level precision.

3.2. Descriptive Analyses on Model-Free Implied Volatility for Individual Equities

We calculate the model-free IV for all stocks, times, and maturities, including such observations where the nest of options has at least one quote for a put and a call. Analogously to the standard VIX calculation, we use out-of-the-money quotes for the near- and next-term contracts. We refer to the different interpolation methods and their respective sets of option contracts by "WK" for the weekly 2014 method and "MN" for the monthly 2003 method. An example of option contracts and resulting MFIV for the WK and MN methods are graphed in Figure 1 and exemplify the differences between the 2003 and 2014 VIX methods, i.e., what we refer to as the WN and MN measures.

At times where weekly and monthly contracts coincide (every third Friday in a month), both methods have one common contract and show strong convergence, e.g., around the peak at time $t = 5000$. However, there are times when no contracts are shared and the MFIV measures differ systematically.

As panel (a) shows, the WK MFIV is always confined by its narrow corridor near- and next-term contracts, whereas the MN MFIV in panel (b) has a wider corridor, depicted by the narrower corridor of the gray lines visible in particular until about half of the depicted time period. In panel (b), it is also visible that, when extrapolated, the MFIV measure (black line) breaches the corridor, i.e., it results in values outside the (gray) corridor. This can lead to substantial differences between the two methods. The idea of our cubic spline interpolation, which we refer to as "SP", is to employ the information of the complete term structure of MFIV in order to reduce these differences and yield an alternative to the MN approach if weekly contracts are not available. Figure A1 in the appendix illustrates the use of contracts as near-term or next-term contracts for an exemplary underlying (AAPL). We developed an R package for these calculations, which is available on https://github.com/m-g-h/R.MFIV, accessed on 20 November 2023.

(**a**)

(**b**)
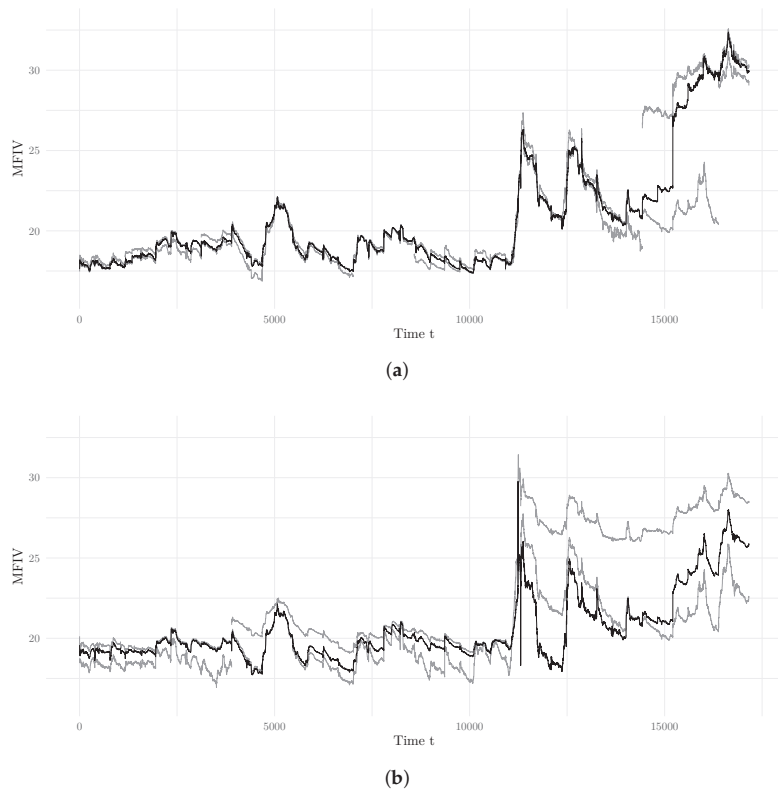
Figure 1. Option contracts used for the MFIV interpolation (gray) and resulting MFIV (black). Panel (**a**) shows the weekly option contracts used for the WK approach, which usually form a narrow corridor. Panel (**b**) shows the set of monthly option contracts used for the MN approach, which form a wider corridor that is sometimes "breached".

We use the amount, range, and spacing of the option strike prices as indicators of data quality with respect to the MFIV calculation. Theoretically, MFIV is based on an option portfolio with a continuum of strike prices ranging from zero to infinity. Empirically, such a continuum is approximated with a set of quoted discrete strike prices. The CBOE (2023) example includes 176 strike-price quotes for out-of-the-money calls and puts on the SPX. Despite the fact that option quotes deep in the tails of the strike-price distribution receive lower weights (proportional to $\frac{1}{K^2}$) and at some point have a negligible effect on the resulting MFIV value, a limited strike-price range potentially incurs a "truncation error", which implies an underestimation of the true implied volatility, according to Jiang and Tian (2005, 2007). As mentioned above, Jiang and Tian (2005) advocated a truncation point of each tail of 3.5 SDs from the at-the-money forward price and spacing between adjacent strike prices below strike-price increments of 0.35 SDs to derive reliable MFIV estimates. Panel (a) in Table 1 provides an aggregated overview of our underlying option data, grouped by the market capitalization of the underlying stocks. On average, the set of all monthly options within a year, which we employ in our SP approach, shows the highest strike-price range (*Min K* and *Max K*), followed by the MN and WK sets. This can be explained by noting that the probability of larger price swings increases with the time to maturity, which, of course, is largest for the SP and smallest for the WK option set. Regarding the strike-price increment, we observe the opposite relationship, with the shorter-term WK set exhibiting the smallest spacing.

As the requirements of the 3.5 SD range and 0.35 SD spacing are not met by all stocks, we report results on the stocks that fulfill them in panel (b) of Table 1 and refer to these set of underlyings as "feasible" stocks. While the MN and SP sets are reduced to below 30% of the original sample size, the WK set retains 138 of the original 178 stocks. Notably, the set of mega-sized stocks is reduced the most and is empty for the SP set. Considering the cross-section of stocks by market capitalization, we observe that the average number of quoted options is decreasing with decreasing market capitalization, which corresponds to the findings of Dennis et al. (2006).

Table 1. Descriptive statistics. The table groups the stocks by their market capitalization (*MC*, quoted in million USD). The average number of call and put options is given by *Calls* and *Puts*. The average range above (*Max K*) and below (*Min K*) the current ATM forward price F_0, as well as the average spacing of the strikes (*dK*), are measured in standard deviations of price. We also report the average price (*P*), return (*R*), standard deviation of return (*R(sd)*), and MFIV. The number of stocks is given by *n*, and "Set" refers to the set of options for the respective interpolation approach.

Size	MC	Calls	Puts	Max K	Min K	dK	n	P	R	R (sd)	MFIV
					WK Set						
Any	101,337	11.0	13.8	1.75	2.31	0.274	178	104.37	1.18×10^{-6}	0.0008	26.82
Mega	485,985	11.4	18.1	1.93	3.11	0.337	21	205.50	2.24×10^{-6}	0.0005	17.84
Large	65,180	11.2	14.0	1.77	2.34	0.275	117	105.33	1.46×10^{-6}	0.0008	25.08
Medium	5692	10.0	11.0	1.61	1.82	0.237	33	51.52	3.60×10^{-7}	0.0010	36.38
Small	1359	10.8	9.8	1.65	1.67	0.238	7	34.04	-2.84×10^{-6}	0.0012	37.73
					MN Set						
Any	101,337	8.2	10.1	2.06	2.74	0.436	178	104.37	1.18×10^{-6}	0.0008	27.02
Mega	485,985	9.1	14.4	2.30	3.84	0.539	21	205.50	2.24×10^{-6}	0.0005	18.15
Large	65,180	8.1	9.9	2.06	2.78	0.447	117	105.33	1.46×10^{-6}	0.0008	25.34
Medium	5692	8.0	8.4	1.93	2.11	0.344	33	51.52	3.60×10^{-7}	0.0010	36.40
Small	1359	8.1	6.7	1.92	1.84	0.374	7	34.04	-2.84×10^{-6}	0.0012	37.50
					SP Set						
Any	101,337	9.0	11.8	3.27	3.80	0.349	178	104.37	1.18×10^{-6}	0.0008	26.97
Mega	485,985	9.6	16.4	3.86	5.34	0.407	21	205.50	2.24×10^{-6}	0.0005	18.05
Large	65,180	8.8	11.8	3.28	3.90	0.357	117	105.33	1.46×10^{-6}	0.0008	25.28
Medium	5692	9.2	9.8	2.90	2.74	0.290	33	51.52	3.60×10^{-7}	0.0010	36.39
Small	1359		7.5	3.02	2.43	0.321	7	34.04	-2.84×10^{-6}	0.0012	37.51

In Table 2, we report descriptive statistics on the calculated intraday MFIV measure per size category. We observe a decrease in the average MFIV with increasing market capitalization, while the average MFIV is highest for the medium-sized category. The differences between the three different approaches (MN, WK, SP) are remarkably small; we do not observe any systematic patterns related to the three approaches.

The 1 min frequency of the MFIV measures allows us to analyze some of the commonly detected stylized facts in high-frequency financial data, as outlined in Cont (2001) or Andersen et al. (2001). Table 3 shows skewness, kurtosis, and correlations for the MFIV measures based on weekly options and the 1, 10, and 60 min frequencies for the different size categories. We observe slightly positive skewed and platokurtic distributions[5]. These findings correspond only partly to those of daily realized volatility measures by Andersen et al. (2001), who also detected slightly positively skewed distributions but observed higher values for kurtosis.

Considering the correlation of the MFIV measures with stock returns, we detect mostly negative correlations, which become more pronounced with decreasing sampling frequency, again corresponding to the stylized facts reported for realized measures by previous studies. With respect to the time series properties of our intraday MFIV measures, Figure A2 in the appendix shows the return series over the whole sample period together with the MFIV autocorrelation functions (ACFs) for a sample stock (Amazon) at 1 min, 10 min, and 60 min frequencies. The high-frequency return series can be seen to exhibit the typical volatility

clustering, however, to a lower degree as commonly observed for daily financial data. The MFIV ACFs show high autocorrelations that decay slowly, in particular at the shortest frequency of 1 min, indicating the presence of long-range dependence, which is commonly also detected for daily volatility measures (Cont 2001; Andersen et al. 2001).

Table 2. Descriptives on MFIV measures. The table presents descriptive statistics on MFIV measures based on monthly options (MN), weekly options (WK), and the interpolation approach (SP) for all samples and the small, medium, large, and mega cap sample firms.

	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis
			All			
MFIV (MN)	25.33	9.59	9.02	88.94	1.57	3.19
MFIV (WK)	25.03	9.74	9.71	86.82	1.58	3.33
MFIV (SP)	25.26	9.65	8.23	92.18	1.57	3.19
			Mega			
MFIV (MN)	22.50	7.35	9.02	88.94	2.27	9.18
MFIV (WK)	22.17	7.51	9.71	82.78	2.27	9.06
MFIV (SP)	22.42	7.38	8.23	92.18	2.24	8.95
			Large			
MFIV (MN)	33.00	7.37	13.88	68.63	0.63	0.58
MFIV (WK)	32.71	7.13	14.33	74.64	0.68	1.02
MFIV (SP)	32.95	7.45	17.60	65.08	0.73	0.88
			Medium			
MFIV (MN)	43.82	12.13	20.86	86.57	0.08	−0.91
MFIV (WK)	43.93	12.43	21.44	86.82	0.17	−0.75
MFIV (SP)	43.89	12.21	21.63	87.26	0.09	−0.91
			Small			
MFIV (MN)	36.82	5.27	29.26	60.25	0.96	0.33
MFIV (WK)	37.77	5.76	26.28	57.44	0.53	−0.93
MFIV (SP)	36.67	5.38	27.79	64.36	1.19	1.37

Table 3. Skewness, kurtosis, and correlations of MFIV measures. The table shows the skewness and kurtosis of MFIV measures based on weekly options as well as their correlation with stock returns based on 1 min, 10 min, and 60 min frequencies.

	Skewness			Kurtosis			Corr(MFIV, R)		
	1 min	10 min	60 min	1 min	10 min	60 min	1 min	10 min	60 min
All	0.61	0.61	0.62	−0.26	−0.25	−0.14	0.000	−0.002	−0.006
Mega	0.6	0.6	0.6	−0.27	−0.28	−0.21	0.000	0.001	0.001
Large	0.62	0.63	0.68	−0.3	−0.21	0.15	−0.001	−0.014	−0.037
Medium	0.68	0.68	0.69	0.18	0.16	0.17	−0.001	−0.004	−0.016
Small	0.54	0.55	0.56	−0.72	−0.71	−0.6	−0.003	−0.007	−0.027

The intraday MFIV also allows us to illustrate potential diurnal patterns. Figure 2 shows the averaged MFIV over a trading day, while Figure A3 in the appendix shows the intraday (1 min) return and MFIV time series for two sample stocks on a randomly selected day. For the sample stocks and random day in Figure A3, the MFIV measures seem to capture the higher fluctuations of the intraday returns at the beginning of the trading day rather well and subsequent decreases with decreasing variability of the returns throughout the day. Averaged over trading days and sample stocks, Figure 2 shows that MFIV increases at the beginning of the trading hours and slightly decreases during the day until markets close. As noted by Chen et al. (2021), this pattern is consistent with the notion that non-trading during the overnight period increases uncertainty, while trading, to some extent, resolves uncertainty. We confirm this pattern for individual equity MFIV and also find the pattern to be similar for the different size categories (detailed results are available upon request).
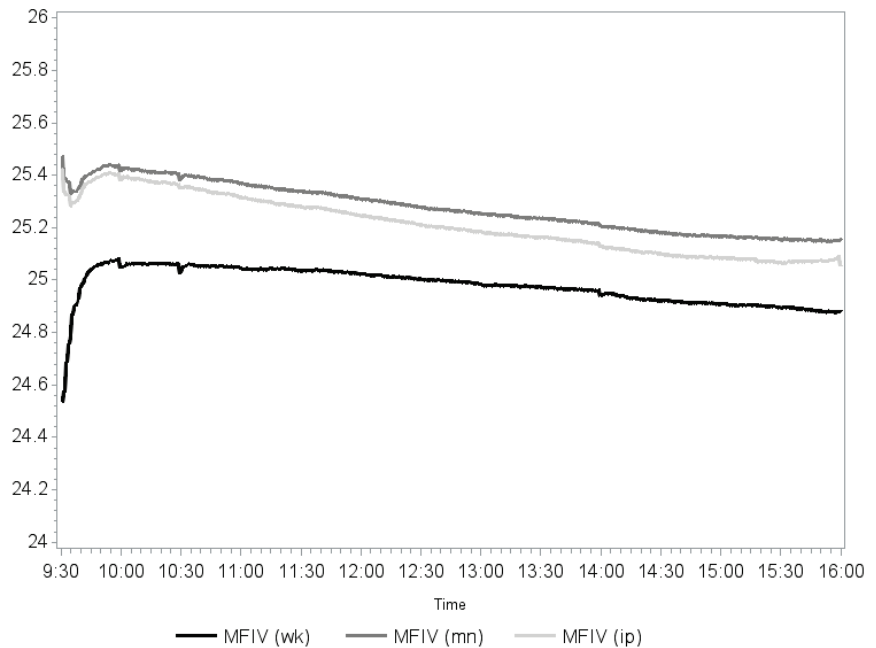
Figure 2. Diurnal pattern of the MFIV. The graph shows the WK, MN, and SP MFIV measures averaged over the sample stocks as well as over the sample days.

4. Analyzing the Intraday Return–Volatility Relationship

The 1 min frequency of the MFIV measures allows us to examine whether specific relationships and stylized facts also hold on an intraday and individual equity level. So far, the return–volatility relationship in financial markets has been analyzed extensively. While theoretically higher risks have to be compensated by higher returns, indicating a positive relationship, empirical studies mostly document a negative and potentially asymmetric relationship between returns and changes in volatility[6]. Concerning equity markets, fundamental theories have been put forward to explain this negative link. The leverage hypothesis suggested by Black (1976) and Christie (1982) explains an increase in equity volatility with a decline in the value of the firm, while the volatility-feedback hypothesis associates an increase in volatility with higher expected returns and lower current prices[7]. As explained by Bollerslev et al. (2006), both theories differ mainly by their implied direction of causality.

More recently, Hibbert et al. (2008) (among others) proposed a behavioral perspective and associated the negative return–volatility relationship with the presence of different groups of investors. These groups may differ with respect to the extent to which they are prone to behavioral biases, such as representativeness, affect, and extrapolation biases. In contrast to the leverage and volatility-feedback hypothesis, which may take time to establish and, therefore, are more plausible at lower frequencies, behavioral biases may very well be present even at intraday frequencies. However, existing studies on the return–volatility relationship at higher frequencies are rare and have focused mostly on equity indices[8]. Therefore, we use our intraday MFIV measures to examine the intraday return–volatility relationship for the individual equities in our sample.

For the subsequent analysis, we show results using the WK MFIV measures and what we refer to as the feasible subset of stocks, as outlined in Section 3.2 over the range of our sampling period. The sample covers the time period from 1 May 2017 to 30 June 2017, and we conduct our analysis on 1 min, 10 min, and 60 min frequencies. Results based on the

other measures are qualitatively similar and are available upon request. Despite the fact that the intraday MFIV measures exhibit only weak diurnal patterns, we conduct some basic stock-wise deseasonalization via standardization based on the time-day averages over all the sample days.

We follow Talukdar et al. (2017), Hibbert et al. (2008), Bekiros et al. (2017), and Badshah (2013) and conduct linear and quantile regressions using intraday stock-specific data on the 1, 10, and 60 min frequencies. We start with the following regressions:

$$\Delta MFIV_t = \beta_0 + \sum_{L=1}^{3} \beta_L \Delta MFIV_{t-L} + \sum_{L=0}^{3} \gamma_L R_{t-L}^+ + \sum_{L=0}^{3} \delta_L R_{t-L}^- + \varepsilon_t$$

$$R_t^+ = \begin{cases} R_t \text{ if } R_t > 0 \\ 0 \text{ if } R_t < 0 \end{cases} \qquad R_t^- = \begin{cases} R_t \text{ if } R_t < 0 \\ 0 \text{ if } R_t > 0 \end{cases}$$

where $\Delta MFIV_t$ indicates the first differences of the MFIV measure for a specific underlying and R_t represents the corresponding (positive and negative) log returns. We run our model on 1, 10, and 60 min frequencies.

Table 4 gives the regression results averaged over all sample stocks. We observe a negative average effect of contemporaneous returns (R_t), which increases in strength with decreasing sampling frequency. The average effect of negative contemporaneous returns is slightly more pronounced compared to the one of positive returns only for the 1 min and 60 min samples. Furthermore, the number of significant negative estimates and the strength of the effect decrease with increasing lag length. These findings constitute some first evidence that the asymmetric return–volatility relationship is evident on an intraday and stock-specific level. They correspond to findings on the index level by, e.g., Talukdar et al. (2017) and Hibbert et al. (2008) in so far as the contemporaneous returns are the most important factor in determining the current changes in volatility. The results are in line with the behavioral explanations mentioned by Talukdar et al. (2017) and Hibbert et al. (2008). However, since the effects of lagged returns are not overall insignificant, leverage or volatility-feedback phenomena cannot be ruled out as well.

To account for the fact that standard regressions may not constitute an adequate approach when applied to commonly leptokurtic returns, we conduct quantile regressions following Koenker and Bassett (1978), Bekiros et al. (2017), Badshah et al. (2016), and Daigler et al. (2014):

$$\Delta MFIV_t = \beta_0^q + \sum_{L=1}^{3} \beta_L^q \Delta MFIV_{t-L} + \sum_{L=0}^{3} \gamma_L^q R_{t-L} + \varepsilon_t \tag{1}$$

where $()^q$ refers to the parameter associated with the quantile regression model equation for the qth quantile. Average results for selected quantiles are shown in Table 5 for the 1 min frequency; results for 10 and 60 min MFIV are provided in the appendix. We observe a significant negative effect of contemporaneous returns on volatility changes for the majority of our sample stocks across all quantiles. We also find this effect to vary in strength across quantiles, being more pronounced in the tails of the distribution.

Figure 3 illustrates the average coefficients over a larger range of quantiles including average confidence bounds for 1 min data (results for other frequencies are available upon request and lead to similar conclusions). The results reveal substantial variation in the effects depending on the quantiles. We observe an inverted U-shaped pattern for the contemporaneous returns, where extreme returns have a substantially stronger negative impact on current volatility compared to those in the center of the distribution. This pattern corresponds to the one detected by Agbeyegbe (2016) for the U.S. stock market indices.

Table 4. Average return–volatility regression results. Regression Results for 1, 10, and 60 min frequencies averaged over the sample stocks. The parameters indicate the effect of positive returns (γ_L), negative returns (δ_L), and first differences in MFIV (β_L) on the first differences in MFIV for lags L of 0 to 3. Numbers in parentheses give the amount of significant negative/significant positive parameter estimates based on a 5% significance level.

	1 min	10 min	60 min
R_t^+	−7.69	−14.79	−15.17
	(66/1)	(65/14)	(69/15)
R_{t-1}^+	−1.22	−1.98	−2.27
	(25/5)	(52/9)	(31/7)
R_{t-2}^+	−1.22	−1.26	−1.58
	(24/1)	(28/6)	(25/4)
R_{t-3}^+	−0.89	−0.23	−1.35
	(21/2)	(12/5)	(20/5)
R_t^-	−9.01	−13.17	−15.35
	(81/1)	(74/14)	(74/14)
R_{t-1}^-	−5.03	−4.31	−4.22
	(52/1)	(47/10)	(42/4)
R_{t-2}^-	−2.95	−2.41	−3.02
	(35/2)	(32/8)	(24/7)
R_{t-3}^-	−2.48	−1.53	−2.37
	(16/3)	(21/6)	(19/6)
$\Delta MFIV_{t-1}$	−0.12	−0.15	−0.13
	(88/0)	(92/0)	(64/0)
$\Delta MFIV_{t-2}$	−0.07	−0.06	−0.05
	(55/2)	(53/1)	(37/1)
$\Delta MFIV_{t-3}$	−0.03	−0.02	−0.03
	(31/2)	(22/0)	(26/2)
β_0	0.00	0.00	−0.01
	(14/25)	(39/26)	(38/30)
$\gamma_0 = \delta_0$	16	60	60
$\gamma_1 = \delta_1$	19	38	26

Table 5. Average quantile regression results using 1 min data Average quantile regression results for individual equity based on Equation (1). Average (over all sample firms) parameter estimates are reported for different quantiles. Numbers in parentheses give the amount of significant negative/significant positive parameter estimates based on a 5% significance level.

	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.975
R_t	−10.167	−8.763	−7.496	−5.528	−3.860	−5.190	−7.405	−9.082	−10.822
	(112/2)	(110/3)	(108/4)	(103/8)	(95/5)	(100/8)	(104/5)	(107/2)	(106/3)
R_{t-1}	−2.695	−2.222	−1.880	−1.361	−0.979	−1.260	−1.735	−2.275	−3.166
	(67/3)	(73/3)	(83/5)	(82/5)	(80/5)	(72/5)	(73/6)	(69/4)	(64/4)
R_{t-2}	−1.692	−1.379	−1.181	−0.887	−0.574	−0.804	−1.245	−1.557	−1.759
	(50/4)	(54/1)	(64/0)	(73/0)	(79/1)	(66/1)	(60/1)	(52/2)	(43/5)
R_{t-3}	−1.033	−0.850	−0.711	−0.598	−0.381	−0.594	−0.939	−1.115	−1.756
	(32/4)	(36/0)	(39/0)	(52/1)	(56/1)	(57/2)	(45/1)	(42/4)	(39/3)
$\Delta MFIV_{t-1}$	−0.06	−0.05	−0.04	−0.03	−0.01	−0.02	−0.03	−0.03	−0.03
	(108/4)	(107/1)	(111/1)	(104/0)	(90/0)	(88/0)	(86/5)	(83/10)	(72/19)
$\Delta MFIV_{t-2}$	−0.030	−0.019	−0.010	−0.004	−0.001	−0.003	−0.005	−0.005	−0.005
	(95/13)	(81/13)	(70/13)	(42/10)	(21/5)	(36/7)	(41/24)	(46/29)	(46/33)
$\Delta MFIV_{t-3}$	−0.014	−0.008	−0.003	0.001	0.001	0.001	0.002	0.004	0.006
	(68/17)	(60/21)	(23/48)	(18/20)	(7/11)	(10/11)	(13/23)	(16/35)	(25/43)
constant	−0.783	−0.527	−0.331	−0.162	−0.037	0.091	0.347	0.608	0.910
	(122/0)	(121/0)	(122/0)	(120/0)	(61/47)	(0/122)	(0/122)	(0/122)	(0/122)

Figure 3. Quantile regression estimates and confidence intervals. The plots show the average estimated parameters based on Equation (1) for different quantiles (solid line) together with their 0.95 confidence bounds.

Over all quantiles, contemporaneous returns are the most important factor consistent with behavioral theories, in particular, those of affect and representativeness.[9] Contrary to Badshah (2013), we find no convincing evidence for loss aversion because extreme changes in the left tail do not seem to have a systematic higher impact than those in the right tail (Kahneman and Tversky 2013). Altogether, we find a significant negative contemporaneous return–volatility relationship on the intraday level for individual equity. The effect increases with decreasing sampling frequency and is more pronounced in the extreme tails of the

distributions. Considering the high frequency and the insignificant or small lagged effects, it is more likely that the asymmetric relationships are caused by behavioral biases rather than the fundamental theories underlying the leverage or volatility-feedback effects.

5. Conclusions

We implement the CBOE VIX method on intraday option data for individual equities. In doing so, we analyze the data quality with respect to the MFIV calculation. Our descriptive data analysis reveals that, for 138 individual equities, weekly options exhibit a sufficient range of strike prices to avoid significant truncation errors according to Jiang and Tian (2005). Descriptive analyses of MFIV sample averages based on weekly options, monthly options, and a cubic spline interpolation approach reveal only marginal differences. We find the intraday individual equity MFIV measures to exhibit a similar diurnal pattern as implied volatility indices, such as the VIX, with higher levels after opening hours and declining levels until closing hours.

We use our estimates to analyze the intraday return volatility–relationship using 1, 10, and 60 min data. We find a negative relationship between returns and volatility changes, which is significant for most of our sample stocks when considering contemporaneous returns. The stronger the link, the lower the frequency. For lagged returns, the effect is less evident. Quantile regressions reveal a clear inverted U-shaped pattern when considering contemporaneous returns, which again becomes less evident for lagged returns. These findings indicate that behavioral biases rather than the fundamental theories of leverage or volatility-feedback effects cause the asymmetric link between returns and volatility. Despite offering a starting point for calculating and analyzing individual equity MFIV at high frequencies, our approach comes with some limitations. The standard VIX calculation method is prone to have a specific cut-off rule to determine what option quotes are used in computing the VIX. As outlined by Andersen et al. (2015), this may bias the calculated MFIV measures. Andersen et al. (2015) offered an improved approach, termed the Corridor Implied Volatility, which alleviates the potential bias of the standard measure. Moreover, Andersen et al. (2015) noted that the relationship between returns and volatility is prone to jumps and co-jumps in both series and may be more complex depending on different regimes. Subsequently, while our empirical analysis of the return–volatility relationship may offer a valid starting point, future research should be conducted accounting for a potential dynamic nature of this relationship. Overall, our individual equity MFIV measure opens up avenues for further research. The forward-looking nature of implied volatility may have merits for volatility prediction on an intraday level, as well as on a daily level for individual equities. Moreover, the derived time series of intraday MFIV measures will allow us to assess the imminent effect of news on volatility relying on time series models. They also offer the opportunity to address the impact of attention and sentiment measures at such high frequencies. It will also be of great interest to extend the sampling period throughout the pandemic period and analyze the impact of lockdown situations on the stock volatilities of different industries. In general, MFIV measures also allow a decomposition into the conditional variance of stock returns and the equity variance premium, where the latter has been shown to possess predictive power for stock returns. The proposed MFIV measures can also be used to decompose volatility into a systematic risk component and idiosyncratic risk, using the VIX index as a measure of market risk. Such idiosyncratic risk measures, then again, may yield interesting insights into investors' risk perceptions and resulting investment decisions, potentially depending on and varying during high and low market volatility regimes.

Author Contributions: The authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: Data sharing is not applicable.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Derivation of Model-Free Implied Volatility

Following Demeterfi et al. (1999), assume the asset price S follows a continuous diffusion process:

$$\mathrm{d}S_t = S_t\big(\mu(t,\dots)\,\mathrm{d}t + \sigma(t,\dots)\,\mathrm{d}Z_t\big) \tag{A1}$$

where the drift μ and volatility σ are functions of time t and Z_t is a Brownian motion. From Ito's lemma, we have

$$\mathrm{d}\ln(S_t) = \sigma\,\mathrm{d}Z_t + \left(\mu - \frac{\sigma^2}{2}\,\mathrm{d}t\right) \tag{A2}$$

as shown in Appendix A. Using Equations (A1) and (A2), we can derive a representation of the incremental variance as

$$\frac{\mathrm{d}S_t}{S_t} - \mathrm{d}\ln(S_t) = \mu\,\mathrm{d}t + \sigma\,\mathrm{d}Z_t - \left(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}Z_t - \frac{\sigma^2}{2}\,\mathrm{d}t\right)$$

$$= \frac{\sigma^2}{2}\,\mathrm{d}t$$

$$\Leftrightarrow \sigma^2\,\mathrm{d}t = 2\left(\frac{\mathrm{d}S_t}{S_t} - \mathrm{d}\ln(S_t)\right) \tag{A3}$$

The realized (integrated) variance $V(t,T)$, with T denoting the time of expiration, is then given by

$$V(t,T) = \frac{1}{T}\int_0^T \sigma^2\,\mathrm{d}t$$

$$= \frac{2}{T}\int_0^T \left(\frac{\mathrm{d}S_t}{S_t} - \mathrm{d}\ln(S_t)\right)$$

$$= \frac{2}{T}\left[\int_0^T \frac{\mathrm{d}S_t}{S_t} - \ln\left(\frac{S_T}{S_0}\right)\right] \tag{A4}$$

where S_0 and S_T refer to the initial and final prices of the asset, respectively.

In the next step, we take risk-neutral expectations of Equation (A4):

$$\mathbb{E}\big[\mathbb{V}\big] = \frac{2}{T}\left\{\underbrace{\mathbb{E}\left[\int_0^T \frac{\mathrm{d}S_t}{S_t}\right]}_{(1)} + \underbrace{\mathbb{E}\left[-\ln\left(\frac{S_T}{S_0}\right)\right]}_{(2)}\right\} \tag{A5}$$

A risk-neutral setting, using the risk-free rate r implies

$$\mathrm{d}S_t = S_t r\,\mathrm{d}t + S_t \sigma\,\mathrm{d}Z_t$$

$$\Leftrightarrow \frac{\mathrm{d}S_t}{S_t} = r\,\mathrm{d}t + \sigma\,\mathrm{d}Z_t$$

The risk neutral expectation of term (1) in Equation (A5) is subsequently given by

$$\mathbb{E}\left[\int_0^T r\,\mathrm{d}t + \sigma\,\mathrm{d}Z_t\right] = rT \tag{A6}$$

The replication of term (2) of Equation (A5) requires a more intricate methodology. First, we decompose the logarithmic payoff by introducing the parameter $S_\star$, which also defines the boundary between put and call options that are used for the replication:

$$-\ln\left(\frac{S_t}{S_0}\right) = \underbrace{-\ln\left(\frac{S_t}{S_\star}\right)}_{(1)} \underbrace{-\ln\left(\frac{S_\star}{S_0}\right)}_{(2)} \tag{A7}$$

While term (2) here is constant, replicating term (1) is where the option methodology comes in. As Carr and Madan (1997) show in their appendix, any twice-differentiable payoff, $f(x)$, can be replicated as

$$f(x) = f(\kappa) - f'(x)\left[(x-\kappa)^+ - (\kappa-x)^+\right]$$
$$+ \int_0^\kappa f''(K)(K-x)^+ \, dK$$
$$+ \int_\kappa^\infty f''(K)(x-K)^+ \, dK$$

For our setting, we use $f(x) = \ln(x)$, $x = S_T$, $\kappa = S_\star$, the functions $P(K,S)$ and $C(K,S)$ to denote option prices with strike K for underlying price S and the put-call parity to write

$$\ln(S_T) = \ln(S_\star) + \frac{d\ln(S_t)}{dS_t}\left[C(S_\star, S_T) - P(S_\star, S_T)\right]$$
$$+ \int_0^{S_\star} \frac{d^2\ln(K)}{d^2K} P(K, S_T) \, dK$$
$$+ \int_{S_\star}^\infty \frac{d^2\ln(K)}{d^2K} C(K, S_T) \, dK$$
$$\Leftrightarrow \ln\left(\frac{S_t}{S_\star}\right) = \frac{1}{S_\star}(S_t - S_\star)$$
$$- \int_0^{S_\star} \frac{1}{K^2} P(K, S_T) \, dK$$
$$- \int_{S_\star}^\infty \frac{1}{K^2} C(K, S_T) \, dK$$

This replication strategy relies on the availability of a continuum of strike prices. Taking the risk-neutral expectation for term (1) in Equation (A7) yields

$$\mathbb{E}\left[-\ln\left(\frac{S_t}{S_\star}\right)\right] = \mathbb{E}\left[-\frac{S_T - S_\star}{S_\star} + \int_0^{S_\star} \frac{1}{K^2} P(K, S_T) \, dK + \int_{S_\star}^\infty \frac{1}{K^2} C(K, S_T) \, dK\right]$$
$$= -\left(\frac{S_0 e^{rT}}{S_\star} - 1\right) + e^{rT}\int_0^{S_\star} \frac{1}{K^2} P_0(K) \, dK + e^{rT}\int_{S_\star}^\infty \frac{1}{K^2} C_0(K) \, dK$$

where P_0, C_0, and S_0 denote the initial put, call, and underlying prices, respectively. Finally, we combine above results with Equations (A6) and (A7), and the replication of the expected variance reads as

$$\mathbb{E}[\mathbb{V}] = \frac{2}{T}\left[rT - \left(\frac{S_0 e^{rT}}{S_\star} - 1\right) - \ln\left(\frac{S_\star}{S_0}\right)\right.$$
$$+ e^{rT}\int_0^{S_\star} \frac{1}{K^2} P_0(K) \, dK \tag{A8}$$
$$\left. + e^{rT}\int_{S_\star}^\infty \frac{1}{K^2} C_0(K) \, dK\right]$$

Following Jiang and Tian (2007), it can be shown that the non-integral terms in Equation (A8) can be restated as

$$\frac{2}{T}\left[rT - \left(\frac{S_0 e^{rT}}{K_0} - 1\right) - \ln\left(\frac{K_0}{S_0}\right)\right] = \frac{2}{T}\left[\ln\left(\frac{F_0}{K_0}\right) - \left(\frac{F_0}{K_0} - 1\right)\right]$$

Applying the Taylor series expansion and ignoring terms higher than second-order yields

$$\ln\left(\frac{F_0}{K_0}\right) = \left(\frac{F_0}{K_0} - 1\right) - \frac{1}{2}\left(\frac{F_0}{K_0} - 1\right)^2$$

which results in the final approximation formula:

$$\frac{2}{T}\left[rT - \left(\frac{S_0 e^{rT}}{S_\star} + 1\right) - \ln\left(\frac{S_\star}{S_0}\right)\right] = -\frac{1}{T}\left(\frac{F_0}{K_0} - 1\right)^2 \tag{A9}$$

Ito's Lemma for $\ln(S_t)$

Applying Ito's lemma to the natural logarithm function $f(S_t) = \ln(S_t)$ yields

$$\begin{aligned}
\mathrm{d}\ln(S_t) &= \frac{\partial \ln(S_t)}{\partial t}\,\mathrm{d}t + \frac{\partial \ln(S_t)}{\partial S_t}\,\mathrm{d}S_t + \frac{1}{2}\frac{\partial^2 \ln(S_t)}{\partial S_t^2}(\mathrm{d}S_t)^2 \\
&= 0 + \frac{1}{S_t}S_t(\mu\,\mathrm{d}t + \sigma\,\mathrm{d}Z_t) - \frac{1}{2}\frac{1}{(S_t^2)}(\mathrm{d}S_t)^2 \\
&= \mu\,\mathrm{d}t + \sigma\,\mathrm{d}Z_t - \frac{1}{2}\frac{1}{S_t^2}\Bigl[\underbrace{(S_t\mu\,\mathrm{d}t)^2}_{\mathrm{d}t^2=0} + \underbrace{2S_t\mu\,\mathrm{d}t\,\sigma\,\mathrm{d}Z_t}_{\mathrm{d}t\,\mathrm{d}Z_t=0} + (S_t\sigma\,\mathrm{d}Z_t)^2\Bigr] \\
&= \mu\,\mathrm{d}t + \sigma\,\mathrm{d}Z_t - \frac{S_t^2\sigma^2(\mathrm{d}Z_t)^2}{2S_t^2} \\
&= \mu\,\mathrm{d}t + \sigma\,\mathrm{d}Z_t - \frac{\sigma^2}{2}\,\mathrm{d}t \qquad \text{since } (\mathrm{d}Z_t)^2 = \mathrm{d}t \\
&= \sigma\,\mathrm{d}Z_t + \left(\mu - \frac{\sigma^2}{2}\,\mathrm{d}t\right)
\end{aligned}$$

Appendix B. Additional Figures and Tables
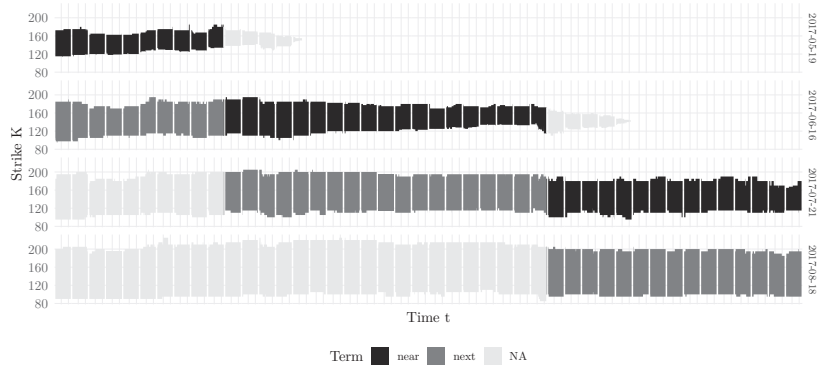
Figure A1. Exemplary option dataset (AAPL) for the MN method. Each bar represents the option contracts for a specific maturity and is colored based on its use as near-term (black) or next-term (dark gray) contract. The width represents the range of offered strike prices each minute. As this range declines sharply when the near-term contract reaches maturity, the set of contracts switches 5 trading days before that date.
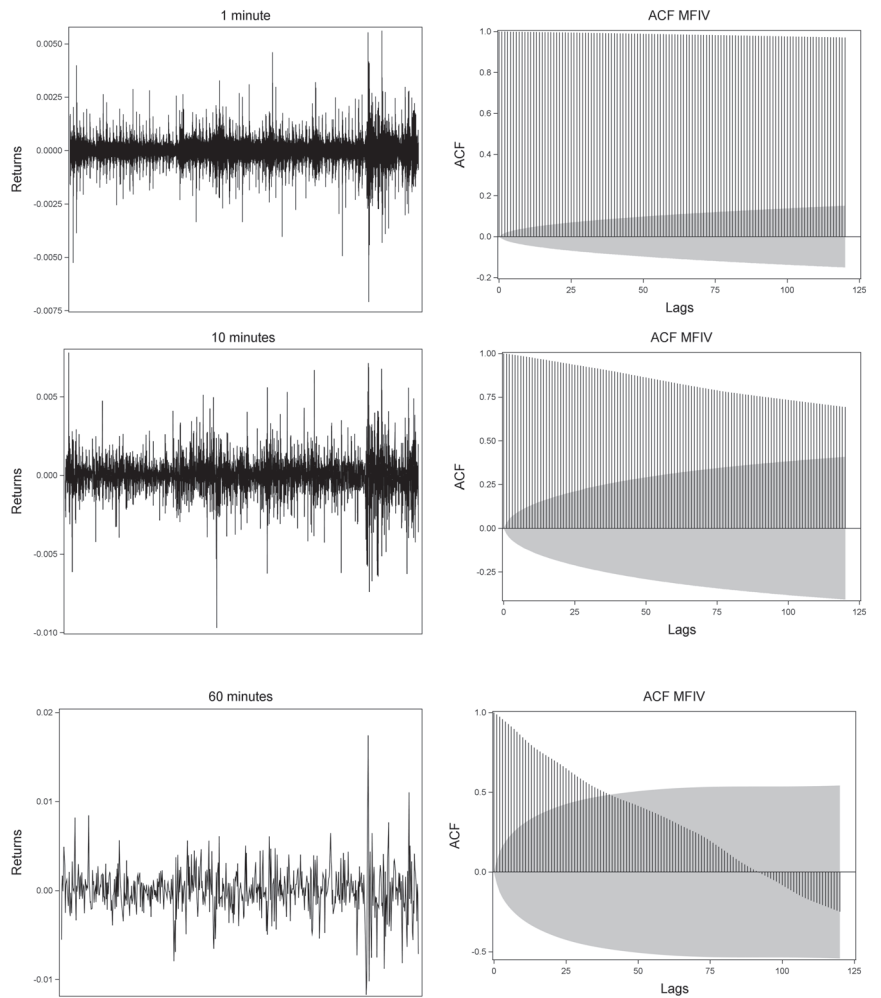
Figure A2. Return series and autocorrelation functions for MFIV measures. The figure shows the stock return series over the whole sampling period at the 1 min, 10 min, and 60 min frequencies and the ACF for the MFIV measures based on weekly options for a sample stock (Amazon).
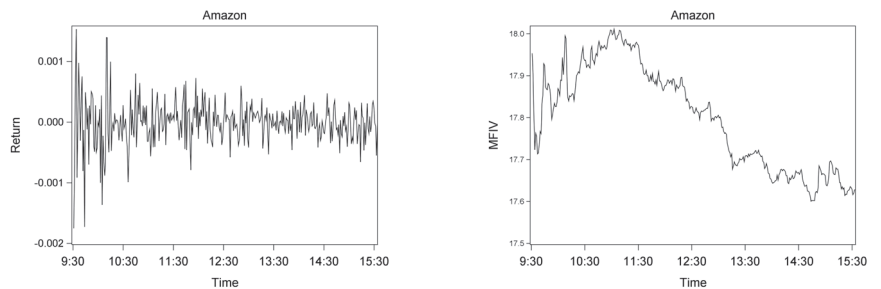
Figure A3. *Cont*.
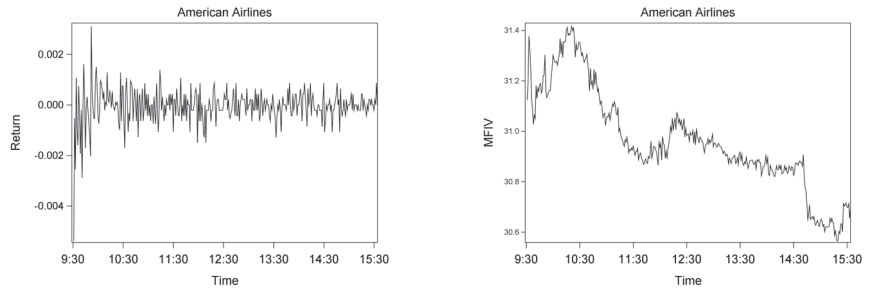
Figure A3. MVIF measure and stock returns. The figure shows the MFIV measure time series based on weekly options for a random day and two stocks for the 1 min frequencies.

Table A1. Average quantile regression results using 10 min data. Average quantile regression results for individual equity based on Equation (1). Average (over all sample firms) parameter estimates are reported for different quantiles. Numbers in parentheses give the numbers of significant negative/significant positive parameter estimates based on a 5% significance level.

	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.975
R_t	−13.764	−11.538	−9.993	−9.201	−9.043	−9.732	−10.825	−12.316	−14.233
	(104/18)	(105/17)	(110/12)	(109/13)	(109/13)	(110/11)	(108/14)	(107/15)	(104/18)
R_{t-1}	−1.470	−1.253	−1.015	−0.988	−0.958	−1.154	−1.926	−2.195	−3.723
	(87/35)	(92/31)	(93/28)	(93/29)	(94/28)	(97/25)	(94/28)	(94/28)	(92/30)
R_{t-2}	−0.438	−0.532	−0.365	−0.182	−0.165	−0.364	−0.763	−1.368	−2.889
	(79/43)	(81/41)	(77/45)	(67/55)	(71/51)	(79/43)	(88/34)	(85/37)	(93/29)
R_{t-3}	−0.202	−0.552	−0.356	−0.188	−0.228	−0.323	−0.695	−1.203	−1.550
	(66/56)	(78/44)	(78/44)	(77/44)	(74/44)	(70/52)	(78/44)	(80/42)	(82/40)
$\Delta MFIV_{t-1}$	−0.12	−0.09	−0.07	−0.05	−0.03	−0.03	−0.05	−0.09	−0.15
	(101/21)	(103/19)	(105/17)	(100/22)	(91/31)	(93/29)	(94/28)	(90/32)	(85/37)
$\Delta MFIV_{t-2}$	−0.063	−0.041	−0.029	−0.016	−0.009	−0.009	−0.020	−0.040	−0.071
	(89/33)	(91/31)	(91/31)	(90/32)	(81/41)	(75/41)	(90/32)	(84/38)	(77/45)
$\Delta MFIV_{t-3}$	−0.025	−0.017	−0.012	−0.005	0.000	0.000	−0.004	−0.013	−0.032
	(89/33)	(87/35)	(83/39)	(83/39)	(79/43)	(74/48)	(81/41)	(76/46)	(74/48)
constant	−0.380	−0.227	−0.135	−0.056	−0.006	0.038	0.132	0.268	0.503
	(122/0)	(122/0)	(122/0)	(122/0)	(108/4)	(0/122)	(0/122)	(0/122)	(0/122)

Table A2. Average quantile regression results using 60 min data. Average quantile regression results for individual equity based on Equation (1). Average (over all sample firms) parameter estimates are reported for different quantiles. Numbers in parentheses give the number of significant negative/significant positive parameter estimates based on a 5% significance level.

	0.025	0.05	0.10	0.25	0.50	0.75	0.90	0.95	0.975
R_t	−13.764	−11.538	−9.993	−9.201	−9.043	−9.732	−10.825	−12.316	−14.233
	(104/18)	(105/17)	(110/12)	(109/13)	(109/13)	(110/11)	(108/14)	(107/15)	(104/18)
R_{t-1}	−1.470	−1.253	−1.015	−0.988	−0.958	−1.154	−1.926	−2.195	−3.723
	(87/35)	(92/31)	(93/28)	(93/29)	(94/28)	(97/25)	(94/28)	(94/28)	(92/30)
R_{t-2}	−0.438	−0.532	−0.365	−0.182	−0.165	−0.364	−0.763	−1.368	−2.889
	(79/43)	(81/41)	(77/45)	(67/55)	(71/51)	(79/43)	(88/34)	(85/37)	(93/29)
R_{t-3}	−0.202	−0.552	−0.356	−0.188	−0.228	−0.323	−0.695	−1.203	−1.550
	(66/56)	(78/44)	(78/44)	(77/44)	(74/44)	(70/52)	(78/44)	(80/42)	(82/40)
$\Delta MFIV_{t-1}$	−0.12	−0.09	−0.07	−0.05	−0.03	−0.03	−0.05	−0.09	−0.15
	(101/21)	(103/19)	(105/17)	(100/22)	(91/31)	(93/29)	(94/28)	(90/32)	(85/37)
$\Delta MFIV_{t-2}$	−0.063	−0.041	−0.029	−0.016	−0.009	−0.009	−0.020	−0.040	−0.071
	(89/33)	(91/31)	(91/31)	(90/32)	(81/41)	(75/41)	(90/32)	(84/38)	(77/45)
$\Delta MFIV_{t-3}$	−0.025	−0.017	−0.012	−0.005	0.000	0.000	−0.004	−0.013	−0.032
	(89/33)	(87/35)	(83/39)	(83/39)	(79/43)	(74/48)	(81/41)	(76/46)	(74/48)
constant	−0.380	−0.227	−0.135	−0.056	−0.006	0.038	0.132	0.268	0.503
	(122/0)	(122/0)	(122/0)	(122/0)	(108/4)	(0/122)	(0/122)	(0/122)	(0/122)

Notes

[1]   We developed an R package for these calculations, which is currently available on https://github.com/m-g-h/R.MFIV, accessed on 20 November 2023.

[2]   Critically assessed, the notion of "model-freeness" is not entirely correct since an assumption about the underlying price process is still made. However, no option pricing model is required, which frees the IV from limitations due to potentially unrealistic model assumptions.

[3]   i.e., multiples of $\sigma_t\sqrt{\tau}$, with σ_t the ATM Black & Scholes IV and τ the time to maturity.

[4]   The VIX method is derived for European-style options only. Individual equity options, however, are American-style options and may be subject to an early exercise premium. As this premium can be assumed to be relatively small for out-of-the-money options and since we do not want to rely on a specific option pricing model for estimation, we do not account for it.

[5]   Jarque-Bera tests indicate the rejection of a normal distribution for all individual equity MFIVs. Detailed results are available upon request.

[6]   Compare Dennis et al. (2006), Fleming et al. (1995), Giot (2005), Hibbert et al. (2008), Carr and Wu (2017), and Talukdar et al. (2017).

[7]   Compare French et al. (1987), Campbell and Hentschel (1992), and Bekaert and Wu (2000).

[8]   Compare Andersen et al. (2021), Kalnina and Xiu (2017), Badshah (2013), and Andersen et al. (2015).

[9]   Among others outlined by Hibbert et al. (2008) and Daigler et al. (2014), investors may view a high return and low risk (decreasing volatility) as representative of a good investment. Combined with an affect heuristic, where investors' decisions may be governed or at least affected by intuition and instincts, they may act on the negative returns and high volatility, which are both negatively labeled and thereby cause a negative return–volatility relationship.

References

Agbeyegbe, Terence D. 2016. Modeling US stock market volatility-return dependence using conditional copula and quantile regression. *Studies in Economic Theory* 29: 597–621. [CrossRef]

Andersen, Torben G., Oleg Bondarenko, and Maria T. Gonzalez-Perez. 2015. Exploring return dynamics via corridor implied volatility. *The Review of Financial Studies* 28: 2902–45. [CrossRef]

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Heiko Ebens. 2001. The distribution of realized stockrreturn volatility. *Journal of Financial Economics* 61: 43–76. [CrossRef]

Andersen, Torben, Ilya Archakov, Leon Grund, Nikolaus Hautsch, Yifan Li, Sergey Nasekin, Ingmar Nolte, Manh Cuong Pham, Stephen Taylor, and Viktor Todorov. 2021. A Descriptive Study of High-Frequency Trade and Quote Option Data*. *Journal of Financial Econometrics* 19: 128–77. [CrossRef]

Badshah, Ihsan Ullah. 2013. Quantile regression analysis of the asymmetric return-volatility relation: Quantile regression analysis. *Journal of Futures Markets* 33: 235–65. [CrossRef]

Badshah, Ihsan, Bart Frijns, Johan Knif, and Alireza Tourani-Rad. 2016. Asymmetries of the intraday return-volatility relation. *International Review of Financial Analysis* 48: 182–92. [CrossRef]

Bekaert, Geert, and Guojun Wu. 2000. Asymmetric Volatility and Risk in Equity Markets. *The Review of Financial Studies* 13: 1–42. [CrossRef]

Bekiros, Stelios, Mouna Jlassi, Kamel Naoui, and Gazi Uddin. 2017. The asymmetric relationship between returns and implied volatility: Evidence from global stock markets. *Journal of Financial Stability* 30: 156–74.

Black, Fischer. 1976. *Studies of Stock Market Volatility Changes*. Irvine: Scientific Research Publishing Inc., pp. 177–81.

Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 81: 637–54. [CrossRef]

Bollerslev, Tim, Julia Litvinova, and George Tauchen. 2006. Leverage and Volatility Feedback Effects in High-Frequency Data. *Journal of Financial Econometrics* 4: 353–84. [CrossRef]

Breeden, Douglas T., and Robert H. Litzenberger. 1978. Prices of state-contingent claims implicit in option prices. *Journal of Business* 51: 621–51.

Britten-Jones, Mark, and Anthony Neuberger. 2000. Option prices, implied price processes, and stochastic volatility. *The Journal of Finance* 55: 839–66. [CrossRef]

Campbell, John Y., and Ludger Hentschel. 1992. No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics* 31: 281–318. [CrossRef]

Carr, Peter, and Dilip Madan. 1997. *Towards a Theory of Volatility Trading*. Cambridge: Cambridge University Press.

Carr, Peter, and Liuren Wu. 2017. Leverage effect, volatility feedback, and self-exciting market disruptions. *Journal of Financial and Quantitative Analysis* 52: 2119–156. [CrossRef]

CBOE. 2023. CBOE VIX Whitepaper. Available online: https://cdn.cboe.com/api/global/us_indices/governance/Volatility_Index_Methodology_Cboe_Volatility_Index.pdf (accessed on 20 November 2023).

Chen, Jingjing, George J. Jiang, Chaowen Yuan, and Dongming Zhu. 2021. Breaking vix at open: Evidence of uncertainty creation and resolution. *Journal of Banking & Finance* 124: 106–60. [CrossRef]

Christie, Andrew A. 1982. The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics* 10: 407–32. [CrossRef]

Cont, Rama. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1: 223–36. [CrossRef]

Daigler, Robert T., Ann Marie Hibbert, and Ivelina Pavlova. 2014. Examining the return–volatility relation for foreign exchange: Evidence from the euro vix. *Journal of Futures Markets* 34: 74–92. [CrossRef]

Demeterfi, Kresimir, Emanuel Derman, Michael Kamal, and Joseph Zou. 1999. A guide to volatility and variance swaps. *The Journal of Derivatives* 6: 9–32. [CrossRef]

Dennis, Patrick, Stewart Mayhew, and Chris Stivers. 2006. Stock returns, implied volatility innovations, and the asymmetric volatility phenomenon. *Journal of Financial and Quantitative Analysis* 41: 381–406. [CrossRef]

Fleming, Jeff, Barbara Ostdiek, and Robert E. Whaley. 1995. Predicting stock market volatility: A new measure. *Journal of Futures Markets* 15: 265–302. [CrossRef]

French, Kenneth R., G. William Schwert, and Robert F. Stambaugh. 1987. Expected stock returns and volatility. *Journal of Financial Economics* 19: 3–29. [CrossRef]

Giot, Pierre. 2005. Relationships between implied volatility indexes and stock index returns. *The Journal of Portfolio Management* 31: 92–100. [CrossRef]

Gonzalez-Perez, Maria T. 2015. Model-free volatility indexes in the financial literature: Areview. *International Review of Economics & Finance* 40: 141–59. [CrossRef]

Green, Richard C., and Robert A. Jarrow. 1987. Spanning and completeness in markets with contingent claims. *Journal of Economic Theory* 41: 202–10. [CrossRef]

Hibbert, Ann Marie, Robert T. Daigler, and Brice Dupoyet. 2008. A behavioral explanation for the negative asymmetric return–volatility relation. *Journal of Banking & Finance* 32: 2254–66. [CrossRef]

Ishida, Isao, Michael McAleer, and Kosuke Oya. 2011. Estimating the leverage parameter of continuous-time stochastic volatility models using high frequency s&p 500 and VIX. *Managerial Finance* 37: 1048–67. [CrossRef]

Jiang, George J., and Yisong S. Tian. 2005. The model-free implied volatility and its information content. *The Review of Financial Studies* 18: 1305–42. [CrossRef]

Jiang, George J., and Yisong S. Tian. 2007. Extracting model-free volatility from option prices: An examination of the VIX index. *The Journal of Derivatives* 14: 35–60. [CrossRef]

Kahneman, Daniel, and Amos Tversky. 2013. Prospect theory: An analysis of decision under risk. In *Handbook of the Fundamentals of Financial Decision Making*. Singapore: World Scientific Publishing, pp. 99–127. [CrossRef]

Kalnina, Ilze, and Dacheng Xiu. 2017. Nonparametric estimation of the leverage effect: A trade-off between robustness and efficiency. *Journal of the American Statistical Association* 112: 384–96. [CrossRef]

Koenker, Roger, and Gilbert Bassett. 1978. Regression quantiles. *Econometrica* 46: 33–50. [CrossRef]

Latané, Henry A., and Richard J. Rendleman. 1976. Standard deviations of stock price ratios implied in option prices. *The Journal of Finance* 31: 369–81. [CrossRef]

Nachman, David C. 1988. Spanning and completeness with options. *The Review of Financial Studies* 1: 311–28. [CrossRef]

Poon, Ser-Huang, and Clive W. J. Granger. 2003. Forecasting volatility in financial markets: A review. *Journal of Economic Literature* 41: 478–539. [CrossRef]

Talukdar, Bakhtear, Robert T. Daigler, and A. M. Parhizgari. 2017. Expanding the explanations for the return–volatility relation. *Journal of Futures Markets* 37: 689–716. [CrossRef]

Taylor, Stephen J., Pradeep K. Yadav, and Yuanyuan Zhang. 2010. The information content of implied vlatilities and model-free volatility expectations: Evidence from options written on individual stocks. *Journal of Banking & Finance* 34: 871–81. [CrossRef]

Whaley, Robert E. 1993. Derivatives on market volatility: Hedging tools long overdue. *The Journal of Derivatives Fall* 1: 71–84. [CrossRef]

Whaley, Robert E. 2000. The investor fear gauge. *The Journal of Portfolio Management* 26: 12–17. [CrossRef]

Article

An Investigation into the Spatial Distribution of British Housing Market Activity

David Paul Gray

Department of Accountancy Finance and Economics, University of Lincoln, Lincoln LN6 7TS, UK;
dgray@lincoln.ac.uk

Abstract: This paper sets out to consider how a simple and easy-to-estimate power-law exponent can be used by policymakers to assess changes in economic inequalities, where the data can have a long tail—common in analyses of economic disparities—yet does not necessarily deviate from log-normality. The paper finds that the time paths of the coefficient of variation and the exponents from Lavalette's function convey similar inferences about inequalities when analysing the value of house purchases over the period 2001–2022 for England and Wales. The house price distribution 'steepens' in the central period, mostly covering the post-financial-crisis era. The distribution of districts' expenditure on house purchases 'steepens' more quickly. This, in part, is related to the loose monetary policy associated with QE driving a wedge between London and the rest of the nation. As prices can rise whilst transactions decline, it may be better for policymakers to focus on the value of house purchases rather than house prices when seeking markers of changes in housing market activity.

Keywords: housing market transactions; house prices; England and Wales; Lavalette's law; convergence and divergence

1. Introduction

Population nodes are observed to follow a regularity characterised by Zipf's law, which is a log–log relationship between the rank-size of cities and their corresponding populations. A direct link with central place theory (Hsu 2012) has been made. Cristelli et al. (2012) observe that Zipf's power law has become a 'universal' expression for measuring scale and size in many fields, including economic convergence (Tang et al. 2016), yet evidence for it is not unequivocal. Perline (2005) is also critical of the widespread use of power laws that may not be the best characterisation of distributions. He argues that some distributions that are believed to follow a power law can be confused with a log-normal distribution if there is substantial truncation.

D'Acci (2023) proposed that the existence of a power law in the distribution of settlement populations should be related to a power law in average house prices, at least in the upper tail. Blackwell (2018) finds limited evidence that house price distributions follow a power law. There is a concession that the tail is fatter than a log-normal one, but not as fat as a 'true' power law in data from housing trades in the County of Charleston, South Carolina from 2001 to 2008. It could be that house price data follow a power law in certain price cycle phases. Ohnishi et al. (2020) find the Tokyo house price dispersion is very close to a log-normal distribution in normal times but fits a power function in a boom. They suggest that the shape of the (size-adjusted) price distribution, especially that of the tail, can be investigated for signalling the existence of a bubble.

Fontanelli et al. (2016) note that empirical data often exhibit good power-law distribution within a limited range. Rather than concentrating on where the power law ceases to hold, they modify a power law by changing the functional form. Lavalette's function is potentially a useful means of describing and quantifying power-law-like behaviours. The

Lavalette distribution yields a very good approximation to the log-normal whilst echoing a standard power function, capable of representing long tails. As such, it could address Perline-Blackwell's critique of applying power functions to log-normal (housing) data.

Van Nieuwerburgh and Weill (2010) find that there is a steeper house price distribution over time. They argue that the driver of spatial house price variations is the city productivity. Behrens et al. (2014) emphasise how productivity affects city size, producing the Zipfian distribution of settlement populations. With productivity also affecting average house prices, a change in the distribution of productivity across space would impact house price inequalities. Van Nieuwerburgh and Weill use a coefficient of variation to assess the steepening spread. The same coefficient is used for sigma-convergence. In the growth literature, this concerns how the distribution (of income) evolves over time (Sala-i-Martin 1996). Gray (2023b) finds that the time profile of the Lavalettean exponents closely tracks that of the coefficient of variation. As both are simple to estimate using, say, Microsoft Excel 2019, the exponent could be quoted alongside the coefficient when presenting cases of growing inequalities to policymakers. This paper considers whether there is 'a steepening' or convergence in district house prices and relates this to other measures of housing trades. It compares the results using the coefficient with the exponent.

The paper is structured as follows: First, there is a discussion of central place theory and convergence. Applications of power laws in the fields of price and affordability spreads follow. The significant change in housing transactions following the financial crash is introduced next, plus work that features transactions.

How house prices are expected to vary across space is reviewed with an emphasis on risk. This is followed by drawing a distinction between price changes and expected housing market participation in hot and cold markets. The data analyses are selected for ease of use with widely available software. This includes simple regression. The focus is a Lavalettean expression. Growth is split into the growth of the exponent and the growth of the median. This is adapted to assess the special case of pro-poor growth. The data sources are outlined.

The results show that price and housing market expenditure distributions steepen but these are not linked to a growth period, at odds with Blackwell-Ohnishi et al. A six-year period of relatively rapid price growth before the crash of 2008 is compared with another after the recovery.

2. Literature

The city size regularity characterised by Zipf's law matches central place theory (Hsu 2012) predictions. Behrens et al. (2014) argue that large cities produce more output per capita than small cities because of a sorting of talented individuals. More talented individuals stand a better chance of becoming highly productive entrepreneurs in larger cities. Correspondingly, there are tougher selection processes in more 'talented' cities. Entrepreneurs and firms have better resources to draw from because of the agglomeration economies, boosting productivity, explaining why cities with higher proportions of those with high levels of human capital are larger in equilibrium. Their model generates a Zipfian relationship for city sizes under plausible parameter values.

Cristelli et al. (2012) argue that many real systems do not show true Zipfian behaviour because they are incomplete or inconsistent with the conditions under which one might expect power laws to emerge. A consequence is that, in general, Zipf's law does not hold for subsets or a union of Zipfian sets. A Zipfian distribution is L-shaped with sizeable outliers at the top end. Perline (2005) points out that it is not uncommon for researchers to truncate the lower tail where the size of the node is small, which could result in some distributions that are believed to follow a power law being confused with log-normal ones.

The notion that a power law in city size has an implication for an associated variable is explored by Rozenfeld et al. (2011), who show that, as well as the population of a node, the footprint of a city also follows a Zipfian distribution. The third leg of the stool, population density, does not. Behrens et al. (2014) predict that, despite urban costs of

higher accommodation and commuting time in larger cities, agents do not apportion a greater share of expenditure on housing.

In the field of house prices, D'Acci (2023) finds that Italian regional house prices have a heavy-tailed distribution for which the maximum likelihood estimator suggests a power-law shape is a plausible function for the majority of cases. He suggests that the link is based on per capita income and spatial equilibrium (Roback 1982). This is at odds with the work of Blackwell (2018) who finds limited evidence that house price distributions follow a power law. He concludes that data from housing trades in the County of Charleston, South Carolina from 2001 to 2008 have a fatter tail than log-normal, but not as fat as a 'true' power law. This 'in-between' possibility is supported when the 'regular' power law is compared with the power law with a cut-off. There is some support for a power law with a cut-off. A proposed candidate for exploring this 'in-between' zone is a Lavalette function.

Fontanelli et al. (2016) review the properties of the Lavalette function. In their Figure 1 (p. 4) they show how various exponents generate different PDFs. A low value (around −0.1) could generate a bell shape whilst over −0.5, what emerges is something akin to a Zipfian distribution. However, in between, the Lavalette rank function generates a PDF indistinguishable from a log-normal distribution. It is a special case of a discrete generalized beta distribution, which entails estimating two exponents rather than one, which in turn presents estimation complexities, making it less than ideal for simple policy analysis. Lavalette's special case entails the two exponents being equal. The formula describes a semi-logarithmic *S*-shape in the cumulative distribution (Chlebus and Divgi 2007). This shape implies that the data should cover the full distribution, not a truncated set. Cerqueti and Ausloos (2015a, 2015b) favour a Lavalettean power law over a Zipfian one for subnational spatial dispersion of Italian tax income. Gray (2022a, 2023b) prefers the Lavalette for subnational inequalities in house prices and affordability ratios over a power law. Using Lavalette's exponent, he also finds a steepening of spatial house prices and the affordability ratio of England and Wales district distributions. The steepening is between 2006 and 2017. Consistent with the sorting argument seen in Behrens et al. (2014), it is argued that lenders are more willing to advance loans to borrowers in areas attractive to talented individuals, which would strongly favour an extended London area in the UK case. This lending bias could be viewed as reflecting risk-adjusted returns to a dwelling purchase (Gray 2023b; Sinai 2010).

It could be that house price distributions vary with price cycle phases. It is argued that house prices tend to grow faster in larger agglomerations beyond that justified by rents, generating excess returns (Amaral et al. 2021). Mian and Sufi (2018) see the credit-driven household demand channel as distinct from traditional financial accelerator models in explaining house price dynamics, primarily due to the centrality of households in explaining the real effects of credit supply expansions. Lenders inflate the wedge between prices and incomes, which is more likely to leave a permanent effect on high-house-priced areas. Evidence for this can be found in Gray (2022b) who reveals that the time paths of British house price–earnings ratios reflect a spatial divide. The ratios generally rose from a low in 1997 to the bubble period of the 2004–2008 peak. Subsequently, for the South of England, there has been a continuation of this increase, whereas for other areas, the picture is one of relative stability. Rising inequality in England and Wales has two dimensions. Firstly, between the North and South, and secondly among the southern districts.

Bogin et al. (2017) conclude that the price acceleration is a signal of a permanent shift in a location's economic fundamentals. As the largest nodes at the top end of price hierarchies offering property investment opportunities for a wealthy, international elite (Fernandez et al. 2016), Dublin and London may have decoupled from the rest of the British Isles (Richmond 2007). This suggests a steepening of the price distribution in both countries.

Ohnishi et al. (2020) find the Tokyo house price dispersion is very close to a log-normal distribution in normal times but fits a power function in a boom. They suggest that the shape of the (size-adjusted) price distribution, especially that of the tail, can be investigated

for signalling the existence of a bubble. So, one might expect the spatial distribution of house prices to expand in a house price boom.
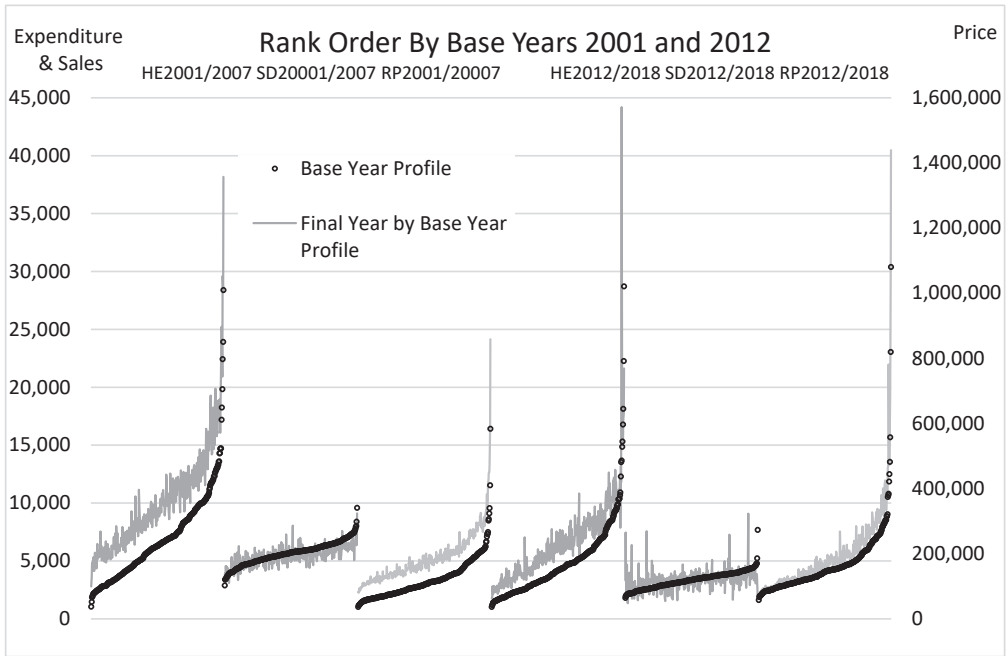
Figure 1. The distribution of housing expenditure, sales and real price.

Hudson and Green (2017) identify that since 2008–2009 there are 400,000 fewer housing transactions taking place each year in the UK compared with the period before the financial crisis, of which 80% could be attributed to a fall in mortgaged home movers. One could argue that this echoes earlier collapses. Andrew and Meen (2003) make a similar point about missing transactions in the 1990s following the 1989 bubble burst. Ortalo-Magné and Rady (2004) suggested that housing market participation in the 1980s among young buyers was unusually high because of credit liberalisation and the rising trend in owner-occupation.

If the housing adjustment following a financial crisis is in transactions, the implications should be of interest to policymakers. Articles featuring house price dispersion *and* transactions are not common. Tsai (2018) finds the ripple effect in four regional housing markets in the U.S. A 'ripple' in transactions was far more evident than that in housing prices. Information is transferred between regional housing markets either through price or volume. The two types of ripple effects are negatively correlated.

Analysing house prices and transactions across European economies, Dröes and Francke (2018) argue that common underlying factors, such as GDP and interest rates, explain part of the price–turnover correlation. The effect of GDP and interest rates mainly operates through turnover. Although a high loan-to-GDP ratio does increase the effect of interest rates and GDP on prices and turnover, it is not considered a key factor in explaining price and turnover dynamics. Similarly, neither population increases, the share of the young population, nor inflation play a central role in this context. They conclude that prices and turnover should be modelled as two interdependent processes. The period 1999–2013 seems unaffected by the drop in participants in 2008, which is not explained.

Clayton et al. (2010) find that price-caused components of prices and volume are negatively correlated. House prices in tight markets are less affected by financial constraints on homebuyers. Dividing the 114 U.S. metropolitan statistical areas into those with high

and low supply elasticities, they find that in markets where supply can easily adjust, transaction volume does not seem to affect future prices.

3. Land and Pricing

An asset price model relates the price of a dwelling to the rental stream and the cost of capital, subject to risk adjustments. DiPasquale and Wheaton (1996) analyse factors that affect rent and land prices. The monocentric urban model characterises a collection of dwellings as part of the same housing market area if there is a tendency towards a stable hierarchy of prices. A standard dwelling closer to the central business district will command a higher price *ceteris paribus*, as owners benefit from the lesser disutility of commuting. The co-movement of prices emerges from 'arbitrage'; buyers switch search behaviour across the commuting space in the face of mispriced local markets. Relative prices change little as the overall market undergoes either cyclic fluctuations or long-term growth (p. 26). Inter-urban differences in house prices are a function of the relative productivity of areas. Spatial arbitrage operates more generally, as individual agents migrate within *and between* population nodes to maximise their utility (Roback 1982). This model implies the driver of house price differences is productivity, adjusted for house characteristics and commuting preferences.

The expected future growth in current rent would affect the current local house price (Amaral et al. 2021; DiPasquale and Wheaton 1996; Sinai 2010). This could be due to population growth (DiPasquale and Wheaton 1996; Glaeser and Gyourko 2005), or productivity growth (Coulson et al. 2013; Van Nieuwerburgh and Weill 2010).

Dual regional economy models (Brakman et al. 2020; McCombie 1988) predict slower or constrained growth in the periphery and its elements should have persistently lower productivity. The city's economic fortunes will be a function of the industries it supports. Martin et al. (2014, 2018) utilise an evolutionary perspective, where agglomeration economies trace out productivity development paths. An ageing economic structure could persistently constrain a city to a poor performance, such as in the northern cities of England that are subject to deindustrialisation (Martin et al. 2018; Pike et al. 2016).

DiPasquale and Wheaton (1996, p. 44) also argue that if the cost of capital falls, this drives up all asset prices. Aimed at mitigating the impact of the financial crisis of 2008, central banks engaged in quantitative easing (QE), which inflated asset prices. The Bank of England's loose monetary policy, which began with its base rate falling to 0.5% in March 2009, down from 5% in the previous October, was implemented to mitigate the impact of the financial crisis. A longer-run view, due to Miles and Monro (2019), is that there was a *sustained* decline in real interest rates between 1985 and 2018. Himmelberg et al. (2005) argue that house prices are more sensitive to changes in real interest rates in rapidly growing cities. Amaral et al. (2021) argue that house prices tend to grow faster in major 'superstar' cities than what is justified by rents. The excess returns are explained by the lower risk associated with the rents. The distributional impact of QE on measured income and wealth between 2008 and 2014 is assessed as minor in proportional terms. In cash terms, London and the South East gained the most, particularly in housing wealth (Bunn et al. 2018). Indeed, it is averred that London pulls away from the rest of England and Wales (Gray 2018; Montagnoli and Nagayasu 2015; Richmond 2007). This suggests that the price distribution broadened in the aftermath of the crisis due to QE.

A relaxation in credit controls should lead to a surge in demand for dwellings. However, vendors could anticipate this and revise their asking (expected) price upwards before the matching rate increases or the number of transactions surges. Allen and Gale (2000) argue higher price levels are supported by the anticipation of further increases in credit and house prices in general.

4. Transaction–Price–Expenditure Nexus

Ortalo-Magné and Rady (2004) place buyers in a hierarchy, with first-time buyers focussing on more modest dwellings, whereas repeat buyers are looking for larger homes.

Smaller dwellings would be traded more frequently than large ones. Smaller dwellings are the first or the only stage in a housing career. Staging could result from a capital constraint or because buyers match their current dwelling with their current space needs. With a variety of divisions of district sizes that are administrative areas rather than markets, the rank order for sales, price levels, and expenditures should not be the same.

Stein's (1995) model of house prices and transactions focuses on repeat buyers and equity. He analyses the impact of rising prices on three mover groups. The first group relates to those already wishing to buy. A price increase enhances their collateral. Not only does this fortify their purchasing power but, as a lower risk, this would grant them greater access to credit, which leads to further price growth.

The second group, not in the market originally, is induced to sell. As they add liquidity in both buyer and seller markets, they enhance housing market activity (Gray 2023a) and they should speed up the matching process. There are two propositions here. First, as they could now better fund a house purchase, an increase in price, and hence equity, affects the second group much like the first. Novy-Marx (2009) proposes that market participation is related to expected returns and high transaction costs. A more active (hot) market, where buyer–seller matching is quicker, would lower the participation costs for the seller. The bargaining position of either party is dependent on the scarcity of the other. A sudden increase in buyer participation speeds up matching, reducing the pool of sellers that remain, enhancing seller bargaining power, and inflating prices. The buyer 'shock' is amplified. Stein asserts that this group does not accelerate prices as they add both demand and supply to the market. In Novy-Marx's scenario, reduced matching time encourages more vendors to join the market. The greater liquidity lowers the risk to lenders of a fire sale if the marginal buyer gets into difficulties.

The third group needs no credit to buy. Wheaton and Lee (2009) and Ortalo-Magné and Rady (2004) add a buyer with no housing equity to the pool of potential purchasers. This first-time buyer (FTB) is possibly currently renting. This fourth group will behave more like normal consumers, responding to higher prices by withdrawing. The participation of the FTB is necessary for a first-time seller to move on. A fifth group, not featured in Ortalo-Magné and Rady, would include downsizers who look to match space with their reduced family size.

Increased participation is either stimulated by or induces higher prices at lower price levels. However, at some point, the asking prices lead to a diminution of interest from potential buyers, which discourages further sellers from joining the market. Housing expenditure would reflect a combination of market participation and price. In a thick market, an increase in participation is accompanied by price. A thin (cold) market could see prices rising whilst participation declines.

The relaxation of credit should have a greater effect on the highest leveraged markets, most likely with higher market prices. However, permitted leverage will be in the hands of the lender. Where the credit lands during a period of credit loosening is based on risk-adjusted expected returns from a dwelling purchase (Amaral et al. 2021; Sinai 2010). If the current structure of prices has the risk assessment baked in, as reflected in house price–earnings ratios (HPER), the proposition that high-priced districts face a tighter credit constraint could be flawed. It is unclear how a relaxation in credit restrictions could favour a local market. Lenders face agency costs when expanding their loan book. Lenders' assessment of local risk may be imperfect and backward-looking, which would favour traditionally hotter markets.

Blackwell-Ohnishi et al. suggest that the tail of a house price distribution would contain information about a price bubble. This implies sigma-divergence in the economic growth sense in a boom. It may be better to look to evidence in transactions and sales values when exploring bubbles rather than prices. Loss aversion may obscure dramatic distributional changes in market sentiment. Power-law exponents have been used to characterise convergence (Tang et al. 2016). The Lavalette distribution yields a very good approximation to the log-normal whilst echoing a standard power function, capable of

representing long tails (Fontanelli et al. 2016), so it could be used to explore housing distributional changes without making assumptions about power laws that only apply in a boom.

5. Method

A rank-size function can be expressed as $P_R = P_1[R]^{-\alpha}$ where P_R is the variable of study (population) and R is the rank score. P_1 is that of the largest value, known as the calibrating value, which is that of the largest city and has a rank score of $R = 1$ with the smallest $R = N$.

Lavalette's ranking power law can be expressed as $P_R = P_1\left[\frac{N \times R}{N-R+1}\right]^{-q}$ (Gray 2022a). The cross-sectional model is estimated using simple OLS as $ln(P_R)_t = -\gamma_t - q_t ln\left(\frac{N \times R}{N-R+1}\right)_t$ for $R = 1 \ldots N$ for T regressions. The exponentiation of the intercept multiplied by the actual highest value provides the expected CV. The projected median at time t can be calculated from $P_{mt} = E(CV_t) \times N^{q_t}$.

Lavalette's formula allows for the prediction of values other than the median. The projected value at percentile ω, $P_{\omega t} = P_t\left(\frac{\omega}{100-\omega}\right)^{q_t}$ has a power relationship with the median and expected calibrating value E(CV). If $q = 0.2$ and E(CV) = 100, the median would be 31.6 and the upper quartile value ($\omega = 75\%$) would be 41.4. The projected growth rate at the median is defined as $\frac{P_{mt+p}}{P_{mt}} = \dot{P}_m$. This, at P_ω, has two components: the movement of the representative value or median, and the spread. Thus, the growth rate is $\dot{P}_\omega = \dot{P}_m\left(\frac{\omega}{1-\omega}\right)^{q_{t+p}-q_t}$. The growth rate at certain points in the distribution depends on whether there is convergence or divergence. It is intuitively obvious that steepening occurs when the growth rate at the upper exceeds that at the lower quartile. Indeed, this could predict the acceleration of prices in the tail as implied by Blackwell-Ohnishi et al. The expression $\dot{Y}_{pp} = Distributional\ Correction \times \dot{Y}_m$, where Y is income, is adapted from Ravallion (2004, p. 6). When poverty reduction is the objective (for which economic growth is one of the instruments) then 'the rate of pro-poor growth (*pp*) defined above is the right way to measure growth consistently with that objective'. The distributional correction corresponds with a narrowing of the spread, indicated by the growth at the lower quartile exceeding that of the median and the upper quartile, resulting in convergence.

6. Data

The Local Authority District house prices, incomes, number of transactions, and number of dwellings are supplied by the UK's Office for National Statistics (ONS) for England and Wales. This covers annual data across 330 districts for the period from 2001 to 2022. The Isles of Scilly are excluded due to intermittent data. All transactions concerning house purchases, whether they entail a loan or not, are captured by these data. The number of transactions reflects the size of the district. To offer some standardisation, this is adjusted by the number of dwellings in the district, generating a sales (or transactions) per dwelling value, which is multiplied by 100,000 (SD). Dröes and Francke (2018) used the same housing stock adjustment to sales. The average house price per district is adjusted by the rate of inflation to provide a real price (RP) based on 2001 levels. The value of sales revenue is the product of the number of transactions and real price. Again, this is weighted by the number of dwellings to provide a measure of housing expenditure (HE).

Figure 1 displays the three variables for four years. The price for 2001 provides the base structure for the displacement of the profile in 2007. The noise in the profile for 2007 relates to the change in order from 2001 to 2007. Generally, low-priced districts in 2001 did not become expensive ones by 2007, yet change is not obviously proportional. Both 2001 and 2007 have the long-tailed *S* that characterises Lavalette's law. This is duplicated in 2012/2018 as well as in housing expenditure (HE). Transactions per dwelling also have an *S* shape, but the long tail is not so pronounced. Also, the number of transactions does not appear to have risen over the first period, and the noise is large relative to the gradient of the profile.

The second period features slower growth. Sales in both 2012 and 2018 appear almost without a gradient, suggesting that if changes in transactions are related to size order this is not much of a claim.

7. Results

There is a concern about misclassifying a distribution when it is quite likely to be log-normal (Perline 2005). The *p*-values of Kolmogorov–Smirnov goodness of fit test results for selected years are displayed in Table 1. The first consideration is whether the data follow a normal and a log-normal distribution and whether a Lavalette function provides a good fit. Both housing expenditure and price are more likely to follow a log-normal distribution. Sales do not appear log-normal but the case for a normal distribution is not strong. One could infer that there is a good Lavalettean fit for housing expenditure and price. Sales do not consistently follow a distribution considered. In addition, R^2 values are reported. A value above 0.98 is linked to a K–S *p*-value of over 0.05.

Table 1. Goodness of fit, medians, and exponents. Obs = observed; Est = estimated; BS L = bootstraps lower band; BS U = bootstraps upper band; *p*-values * sig. at the 5% level ** sig. at the 1% level.

K–S	Housing Expenditure (Price × Sales/D)				Real Price				Sales/Dwelling × 10,000			
	Laval Ette	R^2	Normal	Log-Normal	Laval Ette	R^2	Normal	Log-Normal	Laval Ette	R^2	Normal	Log-Normal
					Kolmogorov–Smirnov							
2001	0.579	0.982	0.00 **	0.014 *	0.645	0.979	0.00 **	0.20	0.26	0.982	0.20	0.00 **
2006	0.774	0.986	0.00 **	0.20	0.645	0.982	0.00 **	0.20	0.46	0.984	0.20	0.00 **
2007	0.579	0.986	0.00 **	0.20	0.774	0.975	0.00 **	0.20	0.30	0.986	0.20	0.007 **
2009	0.710	0.984	0.00 **	0.20	0.049 *	0.967	0.00 **	0.004 **	0.032 *	0.949	0.00 **	0.00 **
2012	0.516	0.983	0.00 **	0.20	0.06	0.959	0.00 **	0.003 **	0.09	0.962	0.036 *	0.00 **
2014	0.456	0.982	0.00 **	0.06	0.109	0.949	0.00 **	0.00 **	0.004 **	0.942	0.00 **	0.00 **
2018	0.516	0.979	0.00 **	0.022 *	0.456	0.964	0.00 **	0.015 *	0.13	0.932	0.034 *	0.00 **
2022	0.456	0.980	0.00 **	0.07	0.516	0.972	0.00 **	0.07	0.002 **	0.879	0.00 **	0.00 **
					Median Values							
	Obs	Est	BS L	BS U	Obs	Est	BS L	BS U	Obs	Est	BS L	BS U
2001	6294	5951	5758	6207	104,810	106,933	102,535	111,409	5730	5566	5466	5667
2006	9755	9509	9275	9770	167,172	168,706	162,741	176,119	5738	5636	5564	5710
2007	9222	9318	9070	9592	172,507	174,176	165,847	184,024	5453	5350	5291	5409
2009	4198	4133	3995	4293	154,134	157,485	148,909	167,391	2692	2624	2534	2729
2012	4185	4052	3874	4255	144,349	146,928	136,039	159,803	2828	2758	2674	2842
2014	5935	5747	5483	6048	146,518	152,864	140,548	166,925	3922	3760	3656	3878
2018	6188	5858	5606	6214	161,986	169,294	156,905	182,115	3596	3460	3341	3602
2022	4688	4569	4372	4784	168,992	169,914	160,339	180,060	2781	2689	2586	2810
					Exponents (-*q*)							
		Est	BS L	BS U		Est	BS L	BS U		Est	BS L	BS U
2001		0.298	0.292	0.305		0.251	0.243	0.259		0.100	0.097	0.103
2006		0.224	0.219	0.228		0.188	0.181	0.195		0.092	0.090	0.094
2007		0.224	0.219	0.229		0.197	0.189	0.207		0.087	0.085	0.089
2009		0.284	0.277	0.291		0.202	0.192	0.212		0.125	0.120	0.131
2012		0.319	0.31	0.328		0.231	0.218	0.245		0.116	0.112	0.121
2014		0.317	0.308	0.326		0.250	0.236	0.266		0.109	0.104	0.114
2018		0.263	0.255	0.273		0.263	0.250	0.276		0.107	0.101	0.114
2022		0.251	0.243	0.260		0.248	0.238	0.259		0.088	0.081	0.096

The mid-section of Table 1 reports the observed medians for the same selected years. The next three columns report the intercept converted into the estimated median, plus a lower and an upper value based on bootstraps 95% confidence intervals. The observed and estimated values are within 6% of each other, and all the observed values are well within the confidence intervals. Indeed, the estimated median almost duplicates the observed geometric mean (not reported). This relationship is found with log-normal distributions.

7.1. Exponents' Time Paths

Tsai (2015) finds a segmentation in housing markets between the northern and southern regions. Following this, northern districts/regions, which are defined as the midlands and North of England and Wales, comprise 159 districts. Southern districts or regions

comprise London, the East of England, the South West, and the South East. This group contains 171 districts.

The time profiles of coefficients of variation are displayed as a reference in Figure 2 on the left for all districts, North and South. The steepening of prices seen in Gray (2023b) is evident in prices for all districts. This is replicated in housing expenditure, but not in transactions (CoV SD). That said, clearly there are observable narrowing periods. There was another increase in the spread around 2019 to 2021, which pre-dated the lockdown.
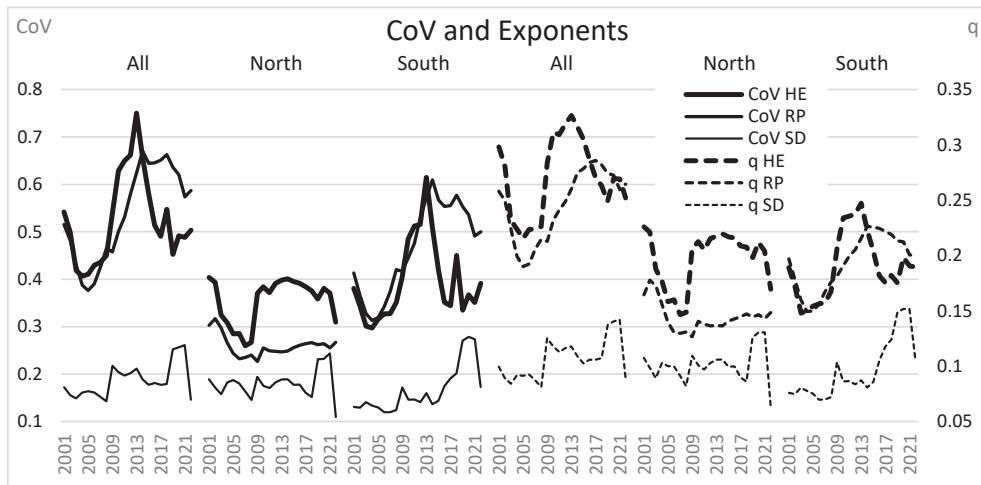
Figure 2. Time paths of spread measures.

Compared with the South, northern districts exhibit a similar range of spread of housing expenditure and transactions, but less variation in price. The profile of all districts' sales per dwelling is relatively flat until 2007. The subsequent crisis period features a slight broadening of the spread.

On the right-hand side of Figure 2, there are three exponent time paths. Although these are similar to the corresponding CoVs, the patterns are smoother. The price path (q RP) displays the *S* shape reported by Gray (2023b) but with slightly different dates. The key steepening phase in price runs from 2006 to 2017. There is a period of convergence from 2001 to 2006, and post-2017. The lower section of Table 1 reports the exponents $-q$ for the selected years, plus a lower and an upper value based on bootstraps 95% confidence intervals. The price and housing expenditure coefficients for 2006 are below others, supporting the claims made above concerning convergence and divergence phases.

The *S* shape is evident in the southern districts' time profile, but the trough and peak occur earlier. A distinctive feature of the northern districts is the stable distribution after 2010. Here, the *S* shape is not evident.

The conclusions drawn about steepening depend on the era and the region one selects. Evidence for a steeper distribution is found in the South, not the North of England and Wales post-2008. The rapid steepening from 2008 puts the spread in district expenditure back to a similar position in 2002. The other two spreads return to the 2002 levels around 2009.

The dramatic fall in housing expenditure inequality across all districts covers the initial years to 2005, which is reversed when there is a notable rise in expenditure dispersion from 2008. This again is reversed around 2013, when the RP and HE trajectories lean in opposite directions. Importantly, when clearly rising, HE has steeper trajectories than RP. This is in line with speculative sellers adding liquidity in both buyer and seller markets, enhancing housing market expenditure.

Observed price and quantity values for England and Wales have two distinct patterns at the 2007/2008 juncture. Mean price, adjusted by the rate of inflation, rose to a peak in 2007. It then fell by 5.5% in 2008. The corresponding measure of Hudson and Green's (2017) missing movers is missing transactions, which entails a propitious drop of 48%. The collapse in expenditure (52%) is greater than sales.

With other consumer durables, one could expect a boost to transactions with a fall in price and lower costs of borrowing. QE's two elements, increased reserves for banks and lower interest rates, may take some time to penetrate the more risk-averse environment. A downswing in a credit cycle (bust) features a severe restriction of lending, and a rise in collateral requirements (Geanakoplos 2010). Banks withdraw mortgage products and impose larger deposit requirements, which would filter out the higher-risk borrowers, particularly affecting those without property. If loans are not available, offers to buy fall through. If buying one property is contingent on selling the existing one, and that buyer fails to secure funding, both contracts fail to be executed.

Those who bought close to the peak of a price cycle could experience loss aversion (Genesove and Mayer 2001). Unprepared to accept a loss, dwellings could just remain in the estate agent's window for longer. The speculative participant could withdraw from the market, in part, because the matching rate had dramatically slowed. As such, housing market activity is adversely affected. Hence, the number of house trades and the spatial distribution would be linked to credit and risk appetite.

Transactions per dwelling rate and average district price are lower in the North than in the South. With the exception of the years from 2009 to 2014, there is a negative correlation between district sales/dwelling and price level, in the South. Cheaper districts are associated with more trades. By contrast, the rank order of the district price level is positively associated with that of sales per dwelling in the North for all years apart from 2003 to 2008. Combined, only in 2003 and 2019 are the relationships not positive for the whole of England and Wales, suggesting that more active markets have higher prices.

In general, as measured by the Spearman coefficient, the rank order of district expenditure is strongly linked to price. As shown in Figure 2, the spatial variation in transactions is small compared with that in price. The similarity in the steepening of both the price and expenditure distributions seen in Figure 2 at the national level could reflect this price dominance. Variation at the national level is not reflected in either the South or the North. Moreover, for much of the post-crisis period, price and expenditure distributions of the northern districts are stable, so the 'steepening' is more likely to reflect a North–South schism, where the South pulls away from the North, plus greater dispersion in the South.

7.2. Expected Median Time Paths

Figure 3 displays the time trajectories of the observed medians of the three measures. There are distinct patterns for price and quantity. The real median price rose to a peak in 2007. It declined by 10% over the period to 2009. This decline continued until 2013. Late in the series, another shock is evident before the COVID-19 lockdown began. It was only then that the price returned to the pre-crisis price level. Up until 2007, there was a general *decline* in the transactions. The equivalent discussion in the context of Hudson and Green (2017) is that there is no recovery in the number of transactions to the pre-2007 levels. The third line associated with housing expenditure traces the price rising to a peak in 2007. The collapse in expenditure is greater than sales, which it traces from then on. The real pre-COVID-19 price peak occurred in 2018, two years after that found in housing expenditure. Housing expenditure peaked in 2006, a year before price in the pre-financial crisis period. Price peaks occur in cooling markets.

The next three sets of lines are medians as derived from Lavalette's function for all districts. The patterns of all three sets for each of the three variables concerned are similar to the E&W measures. Both the time profiles of levels (Figure 3) and spreads (Figure 2) correspond well with observations.
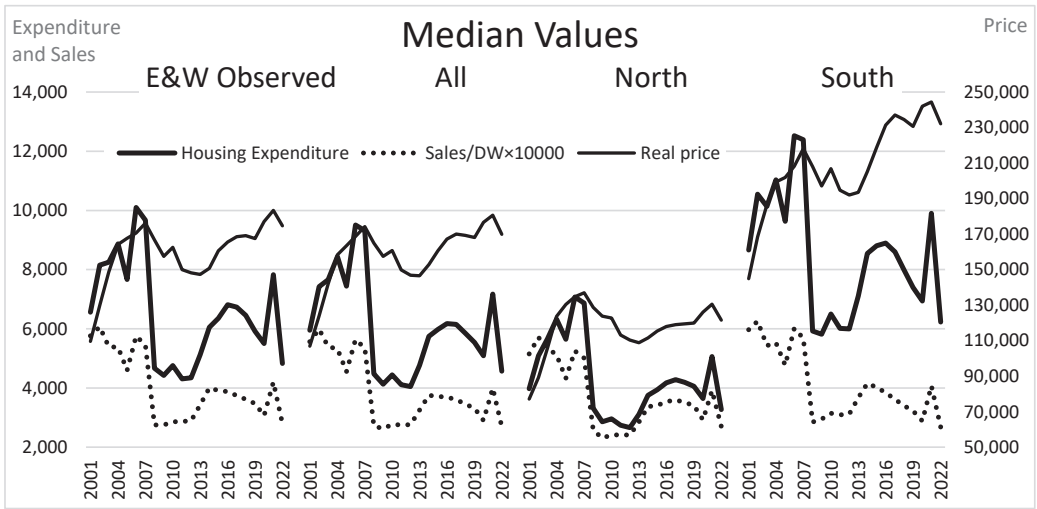
Figure 3. Median values: price transactions and expenditure.

7.3. Stein's Dynamic Framework

For macroprudential purposes, the Bank of England monitors price and affordability (Bank of England 2015). Two six-year periods ending in an inflection point in observed real prices on either side of the financial crisis are analysed. The annualised house price inflation rate at the median from 2012 to 2018 was less than a quarter of the rate from 2001 to 2007 (1.94% < 8.66%). Those districts that had a relatively rapid rise in price in the second period had a slow rate of appreciation in the first (rho = −0.659 [0.000]). By contrast, the growth in expenditure in the first period is positively linked with the second (0.349 [0.000]). This implies that price growth in the second period adversely affected sales growth. Stein's framework suggests that expenditure and sales should rise over a range of price growth. The rank order of growth rates of expenditure and price are strongly associated in the first period (0.769 [0.000]), but not in the second (−0.065 [0.241]). There is a negative relationship between sales and price (−0.249 [0.000]) in the second period and none in the first. Again, this highlights that activity in the second period is not consistent with the first.

Stein's framework has been used to explain the divergence in price in the upswing stage of a cycle, with high house prices pulling away from low (Ortalo-Magné and Rady 2004). Spearman's rho indicates that the growth rates of price over the six years are negatively associated with the price level in the initial year, in the first period (−0.872 [0.000]), but positive in the second (0.502 [0.000]).

Using expenditure data, there is a negative relationship in both the first period (−0.818 [0.000]) and the second (−0.653 [0.000]). This is consistent with the growth literature's beta-convergence in expenditure (Sala-i-Martin 1996). As the credit cycle enters a looser phase, lenders may cast their eyes around to those areas where less creditworthy debtors might be coaxed onto the market. These areas attract investors' attention when other opportunities are limited in traditionally lower-risk markets.

The second period, not consistent with Stein's projection of the co-movement of price and activity could be a result of the distribution of QE funds favouring London (Bunn et al. 2018). Prices in the South reached a point in the post-crisis era where affordability could be beyond what lenders deem as safe, and price is above fundamentals (Duca et al. 2021).

Sigma-convergence entails a narrowing of the distribution over time. Beta-convergence implies that poorer districts grow more quickly than richer ones. Pro-poor growth (Ravallion 2004) implies both. Given a median growth rate, a smaller future Lavalettean exponent implies that the higher housing market values are increasing less quickly than the lower-priced ones, or there is catch-up/pro-poor growth. Table 2 reports observed growth rates

at the median and the upper and lower quartiles. These are compared with projected rates based on $\dot{P}_\omega = \dot{P}_m \left(\frac{\omega}{1-\omega}\right)^{q_{t+p}-q_t}$. In the run-up to 2008, the median price grew by 8.66% annually, which is similar to the Lavalettean projected growth rate (8.47%) within the bootstraps implied confidence interval of 8.34–8.72%. The lower quartile grew faster than the upper (10.45% > 7.08%). The order but not the division is reflected by the expected values (9.54% > 7.41%). All districts and the North and South regions exhibit convergence. This set of results concurs with those using Spearman's coefficient.

Table 2. Annualised growth rates over two growth periods: price, transactions, and expenditure.

2001–2007 (1) 2012–2018 (2)	All Districts			North			South		
	Real Exp/Dw	Sale/Dw	Real Price	Real Exp/Dw	Sale/Dw	Real Price	Real Exp/Dw	Sale/Dw	Real Price
Observed (1)									
Lower Q	9.27	−0.34	10.45	11.72	−0.33	10.56	6.75	−0.97	7.63
Median	6.57	−0.82	8.66	9.54	−0.55	10.46	5.91	−0.50	6.96
Upper Q	5.27	−1.02	7.08	7.80	−0.84	8.91	5.57	−1.14	5.91
Estimated (1)									
Lower Q	9.24	−0.43	9.54	11.12	−0.17	10.72	6.77	−0.73	7.64
Median	7.76	−0.66	8.47	9.51	−0.46	10.02	6.15	−0.84	7.05
Upper Q	6.3	−0.89	7.41	7.93	−0.75	9.32	5.53	−0.95	6.47
Observed (2)									
Lower Q	8.26	4.82	1.50	7.89	6.92	1.01	5.73	1.66	3.29
Median	6.74	4.09	1.94	8.28	6.40	1.32	5.24	2.17	3.38
Upper Q	5.31	3.30	3.52	8.44	5.87	2.00	3.51	2.30	3.13
Estimated (2)									
Lower Q	7.42	4.03	1.79	4.93	6.81	1.14	5.97	0.72	3.12
Median	6.34	3.85	2.39	4.57	6.47	1.34	4.90	1.47	3.39
Upper Q	5.26	3.68	2.99	4.20	6.14	1.54	3.84	2.23	3.65

In the second period, again there is a schism between price and expenditure across England and Wales. Convergence is evident when analysing expenditure. House price spreads are consistent with the correlation results and the South pulling away from the North. However, the observed pattern is not consistent when subdivided by North and South, where the North's distribution narrows whilst the South experiences divergence.

7.4. Stylised Facts

The real price growth rate over the 11 years of the steepening period, 2006–2017, is 0.1% at the median annually and 0.8% at the upper quartile. Expenditure is worse. The annualised contraction rate is 3.2% as opposed to 3.6% at the median. So, the Steinian proposition that higher price growth explains increasing spreads is not supported. It also undermines Blackwell-Ohnishi et al.'s suggestion that the tail of a house price distribution would contain information about a price bubble. Van Nieuwerburgh and Weill (2010) suggest there is a productivity driver of spatial house price variations. Brandily et al. (2022) find that spatial disparities in the UK, although broad, increased slightly up to the financial crisis, but remained generally stable since. Covering 2007–2019, which matches the price steepening period well, Rodrigues and Bridgett (2023) find that London *lags* behind the rest of the UK in productivity growth, in a low-growth era, implying a narrowing of productivity spreads. London's productivity grew by 0.2%/ year in real terms, which they suggest is partly a land price problem. They suggest that rising costs for office space deter firms from choosing a location in the City and crowd out investment. High house prices have weakened London's draw on talented people.

The house price–earnings ratio (HPER) of 4.4 in 2001 rose to 6.96 in 2006. In other words, prices grew by 2.5 annual salaries in a high participation era, or a 'hot' market period. Subsequently, the HPER rose nationally to 7.8 in 2017. This is a much smaller increase over a longer period. The continued decline in affordability was accompanied by an extension of the average mortgage repayment period from 25 to 35 years, plus mortgage interest payments were affected by a historically low Bank of England Base rate, leading to the conclusion that monthly mortgage servicing could be steady despite rising house prices. Debt has become dislocated from incomes (Gregoriou et al. 2014). The distribution

of QE funds appears to have favoured London (Bunn et al. 2018). Its HPER rose from 7.9 to 12.4 in the period from 2006 to 2017. Affordability metrics in northern regions remain unchanged over that period. Combined, there could be a 'pulling away of the South' based on the QE credit dispersion, resulting in the problems highlighted by Gregoriou et al. (2014) and Rodrigues and Bridgett (2023). The QE impact could be spread internationally by a wealthy elite buying up properties in many global centres.

Figure 2 highlights convergence in expenditure across E&W in the post-2013 period whilst prices continued to diverge for the next 4 years. The convergence in expenditure nationally corresponds with markets in the South converging internally plus more rapid growth in (some parts of) the North that catch up. Convergence in the South would not be consistent with a spatial division inflated by a wealthy, international elite, which would focus on London only, reinforcing the QE effect.

7.5. COVID-19

The period from 2019 covering COVID-19 is unusual, as one might expect. The lockdown period began in March 2020 with rapidly declining transactions. Sales in 2020 were at a recent low. Lockdown altered people's locational preferences. As lockdown restrictions were eased, combined with Stamp Duty holidays, which ended in September 2021, there was a flurry of buyer activity favouring greater space, such as gardens or an extra room for working from home (Hammond 2022; Peachey 2021). The surge in expenditure is evident in Figure 3. With less of an emphasis on commuting, housing market metrics should reflect a dash towards districts with amenities, such as parks and water. Oddly, sales spreads remained stable over this period.

There was a spike in 2021 in both the price and the HPER, yet there is little change in the spread. The impact of the COVID-19 period on house trading activity in the North is almost imperceptible. In 2022, prices and transactions declined as the Bank of England base rate rose from 0.75% to 3.5%. It also announced that it would engage in quantitative tightening, reducing the amount of credit in the system, both of which should reduce market participation. The spread indicators feature a strong narrowing of the distributions.

8. Conclusions

This paper set out to consider the 'steepness' of the values of housing transactions as distributed across districts over the period 2001–2022 for England and Wales. The paper finds that the time paths of Lavalette's (1996) exponents compare well with those of the coefficients of variation. Blackwell (2018) finds that real estate data are 'in-between' log-normal and a 'true' power law. The distributions of both price and housing expenditure fit that class of data with a long tail, yet often indistinguishable from the log-normal that Lavalette's law captures. Sales per dwelling do not fit so well. As the exponent is simple to estimate and can provide a meaningful interpretation for data with distributions likely to be found in the worlds of economic inequalities and growth convergence, it has useful properties for the economic policymaker. The exponent produces a time profile akin to the coefficient of variation. This well-used statistic could be inflated as data with a long tail are likely to be skewed and subject to kurtosis, so the exponent could offer a smoother time profile of the dynamics of inequalities.

The paper reveals a Van Nieuwerburgh and Weill 'steepening' in the national house price distribution, but this is in the middle of two convergence periods. The first convergence period is associated with the run-up to the financial crisis. Blackwell (2018) and Ohnishi et al. (2020) propose that the upper tail of a house price distribution could contain information about a price bubble. The results here indicate that the distribution is not steepest at the peak price level. Duca et al. (2021) conclude that real-estate-linked financial crises typically begin with over-valued real estate prices. Using risk as an explanation, it is argued that the over-valuation is observed by lenders in southern markets who switch away to less inflated northern ones. The UK's peak prices nationally occurred as northern price levels 'caught up'.

Even without a spatial effect, the upper tail may not return to normality quickly due to loss aversion. Prices do adjust downwards but not as dramatically as in activity. Prices can rise in thin markets, ones already experiencing a significant drop in financial transactions. As such, it is argued that evidence of volatile housing market activity is better found in housing transaction expenditure, which peaked earlier than prices in both pre- and post-crisis periods. Housing expenditure is found to steepen more rapidly than prices, in a period of low market participation and lower real activity. Again, the upper tail in price would not offer an insight into an over-heated housing market. There is a period when there is convergence in expenditure and divergence in price. The schism between housing expenditure and price, particularly post-2013, would be consistent with a peak price occurring in a cooling market nationally. A thick market in the North could coexist with a thin market but with rising prices in the South.

Macroprudential regulation seeks to limit reckless lending with national metric lending rules (Bank of England 2015) based on price growth and affordability. Rather than productivity, it is averred that lending and QE funds favouring London (Bunn et al. 2018) underpinned the steepening across the UK. The post-financial-crisis price patterns understate the extent of depressed market activity compared with before. QE may have ossified prices at damaging levels in the South, affecting productivity growth itself, the driver of spreads (Van Nieuwerburgh and Weill 2010) elsewhere. Monitoring of housing activity could be improved by analysing housing expenditure, which should be a lead indicator and more in line with lending activity.

Conflicts of Interest: The author declares no conflict of interest.

References

Allen, Franklin, and Douglas Gale. 2000. Bubbles and crises. *The Economic Journal* 110: 236–55. [CrossRef]

Amaral, Francisco, Martin Dohmen, Sebastian Kohl, and Moritz Schularick. 2021. *Superstar Returns*. The New York Federal Staff Reports 999. New York: Federal Reserve Bank of New York. [CrossRef]

Andrew, Mark, and Geoffrey Meen. 2003. House Price Appreciation, Transactions and Structural Change in the British Housing Market: A Macroeconomic Perspective. *Real Estate Economics* 31: 99–116. [CrossRef]

Bank of England. 2015. The Financial Policy Committee's Powers over Housing Tools. Available online: https://www.bankofengland.co.uk/-/media/boe/files/statement/2015/the-financial-policy-committees-powers-over-housing-tools.pdf?la=en&hash=824555A3E0E43904679511189F70706FC5EE5EA2 (accessed on 1 February 2018).

Behrens, Kristian, Gilles Duranton, and Frédéric Robert-Nicoud. 2014. Productive Cities: Sorting, Selection, and Agglomeration. *Journal of Political Economy* 122: 507–53.

Blackwell, Calvin. 2018. Power Laws in Real Estate Prices? Some Evidence. *The Quarterly Review of Economics and Finance* 69: 90–98. [CrossRef]

Bogin, Alexander, William Doerner, and William Larson. 2017. Local House Price Paths: Accelerations, Declines, and Recoveries. *The Journal of Real Estate Finance and Economics* 58: 201–22. [CrossRef]

Brakman, Steven, Harry Garretsen, and Charles van Marrewijk. 2020. *An Introduction to Geographical and Urban Economics: A Spiky World*. Cambridge: Cambridge University Press.

Brandily, Paul, Mimosa Distefano, Hélène Donnat, Immanuel Feld, Henry G. Overman, and Krishan Shah. 2022. Bridging the Gap: What Would It Take to Narrow the UK's Productivity Disparities? *Resolution Foundation*. Available online: https://economy2030.resolutionfoundation.org/wp-content/uploads/2022/06/Bridging-the-gap.pdf (accessed on 1 November 2022).

Bunn, Philip, Alice Pugh, and Chris Yeates. 2018. The Distributional Impact of Monetary Policy Easing in the UK between 2008 and 2014. March Staff Working Paper No. 720. Available online: https://www.bankofengland.co.uk/news/publications (accessed on 1 September 2022).

Cerqueti, Roy, and Marcel Ausloos. 2015a. Cross ranking of cities and regions: Population versus income. *Journal of Statistical Mechanics: Theory and Experiment* 2015: P07002. [CrossRef]

Cerqueti, Roy, and Marcel Ausloos. 2015b. Evidence of economic regularities and disparities of Italian regions from aggregated tax income size data. *Physica A: Statistical Mechanics and Its Applications* 421: 187–207. [CrossRef]

Chlebus, Edward, and Gautam Divgi. 2007. A Novel Probability Distribution for Modeling Internet Traffic and its Parameter Estimation. Paper presented at IEEE Global Telecommunications Conference Global Telecommunications Conference, IEEE GLOBECOM Proceedings: 4670–4075, Washington, DC, USA, 26–30 November; Available online: https://ieeexplore.ieee.org/document/4411796 (accessed on 1 September 2022).

Clayton, Jim, Norman Miller, and Liang Peng. 2010. Price-volume Correlation in the Housing Market: Causality and Co-movements. *The Journal of Real Estate Finance and Economics* 40: 14–40. [CrossRef]

Coulson, N. Edward, Crocker H. Liu, and Sriram V. Villupuram. 2013. Urban economic base as a catalyst for movements in real estate prices. *Regional Science and Urban Economics* 43: 1023–40. [CrossRef]

Cristelli, Matthieu, Michael Batty, and Luciano Pietronero. 2012. There is more than a power law in Zipf. *Scientific Reports* 2: 812. [CrossRef] [PubMed]

D'Acci, Luca S. 2023. Is housing price distribution across cities, scale invariant? Fractal distribution of settlements' house prices as signature of self-organized complexity. *Choas Solitons and Fractals* 174: 113766. [CrossRef]

DiPasquale, Denise, and William Wheaton. 1996. *Urban Economics and Real Estate Markets*. Englewood Cliffs: Prentice Hall.

Dröes, Martijn, and Marc Francke. 2018. What Causes the Positive Price-Turnover Correlation in European Housing Markets? *The Journal of Real Estate Finance and Economics* 57: 618–46. [CrossRef]

Duca, John, John Muellbauer, and Anthony Murphy. 2021. What Drives House Price Cycles? International Experience and Policy Issues. *Journal of Economic Literature* 59: 773–864. [CrossRef]

Fernandez, Rodrigo, Annelore Hofman, and Manuel B. Aalbers. 2016. London and New York as a Safe Deposit Box for the Transnational Wealth Elite. *Environment and Planning A* 48: 2443–61.

Fontanelli, Oscar, Pedro Miramontes, Yaning Yang, Germinal Cocho, and Wentian Li. 2016. Beyond Zipf's Law: The Lavalette Rank Function and Its Properties. *PLoS ONE* 11: e0163241. [CrossRef]

Geanakoplos, John. 2010. Solving the Present Crisis and Managing the Leverage Cycle. *New York Federal Reserve Economic Policy Review* 16: 101–31. [CrossRef]

Genesove, David, and Christopher Mayer. 2001. Loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics* 87: 1233–60. [CrossRef]

Glaeser, Edward L., and Joseph Gyourko. 2005. Urban Decline and Durable Housing. *Journal of Political Economy* 113: 345–75. [CrossRef]

Gray, David. 2018. Convergence and divergence in British housing space. *Regional Studies* 52: 901–10. [CrossRef]

Gray, David. 2022a. Do House Price-Earnings Ratios in England and Wales follow a Power Law? An Application of Lavalette's Law to District Data. *Environment and Planning B: Urban Analytics and City Science* 49: 1184–96. [CrossRef]

Gray, David. 2022b. How Have District-Based House Price-Earnings Ratios Evolved in England and Wales? *Journal of Risk and Financial Management* 15: 351. [CrossRef]

Gray, David. 2023a. Housing market activity diffusion in England and Wales. *National Accounting Review* 5: 125–44. [CrossRef]

Gray, David. 2023b. Power Laws and Inequalities: The Case of British District House Price Dispersion. *Risks* 11: 136. [CrossRef]

Gregoriou, Andros, Alexandros Kontonikas, and Alberto Montagnoli. 2014. Aggregate and regional house price to earnings ratio dynamics in the UK. *Urban Studies* 51: 2916–27. [CrossRef]

Hammond, George. 2022. Is the UK Housing Market at a Turning Point? *Financial Times*, August 12.

Himmelberg, Charles, Christopher Mayer, and Todd Sinai. 2005. Assessing High House Prices: Bubbles, Fundamentals and Misperceptions. *Journal of Economic Perspectives* 19: 67–92. [CrossRef]

Hsu, Wen-Tai. 2012. Central Place Theory and City Size Distribution. *The Economic Journal* 122: 903–32. [CrossRef]

Hudson, Neal, and Brian Green. 2017. Missing Movers: A Long-Term Decline in Housing Transactions? *Council of Mortgage Lenders*. Available online: https://thinkhouse.org.uk/site/assets/files/1756/cmlmissing.pdf (accessed on 1 September 2022).

Lavalette, Daniel. 1996. *Facteur D'impact: Impartialité ou Impuissance?* Orsay: Internal Report INSERM U350 Institut Curie—Recherche, Bât. 112, Centre Universitaire. Available online: http://www.curie.u-psud.fr/U350/ (accessed on 1 September 2022).

Martin, Ron, Ben Gardiner, and Peter Tyler. 2014. *The Evolving Economic Performance of UK Cities: City Growth Patterns, 1981–2011*; Working Paper, Foresight Programme on The Future of Cities. London: UK Government Office for Science, Department of Business, Innovation and Skills.

Martin, Ron, Peter Sunley, Ben Gardiner, Emil Evenhuis, and Peter Tyler. 2018. The city dimension of the productivity growth puzzle: The relative role of structural change and within-sector slowdown. *Journal of Economic Geography* 18: 539–70. [CrossRef]

McCombie, John. 1988. A synoptic view of regional growth and unemployment: II The post-Keynesian theory. *Urban Studies* 25: 399–417. [CrossRef]

Mian, Atif, and Amir Sufi. 2018. Finance and Business Cycles: The Credit-Driven Household Demand Channel. *Journal of Economic Perspectives* 32: 31–58. [CrossRef]

Miles, David, and Victoria Monro. 2019. *UK House Prices and Three Decades of Decline in the Risk Free Real Interest Rate, Staff Working Paper No. 837*. London: Bank of England.

Montagnoli, Alberto, and Jun Nagayasu. 2015. UK house price convergence clubs and spillovers. *Journal of Housing Economics* 30: 50–58. [CrossRef]

Novy-Marx, Robert. 2009. Hot and Cold Markets. *Real Estate Economics* 37: 1–22. [CrossRef]

Ohnishi, Takaaki, Takayuki Mizuno, and Tsutomu Watanabe. 2020. House price dispersion in boom–bust cycles: Evidence from Tokyo. *The Japanese Economic Review* 71: 511–39. [CrossRef]

Ortalo-Magné, François, and Sven Rady. 2004. Housing transactions and macroeconomic fluctuations: A case study of England and Wales. *Journal of Housing Economics* 13: 287–303. [CrossRef]

Peachey, Kevin. 2021. How COVID Has Changed Where We Want to Live. *BBC News*, March 19.

Perline, Richard. 2005. Strong, Weak and False Inverse Power Laws. *Statistical Science* 20: 68–88. Available online: http://www.jstor.org/stable/20061161 (accessed on 1 September 2022). [CrossRef]

Pike, Andy, Danny MacKinnon, Mike Coombes, Tony Champion, David Bradley, Andrew Cumbers, Liz Robson, and Colin Wymer. 2016. *Uneven Growth: Tackling City Decline*. New York: Joseph Rowntree Foundation. Available online: https://www.jrf.org.uk/report/uneven-growth-tackling-city-decline (accessed on 8 August 2023).

Ravallion, Martin. 2004. *Pro-Poor Growth: A Primer. Policy Research Working Paper; No. 3242*. Washington, DC: World Bank. Available online: http://hdl.handle.net/10986/14116 (accessed on 8 August 2023).

Richmond, Peter. 2007. A roof over your head; house price peaks in the UK and Ireland. *Physica A: Statistical Mechanics and its Applications* 375: 281–87. [CrossRef]

Roback, Jennifer. 1982. Wages, Rents, and the Quality of Life. *Journal of Political Economy* 90: 1257–78. Available online: https://www.jstor.org/stable/1830947 (accessed on 1 September 2022). [CrossRef]

Rodrigues, Guilherme, and Stuart Bridgett. 2023. Capital Losses: The Role of London in the UK's Productivity Puzzle. *Centre for Cities' Report*. March 2. Available online: https://www.centreforcities.org/press/londons-productivity-lags-well-behind-global-competitors/ (accessed on 1 September 2022).

Rozenfeld, Hernan, Diego Rybski, Xavier Gabaix, and Hernan Makse. 2011. The Area and Population of Cities: New Insights from a Different Perspective on Cities. *The American Economic Review* 101: 2205–25. [CrossRef]

Sala-i-Martin, Xavier. 1996. Regional cohesion: Evidence and theories of regional growth and convergence. *European Economic Review* 40: 1325–52. [CrossRef]

Sinai, Todd. 2010. Feedback Between Real Estate and Urban Economics. *Journal of Regional Science* 50: 423–48. [CrossRef]

Stein, Jeremy. 1995. Prices and trading volume in the housing market: A model with downpayment effects. *The Quarterly Journal of Economics* 110: 379–405. [CrossRef]

Tang, Pan, Ying Zhang, Belal E. Baaquie, and Boris Podobnik. 2016. Classical convergence versus Zipf rank approach: Evidence from China's local-level data. *Physica A: Statistical Mechanics and its Applications* 443: 246–53. [CrossRef]

Tsai, I-Chun. 2015. Spillover Effect between the Regional and the National Housing Markets in the UK. *Regional Studies* 49: 1957–76. [CrossRef]

Tsai, I-Chun. 2018. The cause and outcomes of the ripple effect: Housing prices and transaction volume. *The Annals of Regional Science* 61: 351–73. [CrossRef]

Van Nieuwerburgh, Stijn, and Pierre-Olivier Weill. 2010. Why Has House Price Dispersion Gone Up? *The Review of Economic Studies* 77: 1567–606. [CrossRef]

Wheaton, William, and Nai Lee. 2009. The Co-Movement of Housing Sales and Housing Prices: Empirics and Theory. *SSRN Electronic Journal*. [CrossRef]

Communication

Information Theory and the Pricing of Contingent Claims: An Alternative Derivation of the Black–Scholes–Merton Formula

Thomas P. Davis

FactSet Research Systems, Broadgate Quarter, London EC2A 2DQ, UK; todavis@factset.com

Abstract: This paper seeks to determine the best subjective probability to use to carry out expectation values of uncertain future cash flows with the smallest number of assumptions. This results in the unique distribution that guarantees no more information is present other than the stated assumptions. The result is a novel derivation of the well-known Black–Scholes equation without the need to introduce high-level mathematical machinery. This formalism fits nicely into introductory courses of finance, where the value of any financial instrument is given by the present value of uncertain future cash flows.

Keywords: Information Theory; derivative pricing; quantitative finance

1. Introduction

Early on in finance, we learn that the fair value of a security is the present value of expected future cash flows:

$$V(t) = \sum_i Z(t, T_i)\mathbb{E}[c(T_i)], \tag{1}$$

where $Z(t, T)$ is the price of a zero-coupon bond at time t, maturing at time T. Many of the formulae for fixed-income security pricing result from Equation (1) when the cash flows $c(T_i)$ are known with certainty Fabozzi and Mann (2021). However, when the cash flows are uncertain, it is natural to ask what probabilities should be used to determine the average implied by the expectation value in Equation (1).

The overwhelmingly accepted answer is, of course, the famous "risk-neutral probabilities" introduced by Black and Scholes (1973), in which many assumptions were made in order to justify the approach. The assumptions were those of the capital asset pricing model (CAPM) Lintner (1965); Sharpe (1964), which depends upon general economic equilibrium and also upon secondary assumptions such as a constant interest rate (or at least non-stochastic interest rate) and the variance of the stock over the lifetime of the option.

Robert Merton did not agree with many of these assumptions, especially the condition of economic equilibrium, saying of the derivation, "the portfolio weights are chosen to eliminate all 'market risk'. By the assumptions of the CAPM, any portfolio with zero ('beta') market risk must have an expected return equal to the risk-free rate. Hence an equilibrium condition is established between the expected return on the option, the expected return on the stock and the risk-less rate" Merton (1973). Merton then set out to relax some of these assumptions by adding stochastic interest rates and, more importantly, proving that the expectations of the investors play no part in the fair value of an option. This insight came from a precise cancellation of terms when solving the equations for the portfolio to be arbitrage-free, that is, the initial capital investment is all that is required to hedge the portfolio without any inflows or outflows; the gains and losses of the portfolio suffice. Thus, the equilibrium assumptions of the CAPM are not required.

These equations are now known as the "market price of risk" and form the basis of dynamic asset pricing theory, as developed by Harrison and Pliska (1981), who developed

the fundamental theorem of asset pricing by specifying the mathematical and economic requirements for arbitrage freedom in a given financial market.

The mathematical machinery needed to price a simple European option is, therefore, quite complex, requiring the knowledge of stochastic calculus, dynamic asset pricing theory and the solution of partial differential equations—very intimidating for a young student in finance. This paper seeks to determine a financially justified probability distribution without relying on any of this advanced mathematical machinery. The frameworks used are those of Information Theory Shannon (1948) and the principle of maximum entropy developed by Jaynes (1957b, 2002).

2. Information Theory and Maximum Entropy

In 1948, Claude Shannon introduced Information Theory in his landmark paper "A Mathematical Theory of Communication" Shannon (1948). Here, Shannon introduced the concept of informational entropy of a discrete probability distribution:[1]

$$H = -\sum_i p_i \ln p_i,$$

(2)

as a measure of information. It has intuitive properties that the information about an event should satisfy, such as being zero when the outcome is certain—i.e., only one of the probabilities is 1.0—and achieves a maximum when all outcomes are equally likely. From this fundamental quantity, he determined optimal ways to encode communication messages into abstract codes, where the environment can adversely affect the message (i.e., what is transmitted is not what is received).

Later, in 1957, Ed Jaynes noted that the definition of entropy in statistical physics could be thought of as informational entropy Jaynes (1957a) and introduced a way to generate a probability distribution that produces the empirical data without adding any additional assumptions, i.e., a distribution that "...has the important property that no possibility is ignored; it assigns a positive weight to every situation that is not absolutely excluded by the given information".

We can apply this concept in finance, as it is an open question whether asset price fluctuations can have an objective probability distribution. The maximum entropy framework does not assume one, whereas an objective probability distribution is required by the Black–Scholes–Merton framework—all models that attempt to go beyond the Black–Scholes–Merton assumptions (Cox and Ross 1976; Dupire 1994; Heston 1993; Merton 1976). Stocks are not physical particles subject to the immutable laws of physics; they are, rather, a human construct with prices traded in a market driven by human emotion. Thus, we seek to obtain a *subjective* probability distribution that not only relies on all of the information that exists in the market, but is "maximally noncommital to missing information" Jaynes (1957a), meaning this distribution, once found, does not (cannot) contain any further assumptions than those that are used in its derivation. In this paper, we show that only two assumptions are required to derive this unique distribution:

1. There is a forward contract on the stock, which is fairly priced in the market;
2. The distribution has a variance; otherwise, statistical measures of risk are difficult to quantify.

Using only these assumptions, the subjective probability can be determined.

In the first section of this paper, we use the method of Lagrange multipliers to apply the constraints and derive the subjective probability. In the second section, we use this unique probability distribution to determine the price of a European call option and show that the Black–Scholes equation is obtained.

3. Subjective Probability

We assume that there is a liquidly traded market in forward contracts for this stock. A forward contract is a contract struck at time *t*, where both parties are obliged to trade

the stock at a later time T for an agreed-upon price, regardless of whether or not the buyer could obtain a lower price in the market at that future time. We denote the forward price by $F(t, T)$.

Suppose that the forward price can be obtained by summing over all possible future values that the stock could possibly attain, weighted by a probability

$$F(t, T) = \sum_\alpha p_\alpha S_\alpha \equiv \mathbb{E}[S(T)]. \tag{3}$$

We can think of this probability as containing the totality of information that all market participants collectively possess. It is this probability that we seek to determine.

The second piece of information reflects the fact that the market participants know, almost certainly, that the stock price will not exactly attain the forward price at time T; there is some uncertainty in the future outcome. They must decide on a measure of uncertainty in the forward price $F(t, T)$, and, more importantly, they must decide how this measure of uncertainty is determined and quoted. Once chosen, the participant can calculate the subjective distribution, which represents the least-biased estimate based on these two pieces of information. By following this prescription, we show how to arrive at the Black–Scholes formula.

To motivate the difference between this approach and the traditional approach, we examine what we mean by a subjective probability distribution rather than an objective one. By subjective, we mean "the sense that it describes only a state of knowledge, and not anything that could be measured in a physical experiment" Jaynes (2002), rather than a proscriptive probability distribution. For instance, one could posit a stochastic differential equation for the underlying stock price:

$$dS(t) = \mu(S, t)dt + \sigma(S, t)dW(t), \tag{4}$$

which a priori describes the dynamics of the stock price and, in principle, gives complete information on the statistics of the movements. A posteriori, these statistics can be empirically tested and (4) can be augmented (or rejected) if need be.

On the other hand, subjective probability distributions do not proscribe any dynamics or constrain the system at all; they are an inferential tool used when the underlying dynamics of the system are too complex to determine (or do not even exist at all).

To begin, we specify that volatility is a measure of the distribution of returns on the underlying asset. Mathematically, we write that the variance is an average of the square of the logarithm of the future underlying price relative to the current price:

$$v(T) = \mathbb{E}\left[\left(\ln \frac{S(T)}{S(t)} \right)^2 \right]. \tag{5}$$

By using the maximum entropy framework to determine the probabilities, we guarantee that the results do not implicitly contain any further assumptions or biases.

The derivation of the subjective distribution proceeds by maximization of the Shannon entropy Shannon (1948):

$$H = -\sum_\alpha p_\alpha \ln p_\alpha, \tag{6}$$

subject to the constraints

$$g_0 = 1 - \sum_\alpha p_\alpha, \tag{7}$$

$$g_1 = \mu(t, T) - \sum_\alpha \ln \frac{S_\alpha}{S(t)} p_\alpha, \tag{8}$$

$$g_2 = v(t,T) - \sum_{\alpha} \left(\ln \frac{S_{\alpha}}{S(t)} \right)^2 p_{\alpha}. \tag{9}$$

These particular constraints enforce normalization of the probability, i.e., that the forward price will be priced correctly (the functional form of μ will be fixed a posteriori to enforce the average (3)) and that the volatility is given by Equation (5).

Functional maximization subject to constraints is carried out by the method of Lagrange multipliers:

$$\delta H + \lambda \delta g_0 + \gamma_1 \delta g_1 + \gamma_2 \delta g_2 = 0, \tag{10}$$

leading to the probability

$$p_{\alpha} = \exp\left(-(1+\lambda) - \gamma_1 \ln \frac{S_{\alpha}}{S(t)} - \gamma_2 \left(\ln \frac{S_{\alpha}}{S(t)} \right)^2 \right). \tag{11}$$

The Lagrange multipliers are determined by the set of constraint functions $\{g_i = 0\}$, $i = 1, 2, 3$. Instead of fixing λ, we introduce the "partition function" $Z \equiv \exp(1+\lambda)$, where

$$Z = \sum_{\alpha} \exp\left(-\gamma_1 \ln \frac{S_{\alpha}}{S(t)} - \gamma_2 \left(\ln \frac{S_{\alpha}}{S(t)} \right)^2 \right). \tag{12}$$

With this definition, the probability in (11) becomes

$$p_{\alpha} = \frac{1}{Z} \sum_{\alpha} \exp\left(-\gamma_1 \ln \frac{S_{\alpha}}{S(t)} - \gamma_2 \left(\ln \frac{S_{\alpha}}{S(t)} \right)^2 \right). \tag{13}$$

In order to make contact with the Black–Scholes derivation, we make the substitution $\sum_{\alpha} \to \int_0^{\infty} dx$. The partition function now has a closed-form solution:

$$Z = S(t) \sqrt{\frac{\pi}{\gamma_2}} \exp\left(\frac{1}{4} \frac{(\gamma_1 - 1)^2}{\gamma_2} \right). \tag{14}$$

The further two constraints $\{g_i = 0\}$, $i = 1, 2$ are enforced by choosing the γ_is such that

$$\mu(t,T) = -\frac{\partial}{\partial \gamma_1} \ln Z, \tag{15}$$

$$v(t,T) - \mu(t,T)^2 = -\frac{\partial}{\partial \gamma_2} \ln Z - \left(\frac{\partial}{\partial \gamma_1} \ln Z \right)^2. \tag{16}$$

These equations have solution

$$\gamma_1 = 1 - \frac{\mu(t,T)}{v(t,T)}, \tag{17}$$

$$\gamma_2 = \frac{1}{2v(t,T)}. \tag{18}$$

With the constraints enforced, the probability in (11) becomes

$$p(x,t,T) = \frac{1}{x\sqrt{2\pi v(t,T)}} \exp\left(-\frac{1}{2v(t,T)^2} \left(\ln \frac{x}{S(t)} - \mu(t,T) \right)^2 \right). \tag{19}$$

The next step is to identify the averages $v(t, T)$ and $\mu(t, T)$ with their corresponding financial parameters. The realized volatility is identified as the integral of the volatility term structure that the market participants wish to use to model the underlying

$$v(t, T) = \int_t^T \sigma^2(s)ds; \tag{20}$$

the maximum entropy distribution does not limit this choice. [2]

The average quantity $\mu(t, T)$ is fixed by the forward price (3):

$$F(t, T) = \int_0^\infty x p(x, T)dx, \tag{21}$$

which is an implicit equation fixing $\mu(t, T)$ in terms of the forward price $F(t, T)$ with solution

$$\mu(t, T) = \ln \frac{F(t, T)}{S(t)} - \frac{1}{2} v(t, T). \tag{22}$$

We are now in a position to price any contingent claim armed with the probability distribution

$$p(x, t, T) = \frac{1}{x\sqrt{2\pi v(t, T)}} \exp\left(-\frac{1}{2v(t, T)}\left(\ln \frac{x}{F(t, T)} + \frac{1}{2}v(t, T)\right)^2\right). \tag{23}$$

In particular, we can now price the European call option by explicitly calculating the present value of the uncertain future cash flow using the probability distribution that is guaranteed to not have any more information, or assumptions, than the two that we imposed: it prices forward prices exactly, and the distribution has a variance.

4. Re-Deriving the Black–Scholes Formula

We look at the simplest security with uncertain cash flows, that of a European call option on a stock with a current traded market price $S(t)$. The option gives the buyer the right, but not the obligation, to purchase the stock at the strike price K at some point in the future time $T > t$. That is, we seek to determine the unique value of $C(t, T)$ that is the present value of the uncertain future cash flow:

$$C(t, T) = Z(t, T)\mathbb{E}[\max(S - K, 0)]. \tag{24}$$

Inserting the subjective probability (23) into this payoff formula, i.e.,

$$C(t, T) = Z(t, T)\left(\int_K^\infty x p(x, t, T)dx - K \int_K^\infty p(x, t, T)dx\right) \tag{25}$$

the price can be calculated exactly:

$$C(t, T) = Z(t, T)(F(T)\mathcal{N}(d_+) - K\mathcal{N}(d_-)), \tag{26}$$

where $\mathcal{N}(x)$ is the cumulative normal distribution and

$$d_\pm = \frac{\ln \frac{F(t, T)}{K} \pm \frac{1}{2}v(t, T)}{\sqrt{v(t, T)}}. \tag{27}$$

Equation (26) is the celebrated Black–Scholes formula for the price of a European call option with a volatility term structure.

5. Conclusions

The standard way of teaching the pricing of contingent claims in finance classes relies on the Black–Scholes–Merton framework, stochastic calculus and dynamic asset pricing theory. While mathematically sound, this approach has two major pedagogical drawbacks. First, the student may not have been exposed to such advanced mathematics, and second, more fundamentally, a stock price may not have an objective probability distribution such as, say, a diffusing pollen particle. Stocks are human creations and their trading depends on human emotion. Further, if a model becomes accepted in the market, this can fundamentally alter the way an asset trades.

In this paper, we have derived the same equation without the use of stochastic calculus or dynamic asset pricing theory; therefore, this approach could be better suited to early finance courses. Furthermore, the Black–Scholes equation is derived with only two assumptions: that a forward contract is traded in the market, and that the probability distribution used has some measure of dispersion and, importantly, a guarantee that no other information has been used. A recent paper extends the analysis in this paper by using the maximum entropy formulation and all available option prices to infer higher moments of the distribution Ardakani (2022).

Although we do not believe this supplants the standard treatment of pricing-contingent claims, this method offers a new perspective on the age-old problem, and one that fits easily into an introductory framework of finance without the need to complicate matters by the introduction of advanced concepts such as filtrations, stochastic calculus and numeraires.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The author declares no conflict of interest.

Notes

[1]   Shannon used the logarithm of base 2 in the paper, as he was concerned with error-correcting codes of binary digits. Since we are interested in the continuous case, we use the natural logarithm.

[2]   In the standard treatment of Black–Scholes, the volatility of the underlying asset is assumed to be a constant, resulting in the realized variance, as given by the formula volatility, becoming $v(t, T) = \sigma^2 T$.

References

Ardakani, Omid M. 2022. Option pricing with maximum entropy densities: The inclusion of higher-order moments. *Journal of Futures Markets* 42: 1821–36. [CrossRef]

Black, Fischer, and Myron S. Scholes. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81: 637–54. [CrossRef]

Cox, John C., and Stephen A. Ross. 1976. The valuation of options for alternative stochastic processes. *Journal of Financial Economics* 3: 145–66. [CrossRef]

Dupire, Bruno. 1994. Pricing with a smile. *Risk* 7: 18–20.

Fabozzi, Frank J., and Steven V. Mann. 2021. *The Handbook of Fixed Income Securities*, 9th ed. New York: McGraw-Hill Education.

Harrison, J. Michael, and Stanley R. Pliska. 1981. Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications* 11: 215–60. [CrossRef]

Heston, Steven L. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* 6: 327–43. [CrossRef]

Jaynes, Edwin T. 1957a. Information theory and statistical mechanics. *Physical Review* 106: 620–30. [CrossRef]

Jaynes, Edwin T. 1957b. Information theory and statistical mechanics. ii. *Physical Review* 108: 171–90. [CrossRef]

Jaynes, Edwin T. 2002. *Probablity Theory: The Logic of Science*. Cambridge: Cambridge University Press.

Lintner, John. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics* 47: 13–37. [CrossRef]

Merton, Robert C. 1973. Theory of Rational Option Pricing. *Bell Journal of Economics* 4: 141–83. [CrossRef]

Merton, Robert C. 1976. Option Pricing When Underlying Stock Returns are Discontinuous. *Journal of Financial Economics* 3: 125–44. [CrossRef]

Shannon, Claude. 1948. A mathematical theory of communication. *The Bell Systems Technical Journal* 27: 379–423. [CrossRef]

Sharpe, William F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19: 425–42.

Article

Separating Equilibria with Search and Selection Effort: Evidence from the Auto Insurance Market

David Rowell [1,*] and Peter Zweifel [2]

[1] Center for the Business and Economics of Health (CBEH), School of Economics, The University of Queensland, St Lucia, Brisbane 4067, Australia

[2] Department of Economics, University of Zurich, Rämistr. 71, 8006 Zurich, Switzerland; peter.zweifel@econ.uzh.ch

[*] Correspondence: heterodox62@gmail.com; Tel.: +61-042-132-1944

Abstract: The objective of this paper is to assess the behavior of policyholders and insurance companies in the presence of adverse selection by accounting for costly search and selection efforts, respectively. Insurers seek to stave off high-risk types, while consumers are hypothesized to maximize coverage at a given premium. Reaction functions are derived for the two players giving rise to Nash equilibria in efforts space, which are separating almost certainly regardless of the share of low risks in the market. Empirical evidence from the Australian market for automobile insurance is analyzed using Structural Equation Modeling. Convergence has been achieved with both the developmental and test samples. Both consumer search and insurer selection are found to be positively correlated with risk type, providing a good measure of empirical support for the theoretical model.

Keywords: adverse selection; separating equilibria; consumer search effort; insurer selection effort; automobile insurance

1. Introduction

Ever since the seminal article by Rothschild and Stiglitz (1976), hereafter abbreviated as RS, both economists and policy-makers have been concerned about the effects of asymmetric information on insurance markets. Since an equilibrium pooling of high- and low-risk types cannot be sustained according to RS, an insurance company (IC henceforth) enrolling both types can be challenged by a competitor who launches a policy with limited coverage but a low premium that attracts only low-risk types. The incumbent IC may respond by launching separating contracts, one offering full coverage at a high premium (which appeals to the high-risk types), and the other offering limited coverage at a low premium (which appeals to the low-risk types only). Yet these separating contracts can still be challenged by an (unsustainable) pooling contract, provided the share of low-risk types is high enough, which potentially raises the specter of the nonexistence of equilibrium in insurance markets (Mimra and Wambach 2014).

However, to the best knowledge of the authors, the literature building on RS has accepted the implicit assumptions that the challenging IC does not incur any risk selection expense while low-risk types find the policy suiting them without undertaking costly effort. Both assumptions are far from reality. On the part of the ICs, risk selection involves the creation, marketing, and monitoring of policies—all costly activities. As to consumers, while the Internet abounds with sites designed to make their search easier [Choice (2024) and Consumer Reports (2019)], a survey suggests that many of them have difficulty finding a policy suited to their needs (Liferay 2019).

The objective of this contribution is to answer the following research question: could a separating equilibrium as described by RS be shown to exist, theoretically and empirically, in a market for insurance where policyholders and ICs engage in costly search and risk selection, respectively? Against this background, this contribution introduces the first costly

search effort on the part of consumers and the risk selection effort on the part of ICs. In a competitive market, ICs set their selection effort, which is found to increase with consumers' search effort. Second, consumers choose the policy granting them maximum coverage for the given premium, with high-risk types exerting more search effort than low-risk ones. In the Nash equilibria, they end up paying a higher premium while obtaining a higher degree of coverage. In contradistinction to RS, the existence of separating equilibria is almost certain and does not depend on the share of low risks in the market. Also, taking into account efforts is shown to generate new testable predictions. In particular, high-risk types undertake high search effort matched by high selection effort; conversely, in the case of low-risk types, low search effort combines with low selection effort. Third, this theoretical finding is tested using a rather comprehensive dataset on Australian auto insurance. Since both types of effort are not directly observable, Structural Equation Modeling (SEM) is applied, which permits distinguishing multiple indicators with their measurement errors from type-specific efforts as the latent variables making up the structural core. The hypothesized relationships between IC selection effort and risk type, on the one hand, and consumer search effort, on the other, receive a good measure of confirmation.

The remainder of this paper is structured as follows: Section 2 of the Literature Review provides a review of both the theoretical and empirical literature relating to the RS model. In Section 3 Materials and Methods, the interaction between an IC optimizing its risk selection effort and a consumer searching for a suitable policy (i.e., one offering a maximum amount of coverage for a given premium) is modeled. The resulting Nash equilibria are first characterized in efforts space and then projected into conventional RS wealth levels space. In Section 4 Empirical Analysis, a dataset containing indicators of both consumer search and IC selection efforts in the Australian auto insurance market is used to test these predictions using SEM. Section 5 offers a summary and concluding remarks.

2. Literature Review

2.1. Theoretical Literature

In 1976, RS presented a static model of a market for insurance, which relaxed the assumption of homogeneous loss probabilities and perfect information. High- and low-risk consumers exist and possess private information regarding their risk type. RS hypothesized the possibility of a separating equilibrium where high- and low-risk types accept different premium-coverage contracts. Their concept of non-linear pricing without cross-subsidization challenged earlier models of insurance markets with linear pricing, making policyholders pay the same average price for insurance and resulting in cross-subsidization [(Arrow 1970; Pauly 1974)].

Much of the analysis that followed has used game theory to more precisely define the nature of the interaction between insurance companies and customers (Rothschild and Stiglitz 1997). Immediately after the publication of RS, several theoretical papers sought to demonstrate the existence of an equilibrium in insurance markets by including IC behavior in their models.[1]

Wilson (1977) stated that while no equilibrium may exist if the incumbent IC has static expectations of challenger ICs, a pooling equilibrium may exist if expectations can be revised. Spence (1978) extended Wilson's (1977) analysis to include a menu of contracts and derived an equilibrium with separating, cross-subsidizing contracts. Jaynes (1978) relaxed the assumption that contracts are exclusive and ICs do not share information. Firms that share information offer a pooling contract, while those that abstain underwrite contracts for high-risk policyholders. Riley (1979) posited that if a challenger can respond with a new contract, a separating equilibrium is possible. Engers and Fernandez (1987) generalized Riley's (1979) reactive equilibrium by considering the possibility of adding multiple new contracts.

Hellwig (1987) recast the RS model in the mold of a two-stage game where, in the first stage, uninformed ICs offer contracts and, in the second stage, informed consumers choice of contracts. Realism is added to the model by including a third stage, where ICs

can reject consumers' applications, in contrast to Wilson (1977), who proposed that loss-making contracts are not necessarily withdrawn to create a sustainable pooling equilibrium. He noted quite generally that the exact formulation of the game may change predictions substantially. Asheim and Nilssen (1996) varied the conditions of the game by allowing ICs to renegotiate contracts with their policyholders, such that the revised contract is universally offered to all policyholders, while Netzer and Scheuer (2014) allow the IC to exit from the market altogether. Both models predict a separating equilibrium.

A recent focus of the RS literature [Ales and Maziero (2014), Attar et al. (2011, 2014, 2016, and 2020)] has been to explore the implications for equilibrium under adverse selection when insurers do not know whether or not to sell their contracts exclusively. In this situation, the predicted outcomes are (i) no coverage of low-risk types or (ii) an absence of equilibrium, although Attar et al. (2011) have argued that some pooling and hence coverage of low-risk types could also exist.

Research that was initially published as a working paper by Stiglitz et al. (2017) and subsequently revisited by Kosenko et al. (2023) models a market for insurance that incorporates information revelation strategies by consumers and insurers. Kosenko et al. (2023) introduce bilateral endogenous information disclosure about insurance purchases. They assume non-exclusivity in that consumers buy from multiple sellers while insurers offer contracts to consumers not observed by competitors. Each consumer and insurer can make strategic decisions about what information to disclose to whom. The authors find that there always exists an equilibrium outcome, which entails partial pooling. According to Kosenko et al. (2023), their contribution differs from those of Jaynes (1978), Jaynes (2011), and Hellwig (1987) because it considers information revelation by consumers as well as between insurers.

As will be described in greater detail below, this paper also models the interaction between insurer and policyholder. However, it does not assume that the two players passively process information that has been strategically revealed to them. Rather, they actively seek out, at non-zero cost, their preferred policy and undertake a selection effort. It is only through the two players' interaction in the Nash equilibrium that the risk types are revealed.

2.2. Empirical Literature

Kosenko et al. (2023) conclude their paper by identifying a need for empirical research. We hope that our results provide an impetus for further policy and empirical applications, with insights into why certain markets take the form they do and how one might improve the design of markets with asymmetric information (Kosenko et al. 2023, p. 146).

Mimra and Wambach (2014) had already noted the paucity of empirical evidence. Curiously, although there is by now substantial empirical literature investigating whether adverse selection is prevalent and important in insurance markets[2], the question of whether the allocation in these markets is of the RS-type or the Miyazaki-Wilson-Spence (MWS) type has so far been neglected. (Mimra and Wambach 2014, p. 15).

Indeed, the authors of this contribution could find only two research papers that explicitly tested for evidence of a separating equilibrium in an insurance market. The first, written by Dionne and Doherty (1994), importantly introduced experience rating into the RS model. The authors modeled the effect of semi-commitment with renegotiation (defined as insurance with an option to renew with pre-specified conditions) and contrasted its implications with single-period and no-commitment models. Under competitive conditions, an IC offers a pooling policy with partial coverage in the first period and an experience-rated, separating set of policies in the second period. They tested their theoretical predictions using aggregated Californian automobile insurance data. They report that some automobile insurers use commitment to attract low-risk policyholders, while others attract high-risk policyholders, which is presented as evidence of a separating equilibrium.

The second paper, by Puelz and Snow (1994), used claims data from an automobile crash insurer in Georgia to test for evidence of a separating equilibrium. They claimed

their analysis supports the hypothesis of adverse selection with a separating equilibrium. Despite criticism that the test for adverse selection did not control for *ex ante* moral hazard (Chiappori 1999; Chiappori and Salanié 2000; Dionne et al. 2001), their paper still offers a credible test of the proposition contained in the RS paradigm.

A third identified paper published by Dionne et al. (2013) does not report an explicit test for separating equilibrium. However, arguably their evidence regarding *ex ante* moral hazards in the French market for automobile insurance using longitudinal data suggests the emergence of a separating equilibrium. The authors distinguish between a liability-only (*responsabilite civile*) and a comprehensive optional (*assurance tous risques*) contract, both experience-rated. Their analysis based on parameters characteristic of the French market shows that the probability of a high-risk type having a comprehensive policy exceeds that of a low-risk type, with the difference in probabilities increasing rather than diminishing over time. This suggests that separating contracts emerges over time through learning by both consumers and insurers.

While remaining close to the RS paradigm for facilitating comparison, this contribution differs from the received literature in three ways. First, it introduces costly searches on the part of consumers and costly selection efforts on the part of ICs. Second, it derives Nash equilibria in efforts space along with several new testable predictions. Finally, it benefits from a large array of indicators of Australian consumers' search effort and ICs' selection effort for testing a core prediction.

3. Materials and Methods

The interaction between an IC optimizing its risk selection effort and a consumer searching for a suitable policy (i.e., one offering a maximum amount of coverage for a given premium) is modeled. The resulting Nash equilibria are first characterized in effort space and then projected into conventional RS wealth level space.

3.1. A Game-Theoretic Model with Consumer Search Effort and IC Selection Effort

Both the extant theoretical and empirical literature neglect an important fact: both high-risk (c^H) and low-risk (c^L) consumers engage in costly search efforts (c) to find insurance policies that best suit them. In turn, ICs engage in a costly selection effort (e) designed to attract low-risk and avoid high-risk consumers without being able to distinguish between them initially.

In this section, a simple game-theoretic model is developed to determine Nash equilibria for high- and low-risk types in effort space. Note that both types of effort are implicit in the RS model (otherwise, there would never be a challenger of the incumbent IC, and high-risk types would not infiltrate the contract designed for the low-risk ones). In the present model, search effort and risk selection effort are the decision variables controlled by the respective players. Figure 1 shows the stages of the game.

1   IC sets common selection effort level e in anticipation of consumer search effort c; $de/dc > 0$ derived from comparative statics

2   High-risk type and low-risk type consumers set levels of search effort c^H, c^L, with $c^H > c^L$ and $dc^H/de < 0$, $dc^L/de < 0$ derived from comparative statics

3   The interaction between IC and consumers results in Nash equilibria with $e^H > e^L$ and $c^H > c^L$ which are almost certainly of the separating type.

Figure 1. Stages of the game.

Stage 1: Insurers

Insurers are viewed as expected profit maximizers,

$$E\Pi_{e} = \pi(e)\left[P^{H}(e) - EI^{H}\left(c^{H}\right)\right] + (1 - \pi(e))\left[P^{L}(e) - EI^{L}\left(c^{L}\right)\right] - e, \qquad (1)$$

with $E\Pi$ denoting expected profit, $\pi(e)$, the probability of enrolling a high-risk type depending on risk selection effort e (at unit cost of one for simplicity) with $\partial\pi/\partial e < 0$ and $\partial\pi^{2}/\partial e^{2} > 0$ indicating decreasing marginal effectiveness. The notation emphasizes the fact that when launching a contract, the IC cannot identify risk types and has to set the selection effort at a single value e. Premiums $P^{H}(e)$ $\left[P^{L}(e)\right]$ are market-determined (see stage 3) but must cover both the expected value of claims $EI^{H} = \overline{\rho}I^{H}\left(c^{H}\right)$ and $EI^{L} = \overline{\rho}I^{L}\left(c^{L}\right)$, respectively based on the known population average of loss probability $\overline{\rho}$ as well as the cost of selection effort. The first-order condition for an interior optimum reads,

$$\frac{dE\Pi}{de} = \partial\pi/\partial e \cdot \left\{\left[P^{H} - EI^{H}\left(c^{H}\right)\right] - \left[P^{L} - EI^{L}\left(c^{L}\right)\right]\right\} - 1 = 0. \qquad (2)$$

This shows that selection effort has a positive marginal return if the expected margin on the high-risk types $\left[P^{H} - EI^{H}(c^{H})\right]$ is smaller than that on the low-risk types $\left[P^{L} - EI^{L}(c^{L})\right]$. The difference between the two margins is especially marked if $P^{L} - EI^{L}\left(c^{L}\right) < 0$, as is often the case under community rating [which has been argued to induce risk selection in health insurance by Pauly et al. (2007)].

Through the marginal effectiveness of consumers' search efforts, the IC's reaction function in principle depends on the risk type it is confronted with [see Equation (A3) of Appendix A.1]. However, since the IC cannot distinguish between risk types prior to the determination of the Nash equilibria (which depend on the consumers' reaction functions), only one IC reaction function is shown in Figure 2, with

$$\frac{de}{dc} > 0. \qquad (3)$$
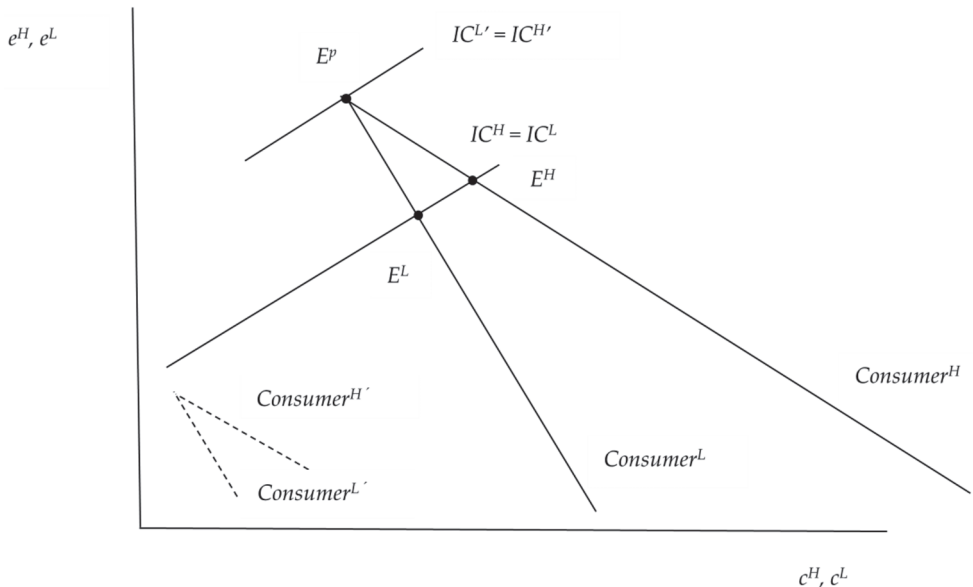
Figure 2. Reaction functions and Nash equilibria in efforts space.

Stage 2: Consumers

Consumers are seen as expected utility maximizers who undertake search efforts to secure a maximum amount of coverage at the going premium:[3]

$$\underset{c^H}{EU^H} = \rho^H v^H \left[W_0 + I^H\left(c^H, e^H\right) - L - P^H(e)\right] + \left(1 - \rho^H\right) v^H \left[W_0 - P^H(e)\right] - c^H; \quad (4a)$$

$$\underset{c}{EU^L} = \rho^L v^L \left[W_0 + I^L\left(c^L, e^L\right) - L - P^L(e)\right] + \left(1 - \rho^L\right) v^L \left[W_0 - P^L(e)\right] - c^L. \quad (4b)$$

Here, $EU^H\left(EU^L\right)$ denotes the expected utility of a high- (low-) risk type, $v^H\left(v^L\right)$, VNM (Von Neumann and Morgenstern) risk utility functions with $v^{H'} > 0\left(v^{L'} > 0\right)$ and $v^{H''} < 0\left(v^{L''} < 0\right)$, W_0, exogenous initial wealth, $I^H\left(c^H, e\right)$ $[I^L\left(c^L, e\right)]$ the degree of coverage, which depends on search effort with $I^H[0, \cdot] = 0$, $I^L[0, \cdot] = 0$, $\partial I^H / \partial c^H > 0$, $\partial I^L / \partial c^L > 0$, and $\rho^H(\rho^L, \rho^l < \rho^H)$ the loss probabilities[4]. For simplicity, search effort by consumers is assumed to have a unit cost of one.

However, insurance coverage also depends on the IC's selection effort e. Arguably, selection effort lowers the effectiveness of consumer search, implying $\partial^2 I^H / \partial c^H \partial e < 0$, $\partial I^L / \partial c^L \partial e < 0$. The reason is that it burdens consumers with transaction costs, e.g., extra documentation. Recall that selection effort initially has a common value because the IC cannot distinguish between risk types (however, values of e differ in the Nash equilibria due to differing consumer responses).

The first-order conditions for an interior optimum[5] are given by

$$\frac{dEU^H}{dc^H} = \rho^H v^{H'}\left[W_0 + I^H\left(c^H, e\right) - L - P^H(e)\right] \cdot \partial I^H / \partial c^H - 1 = 0; \quad (5a)$$

$$\frac{dEU^L}{dc^L} = \rho^L v^{L'}\left[W_0 + I^L\left(c^L, e\right) - L - P^L(e)\right] \cdot \partial I^L / \partial c^L - 1 = 0. \quad (5b)$$

Note that unless the derivatives of $I(\cdot)$ functions differ substantially (for which there is no apparent reason), the high-risk types are predicted to undertake more effort than the low-risk ones. First, $\rho^H > \rho^L$; second, given risk aversion and identical initial wealth, this implies $v^{H'}[W] > v^{L'}[W]$; third, this difference is not neutralized because the high-risk type's amount of coverage is matched by a higher premium (see Stage 3 below). Thus, the marginal benefit of search is higher for the high-risk types than the low-risk ones, while its marginal cost is the same by assumption, inducing more search effort.

The derivation of the consumers' reaction functions is relegated to Appendix A.2 [see Equations (A4) and (A5)]; their slopes are

$$\frac{dc^H}{de} < 0, \frac{dc^L}{de} < 0, \text{ with } \left|\frac{dc^H}{de}\right| > \left|\frac{dc^L}{de}\right| \quad (6)$$

In Figure 2, the reaction functions are drawn as straight lines (with dc^H / de running flatter since c^H and c^L are depicted on the horizontal axis) since nothing can be said about their curvature, which depends on the third derivatives of the functions $I^H\left(c^H, e\right)$ and $I^L\left(c^L, e\right)$, respectively. However, the reaction function of the high-risk type is farther out in the relevant domain because the respective probabilities are multiplied with first-order derivatives in Equations (5a) and (5b), which must dominate the second-order ones lest they change sign from positive to negative, contradicting assumptions.

Stage 3: Nash equilibria in efforts and wealth levels space

Given the reaction functions, the resulting Nash equilibria can now be characterized; in effort space and are represented by E^H and E^L in Figure 2. It shows that, from the interaction with consumers, the IC can now distinguish between the two risk types. Even if

it is unable to replace the common value of loss probability $\bar{\rho}$ by ρ^H and ρ^L, respectively, it will charge premiums P^H and $P^L < P^H$ to recover its costly risk election efforts e^H (with $e^L < e^H$), presumably in the guise of a proportional loading. Evidently, a separating equilibrium in the market is almost certain to exist. Nonexistence would require consumers to exert almost no search effort regardless of the IC's selection effort (indicated by the two dashed lines that do not intersect with the IC's reaction function), contrary to evidence, especially in the context of renewals of auto insurance policies (Mathews 2022).

Conversely, the likelihood of a pooling equilibrium occurring (E^p) is also very low. The two consumer types would have to exert exactly the same amount of effort in response to the selection effort by the IC. Moreover, pooling equilibria beyond E^p can be excluded because they contradict first-order conditions (4a) and (4b), calling for high-risk types to exert more effort than low ones. Finally, the separating equilibrium is sustainable because it does not depend on the share of low-risk types in the population and cannot be challenged by a competing contract, in contradistinction with the conventional RS framework.

3.2. Theoretical Findings
3.2.1. Results in Efforts Space

Figure 2 shows a separating equilibrium modeled in effort space. High-risk consumers are predicted to exert high search effort, which is matched by high selection effort on the part of the IC, while low-risk ones exert little search effort combined with low selection effort.

A Testable Prediction. The interaction of risk-selecting insurers with consumers searching for maximum coverage given the premium is predicted to result in a separating Nash equilibrium (which is almost certain to exist) that is characterized by high selection effort combined with high consumer search effort in the case of high-risk types (E^H) and low selection effort combined with low search effort in the case of low-risk ones (E^L).

Other theoretical insights implied by Figure 2, which are not available in the conventional RS approach, include:

- On the IC's side, information, e.g., concerning miles driven per year, quality of roads typically traveled, and crime incidence in the area of residence, may make the IC's risk selection effort more effective in the case of auto insurance. This increases $\partial\pi/\partial e$ in absolute value, causing the slope of the IC's reaction function to increase according to Equations (A3) of Appendix A.1. The result is a greater difference between e^H and e^L (facilitating the separation of equilibria) combined with a smaller difference between c^H and c^L (see Figure 2).

- The same effects are predicted *ceteris paribus* if consumers' search effort becomes more effective, e.g., due to the Internet, media such as *Consumer Reports*, and public regulation designed to enhance transparency. In Equations (A4) and (A5) of Appendix A.2, the terms $\partial EI^H/\partial c^H > 0$ and $\partial EI^L/\partial c^L > 0$ increase, and with them, IC's reaction function in Figure 2 becomes more responsive to consumers' search efforts.

- The *ceteris paribus* clause above cannot be neglected because the consumers' reaction functions would be affected as well. In Equations (A4) and (A5), the terms $\partial^2 I^H/\partial c^H \partial e < 0$ and $\partial^2 I^L/\partial c^L \partial e < 0$ go towards zero, indicating that the IC's risk selection effort does not counterbalance consumers' search effort to the same extent when they are better informed. In Figure 2, the reaction function labeled *Consumer*^{H} in particular becomes more responsive to IC's selection effort since the term $\partial^2 I^H/\partial c^H \partial e$ is multiplied by $\rho^H > \rho^L$, causing the differences between e^H and e^L as well as c^H and c^L to increase.

- Differences in risk aversion [indicated by $RA^H = -v^{H\prime\prime}/v^{H\prime} > 0$ and $RA^L = -v^{L\prime\prime}/v^{L\prime} > 0$ in Equations (A4) and (A5)] have an impact on consumers' reaction functions. For instance, let RA^H increase relative to RA^L; a possible reason is that high-risk types happen to coincide with higher age, which is associated with increased risk aversion (Halek and Elisenhauer 2001). This has the effect of making the high risk's response to IC selection effort more marked, resulting in a flatter *Consumer*^{H} line of Figure 2 and hence a larger difference between e^H and e^L as well as c^H and c^L.

3.2.2. Results in Wealth Space

In view of the deeply entrenched RS approach, it was important to explore whether and how the prediction from effort space (see Figure 2) carries over to the two-state wealth space (W_1, W_2) described in the conventional RS model (see Figure 3). The projection in Figure 3 reveals several differences from the RS model:

- Since the IC makes a risk selection effort, the cost, which typically gives rise to a pro-portional loading, a (marginally) fair premium is excluded from the onset. Therefore, at C^{*L}, high-risk types necessarily opt for partial coverage.
- Even though the IC is not able to infer the true loss probabilities, forcing it to continue using the average value $\bar{\rho}$, the insurance line labelled $P^H(e^H)$ has a lower slope than $P^L(e^L)$, reflecting the IC's higher amount of risk selection effort in its interaction with a high-risk type in stage 3.
- Because high-risk types are predicted to invest relatively more effort in seeking out the contract that maximizes coverage for a given premium, they bear a higher initial transaction cost, c^H, which shifts the origin of their insurance line from A_0 to $I^H = 0$. Thus, the probability of $I^H = 0$ constituting the optimum is far greater than in the RS approach. This provides an explanation for the observation that it is the widely discussed inability of high-risk types to obtain insurance coverage that constitutes a policy issue rather than the rationing of low-risk types' coverage at Q^L because of the need to maintain a separating equilibrium.
- The location of the optimum C^{*H} in Figure 3 depends on the parameters appearing in Equation (A4), viz. $v^{H'}, RA^{H'}, \partial EI^H/\partial c^H$, and importantly on the IC's amount of selection effort e and hence $\partial \pi/\partial e$ in Equation (2).
- In the RS modeling, the pooling contract X (see Figure 3) can undermine a separating equilibrium provided the share of low-risk types in the population is sufficiently high (the pooling insurance line must run close to that labeled $P^L(e^L)$). Yet when consumer search and insurer selection efforts are considered, a pooling equilibrium can be excluded almost with certainty, which implies that the separating equilibrium cannot be undermined. (see E^p in Figure 2 again).
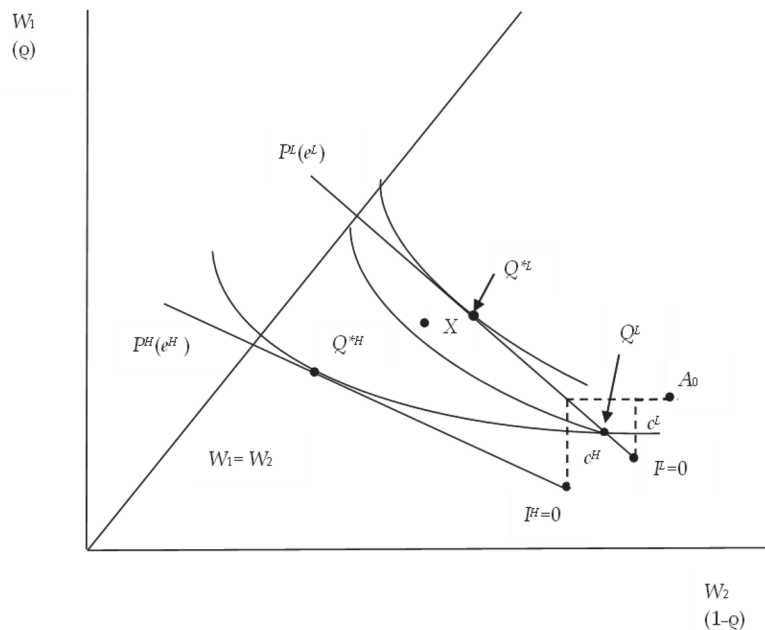
Figure 3. Projecting Nash equilibria into (W_1, W_2)-levels space.

Since the findings derived from the efforts space (see Figure 2 again) carry over to the wealth-levels space of Figure 3, the comparison of the present analysis with the conventional RS model can be summed up as follows: The interaction of consumers searching for maximum coverage given the premium and the risk-selecting insurer is more likely to result in a separating Nash equilibrium but also to involve no insurance coverage for high-risk types than in the RS framework. Moreover, contrary to RS, it cannot be undermined by a pooling contract, regardless of the share of low-risk types in the population.

4. Empirical Analysis

Data from the Australian market for automobile insurance (as described in Section 4.1) will be used to empirically test (as specified in Sections 4.2 and 4.3) the central prediction from Figure 2 (effort space). That is, high-risk consumers are predicted to exert high search effort, which is matched by high selection effort on the part of the IC, while low-risk ones exert little search effort combined with low selection effort. Unfortunately, the additional predictions that follow from Figure 3 (wealth space) are not tested empirically because (i) the consumers' degrees of risk aversion are not reported or known, and (ii) the data are cross-sectional and changes in behavior across time are largely unreported.

4.1. Data

Automobile insurance data are suited to testing the theoretical model because the risk rating of policyholders is less regulated than in other lines of insurance (e.g., health). This renders the ICs' selection effort potentially more readily observable [at least through a set of indicators (see below)]. It has been stated that

"[e]mpirical models of insurance markets would greatly enhance our ability to understand policy-relevant questions. Yet they are still quite rare. ...While much progress has been made in recent years in our understanding of insurance demand in particular, the most crying need is for market-wide data" (Salanié 2017).

The analysis of a dataset representative of the Australian market may therefore be of interest. Data are drawn from two sources: (i) a household survey of vehicle owners collected by the market research firm IMRAS Consulting (henceforth referred to as the IMRAS dataset) and (ii) insurance surveys published by the consumer advocacy group Choice.

4.1.1. Insurance Policies

In Australia, every vehicle must carry compulsory third-party (CTP) insurance to partially cover the cost of treating third-party injuries. Comprehensive insurance, which indemnifies the policyholder against the costs of damage to their own or another party's vehicle, is optional. Approximately 80% of vehicles in the survey were comprehensively insured. The IMRAS survey reported the name[s] of the respondent's CTP and comprehensive automobile insurer. The premium and amount of comprehensive coverage purchased are reported; however, the excess (i.e., deductible) is not reported. Policyholders were also asked to report their no-claim bonus (NCB), which typically ranges from 0% to 60% depending on the NCB scheme and claim history.

4.1.2. Insurers

The IMRAS dataset contains no information about the composition of individual insurance policies or the underwriting strategies of ICs. However, the journal *Choice* regularly compares many goods and services, including comprehensive insurance, to inform the purchasing decisions of its readership. Measures of insurer behavior were obtained through three reports. The first, a special report *Car Insurance*, published in 1997, compared premiums for three insurance vignettes (a high-risk scenario, a medium-risk scenario, and a low-risk scenario) within two regions (a high-risk region and a low-risk region) across six states. Some areas are risk-rated more highly than others because the risks of theft and accidents vary, as does the cost of repairs. Generally, urban areas are

rated as high-risk, and regional areas are rated as low-risk. The second report (Australian Consumer Association (ACA) 1997) compared premiums using a 5-star scale ranging from cheapest to most expensive (see Table A1 in Appendix B for details).

The third source is the report, *Your Car Insurance Toolkit* (Australian Consumer Association (ACA) 1999). It differentiates comprehensive insurance policies on the basis of three policy characteristics: (i) adjustment to the NCB following a claim; (ii) the option to protect the NCB following a claim; and (iii) the option of reducing the excess (see Table A2 in Appendix B for details).

These data were matched to respondents in the IMRAS dataset using the name of the comprehensive insurer. The result is a rich dataset providing information on 4005 vehicle owners but covering the year 1999 only. In addition to the market leaders (NRMA, AMI, RACV, and Suncorp), many smaller insurers are also in the dataset. Figure 4 reports the number of policies underwitted by each IC as well as the proportion of policyholders who reported a road traffic crash (RTC) from 1997 to 1999. This proportion is seen to vary substantially, providing a first indication that Australian ICs may differ in their selection efforts.
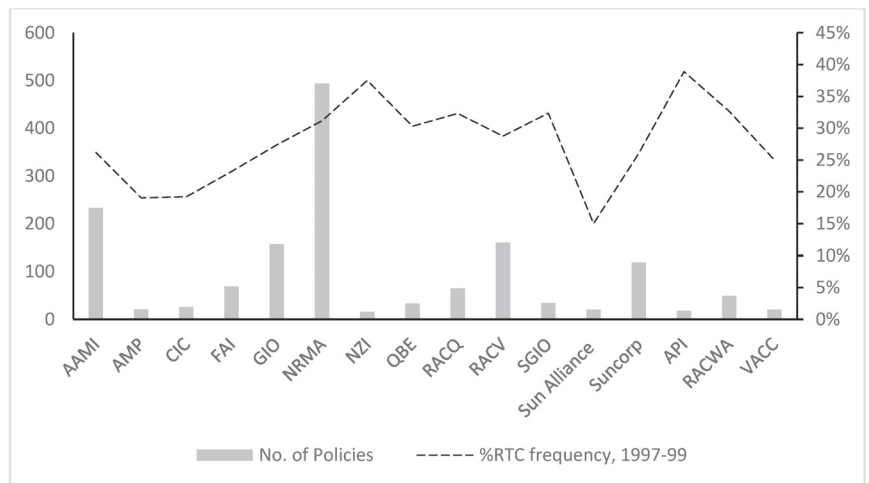
Figure 4. Market for automobile insurance in Australia, 1997–1999. Note: Insurers with < 15 policies in the IMRAS dataset are excluded from the graph.

4.1.3. Consumers

During a six-week period commencing in October 1999, market research was commissioned by IMRAS Consulting to analyze community attitudes toward the Australian smash (collision) repair market. Computer-assisted telephone interviews (CATI) were used to contact 37,833 rural and metropolitan households in four Australian states (New South Wales, Victoria, Queensland, and Western Australia). The response rate of 16.9 percent enabled data to be collected on 4006 households who provided policyholder characteristics (age, gender, and postcode), vehicle type (make, model, and vehicle age), and RTC history from 1994 to 1999. Although the data are now over twenty years old, they have an important advantage. The CATIs were conducted prior to the widespread use of mobile phones, which offer opportunities for recipients to screen calls. Arguably, this improves data quality.[6]

4.1.4. Evidence of Adverse Selection

Evidence of adverse selection is a necessary but not sufficient condition for the existence of a separating equilibrium (Puelz and Snow 1994). In 2017, Rowell et al. published an empirical analysis of the IMRAS dataset that tested for *ex ante* moral hazard in the

Australian automobile insurance market. The authors adapted a recursive model proposed by Dionne et al. (2013), which used a lagged measure of RTCs as opposed to claims to control for adverse selection. The rationale was that RTCs that did not result in a claim constitute insured motorists' private information about their risk type that is not available to the insurer. The statistically significant coefficient on lagged RTCs reported by Rowell et al. (2017) provides *prima facie* evidence of adverse selection in this market.

4.2. Model Specification

Since neither ICs' risk selection effort nor consumers' search effort is directly observable, they are treated as latent variables reflected by a set of indicators. The term "indicator" implies that (1) it need not vary in 1:1 proportion with the latent variable it represents, and (ii) it may contain measurement error with respect to the latent variable. Work with multiple indicators was pioneered by Jöreskog and Goldberger (1975); their approach has become known as "Structural equation modeling" (SEM) (Fan et al. 2016). SEM enables the analysis of relationships between one or more independent variables (continuous or discrete) and one or more dependent variables (continuous or discrete). Both the independent and dependent variables can also be measured directly, as in conventional regression analysis (Ullmann and Bentler 2004). In the present context, the advantage of SEM is that it allows for testing for the postulated causal relationship between ICs' risk selection effort and consumers' search effort using correlations between observed indicator variables (Kline 2016).

According to the Testable Prediction of Section 3.2.1, the interaction between consumers and ICs results in a Nash equilibrium, which is characterized by high consumer search and IC selection effort for high-risk types and low consumer search and IC selection effort for low-risk types. The dataset described above (see Table 1 for variable definitions) features several indicators of latent quantities. Equation (7) defines the structural core, which is composed of three latent variables: consumer search effort (*CSE*), insurer selection effort (*ISE*), and increasing risk type (*RT*$^+$).

$$ISE = \alpha_1 RT^+ + \varphi_1;$$
$$CSE = \alpha_2 RT^+ + \varphi_2, \text{ with} \tag{7}$$
$$Var(RT^+) = 1, E\varphi_1 = 0, \ E\varphi_2 = 0, Var(\varphi_1) = \sigma_1^2, Var(\varphi_2) = \sigma_2^2, E(\varphi_1, \varphi_2) = \sigma_{12}.$$

Since the distinction between high- and low-risk types in Section 3 would be difficult to implement, RT^+ is continuous rather than dichotomous[7]. In the path diagram of Figure 5 below, α_1 and α_2 are symbolized by arrows linking RT^+ and *ISE* and *CSE*, respectively. According to the Testable Prediction (Section 3.2.1), both coefficients are positive, *ceteris paribus*.
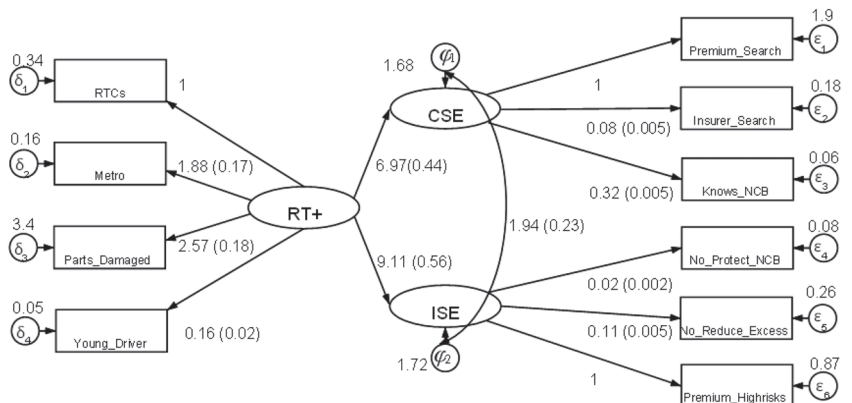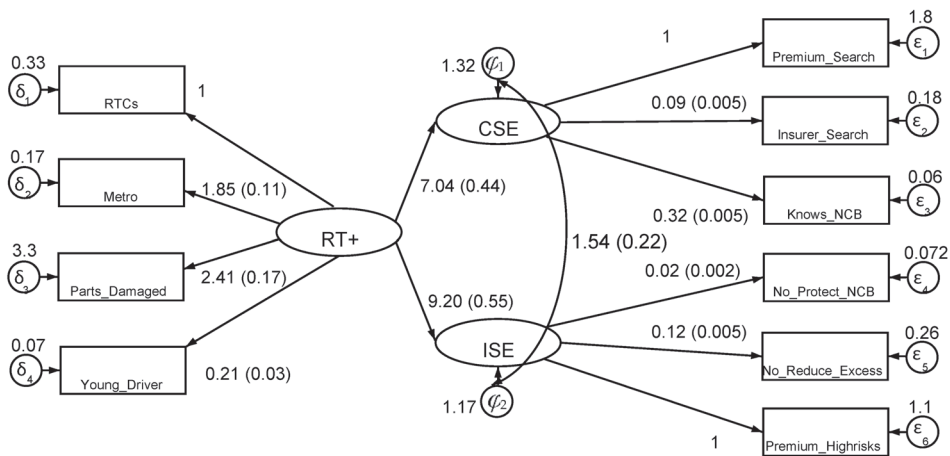

Figure 5. Structural equation model (developmental dataset, 1999, n = 2000). Note: LR test of model vs. saturated: $\chi^2(42) = 4936.82$, Probability $> \chi^2 < 0.01$, (n = 1033).

Table 1. Variable descriptions, summary statistics, and rationale for hypothesized correlations (full sample).

Variable	Mean [e]	S.D.	Skew	Rationale for Predicted Correlation with Latent Variable
Consumer Search Effort (CSE)				
Premium_Search (coverage/premium ratio reported as quintiles)	2.98	1.14	0.02	The coverage/premium ratio is positively correlated with *CSE*. Consumers are hypothesized to seek maximum coverage for a given premium.
Insurer_Search (=1 if CTP and Comprehensive insurers are not equal, =0 if otherwise)	0.27	0.44	1.05	Buying CTP and comprehensive insurance from the same firm is indicative of a low CSE; buying comprehensive insurance from an alternative firm is indicative of higher *CSE*. [a]
Knows_NCB (=1 if the consumer knows NCB, =0 otherwise)	0.89	0.31	−2.53	A policyholder knowing their NCB is indicative of a higher CSE.
Insurer Selection Effort (ISE)				
Premium_Highrisks (Premiums for high-risk policyholders in high-risk areas: 1 to 5 max.) [c]	3.68	0.70	−2.10	Higher premiums for the highest-risk policies are indicative of a higher ISE
No_Protect_NCB (=1 if NCB protection is not offered, =0 otherwise)	0.08	0.30	3.12	Prohibiting NCB protection is indicative of high ISE. [b]
No_Reduce_Excess (=1 if the consumer cannot reduce excess)	0.40	0.49	0.50	Prohibiting excess reduction is indicative of high *ISE*.
Both CSE and ISE				
Rejected_C (=1 if the consumer changed the IC after the incident, =0 otherwise)				In principle, the change is interpreted as reflecting an action by the IC because consumers rarely wish to change insurers after an incident, but as an exception, this may occur.
High-Risk-Type (RT⁺)	0.05	0.21	0.04	
RTCs (number of accidents reported; 0, 1 or ≥2)	0.31	0.60	1.72	Reporting many RTCs (1994–1999) is indicative of a high risk-type
Parts_Damaged (count of car parts damaged)	0.07	1.72	4.63	Number of damaged car parts is positively correlated with a high-risk type
Metro (=1 if someone lives in a metropolitan region, =0 if otherwise)	0.62	0.48	−0.53	Due to the increased cost of repairs and the probability of theft, metropolitan regions are correlated with higher risk type. [d]
Young_Driver (=1 if aged <25 years, =0 if otherwise)	0.09	0.29	2.87	Young drivers are associated with more RTCs, hence the high-risk type

Notes: [a] CTP insurance is compulsory, but comprehensive insurance is optional. Many insurers offer both types of policies. [b] Insurers can design their own NCB scheme, and hence the rules and rates vary between them. [c] See Table A1, Col. 2, for data. [d] See Australian Consumer Association (ACA) (1997) pp. 6–13 for further discussion. [e] The mean of a dichotomous variable indicates the share of cases where the characteristic in question is observed (=1).

The measurement equations linking the indicators to the latent variables are given by Equations (8)–(10). Equation (8) specifies the indicators pertaining to ICs' risk selection effort *ISE* (for their explanation, see Section 4.3.1). Their so-called loadings $(\kappa_4, \kappa_5, 1)$ are all positive, with the first normalized to one to ensure identification. Their measurement errors have zero expected value and constant variance throughout, and they are assumed to be uncorrelated among themselves as well as with measurement errors pertaining to the indicators of *ISE* as well as RT^+,

$$
\begin{aligned}
&(No\_Protect\_NCB,\ No\_Reduce\_Excess,\ Premium\_Highrisks)' \\
&= (\kappa_4, \kappa_5, 1)' \cdot ISE + \left(\varepsilon_4, \varepsilon_5, \varepsilon_6\right)', \\
&\text{with} \\
&E\varepsilon_j = 0, Var\left(\varepsilon_j\right) = \theta_j^2, E\left(\varepsilon_j, \varepsilon_{i \neq j}\right) = 0, E\left(\varepsilon_j, \delta_k\right) = 0; i = 1, 2, 3; j = 1, \ldots, 3; k = 1, \ldots, 4.
\end{aligned}
\tag{8}
$$

Analogous specifications hold for the indicators of *CSE* in Equation (9) (explained in Section 4.3.2) as well as higher consumer risk RT^+ in Equation (10) (explained in Section 4.3.3). They are standard in SEM, along with the assumption that the indicators vary in a linear fashion with the latent variable, except for measurement error. This restriction can be justified by noting that if the three dummy variables (*Premium_Search*, *Insurer_Search*, *Knows_NCB*) in Equation (9) all take on the value of zero, it would be strange to argue that *CSE* nevertheless is positive.

$$
\begin{aligned}
&(Premium\_Search,\ Insurer\_Search,\ Knows\_NCB)' \\
&= (1, \kappa_2, \kappa_3)' \cdot CSE + \left(\varepsilon_1, \varepsilon_2, \varepsilon_3\right)', \\
&\text{with} \\
&E\varepsilon_i = 0, Var\left(\varepsilon_i\right) = \varsigma_i^2, E\left(\varepsilon_i, \varepsilon_{j \neq i}\right) = 0, E\left(\varepsilon_i, \delta_k\right) = 0; i = 1, 2, 3; j = 1, \ldots, 3; k = 1, \ldots, 4;
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
&(RTCs,\ Metro,\ Parts\_Damaged,\ Young\_Driver)' \\
&= (1, \lambda_2, \lambda_3, \lambda_4)' \cdot RT^+ + \left(\delta_1, \delta_2, \delta_3, \delta_4\right)', \\
&\text{with} \\
&E\delta_k = 0, Var\left(\delta_k\right) = \chi_k^2, E\left(\delta_{k \neq \updownarrow}, \delta_\updownarrow\right) = 0, E(\delta_k \varepsilon_i) = 0; E(\delta_k \varepsilon_j) = 0; \\
&k = 1, \ldots, 4; i = 1, 2, 3; j = 4, 5, 6.
\end{aligned}
\tag{10}
$$

Most of the available indicators are binary, so they depart from the normality assumption used in Maximum Likelihood (ML) estimation. Nevertheless, the ML function converged after a few iterations. To prevent overfitting and potentially committing a Type I error, the data are divided into two parts. The first (n = 2000) is used for model development, while the second (n = 2006) is reserved for an out-of-sample test. Statistics for the full dataset are reported in Table 1 (they do not differ to a noticeable degree between the two subsets).

4.3. Indicator Variables

4.3.1. Indicators of Insurer Selection Effort (ISE)

Premium-Highrisks has five levels, indicating the premium for the highest risk category relative to the lowest charged by an IC. A high value arguably reflects the IC's risk selection effort. Being quasi-continuous, this indicator qualifies as the benchmark indicator with its loading set to one.

No_Protect_NCB is a dummy variable that takes the value of one if the IC does not offer the option of protecting the no-claims bonus in the event of an accident, thus preserving the effect of the bonus to attract favorable risks.

No_Reduce_Excess is a dummy variable that takes the value of one if the IC does not offer the option of reducing the deductible, thus preserving its effect of attracting favorable risks in exchange for a low premium.

4.3.2. Indicators of Consumer Search Effort (CSE)

Premium_Search is the amount of coverage relative to the premium paid. According to the model in Section 3.1, a high value of this ratio reflects a high *CSE*. Being reported in quintiles, this indicator comes close to a continuous variable, so it qualifies as the benchmark indicator of *CSE* with its loading constrained to 1.

Insurer_Search is a dummy variable that takes the value of one if the consumer purchased comprehensive coverage from a different IC than for mandatory coverage. This entails a certain amount of searching.

Knows_NCB is a dummy variable that takes the value of one if the policyholder knows the amount of his/her no-claims bonus. This is likely to reflect the search for optimal coverage.

4.3.3. Indicators of High-Risk Type (RT+)

This variable is important because both Predictions 1 and 2 regarding *ISE* and *CSE* are conditional upon risk type. However, contrary to the theoretical argument, which distinguishes two types only for simplicity, *RT+* is continuous, with variance normalized to one. Four indicators of high-risk type were identified in the data, three of which (driver age, location, and RTC history) are frequently found in empirical analyses of asymmetric information in automobile insurance to reflect the insurer's information set, as e.g., in Chiappori and Salanié (2000) or Dionne et al. (2013).

RTCs count the number of accidents reported by the policyholder from 1994 to 1999. Being quasi-continuous (0, 1, and $\geq$2), it serves as the benchmark indicator.

Parts_Damaged counts the number of parts damaged; it arguably also reflects higher risk on the part of the driver.

Metro is a dummy variable that takes the value of one if the policyholder lives in a metropolitan area. It reflects the IC's experience that accidents happen with a higher frequency there.

Young_Driver is a dummy variable that takes the value of one if the policyholder is 25 years old or younger. It also reflects the IC's loss experience.

A simple rule of thumb proposed by Kenny (2020) states that there should be at least two indicators per latent variable. This condition is satisfied by the proposed model.

4.4. Empirical Results

The specified SEM is over-identified and therefore can be estimated using Stata's maximum likelihood function. Standard errors are assumed to be uniform across ICs and member states, taking advantage of the fact that markets for comprehensive automobile insurance are broadly homogenous across Australia (Compare the Market 2020). The correlation matrix reports a substantial number of weak but statistically significant correlations between the indicators (see Table A3). Nevertheless, convergence was achieved with both the developmental and the test samples.

The estimates derived from the developmental sample are reported in Figure 5. Starting with the theoretical core, one notes that both *CSE* and *ISE* increase significantly with *RT+.* This vindicates the crucial the Testable Prediction (Section 3.2.1), which states that *CSE* and *ISE* are high for high-risk types and low for low-risk types. As to the measurement part, all three indicators of *ISE* (*No_Protect_NCB*, *No_Reduce_Excess*, *Premium_Highrisks*) have loadings that are significantly positive; however, the measurement error contained in the benchmark indicator *Premium_Highrisks* is the highest, contrary to expectations. The three indicators of *CSE* (*Premium_Search*, *Insurer_Search*, *Knows_NCB*) also have a significant positive relationship with the latent variable, as expected. However, *Premium_Search*, which arguably should be the closest reflection of *CSE* and whose loading is therefore constrained to one, displays the highest measurement error. As to the indicators of RT+, higher risk is reflected by the four indicators (*Prior_RTC, Parts_Damaged, Metro, Young_Driver*), with the benchmark one (*Prior_RTC*, number of road traffic crashes) exhibiting a measurement error that is in line with the others. Interestingly, *Young_Driver*, which is used routinely by

ICs, turns out to be a rather weak indicator with a loading well below one; in return, its measurement error variance is very small, at least in the context of the present model.

In view of the substantial correlation coefficient between the structural error terms φ_2 and φ_1[8], there may be important determinants of *ISE* and *CSE*, respectively, that are left unaccounted for. Still, a robustness check involving different choices of the benchmark indicator does not affect the estimated relationship between *RT+*, *ISE*, and *CSE* in a material way. However, goodness of fit is poor. The comparative fit index (CFI) is zero, and the root mean square error of approximation (RMSEA) is 0.336. Furthermore, the χ^2 statistic clearly suggests rejection of the null hypothesis that the estimated model fits the data. Yet according to Kenny (2020), the χ^2 statistic is almost always significant for n > 400.

Turning to the test dataset (n = 2006), one may notice that the estimates presented in Figure 6 are very similar to those of Figure 5. In particular, the model core looks robust. In both estimates, the coefficients pertaining to the relationship between *RT+* and *ISE* and *RT+* and *CSE* are approximately 9 and 7, respectively.

Figure 6. Structural equation model (test dataset, 1999, n = 2006). Note: LR test of model vs. saturated: $\chi^2(42) = 4681.91$, Probability $> \chi^2 < 0.01$, (n = 1024).

The estimated relationship between a policyholder's risk status and consumer search effort as well as insurer selection effort confirms the Testable Prediction (Section 3.2.1). Using indicators derived from Australian auto insurance data and applying structural equation modeling, higher risk status is indeed found to be associated with increased consumer search as well as increased insurer selection effort.

5. Conclusions

5.1. The Theoretical Contribution

To the best of the authors' knowledge, the literature building on the RS model has accepted the implicit assumptions that the challenging IC does not incur any risk selection expense, while low-risk policyholders can identify preferred insurance policies without undertaking costly effort. The theoretical model developed in this paper relaxes both of these unrealistic assumptions. Although intuitively promising, the model is subject to several limitations. First, consumers are modeled as expected utility maximizers, which may serve as long as one is willing to concede that their decision-making may be beset by error (Hey 2002). Second, a one-period model of insurer behavior likely fails to fully depict the complexity of monitoring and structuring the insured population. In particular, when discarding a consumer categorized as a high-risk type, the IC has no guarantee to find a low-risk replacement, contrary to the simplified model. Finally, the existence of a separating equilibrium is taken as granted, although according to the theoretical model,

there is a very low probability that it fails to exist. Despite these limitations, pursuing the extension of the RS model put forward here may be worthwhile, paving the way to a more in-depth exploration of the RS paradigm than has hitherto been undertaken.

5.2. The Empirical Contribution

The use of structural equation modeling (SEM) for estimation is well suited to the present context. Both consumer search effort and insurer selection effort arguably constitute latent variables that are reflected by indicators, which, however, need not vary in 1:1 proportion with them and are subject to measurement errors. Rather than trying one indicator after another, as is typical in regression analysis, the SEM approach is full-information in that it permits exploiting all available indicators simultaneously. The Testable Prediction (Section 3.2.1), states that higher risk status is associated with an increase in both consumer search and insurer risk selection efforts, is supported by the evidence.

However, a limitation is that the existence of a separating equilibrium, while highly credible in view of the theoretical analysis, is not tested for. Moreover, the data analyzed is now almost 25 years old. One obvious change that has occurred since is the growth of the internet. This could have reduced policyholders' search costs but also insurers' risk selection costs. To the extent that these changes have increased the effectiveness of consumer search and/or the effectiveness of insurers' selection efforts, the estimates presented here are biased downward. Hence, contemporary markets for automobile insurance may well be characterized by an even more marked separation of risks than found here, and no direct conclusions for current public policy should be drawn.

For all its potential shortcomings, this work illustrates the value of using market-level data that captures the behaviors of policyholders and insurance firms rather than relying on claim data obtained from a single insurer. Yet future empirical research would benefit from measurements that are more closely related to the latent variables of this study. Consumer surveys reporting time spent in search of the chosen insurance policy would be valuable, as would be industry surveys reporting more detail on insurers' selection strategies. Finally, more refined indicators of risk status might allow us to directly determine the two risk types distinguished in the theoretical analysis.

Appendix A.

This Appendix is devoted to the derivation of the reaction functions displayed in Figure 2 of the text.

Appendix A.1. The Insurer's Reaction Function

Let the optimum condition (2) of the text be disturbed by an increase in consumers' search effort $dc > 0$. Note that it is sufficient to derive only one reaction function because the IC cannot distinguish between risk types; this becomes only possible due to the consumers' type-specific reaction functions resulting in different Nash equilibria. This gives rise to the comparative static equation (applying the implicit function theorem),

$$\frac{\partial^2 E\Pi}{\partial e^2} de + \frac{\partial^2 E\Pi}{\partial e \partial c} dc = 0, \tag{A1}$$

which can be solved to obtain

$$\frac{de}{dc} = -\frac{\partial^2 E\Pi / \partial e \partial c}{\partial^2 E\Pi / \partial e^2} \tag{A2}$$

From Equation (2), the solutions to the comparative-static equation are given by

$$\frac{de}{dc} \propto \frac{\partial^2 E\Pi}{\partial e \partial c} = \partial \pi / \partial e \cdot \left[-\frac{\partial EI}{\partial c} \right] > 0. \tag{A3}$$

The IC reaction function is exhibited in Figure 2. It is drawn linear for simplicity because on the one hand $|\partial \pi / \partial e|$ decreases with e, implying a decreasing positive slope; on the other hand, $\partial^2 I / \partial c^2 > 0$ is a possibility, which *per se* would imply an increasing positive slope.

Appendix A.2. Consumers' Reaction Functions

Here, the exogenous shock is $de > 0$, an increase in the IC's risk selection effort. In analogy to Equation (A1), one obtains from Equations (4a) and (4b) of the text,

$$\frac{dc^H}{de} \propto \frac{\partial^2 EU^H}{\partial c \partial e} = \rho^H \left[v^{H''} \frac{\partial I^H}{\partial c^H} + v^{H'} \frac{\partial^2 I^H}{\partial c^H \partial e} \right] = \rho^H v^{H'} \left[\frac{v^{H''}}{v^{H'}} \frac{\partial I^H}{\partial c^H} + \frac{\partial^2 I^H}{\partial c^H \partial e} \right] < 0; \tag{A4}$$

$$\frac{dc^L}{de} \propto \frac{\partial^2 EU^L}{\partial c \partial e} = \rho^L \left[v^{L''} \frac{\partial I^L}{\partial c^L} + v^{L'} \frac{\partial^2 I^L}{\partial c^L \partial e} \right] = \rho^L v^{L'} \left[\frac{v^{L''}}{v^{L'}} \frac{\partial I^L}{\partial c^L} + \frac{\partial^2 I^L}{\partial c^L \partial e} \right] < 0. \tag{A5}$$

It can be realistically assumed that the marginal effectiveness of consumer search is lowered by the IC's selection effort, implying $\partial^2 I^H / \partial c^H \partial e < 0$, $\partial^2 I^L / \partial c^L \partial e < 0$. In addition, the low-risk type's coefficient of absolute risk aversion, $RA^L = -v^{L''} / v^{L'} > 0$ is unlikely to be smaller than that of the high-risk type $RA^H = -v^{H''} / v^{H'}$; therefore, one has $|dc^H / de| < |dc^L / de|$ since $\rho^H > \rho^L$. Regardless of risk type, consumers are predicted to decrease search effort because they are burdened with additional transaction cost (e.g., the IC may require more forms regarding risk status), with the response of the high-risk type less marked than that of the low-risk type.

Appendix B.

Table A1. Insurer selection effort: Pricing of high, medium and low-risk Scenarios.

Insurers: New South Wales	High-Risk Scenario		Medium-Risk Scenario		Low-Risk Scenario	
	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area
AAMI	2	3	4	5	2	3
ANSVAR	3	3	2	3	3	3
Australian Alliance	.	.	2	3	4	5
Commercial Union	3	5	3	5	3	4
Direcdial	4	3	3	2	3	2

Table A1. *Cont.*

Insurers: New South Wales	High-Risk Scenario		Medium-Risk Scenario		Low-Risk Scenario	
	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area
FAI car	2	2	3	2	2	2
Mercantile Mutual	4	2	5	3	5	2
NRMA	2	2	3	3	3	2
NZI Comprehensive	3	3	3	3	2	3
NZI Top Cover	3	3	3	4	3	4
QBE	5	3	4	3	4	3
Suncorp	2	2	2	2	3	2
SWANN Agreed value	4	4	2	2	2	3
TII	3	4	.	2	.	3
Zurich Personal Assistance	2	2	4	3	3	3

Insurers: Queensland	High-Risk Scenario		Medium-Risk Scenario		Low-Risk Scenario	
	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area
AAMI	3	4	5	5	3	3
ANSVAR	2	2	2	2	3	3
Australian Alliance	.	.	2	2	2	2
Commercial Union	4	5	4	4	5	5
Direcdial	4	3	2	2	3	2
FAI car	2	2	3	2	2	2
Mercantile Mutual	2	3	3	3	2	3
NRMA	2	2	3	3	2	2
NZI Comprehensive	3	3	3	3	3	3
NZI Top Cover	3	3	4	4	5	5
QBE	5	5	4	4	3	4
RACQ	2	2	2	2	3	3
Suncorp	2	2	3	3	3	3
SWANN Agreed value	4	4	2	2	4	3
TII	3	3	2	2	2	2
TIO	.	.	.	.	.	.
Zurich Personal Assistance	2	2	4	3	4	3

Insurers: Victoria	High-Risk Scenario		Medium-Risk Scenario		Low-Risk Scenario	
	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area
AAMI	4	4	2	2	3	2
ANSVAR	2	2	2	2	2	3
Australian Alliance	.	.	.	.	2	2
Commercial Union	3	3	4	4	4	3
Direcdial	4	3	3	2	3	2
FAI car	2	2	4	3	3	2
HBF	.	.	.	.	.	.
Mercantile Mutual	2	2	3	3	2	2
NRMA	2	2	4	4	3	3
NZI Comprehensive	3	3	4	3	3	2
NZI Top Cover	3	3	5	5	4	4
QBE	3	4	3	3	2	3
RACV Fair Deal	2	3	2	3	4	4
RAC (WA) Motorguard	.	.	.	.	.	.
SIGO	.	.	.	.	.	.
SWANN Agreed value	5	5	4	4	5	5
TII	3	4	2	3	2	4
Western QBE	.	.	.	.	.	.
Zurich Personal Assistance	2	2	4	4	2	3

Insurers: Western Australia	High-Risk Scenario		Medium-Risk Scenario		Low-Risk Scenario	
	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area
AAMI	.	.	.	.	.	.
ANSVAR	3	3	3	3	3	3
Australian Alliance	.	.	2	3	4	4
Commercial Union	4	5	5	5	5	5
Direcdial	4	3	2	2	2	2
FAI car	2	2	3	3	2	2
HBF	2	2	2	3	3	3
Mercantile Mutual	2	3	3	4	2	3

Table A1. *Cont.*

Insurers: Western Australia	High-Risk Scenario		Medium-Risk Scenario		Low-Risk Scenario	
	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area	High-Risk Area	Low-Risk Area
NRMA	.	.	.	.	.	.
NZI Comprehensive	3	3	4	4	3	3
NZI Top Cover	4	3	5	5	5	5
QBE	.	.	.	.	.	.
RACV Fair Deal	.	.	.	.	.	.
RAC (WA) Motorguard	3	3	3	3	4	3
SIGO	2	2	2	2	2	2
SWANN Agreed value	5	5	2	2	3	3
TII	4	3	2	2	3	2
Western QBE	2	2	2	2	2	2
Zurich Personal Assistance	2	2	4	3	3	2

Note: Policies were rated for affordability from 1 star (most expensive) thru to 5 stars (cheapest). Source: Australian Consumer Association (ACA) (1997).

Table A2. Insurer selection effort: Policy exclusions.

Insurance Company	States Available	Reduction of NCB	Protection of NCB	Reduce Excess
Australian Alliance	All but NT	1	1	1
Australian Pensions	All but NT	1	1	1
RACT	Tas.	2	1	1
NRMA	ACT, NSW, Vic.	2	1	1
CGU	All but NT	2	1	1
FAI	All but NT	2	1	Not in Qld.
RACQ-GIO	Qld.	1	1	2
TII	All but NT	2	1	1
AAMI	All but WA	1	1	2
EIG-ANSVAR	All	2	1	1
RAA-GIO	SA	1	1	1
COTA	All	1	1	Not in Qld.
HBF	WA	2	1	1
Suncorp-Metway	Qld.	1	2	1
SWANN	All	2	1	2
Mutual Community	SA	2	1	1
Western QBE	All but NT	2 or 1 if < USD 1000	1	1
Directdial	All but NT	2	1	2
HBA	Vic.	Depends on NCB	1	1
GIO	Vic.	2	1	1
SGIC	SA	2	1	1
AMP	All	2	1	1
TIO	NT	2	1	1
RACV (E.)	Vic.	2	1	2
RAC Motor guard	WA	2	1	1
TGIO	Tas.	2	1	1
GIO	ACT, NSW	2	1	1
GIO	NT	2	2	2
SGIO	WA	3	1	1
GIO	WA	2	1	1
AMP car insurance Options	All	2	1	1
GIO Rode Cover Basic	Vic.	2	1	2
		=1 level.	1 = Yes	1 = Yes
		=2 levels	2 = No	2 = No
		=3 levels		

Source: Australian Consumer Association (ACA) (1999).

Table A3. Correlation matrix, full sample (n = 4006).

		CSE				ISE			Risk Type		
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
CSE	(1) Premium_Search	1.000									
	(2) Insurer_Search	0.030 *	1.000								
	(3) Knows_NCB	0.043 *	−0.015	1.000							
	(4) Premium_Highrisks	−0.038 *	0.020	0.034 *	1.000						
ISE	(5) No_Protect_NCB	0.082 *	−0.040 *	−0.049 *	0.137 *	1.000					
	(6) No_Reduce_Excess	0.035 *	0.048 *	−0.056 *	−0.351 *	−0.229 *	1.000				
	(7) RTCs	−0.021	−0.022	0.009	0.026	−0.006	−0.024	1.000			
Risk Type	(8) Parts_Damaged	−0.028	0.004	0.009	0.017	0.012	−0.028	0.487 *	1.000		
	(9) Metro	−0.136 *	−0.111*	0.002	−0.039 *	−0.154 *	0.044 *	0.080 *	0.082 *	1.000	
	(10) Young_Driver	−0.134 *	−0.004	−0.129 *	0.003	0.051 *	0.025	0.060 *	0.045 *	0.030 *	1.000

Notes: * denotes *p*-value < 0.1.

Table A4. Structural equation model (full dataset).

		Coeff.	Std. Err.	z	P > |z|	[95% CI]	
Structural Model							
CSE							
	HIGH_RT	7.001	0.286	24.49	<0.01	6.444	7.565
ISE							
	HIGH_RT	9.159	0.360	25.45	<0.01	8.453	9.864
Measurement Model							
Insurer_Search							
	CSE	0.083	0.003	24.94	<0.01	0.079	0.090
	constant	0	constrained				
Premium_Search							
	CSE	1	constrained				
	constant	0	constrained				
Knows_NCB							
	CSE	0.319	0.004	82.72	<0.01	0.312	0.327
	constant	0	constrained				
No_Protect_NCB							
	ISE	0.021	0.002	12.29	<0.01	0.017	0.024
	_cons	0	constrained				
No_Reduce_Excess							
	ISE	0.115	0.003	36.52	<0.01	0.109	0.122
	constant	0.000	constrained				
Premium_Highrisks							
	ISE	1	constrained	12.2	<0.01	40.612	56.161
	constant	0	constrained				
RTCs							
	HIGH_RT	1	constrained				
	constant	0	constrained				
Metro							
	HIGH_RT	1.866	0.073	25.5	<0.01	1.722	2.009
	constant	0	constrained				
Parts_Damaged							
	HIGH_RT	2.494	0.141	17.72	<0.01	2.21	2.77
	constant	0	constrained				
Young_Driver							
	HIGH_RT	0.183	0.016	11.13	<0.01	0.151	0.215
	constant	0	constrained				
var(e.Insurer_Search)		0.178	0.006			0.167	0.189
var(e.Premium_Search)		1.859	0.065			1.735	1.991
var(e.Knows_NCB)		0.063	0.004			0.056	0.070
var(e.No_Protect_NCB)		0.076	0.002			0.071	0.081
var(e.No_Reduce_Ecess)		0.258	0.008			0.242	0.274
var(e.Premium_Highrisks)		1.001	0.107			0.811	1.235
var(e.Prior_RTC)		0.129	0.004			0.121	0.137
var(e.Metro)		0.167	0.012			0.146	0.192
var(e.Parts_Damaged)		3.472	0.111			3.261	3.698
var(e.Young_Driver)		0.059	0.002			0.055	0.063
var(e.CSE)		0.010	0.001			0.007	0.013
var(e.ISE)		0.001	0.000			0.000	0.001
var(HIGH_RT)		0.029	0.003			0.024	0.036
cov(e.CSE,e.ISE)		0.002723	0.0004365	6.24	0	0.001868	0.003579

Notes

[1] Mimra and Wambach (2014) provide an excellent summary of the literature that has reviewed by Rothschild and Stiglitz (1976).

[2] For example, in markets for health insurance empirical research has reported that ICs are able to control adverse selection (Pauly et al. 2007; Marton et al. 2015). However, Cutler and Reber (1998) found that comprehensive health insurance coverage sponsored by Harvard University had to be withdrawn from the market; they interpreted this as evidence of a "death spiral" Frech III and

Smith (2015) do find evidence suggesting a "death spiral"; however, the spiral moves so slowly as to give ICs plenty of time to withdraw loss-making contracts.

[3] Whereas search theory deals with optimal stopping rules in the presence of imperfect information and its implications especially in labor markets [see e.g., Shi (2008)], the focus here lies on the outcome of search in terms of a favorable premium-coverage ratio.

[4] The notation is in accordance with Stage 2 of the game (see Figure 1), where consumers are still confronted with one level of IC risk selection effort.

[5] Conceivably, the marginal benefit of search effort could fall short of its marginal cost of one right away, resulting in no purchase of insurance.

[6] For a more detailed description of the IMRAS data set, interested readers are directed to the papers "Two tests for *ex ante* moral hazard in a market for automobile insurance" (Rowell et al. 2017) and "Empirical tests for *ex post* moral hazard in a market for automobile insurance" (Rowell et al. 2022).

[7] Making risk type dichotomous would call for latent class modeling, which however would put a heavy extra burden on SEM both in terms of identification and estimation (see e.g., Clark (2022)).

[8] In Stata, the estimate of $\beta = 1.94$ relates to a regression of *ISE* on *CSE*. Using the formula, $\rho_{x,y} = \sigma_{x,y}/(\sigma_x \cdot \sigma_y) = \sigma_{x,y}/\sigma_x^2(\sigma_x/\sigma_y) = \beta(\sigma_x/\sigma_y)$ and noting that $\hat{E}\sigma(ISE) = 9.11 \cdot \hat{E}\sigma(RT^+) = 9.11$ because $\sigma(RT^+)$ is normalized to one, the estimated value of $\rho_{x,y}$ becomes $1.94/9.11 = 0.28$.

References

Ales, Laurence, and Pricilia Maziero. 2014. *Adverse Selection and Non-Exclusive Contracts*. Working Paper. Pittsburgh: Carnegie Mellon University, Temper School of Business.

Arrow, Kenneth. 1970. Political and economic evaluation of social effects and externalities. In *The Analysis of Public Output*. Edited by Julius Margolis. Cambridge, MA: NBER, pp. 1–30.

Asheim, Geir B., and Tore Nilssen. 1996. Non-discriminating renegotiation in a competitive insurance market. *European Economic Review* 40: 1717–1736. [CrossRef]

Attar, Andrea, Thomas Mariotti, and François Salanié. 2011. Nonexclusive Competition in the Market for Lemons. *Econometrica* 79: 1869–1918. [CrossRef]

Attar, Andrea, Thomas Mariotti, and François Salanié. 2014. Nonexclusive competition under adverse selection. *Theoretical Economics* 9: 1–40. [CrossRef]

Attar, Andrea, Thomas Mariotti, and François Salanié. 2016. *Multiple Contracting in Insurance Markets*. TSE Working Paper n. 14-532. Toulouse: Toulouse School of Economics.

Attar, Andrea, Thomas Mariotti, and François Salanié. 2020. The Social Costs of Side Trading. *The Economic Journal* 130: 1608–1622. [CrossRef]

Australian Consumer Association (ACA). 1997. Car insurance. *CHOICE Magazine*, July. 6–13.

Australian Consumer Association (ACA). 1999. Your car insurance toolkit. *CHOICE magazine*, August. 18–29.

Chiappori, Pierre-André. 1999. Asymmetric information in automobile insurance: An overview. In *Automobile Insurance: Road Safety, New Drivers, Risks, Insurance Fraud and Regulation*. Boston/Dordrecht/London: Kluwer Academic Publishers, pp. 1–11.

Chiappori, Pierre-André, and Bernard Salanié. 2000. Testing for asymmetric information in insurance markets. *Journal of Political Economy* 108: 56–78. [CrossRef]

Choice. 2024. CHOICE. Available online: https://www.choice.com.au/ (accessed on 15 January 2024).

Clark, M. 2022. Structural Equation Modeling; Mixture Models. Available online: https://m-clark.github.io/sem/mixture-models.html (accessed on 20 March 2024).

Compare the Market. 2020. Car insurance in Australia. Available online: https://www.comparethemarket.com.au/car-insurance/ Australia/ (accessed on 20 March 2024).

Consumer Reports. 2019. Homeowners Insurance. Available online: https://www.consumerreports.org/cro/homeowners-insurance. htm (accessed on 20 March 2024).

Cutler, David M., and Sarah J. Reber. 1998. Paying for Health Insurance: The Trade-Off between Competition and Adverse Selection. *The Quarterly Journal of Economics* 113: 433–466. [CrossRef]

Dionne, Georges, and Neil A. Doherty. 1994. Adverse Selection, Commitment, and Renegotiation: Extension to and Evidence from Insurance Markets. *Journal of Political Economy* 102: 209–235. [CrossRef]

Dionne, Georges, Gouriéroux Christian, and Charles Vanasse. 2001. Testing for Evidence of Adverse Selection in the Automobile Insurance Market: A Comment. *Journal of Political Economy* 109: 444–453. [CrossRef]

Dionne, Georges, Pierre-Carl Michaud, and Maki Dahchour. 2013. Separating moral hazard from adverse selection and learning in automobile insurance: Longitudinal evidence from france. *Journal of the European Economic Association* 11: 897–917. [CrossRef]

Engers, Maxim, and Luis Fernandez. 1987. Market Equilibrium with Hidden Knowledge and Self-Selection. *Econometrica* 55: 425. [CrossRef]

Fan, Yi, Jiquan Chen, Gabriela Shirkey, Ranjeet John, Susie R. Wu, Hogeun Park, and Changliang Shao. 2016. Applications of structural equation modeling (SEM) in ecological studies: An updated review. *Ecological Processes* 5: 1–12. [CrossRef]

Frech III, H. E., and Michael P. Smith. 2015. Anatomy of a Slow-Motion Health Insurance Death Spiral. *North American Actuarial Journal* 19: 60–72. [CrossRef]

Halek, Martin, and Joseph G. Elisenhauer. 2001. Demography of Risk Aversion. *The Journal of Risk and Insurance* 68: 1–24. [CrossRef]

Hellwig, Martin. 1987. Some recent developments in the theory of competition in markets with adverse selection. *European Economic Review* 31: 319–325. [CrossRef]

Hey, John D. 2002. Experimental Economics and the Theory of Decision Making Under Risk and Uncertainty. *The Geneva Risk and Insurance Review* 27: 5–21. [CrossRef]

Jaynes, Gerald David. 1978. Equilibria in monopolistically competitive insurance markets. *Journal of Economic Theory* 19: 394–422. [CrossRef]

Jaynes, Gerald D. 2011. *Equilibrium and Strategic Communication in the Adverse Selection Insurance Model*. Economics Department Working Paper No. 91. New Haven: Yale University. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1865367 (accessed on 20 March 2024).

Jöreskog, Karl G., and Arthur S. Goldberger. 1975. Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association* 70: 631–9.

Kenny, David A. 2020. Measuring Model Fit. Available online: http://davidakenny.net/cm/fit.htm (accessed on 20 March 2024).

Kline, Rex B. 2016. *Principles and Practice of Structural Equation Modeling*, 4th ed. New York: New York Guilford Press.

Kosenko, Andrew, Joseph Stiglitz, and Jungyoll Yun. 2023. Bilateral information disclosure in adverse selection markets with nonexclusive competition. *Journal of Economic Behavior & Organization* 205: 144–168. [CrossRef]

Liferay. 2019. 20 Must-know stats for Insurers in 2022. Available online: https://www.liferay.com/blog/customer-experience/20-must-know-stats-for-insurers-in-2022 (accessed on 20 March 2024).

Marton, James, Patricia G. Ketsche, Angela Snyder, E. Kathleen Adams, and Mei Zhou. 2015. Estimating Premium Sensitivity for Children;s Public Health Insurance Coverage: Selection but No Death Spiral. *Health Services Research* 50: 579–98. [CrossRef]

Mathews, Martyn. 2022. Insurance Quote Manipulation Flagged as Research Reveals Half of U.K. Consumers Think It's Fine to Fib. Available online: https://risk.lexisnexis.co.uk/about-us/press-room/press-release/20220511-quote-intelligence (accessed on 20 March 2024).

Mimra, Wanda, and Achim Wambach. 2014. New Developments in the Theory of Adverse Selection in Competitive Insurance. *The Geneva Risk and Insurance Review* 39: 136–152. [CrossRef]

Netzer, Nick, and Florian Scheuer. 2014. A game theoretic foundation of competitive equilibria with adverse selection. *International Economic Review* 55: 399–422. [CrossRef]

Pauly, Mark V. 1974. Overinsurance and Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection. *The Quarterly Journal of Economics* 88: 44–62. [CrossRef]

Pauly, Mark V., Olivia S. Mitchell, and Yuhui Zeng. 2007. Death Spiral or Euthanasia? The Demise of Generous Group Health Insurance Coverage. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing* 44: 412–27. [CrossRef] [PubMed]

Puelz, Robert, and Arthur Snow. 1994. Evidence on Adverse Selection: Equilibrium Signaling and Cross-Subsidization in the Insurance Market. *Journal of Political Economy* 102: 236–257. [CrossRef]

Riley, John G. 1979. Informational equilibrium. *Econometrica* 47: 331–59. [CrossRef]

Rothschild, Michael, and Joseph Stiglitz. 1976. Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *The Quarterly Journal of Economics* 90: 629. [CrossRef]

Rothschild, Michael, and Joseph E. Stiglitz. 1997. Competition and insurance twenty years later. *The Geneva Papers on Risk and Insurance-Theory* 22: 73–79. [CrossRef]

Rowell, David, Son Nghiem, and Luke B Connelly. 2017. Two Tests for *Ex Ante* Moral Hazard in a Market for Automobile Insurance. *Journal of Risk and Insurance* 84: 1103–26. [CrossRef]

Rowell, David, Son Nghiem, and Luke B Connelly. 2022. Empirical Tests for *Ex Post* Moral Hazard in a market for Automobile Insurance. *Annals of Actuarial Science* 16: 243–260. [CrossRef]

Salanié, Bernard. 2017. Equilibrium in Insurance Markets: An Empiricist's View. *The Geneva Risk and Insurance Review* 42: 1–14. [CrossRef]

Shi, Fei. 2008. Endogenous Timing with Demand Uncertainty. Working paper No. 30. Thurgau Institute of Economics and Department of Economics, University of Konstanz. Konstanz. Available online: http://nbn-resolving.de/urn:nbn:de:bsz:352-opus-59885 (accessed on 20 March 2024).

Spence, Michael. 1978. Product differentiation and performance in insurance markets. *Journal of Public Economics* 10: 427–47. [CrossRef]

Stiglitz, Joseph, Jungyoll Yun, and Andrew Kosenko. 2017. Equilibrium in a Competitive Insurance Market under Adverse Selection with Endogenous Information. Working Paper No. w23556. National Bureau of Economic Research. Cambridge, Massachusetts. Available online: http://www.nber.org/papers/w23556 (accessed on 20 March 2024).

Ullmann, Jodie B., and Peter M. Bentler. 2004. Structural Equation Modeling. In *Handbook of Data Analysis*. Edited by M. Hardy and A. Bryman. London: Sage Publications, pp. 431–58.

Wilson, Charles. 1977. A model of insurance markets with incomplete information. *Journal of Economic Theory* 16: 167–207. [CrossRef]

Article

Using the Capital Asset Pricing Model and the Fama–French Three-Factor and Five-Factor Models to Manage Stock and Bond Portfolios: Evidence from Timor-Leste

Fernando Anuno [1,2,*], Mara Madaleno [2] and Elisabete Vieira [3]

[1] Faculty of Economics and Management, National University of Timor Lorosa'e (UNTL),
Díli 10000, Timor-Leste
[2] GOVCOPP—Research Unit on Governance, Competitiveness and Public Policies, Department of Economics,
Management, Industrial Engineering and Tourism(DEGEIT), University of Aveiro, Campus Universitário de
Santiago, 3810-193 Aveiro, Portugal; maramadaleno@ua.pt
[3] GOVCOPP Unit Research, Aveiro Institute of Accounting and Administration, University of
Aveiro (ISCA-UA), Campus Universitário de Santiago, 3810-902 Aveiro, Portugal; elisabete.vieira@ua.pt
* Correspondence: fernando.anuno@ua.pt

Abstract: Timor-Leste is a new country still in the process of economic development and does not yet have a capital market for stock and bond investments. These two asset classes have been invested in international capital markets such as the US, the UK, Japan, and Europe. We examine the performance of the capital asset pricing model (CAPM) and the Fama–French three-factor and five-factor models on the excess returns of Timor-Leste's equity and bond investments in the international market over the period 2006 to 2019. Our empirical results show that the market factor (MKT) is positively and significantly associated with the excess returns of the CAPM and the Fama–French three-factor and five-factor models. Moreover, the two variables Small Minus Big (SMB) as a size factor and High Minus Low (HML) as a value factor have a negative and significant effect on the excess returns in the Fama–French three-factor model and five-factor model. Further analysis revealed that the explanatory power of the Fama–French five-factor model is that the Robust Minus Weak (RMW) factor as a profitability factor is positively and significantly associated with excess returns, while the Conservative Minus Aggressive (CMA) factor as an investment factor is insignificant.

Keywords: CAPM; Fama–French three-factor model; Fama–French five-factor model; emerging market

1. Introduction

Modern portfolio theory was initiated by Markowitz (1952) by developing important ideas about portfolios, risk, and diversification concerning different asset classes. Based on this principle, Lintner (1965), Mossin (1966), and Sharpe (1964) are considered pioneers and developers of the concept of the CAPM. The model explains the linear relationship between the systematic risk coefficient, beta, and expected stock returns (Wang et al. 2017; Anjum and Rajput 2021; Taussig 2022). Moreover, the basic concept of the CAPM is a metric that explains expected excess return, beta risk, and the market risk premium by calculating the difference between an asset's return and the risk-free rate. Therefore, beta risk is generally estimated using a linear regression model (Yamaka and Phadkantha 2021).

Over the past two decades, researchers have used the Fama–French model to estimate cross-sectional stock returns using market premium risk factors, size factors, and value factors (Fama and French 1993). Also, Fama and French (2015) proposed a five-factor model that adds two new factors to the three-factor model to capture profitability and investment. Their study was an important contribution to the development of a multifactor model for asset valuation. However, their research results remain inconsistent in explaining the abnormal excess return associated with firm size, book-to-market ratio, liquidity, price–earnings ratio, cash flow–price ratio, return on equity, volatility, and return reversal (Zhang

and Lence 2022). However, Fama and French (2017) found that average stock returns for North America, Europe, and Asia–Pacific improved with the book-to-market (B/M) ratio and profitability and were negatively related to investment. For Japan, the relationship between average returns and the B/M ratio is strong, but average returns show little relationship with profitability and investment. Furthermore, the inclusion of a momentum factor was proposed by Fama and French (2018) as an extension of a six-factor model used to assess investment risk. The researchers conducted an analysis of several metrics used to evaluate asset pricing models and identified three specific challenges associated with the six-factor model. These challenges include (1) the dilemma of choosing either cash profitability or operating profitability as variables for constructing profitability factors, (2) the decision process in choosing between long–short spread factors and excess returns, and (3) the comparison between factors based on small or large stocks and factors that include both categories. In addition, after more than two decades, in a recent paper, Fama and French (2020) used the Fama–MacBeth cross-sectional factor, originally introduced by Fama and MacBeth (1973) in the context of a time series model developed in the field of asset price research. Fama and French (2020) argue that the inclusion of cross-sectional factor returns in a time series factor model leads to better results than a time series factor return model.

Given the gaps in the CAPM model and the Fama–French model, it is, therefore, an incentive for researchers to continue to conduct extensive research, such as the study by López-García et al. (2021), which extended Fama and French (1993) and Fama and French (2015) and found that the significance level is similar to that of the capitalization factor (SMB) and that the book-to-market factor (HML) is even larger than the momentum factor (MOM). Thus, market factors with equally weighted portfolios are very significant in the model, while market factors calculated as capital-weighted portfolios (in this case, the S&P500 index) are almost irrelevant in the model. Moreover, Jareño et al. (2020) concluded that the CMA and RMW factors have a negative sign across all periods and quantiles, thus negatively affecting financial institution returns. In addition, Mosoeu and Kodongo (2020) documented the following observations: (1) There is a variable relationship between average returns and SMB, value (B/M), profitability (P), and investment (INV) by market, although the factors are not consistent across portfolio types. (2) Overall, factors in the market and factors in different markets tend to have low correlation. However, there is a high correlation between the same pairs of factors constructed using different portfolio types, suggesting that different portfolio sorting strategies provide consistent information. (3) In addition, large companies tend to have better average stock returns than small companies, and aggressive companies that buy more assets tend to have better returns than cautious companies.

In addition, Hung et al. (2019) showed that the relationship between size risk and stock returns for small companies was also significantly positive, which was in contrast to the negative relationship for large companies. Moreover, the HML factor was negatively and positively correlated with returns for large and small companies, respectively. Bank and Insam (2019) found that the contribution of risk premium is not correlated with the excess return factor and captures the isolated compensation of a particular risk factor. Likewise, the contribution of risk premium shows a negative shift after 1993. Additionally, Shaikh et al. (2019) indicated that size premium positively and significantly explains the cross-section of stock returns of small companies, while value premium positively and significantly explains the cross-section of returns of quality companies.

The CAPM, the Fama–French three-factor model, and the Fama–French five-factor model in portfolio investment studies are empirically less effective in explaining maximum return investments. Thus, our motivation for this paper is to (i) fill the gaps in the existing literature on stock and bond investing using Fama and French's (1993) three-factor model and Fama and French's (2015) five-factor model. The present article also aims to add additional knowledge to other studies that have used capital market data from four regions, (1) North America, (2) Japan, (3) Asia–Pacific, and (4) Europe, which have documented

their empirical research findings in different regions with different results, such as the study by Nichol and Dowling (2014), showing that Fama–French five-factor profitability offers the greatest potential when implemented in the market of the United Kingdom. In addition, Chai et al. (2019) found that the SMB and HML factors are insignificant, but the HML factor is excessively high for the United States. However, these factors are also important for stock prices in Australia, suggesting that the five-factor model should at least be considered as a reference model for the Australian market. Similarly, Pandey and Joshi (2021) stated in their results that the CAPM seems to be a good model for explaining the returns for Italy and Spain. The Fama–French three-factor model and the Fama–French five-factor model seem to better describe returns in Germany, while the multifactor model plays a limited role in explaining returns in France. Meanwhile, multifactor models play a role in explaining returns for the Western European market, with the sole exception of France, where they appear to be ineffective in explaining returns. In addition, Roy (2021) found that six factors produced better estimates, outperforming Fama and French's three-factor model, Carhart's four-factor model, and Fama and French's five-factor model alike. However, Fama and French (2012) stated that integrated pricing across regions did not find strong support in their tests. Nonetheless, only local models using local explanatory returns describe the average return for a portfolio by size and value versus growth. For example, further research by Fama and French (2017) found that the average returns of stocks in North America, Europe, and Asia–Pacific show a positive correlation with the book-to-market ratio (B/M) and profitability while showing a negative relationship with investment. In addition, the underperformance of small stocks, characterized by low profitability but high investment, is not taken into account. (ii) As a developing country, Timor-Leste has assets such as stocks and bonds that it can invest in the local capital market. However, when Timor-Leste did not have a local capital market, it had to invest its capital in the international markets, particularly in the form of shares of 1775 companies that invested in petroleum funds in the United States, Europe, the United Kingdom, Japan, Australia, Canada, and Norway. (iii) The total investment in equities amounted to USD 6541 million (Timor-Leste Ministry of Finance 2019), which is associated with high investment risk. Therefore, as emerging economies have not reached their true economic potential through diversification, they rely on foreign investors and the transfer of inflows to these economies (Lone et al. 2021). (iv) The capital investment of oil funds is found to be the maximum cumulative return to bring further capital into economic diversification, especially the contribution of revenues from the non-oil-and-gas sector due to the dependence on oil and gas.

Timor-Leste is a new country preparing for economic development through portfolio investment. Timor-Leste's main income comes from oil and gas. The Petroleum Fund was established in 2005 to collect Timor-Leste's petroleum revenues from the Timor Sea. As a sovereign wealth fund, the Petroleum Fund is therefore subject to legal restrictions. In addition, the petroleum funds are invested in the bond and equity markets. These portfolio investments generate profits and increase the Petroleum Fund's income in addition to domestic revenues (Doraisami 2018; John et al. 2020; Zaimovic et al. 2021).

In portfolio investment, the objective is to make a profit, but the investor (the government) is exposed to risk. This risk means that the government must understand the concept of the Fama–French model to make investment decisions efficiently. This is because investment decisions must be based on the principle of optimizing profit and minimizing risk. This risk requires the government to exercise prudence in optimizing profits through portfolio diversification. Thus, the purpose of portfolio diversification is to minimize risk by spreading assets across different asset classes (Zaimovic et al. 2021; Sahabuddin et al. 2022).

The concepts of the CAPM and the Fama–French model provide an important understanding for managers (the government) to determine the performance of small-company stocks and large-company stocks (SMB). Similarly, it is important to understand the performance of stocks with the highest value (value) and stocks with the lowest value (growth) (HML). In addition, the performance factor is shares of companies with good profits (ro-

bust) and shares of companies with weak operating profits (RMW). Similarly, the factors reflect the equity performance of companies with conservative and aggressive investment policies (CMA) (Fama and French 1993, 2015, 2017; Ali et al. 2021; Ryan et al. 2021; Taib and Benfeddoul 2023).

Thus, understanding the performance of the CAPM and the Fama–French model helps the government of Timor-Leste to identify potential investment opportunities to enhance portfolio returns. In this way, the government can make efficient investment decisions based on accurate data and information on optimal portfolio diversification. In addition, portfolio investments ensure the country's long-term financial stability, provide effective risk management, and contribute to the country's long-term development and prosperity through economic diversification (Lopes 2021; Scheiner 2021).

This paper makes two important contributions to the finance literature, particularly in portfolio investment. Our first contribution is to provide new empirical evidence to fill the gap in the finance literature that has existed since the development of the CAPM by Sharpe (1964) and Lintner (1965) and its subsequent evolution into the three-factor model of Fama and French (1993) and the five-factor model proposed later by Fama and French (2015). Second, to the best of our knowledge, this paper is the first attempt to use a new dataset from a new country still in the process of economic development, namely Timor-Leste, to examine the performance of the CAPM, Fama and French's three-factor model, and the five-factor model on the excess returns in the context of equity and bond investment in the international market. The results of Fama and French's three-factor model and five-factor model suggest that the market factor (MKT) risk has a positive effect on assets excess return considering the CAPM model, the three-factor model, and Fama and French's five-factor model. Meanwhile, the size and investment value have a significant negative effect on the excess return in the three-factor model and Fama and French's five-factor model. The probability factor has a significant positive effect on the excess return, while the investment factor has a negative but insignificant effect on excess returns. Thus, the key question of the present study is: Do the CAPM model, the Fama–French three-factor model, and the Fama–French five-factor model show significant differences in assessing the risks and potential returns of Timor-Leste equity and bond investments in international markets?

The remainder of this article is organized as follows: Section 2 presents a brief literature review; Section 3 exposes the data and methodology; Section 4 is dedicated to the empirical results; and Section 5 is the results discussion. Finally, Section 6 provides the conclusions and policy implications and suggests future research.

2. Literature Review

2.1. CAPM Model

The goal of investors in asset allocation is to maximize profits while minimizing risk. According to Saiti et al. (2020), investment consists of allocating financial resources among different classes of assets, including commodities, real estate, stocks, and bonds in domestic and international markets, benefiting from diversification strategies. These investments aim to increase wealth as the most important resource. In addition, investors have decision-making principles to minimize investment risk. One of the most important strategies of fund managers is the diversification of investments to reduce this risk. Thus, modern portfolio theory states that the portfolio option prioritizes expected returns over risk mitigation.

Several empirical studies by Sharma and Vipul (2018), Silva et al. (2020), Yunus (2020), and Dichtl et al. (2021) on the allocation of financial assets (e.g., gold, stocks, bonds, and real estate) show that funds can be fixed income, stocks or net asset values, multiple markets or currencies, and commodities. Researchers are increasingly trying to figure out what impact stocks have on other asset classes such as currencies, fixed income, and commodities. This proves that there is a value and momentum premium in currencies, government bonds, and commodities as well as equities to predict trading returns in global equities, global bonds, currencies, and commodities (Bartram et al. 2021).

In measuring the performance of funds, asset management is very important to understand the systematic risk factors and actively manage the funds. For example, a traditional stock/bond portfolio generates the same average return and contains a portfolio with a much lower risk factor. Therefore, the benefits of diversification, which significantly reduces the risk of an increase in excessive returns, are high (Bessler et al. 2021). In addition to the systematic risk factors, there are other factors, namely the Fama–French model factors, such as beta in conjunction with market factors, size, value, momentum, investment, and profitability, which are discussed in the study by Nazaire et al. (2020) to examine which factor exposures (betas) and characteristics provide independent information for US stock returns in a multifactor context and to identify betas associated with unweighted market factors, size, value, momentum, investment, and profitability. In contrast, firm characteristics associated with size, value, investment, and profitability have significant and independent explanatory power, suggesting that they are important in determining expected returns. Moreover, asset allocation is a problem for investors. Therefore, investors need to estimate expected returns when constructing an optimal portfolio. Thus, a profitable portfolio combination is a combination of stocks, bonds, and commodity classes compared to a combination of simple estimates, equally weighted portfolios, or portfolios based on historical averages (Kynigakis and Panopoulou 2022).

The work of Markowitz (1952) was seminal for modern portfolio theory. On this basis, Lintner (1965), Mossin (1966), and Sharpe (1964) developed an important financial model that establishes a simple relationship between the returns of an asset and its risk: the CAPM. The CAPM is one of the main pillars of modern finance. It empirically proves that not every investor avoids risk in portfolio investments absolutely and globally (Levy 2022).

Beta in the CAPM model is interesting to measure stock returns during stock market movements. CAPM beta is used to measure the financial performance of an investment, which can estimate the performance of management funds, cost of capital, and securities as the determinant of beta value (Liu et al. 2022). In addition, the CAPM is a financial market risk measurement model that cannot necessarily explain the relationship between risk characteristics and investment returns. Therefore, CAPM always makes a negative and inconsistent contribution to financial theory (Maneemaroj et al. 2021). Moreover, the study by Hundal et al. (2019) analyzed only secondary data for the period 2012–2016 with a sample of 90 stocks listed on the Helsinki Stock Exchange. The results suggest that the relationship between risk and return is synchronous and that the stock returns of the sample companies are less volatile than the market index.

The CAPM model is the first model for portfolio investment management, although its empirical validity shows a weak risk–return relationship. Therefore, the relationship between risk and return has long been the backbone of portfolio management (Kazmi et al. 2021). In this context, Fama and French (2015) extend the CAPM model to include investment and profitability factors to determine the factors associated with average returns in optimizing investment decisions. In addition, the error rate of the CAPM in pricing has decreased significantly compared to previous results in the empirical literature. Moreover, the beta model, which varies over time, has a similar performance to the Fama–French model in most cases. This result is consistent with increased trading activity reducing arbitrage opportunities and thus increasing market efficiency (Rojo-Suárez et al. 2022). CAPM betas positively predict portfolio and individual stock returns when market returns are expected to be high, which is about 50% of the time. Consequently, the product of beta and expected market return (CAPM) predicts out-of-sample asset returns, and the predictive power of CAPM exceeds that of alternative factor models. Strategies that exploit the joint predictive power of beta and market return prediction have average returns that increase with beta and Sharpe ratios that are up to twice those of the corresponding buy-and-hold strategies (Hasler and Martineau 2022).

Boussaidi and AlSaggaf (2022) found that the CAPM was unable to capture the off-setting gains in most Middle East and North Africa (MENA) equity markets, so the gains cannot be explained by investment risk. Moreover, the hypothesis that the representa-

tiveness heuristic causes investors to overreact does not hold for all stock markets. In contrast to the representativeness heuristic, the authors extended the five-factor model to include factors based on similar past earnings shocks and found that offsetting gains in most MENA markets are not captured by short zero-investment portfolios on portfolios with a series of shocks and positive and long gains on a portfolio with a series of negative earnings surprises.

The relationship between a security's market line (SML) and the CAPM persists if betas are appropriately adjusted before investors analyze the level of market risk among different investment securities. Therefore, the adjusted CAPM is used to show the behavior of non-average variance in explaining the CAPM anomaly, e.g., the low beta anomaly when investors with unequal variance underweight high beta (low beta) assets. Thus, the empirical analysis shows that two-thirds of investors must deviate from the mean-variance analysis to explain the low beta anomaly (Hens and Naebi 2021).

Investments always involve risks that differ from one investment market to another in the form of systematic, cross-sectional, and time-varying risks. Nonetheless, the CAPM provides an excellent risk–return framework, and market beta can reflect the risks associated with risky investments. However, there are opportunities for investors to exploit dimensional and time anomalies to improve investment returns. Since stock returns exhibit positive autocorrelation in the short-to-medium term, stocks that have performed well in the past tend to perform well in the future, while stocks that have performed poorly in the past tend to perform poorly. For this reason, Mohanty (2019) found significant differences in explaining the sources of risk, where each market is unique in terms of the characteristics of risk factors, and market risk as described by the CAPM is not a true measure of risk, which contradicts the risk–return efficiency framework. For example, lower market risk leads to higher excess returns in 19 of the 22 developed markets, which is a significant anomaly. However, the Asness, Frazzini, and Pederson (AFP) model also leads to lower market risk (15 countries) and higher alpha (11 countries) in most markets. It is also interesting to note that the CAPM is a model that leads to excess returns in developed markets. However, beyond that, each market is unique in its composition and trends, even over long periods, so a general asset allocation approach cannot be applied to all markets.

2.2. Fama–French Model

The asset pricing model is a financial theory concept that contributes to popular research in the finance literature. The concept of finance theory reveals the most commonly used asset pricing models in the financial world, such as the CAPM, arbitrage pricing theory (APT), or the Fama–Francis model. Fama and French (1992) used data on average stock returns on the New York Stock Exchange (NYSE), the American Express (AMEX), and the National Association of Securities Dealers Automated Quotations (NASDAQ) for the period 1963–1990. The empirical results of two easily quantifiable variables, market equity (ME) and the ratio of book equity to market equity (BE/ME), capture much of the average stock returns associated with size and earnings–price ratios (E/P), book capital, and leverage.

After Fama and French (1993) discovered three risk factors for portfolio investments, namely the SMB factor, the HML factor, and the low B/M, Fama and French (2015) added two more factors to the three risk factors, namely profitability and investment, to form five factors that capture the average return pattern of stocks in the investment portfolio. Moreover, the main problem of the five-factor model is that it is not able to capture the low average returns of small stocks, whose returns behave especially poorly for companies that invest in low profitability (Fama and French 2015). Hence, the results of Fama and French (2015) showed that HML is an over factor in the sense that the high average returns are fully captured by its exposure to RM-RF, SMB, and particularly RMW and CMA. Therefore, better stock returns can be expected.

Fama and French (2012) examined international stock returns in North America, Europe, Japan, and the Asia–Pacific region to detail the size, value, and momentum patterns

of average returns for developed country markets. They then examine how well they capture average returns for a portfolio of size and value or size and momentum. The results suggest that there is a premium in average returns in North America, Europe, Japan, and Asia–Pacific and that there is strong return momentum in all regions except Japan, with no sign of momentum returning in any size group. In addition, there is new evidence on how international value and momentum returns vary with company size. Except in Japan, the value premium is larger for small stocks.

Based on a sample of 500 non-financial firms from the Bombay Stock Exchange for the period 2003–2019, the data suggest the superiority of the Fama–French three-factor model over the CAPM. Sehrawat et al. (2020) demonstrated that there is evidence of market segmentation in the first half of the sample period (2003–2010). However, the second subperiod (2011–2019) showed weak signs of market integration, supported by the Johansen cointegration test, suggesting that the Indian market is gradually integrating with global markets. In addition, Lalwani and Chakraborty (2020) used multifactor asset pricing models in emerging and developed markets to compare the performance of different multifactor asset pricing models in ten emerging and developed markets. The final country selection consists of Australia, Canada, Japan, the United Kingdom, and the United States as developed markets, and China, India, Malaysia, South Korea, and Taiwan as emerging markets. They find that the FF5 model (the Fama–French five-factor model) improves the pricing of stocks in Australia, Canada, China, and the United States. Price formation in these countries appears to be more integrated. However, the superior performance in these four countries is not consistent across a wide range of test values, and the magnitude of the reduction in pricing errors relative to three- or four-factor models is often economically insignificant. For other markets, the simple three-factor model or its four-factor variants appear to be more appropriate.

Ekaputra and Sutrisno (2020) tested the Fama–French three-factor model and the Fama–French five-factor model in contrast to previous studies. They concluded that the Fama–French five-factor model does not perform better than the Fama–French three-factor model in explaining excess portfolio returns in either market. In contrast to the US market, they found that the HML factor is not redundant in either market. The results are robust for both equally weighted and value-weighted portfolios and also for different factor construction methods. For the Johannesburg Securities Exchange (JSE), Cox and Britten (2019) examined in detail the effectiveness of the FF5 model in explaining returns for the period 1991 to 2017. Their results confirmed that the three-factor models of size-value and size-profitability best describe the returns of the time series when comparing the models. The five-factor model best explains the cross-section of returns. Overall, the results show a significant inverse size premium and a negative relationship between beta and returns but also a significant value premium. The additional factors of profitability and investment help explain returns on the JSE, but profitability is more consistent than investment.

The economic environment becomes a challenge in investing assets. Therefore, investors need to evaluate the price of assets in anticipation of risk and return. Thus, investors need to evaluate the efficiency of the firm when making investment decisions. Based on this assumption, efficiency is considered an additional factor when evaluating security returns. Therefore, the study by Aygoren and Balkan (2020) investigated the role of efficiency in capital asset pricing the stocks of NASDAQ. The results show that all factors in the models are found to be valid in asset pricing. Moreover, the paper provides evidence that the explanatory power of the proposed four-factor model exceeds the explanatory power of the CAPM and the Fama–French three-factor model.

The Fama–French model makes the basic assumption that investment returns are influenced by the unique risk variables associated with an asset. This model is based on the assumption that investment returns are influenced by factors other than the market risk described by the CAPM. The market size factor and the value factor are the two most important determinants in this model. When the Fama–French model was further developed into five factors, two more factors were added, namely the investment factor

and the profitability factor. After careful examination of the above risk variables concerning investments in stocks and bonds, the empirical hypotheses were formulated as follows:

Hypothesis 1. *There is a significant influence of the market factor on investment returns.*

Hypothesis 2. *There is a significant influence of the size factor on investment returns.*

Hypothesis 3. *There is a significant influence of the value factor on investment returns.*

Hypothesis 4. *There is a significant influence of the profitability factor on investment returns.*

Hypothesis 5. *There is a significant influence of the investment factor on investment returns.*

3. Data and Methodology

3.1. Data Collection

Our study uses data on returns on investments in petroleum funds in the form of stocks and bonds collected by the Ministry of Finance of Timor-Leste. Monthly data on stock and bond returns are provided by the Petroleum Fund Policy and Management Office in the form of raw Excel data. The objective of our study is to identify the CAPM model, the three factors of the Fama–French model, and the five factors of the Fama–French model in determining the return of stocks in the international stock market. The research approach used is to test the effects of the variables of the three-factor model and the five-factor model on the excess return of the oil fund investment portfolio using monthly data for 2006–2019. The authors analyze five explanatory variables in regression equations (2) and (3), including market, SMB, HML, RMW, and CMA factors, using data from French's data library, accessible at http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (accessed on 14 January 2022). Table 1 also provides a brief description and definition of the explanatory variables. In addition, Figure 1 shows the time evolution of excess returns on stock and bond investments with variations in the investment risk factors of the Fama–French model over the sample period.

Table 1. Variable definitions and data specification.

Measure	Definition	Data Source
Excess returns	A return earned by an investment in excess of a risk-free investment.	Ministry of Finance of Timor-Leste
MKT (market factor)	Return investment minus risk-free rate is the excess return on Timor-Leste portfolio investment.	
SMB (size)	Small Minus Big is the difference between the average returns of companies in small equity portfolios and companies in large equity portfolios.	http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html (accessed on 7 November 2023)
HML (value)	High Minus Low is the difference between the average return on the value portfolio and the growth portfolio.	
RMW (profitability)	The difference between the returns of companies with robust (high) and weak (low) operating profitability.	
CMA (investment)	The difference between the returns of companies that invest conservatively and companies that invest aggressively.	
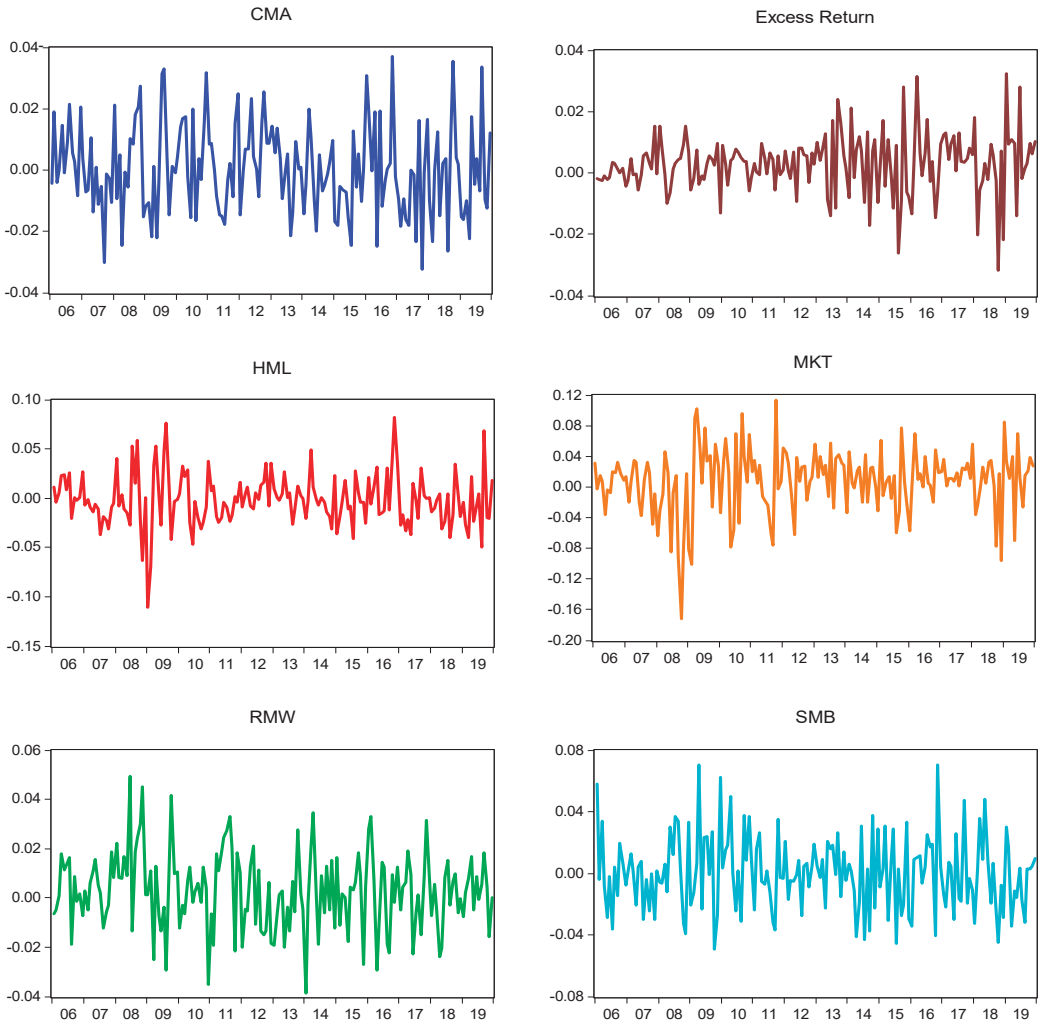
Figure 1. Time trend of excess return on equity and bonds and the Fama–French five-factor model. Source: own elaboration. This figure is the monthly value of portfolio investments since 2005, i.e., the monthly excess returns plus five investment factors such as market premium (MKT), size (SMB), value (HML), profitability (RMW), and investment (CMA) in the Fama–French model, which may be accessed via the Kenneth French Web. The vertical axis and the horizontal axis refer to the period from January 2006 to January 2019.

3.2. Empirical Approach

To determine the optimal return for stock and bond investments in Timor-Leste, this study uses the CAPM, the Fama–French three-factor model, and the Fama–French five-factor model. Below is a brief explanation of the testing procedures for using this model in asset valuation.

3.2.1. Capital Asset Pricing Model

The standard algebraic form of the CAPM is as follows:

$$E(R_i) = R_f + (R_m - R_f)b_i \tag{1}$$

Here, $E(R_i)$ is expected return on capital asset "*i*", R_f is a risk-free rate of return, R_m is the return on the market portfolio, and b_i is the index of systematic risk.

3.2.2. Fama–French Three-Factor Model

Fama and French (1993) defined three portfolios to capture risk: MKT (return on the market portfolio minus the risk-free rate), SMB (return on the portfolio of small stocks minus the return on the portfolio of big stocks), and HML (return on the portfolio of stocks with high book-to-market ratios minus the return on the portfolio of stocks with low book-to-market ratios). In portfolio investing, the asset pricing model is empirically tested to determine the function of the risk factor as the independent variable and the return of the assets as the dependent variable. Thus, the risk factors and asset returns are used in multiple regression to determine the portfolio investment formulation. Thus, Fama and French's (1993) three-factor model can be estimated using the following regression formula:

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + e_{it} \tag{2}$$

where $R_{it} - R_{Ft}$ is the excess return over the risk-free return of the portfolio *i* at *t* time, R_{Mt} is the return on the value-weight (VW) market portfolio, SMB_t is the return on a diversified portfolio of small stocks minus the return on a diversified portfolio of big stocks, HML_t is the difference between the returns on diversified portfolios of high and low B/M stocks, and e_{it} is a zero-mean residual.

3.2.3. Fama–French Five-Factor Model

After Fama and French (1993) introduced three risk factors, namely firm size, book-to-market value, and excess market return, Fama and French (2015) introduced a five-factor asset pricing model with two new factors: profitability and investment. The econometric model used for estimation is as follows:

$$R_{it} - R_{Ft} = a_i + b_i(R_{Mt} - R_{Ft}) + s_i SMB_t + h_i HML_t + r_i RMW_t + c_i CMA_t + e_{it} \tag{3}$$

where RMW_t is the factor related to firm profitability, i.e., the difference between the returns of portfolios of firms with robust (high) profitability and those with weak (low) profitability, and CMA_t is the factor related to investment, i.e., the difference between the returns of conservative (low) and aggressive (high) investment portfolios.

4. Empirical Results

4.1. Descriptive Statistics

Based on the collected data and the previously created indicators, we were able to perform a descriptive statistical analysis. The corresponding data processing was performed in STATA 14 and EViews 12.

Table 2 shows the summary statistics of all return factors of Timor-Leste stocks and bonds in the international stock market, monthly data from 2006 to 2019. The mean MKT for the return of Timor-Leste stocks and bonds is 0.75% per month, and the mean SMB premium and the mean HML premium are 0.03% and −0.19%, respectively. The monthly premiums for RMW and CMA have a value of 0.27% and 0.03% over the period 2006–2019. In addition, Table 2 shows that the MKT factor has the highest standard deviation (4.21%), and this factor has the highest risk. In addition, the CMA factor has the lowest standard deviation (1.45%) and is the factor with the lowest risk.

Table 2. Summary statistics for Fama–French factors of equity and bonds of Timor-Leste in the international stock market.

Factors	Obs.	Mean	Std. Dev.	Min	Max
MKT	168	0.0075	0.0421	−0.1723	0.1135
SMB	168	0.0003	0.024	−0.0492	0.0704
HML	168	−0.0019	0.0262	−0.1111	0.0821
RMW	168	0.0027	0.0154	−0.0388	0.0494
CMA	168	0.0003	0.0145	−0.0323	0.037
Excess Monthly Return	168	0.0028	0.0095	−0.0317	0.0323

Source: own elaboration. Note: The table provides summary statistics for the five Fama–French factors, i.e., monthly excess returns plus five investment factors: market risk premium or "portfolio return minus risk-free rate" (MKT), size or "Small Minus Big" (SMB), value or "High Minus Low" (HML), profitability or "Robust Minus Weak" (RMW), and investment or "Conservative Minus Aggressive" (CMA). The table includes statistics for the mean, standard deviation (Std. Dev.), maximum (max), and minimum (min).

4.2. Correlation Matrix

Table 3 presents the correlation coefficients between the variables included in the study. The portfolio return is positively and significantly related to the market risk premium (0.40) but negatively related to the CMA factor (−0.16). Moreover, the MKT factor is positively and significantly related to the SMB factor and the HML factor but negatively related to the RMW factor. In addition, there is a positive and significant relationship between the SMB factor and the HML factor but a negative relationship with the RMW factor. Finally, the HML factor is negatively related to the RMW factor but positively related to the CMA factor.

Table 3. Correlation among parameter and portfolio returns.

	Excess Monthly Return	MKT	SMB	HML	RMW	CMA
Excess Monthly Return	1					
MKT	0.401 ***	1				
SMB	−0.0218	0.400 ***	1			
HML	−0.104	0.258 ***	0.284 ***	1		
RMW	0.0155	−0.396 ***	−0.374 ***	−0.175 *	1	
CMA	−0.164 *	−0.110	0.0638	0.523 ***	0.0475	1

Source: own elaboration. Note: The table shows the correlation matrix for the five Fama–French factors, which consists of monthly excess returns plus five investment factors such as market risk premium or "portfolio return minus risk-free rate" (MKT), size or "Small Minus Big" (SMB), value or "High Minus Low" (HML), profitability or "Robust Minus Weak" (RMW), and investment or "Conservative Minus Aggressive" (CMA). * $p < 0.1$, *** $p < 0.01$ indicate significance levels 10% and 1%, respectively.

4.3. Regression Multipliers

The regression results for the CAPM, the Fama–French three-factor model, and the Fama–French five-factor model are presented in Table 4. The results presented in this study provide a better understanding of the elements associated with the MKT factor, the SMB factor, the HML factor, the RMW factor, and the CMA factor. The table shows that the coefficient values for the CAPM, the Fama–French three-factor model, and the Fama–French five-factor model show statistical significance or positive correlations at a 1% significance level. This means that a one-unit increase in the positive direction is associated with a positive return for both the equity and bond portfolios. The results are consistent with the research conducted by Ali et al. (2018).

Table 4. Regression results for the CAPM model, the Fama–French three-factor model, and the Fama–French five-factor model.

Factors	CAPM	Fama–French Three-Factor Model	Fama–French Five-Factor Model
MKT	0.0906 ***	0.116 ***	0.127 ***
	(0.0161)	(0.0170)	(0.0183)
SMB		−0.0673 **	−0.0543 *
		(0.0303)	(0.0310)
HML		−0.0784 ***	−0.0652 **
		(0.0257)	(0.0317)
RMW			0.0964 **
			(0.0474)
CMA			−0.00489
			(0.0550)
Constant (α)	0.00211 ***	0.00179 ***	0.00147 **
	(0.000685)	(0.000666)	(0.000686)
Observations	168	168	168
R-squared	0.1607	0.2302	0.2511
Adj. R-squared	0.1560	0.2161	0.2280
F-statistic	31.7882	16.3428	10.8647
Prob. (F-statistic)	0.0000	0.0000	0.0000

Source: own elaboration. Note: The table shows the regression multipliers for the CAPM model, the three-factor model, and the five-factor model of Fama and French, i.e., monthly excess returns plus five investment factors: market risk premium or "portfolio return minus risk-free rate" (MKT), size or "Small Minus Big" (SMB), value or "High Minus Low" (HML), profitability or "Robust Minus Weak" (RMW), and investment or "Conservative Minus Aggressive" (CMA). Significance at the 1%, 5%, and 10% levels is indicated by *** $p < 0.01$, ** $p < 0.05$, and * $p < 0.1$.

The beta coefficients of the three models for the market factor show a positive and statistically significant correlation at the 1% level. This result empirically supports the hypothesis that market factors have a significant impact on investment returns. This provides empirical evidence for our hypothesis H1. The correlation between the return on equity and bond investments and the level of risk can be seen as indicating a positive relationship, with a higher level of risk usually being associated with higher returns. Conversely, it is a common phenomenon that investors are willing to pay excessive prices for investment opportunities that are associated with lower risk. The market factor is therefore an important factor that can shed light on the results of a portfolio. According to the Fama–French three-factor model, the size factor beta has a statistically significant negative value at a 5% significance level. The results of this study provide empirical support for the hypothesis that factor size has a discernible influence on investment returns. The above results provide empirical support for hypothesis H2. In addition, the Fama–French five-factor model also shows a statistically significant negative value for the size factor beta, but at a slightly higher significance level of 10%. This result shows that the size factor exerts a statistically significant negative influence on the average return of the portfolio. The inverse correlation between the size effect and the average stock return is also evident. The results presented in this study are consistent with previous research by Banz (1981) and Fama and French (1992), which showed that smaller stocks have higher risk-adjusted returns compared to larger companies. The results of this study suggest that there is a higher risk associated with the stock returns of smaller portfolios, so investors must earn a correspondingly higher compensating return compared to larger companies.

The value factor has a statistically significant negative coefficient at the 1% level in the three-factor Fama–French model and the 5% level in the five-factor Fama–French model. The results of this research study provide empirical support for the hypothesis that the value factor exerts a discernible influence on investment returns. The above results provide empirical support for hypothesis H3. This suggests that stocks with value characteristics, sometimes referred to as value stocks, are likely to experience a decline in expected returns. This refers to the distinction between stocks with a low price-to-book ratio, which stands for

value, and stocks with a high price-to-book ratio. Furthermore, the study shows a negative correlation, suggesting that companies with a high price-to-book ratio tend to have lower average returns. This result contradicts the conclusions of Fama and French (1992), who found a positive and statistically significant correlation between average returns and the book value of equity. The strength of this correlation exceeds that of size, debt, earnings, and price as determinants of average stock returns.

The profitability factor has a statistically significant positive coefficient at the 5% level in the Fama–French five-factor model. The results of this study provide empirical evidence for the hypothesis that the profitability factor has a significant impact on investment returns. The above results provide empirical support for our hypothesis H4. This result is supported by Ali et al. (2021), and Horváth and Wang (2021), who found that the profitability factor significantly increases the description of the average return, which is in contrast to the results of Alqadhib et al. (2022), which in turn conclude that the profitability factor has a significant negative relationship with fund returns. The investment factor, on the other hand, has an insignificant effect on the excess return. This is strong evidence for the acceptance of the fifth null hypothesis (H5) and the rejection of the fifth alternative hypothesis (H5). This result is in good agreement with existing studies by Horváth and Wang (2021). However, our study does not support the recent study by Kaya (2021) that the CMA coefficient is negative and significant in eight of the twelve portfolios, and the mean return shows a strong investment pattern in the regression estimation.

The CAPM, the three-factor model, and the five-factor model of Fama and French are widely accepted models for determining the average return of a portfolio. Table 4 shows the results of the CAPM, Fama and French's three-factor model, and Fama and French's five-factor model in terms of the R^2 value of the investment portfolio, which is 21.40% on average. This means that the valuation of the change is explained by the market premium associated with the risk-free interest rate. Furthermore, it should be noted that the adjusted R^2 value of the Fama–French five-factor model, namely 0.2280, exceeds the adjusted R^2 values of both the Fama–French three-factor model (0.2161) and the CAPM model (0.1560). The observed F-statistic is statistically significant at the 1% level.

Figure 2 shows a graphical overview of the evolution of the factors over time. The market risk premium exhibits higher cumulative fluctuations compared to other portfolio investment risk factors. However, the trend of the market risk premium factor first developed positively until the end of 2008 and then negatively until mid-2011. Thereafter, the positive trend continued until the end of 2011, when it turned negative again. After mid-2011, it slowly increased in a positive direction until it reached its highest level in 2019. Thus, the market risk premium factor is always highest when it succeeds in predicting the return on investment when it is profitable. The valuation factor underperformed from 2007 until mid-2010. It then returned to a positive trend until the end of 2019, when it fell back into negative territory. This shows that small companies perform better than larger companies in the long run. The value factor only performed well from 2006 to 2007. After that, it developed negatively until 2019. This means that there is a difference in the value premium between the return of a high book-to-market portfolio and a low book-to-market portfolio, so it continues to generate negative returns (Ryan et al. 2021). The RMW factor had the best performance from the beginning of 2006 to 2019. This means that a positive value of the RMW factor indicates that the company has higher profitability and continues to exceed over the investment period of the portfolio. In addition, the CMA factor shows a decrease in investment at the end of 2008, then an increase and then a decrease in 2009, and only during the 2008/2009 financial crisis (Dirkx and Peter 2020). Finally, the over-return factor declined negatively only from 2006 to 2007, and then the contribution of the over-return rate increased significantly until 2019.
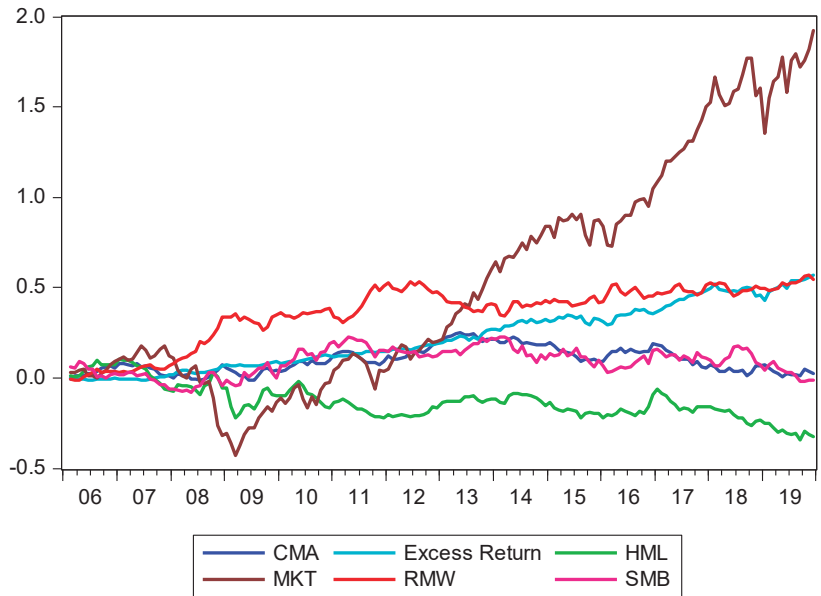
Figure 2. Cumulative value of the five factors and excess return. Source: own elaboration. This figure represents the monthly cumulative value of portfolio investments since 2005, which consists of monthly excess returns plus five investment factors such as market premium (MKT), size (SMB), value (HML), profitability (RMW), and investment (CMA) in the Fama–French model. Cumulative values are on the vertical axis, and the horizontal axis represents the period of January 2006 to January 2019.

5. Results Discussion

Regarding the correlations between factors, this result is consistent with the findings of Fama and French (2015) and Ryan et al. (2021) that the RMW factor is negatively correlated with all factors of portfolio investment. Moreover, these results confirm the findings of the earlier work by Ryan et al. (2021). There is a high and positive correlation between HML and CMA, suggesting that companies with high B/M tend to be companies with low investment. Consistent with the previous study by Carvalho et al. (2022), the current results also show that the factors HML and market (Rm-Rf) have a significant positive correlation. Moreover, this result contradicts the findings of Fama and French (2017), who found that the RMW factor is negatively correlated with investment.

Meanwhile, our results for the Timor-Leste economy are not consistent with the findings of Zaremba et al. (2019), who found that the MKT factor has a significant negative relationship with the firm size factor. However, our results are consistent with the HML factor being negatively related to the RMW factor.

The current results show that the CAPM has successfully captured the effects of MKT risk. Extending the CAPM, the three-factor model of Fama and French (1993) introduces two additional factors, the SMB factor, and the HML factor. This result was also reported by Huang (2019). Moreover, these results for the SMB factor do not agree with the results of Huang (2019), but the results are negative and significantly the same for the value factor. For the five Fama–French factors, the results are consistent for the MKT risk factor and the HML factor, while they are not consistent for the SMB factor and the RMW factor (Huang 2019).

Likewise, these results are consistent with the application of the Fama–French model of Fang et al. (2021) with three factors in the Chinese stock market, namely that the MKT risk factor has a positive and significant effect, as well as the SMB factor and the HML factor. This result is in contrast to Kubota and Takehara (2018), who found that MKT is

significant only for the CAPM with a negative coefficient. Similarly, the HML factor was significant with positive coefficients for Fama and French's three-factor and five-factor models. The results are consistent with the findings of Richey (2017), who found that the CAPM, Fama and French's three-factor model, and Fama and French's five-factor model have a positive and significant effect. Moreover, it is consistently positive and significant only for the MKT risk factor. The SMB factor is significantly positive, and it is inconsistent with the HML factor. However, the Fama and French (2015) five-factor model shows that all factors are consistent except for the CMA factor.

It is interesting to compare these results with those of Kostin et al. (2022) on multifactor asset pricing and factor models during pandemic situations in developed and emerging markets. Timor-Leste is one of the new/emerging markets that invest their assets in the form of stocks and bonds in the markets of developed countries such as the United States (US), the United Kingdom (UK), Japan, and Germany. For these developed countries, equity investments amounted to USD 4103 million (62.7%) for the United States, USD 375 million (5.7%) for the United Kingdom, USD 540 million (8.3%) for Japan, and USD 686 million (10.5%) for Germany, which in this case is part of the European Union. Similarly, total bond investments amounted to USD 7914 million (83.1%) in the United States, USD 153 million (1.6%) in Japan, USD 164 million (1.7%) in the United Kingdom, and USD 37 million (0.4%) in Germany (Timor-Leste Ministry of Finance 2019). Thus, Timor-Leste's total equity investments contribute to corporate returns in the capital markets of these developed countries. Therefore, these results provide empirical evidence that the multifactor performance of the Fama and French (2015) five-factor model of Kostin et al. (2022) is positive and significant for all countries out of the US, UK, and Japan for MKT risk factor, except for Germany. Similarly, the SMB factor is significantly negative in line with the UK and in contrast to Japan, which is significantly negative. In addition, the HML factor is consistently significantly negative for the United States and inconsistent for Germany. In addition, the RMW factor is inconsistent for the United States and Germany, while the CMA factor is consistent in all countries, being, respectively, positive and insignificantly negative.

It is interesting to note that Timor-Leste has a Petroleum Fund, established in 2005, whose source of revenue is oil and gas. The income from this fund is invested in international capital markets such as the NASDAQ, the NYSE, the London Stock Exchange (LSE), and the Tokyo Stock Exchange. Why are these funds invested in the international capital markets? Timor-Leste is a country that does not yet have a national capital market and is therefore currently focusing on economic development. With this investment, you obtain a return every month, which fluctuates. Apart from that, petroleum money is also a source of funding for the national budget every fiscal year.

Therefore, the investment returns become a source of data for conducting research. In addition, the Fama–French dataset is used to investigate the extent to which risk factors (loading factors) affect excess returns.

For example, the results of the Fama–French five-factor model, namely the CMA factor, show that the CMA factor has a negative and non-significant influence on the excess return.

Every investor faces risk in their investments, but portfolio diversification can balance risk and return. This is because understanding risk management can help investors manage risk well, and it is the most important key to ensuring investment sustainability and mutual fund performance.

However, the results of the five Fama–French factors show that the CMA factor has a negative and insignificant impact on excess returns. However, other factors such as SMB, HML, and RMW have a significant impact and have a greater effect on investment performance. Other factors also have a varying impact on investment performance, such as the impact of global markets and global economic policies. In addition, investors analyze historical data more thoroughly by consulting financial experts or investment advisors. Thus, governments (investors) use information based on factors that ensure the sustainability of investments when making decisions. This is because investment

is becoming a source of income for economic diversification, as Timor-Leste is heavily dependent on oil and gas.

The results of this investigation will serve as a reference source for other researchers for further investigations (research gaps). This is because only sample data from Timor-Leste were used in this research. Therefore, future research can be compared with other SWFs. For example, the sample data from Timor-Leste can be compared with the sample data from other SWFs. Apart from this, the strategy of portfolio diversification in asset allocation is very different. This difference is interesting for the question of whether portfolio construction is appropriate or not. The differences in economic size and the types of funds used, such as development funds, reserve investment funds, and pension reserve funds, are also interesting. All these funds depend on the type of fundraising and sources (non-commodities, oil and gas, or minerals) of each country. All these sources depend on the individual countries. In addition, Timor-Leste's investment portfolio is unique in that it still depends on equity investments in international markets, as Timor-Leste does not yet have a national capital market. This was done to accumulate profits and diversify Timor-Leste's economy.

6. Conclusions and Policy Implications

Based on the results of the analysis and discussion of the data in the previous section, the following conclusions were drawn. This study empirically examines the CAPM, Fama and French's three-factor model, and Fama and French's five-factor model for the excess returns of Timor-Leste's equity and bond investments in the international market. The sample used includes 156 monthly excess returns over the period from 2006 to 2019. The results of the CAPM model test show that market returns have a positive and significant impact on excessive stock returns. The empirical evidence supports hypothesis H1, which states that the market factor has a positive and statistically significant impact on the excess return on equity. However, the MKT factor has a positive and significant impact on the rate of excessive stock returns in both models, i.e., the three-factor model of Fama and French and the five-factor model of Fama and French. The available empirical data support hypothesis H1, which states that the market component exerts a positive and statistically significant influence on the excess return on equity. In addition, the test results of Fama and French's three-factor model show that the SMB factor and the HML factor have a negative and significant influence on the rate of excessive stock returns. The previous findings offer empirical evidence in favor of hypotheses H2 and H3. Furthermore, the test results of Fama and French's five-factor model show that only two of the five model variables, namely the SMB factor and the HML factor, have a negative and significant impact on the rate of excess stock returns. The aforementioned findings offer empirical evidence in support of hypotheses H2 and H3. On the other hand, both the MKT risk and RMW factor variables are positively and significantly associated with excess returns. This result provides empirical support for our hypothesis H4. The CMA factor, on the other hand, has a negative and insignificant effect on excess returns. The available evidence supports the rejection of the fifth hypothesis (H5).

In summary, this analysis highlights the complexity of risk factors for excessive returns. The results show that market risks such as size and value play a crucial role in determining the excess return of Timor-Leste's portfolio investments. Therefore, the government needs to consider these aspects in their investment decisions.

The empirical results presented above are interesting and certainly have important implications for the excess return of Timor-Leste's portfolio as an emerging stock market. The significant positive and negative effects of stock and bond investments clearly show the attention of investors in investing. This also means that it will be a major investment challenge to achieve the goal of increasing the maximum equity allocation, with an expected allocation target of 40% equities and 60% bonds to achieve the target of 3% real return with reasonable probability (Timor-Leste Ministry of Finance 2019).

Investments in Timor-Leste stocks are contributions from oil revenues pooled in the Petroleum Fund and then managed for investment in a portfolio of stocks in the international market. This is because the SMB and HML factors can predict gross domestic product (GDP) growth when investing in equities, and these factors forecast future investment opportunities. Thus, this portfolio investment provides a higher return during fluctuations in economic performance (Carson 2022). To this end, oil fund managers need to diversify their portfolios into different asset classes to reduce investment risk.

Fund managers managing petroleum fund portfolios in the form of stocks and bonds need to pay attention to the SMB factor in terms of investment performance, where small companies outperform larger companies over the long term. This phenomenon suggests that the performance of small companies in the stock market may be a predictor of the future performance of a low-beta-against-beta strategy. Thus, it is the short-term performance of small companies and funding liquidity that affects the profitability of the low-beta strategy, which ultimately leads to low or negative returns for the low-beta security class (Zaremba 2020). The same was also found by Ji et al. (2020), who state that the size effect is that the returns of small listed companies are on average much higher than those of large companies. Still, for the same researchers, the effect of BM shows that stock returns have a positive relationship with the book-to-market ratio of the company. A higher book-to-market ratio can lead to a higher stock return. In addition, Hu et al. (2019) found that a strong size effect means that smaller companies have higher returns on average than larger companies.

Managers need to understand information about portfolio market activity when managing investments and macroeconomic risks to deal with increasing unsystematic and systematic risks in portfolio investments. Therefore, managers need to understand the relationship between stock market volatility and macroeconomic forces in policy making. For example, stock price movements in economic activity, especially in portfolio investing, are influenced by macroeconomic variables such as inflation in predicting excess returns, especially directional relationships with variables that interact with each other (de Jesus et al. 2020). In addition, investment managers need to understand the balance of values of expected cash flows when forecasting interest rate fluctuations that affect changes in stock prices. This is because high interest rates affect excessive returns on portfolio investments, i.e., when cash flows are capitalized. It is interesting to make income securities an alternative investment necessary for holding equity investments. Similarly, high interest rates can affect investment costs, making investors less willing to borrow and make portfolio investments. This also affects the value of future cash flows and ultimately leads to a decline in stock prices (Tiwari et al. 2022). In addition, investment managers need to understand monetary policy, even though Timor-Leste is still dependent on US monetary policy, especially the official use of the dollar in the economy, where the contribution of the policy of rising interest rates affects the rise of the stock market, which in turn has ultimately disrupted economic activity due to greater inflationary pressures, such as the current war between Russia and Ukraine.

Based on our findings described in the conclusion and policy implications, we summarize and recommend the following policy actions as a good basis for portfolio investment decision making:

1. The Petroleum Fund invests in Timor-Leste bonds and equities on the international markets intending to accumulate capital. This capital is used for economic diversification to increase GDP growth. Therefore, to reduce investment risks, the government needs to diversify its portfolio into different asset classes.
2. Fund managers should consider the SMB factor for petroleum fund portfolios, as smaller companies often outperform larger companies, possibly indicating the future performance of a low-beta-against-beta approach.
3. To properly manage assets, managers must have a comprehensive understanding of market activity and macroeconomic risks in the portfolio.

4. Fund managers need to have insights into the correlation between stock market volatility and macroeconomic factors, which are essential for policy decisions, especially for predicting excess returns.

5. Investment managers need to consider the balance of expected cash flows when forecasting interest rate fluctuations, as high-interest rates can affect excessive returns on portfolio investments and reduce investors' appetite for portfolio investments.

6. Investment managers need to understand the monetary policy in Timor-Leste, which is heavily influenced by US monetary policy, specifically the dollar. This has resulted in higher inflationary pressures and disruptions in economic activity.

Timor-Leste, as a recent or new economy (emerging country), participates in equity investments in international markets. This is the first article that uses the Fama and French (2015) five-factor model and the three-factor model of Fama and French (1993) in equity portfolio investment for the country. For this reason, this research is limited to using the excess returns of equity investments in the form of stocks and bonds as invested in model markets in various countries such as the US, the UK, Japan, and Australia. It is hoped that further researchers will add other variables such as momentum and quality and use the names of listed companies and make Timor-Leste's investment portfolios comparable between developing and developed countries since Timor-Leste itself does not yet have a national capital market.

Author Contributions: F.A. was responsible for conceptualization, data curation, formal analysis, investigation, methodology, and roles/writing—original draft. M.M. was responsible for investigation, methodology, project administration, resources, software, supervision, validation, visualization, roles/writing—original draft, and writing—review and editing. E.V. was responsible for resources, software, supervision, validation, visualization, roles/writing—original draft, and writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: We confirm that the data sources described in the methodology were used. Monthly data on equity and bond returns are provided by the Office of Policy and Management of Petroleum Funds, Ministry of Finance of Timor-Leste, where the data can be accessed in Excel and the Fama–French and CAPM models from Kenneth R. French—Data Libraries. The Excel data are available upon request from the first author.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AFP	Asness, Frazzini, and Pederson model
AMEX	American Express
B/M	Book-to-market ratio
BE/ME	Book equity to market equity
CAPM	Capital asset pricing model
CMA	Conservative Minus Aggressive
FF5	Fama–French five-factor model
GDP	Gross domestic product
HML	High Minus Low
INV	Investment by market

ME	Market equity
MENA	Middle East and North Africa
MKT	Market risk premium
MOM	Momentum factor
MPT	Modern portfolio theory
NASDAQ	National Association of Securities Dealers Automated Quotations
NEP	Non-negative equity premium
NYSE	New York Stock Exchange
OLS	Ordinary least squares
P	Profitability
RMW	Robust Minus Weak
SMB	Small Minus Big
SSA	Sub-Saharan African countries
UK	United Kingdom
US	United State
USD	United States dollar

References

Ali, Fahad, Muhammad U. Khurram, and Yuexiang Jiang. 2021. The Five-factor asset pricing model tests and profit-ability and investment premiums: Evidence from Pakistan. *Emerging Markets Finance and Trade* 57: 2651–73. [CrossRef]

Ali, Fahad, Rongrong He, and Yuexiang Jiang. 2018. Size, value and business cycle variables. the three-factor model and future economic growth: Evidence from an emerging market. *Economies* 6: 14. [CrossRef]

Alqadhib, Haidar, Nada Kulendran, and Lalith Seelanatha. 2022. Impact of COVID-19 on mutual fund performance in Saudi Arabia. *Cogent Economics & Finance* 10: 2056361. [CrossRef]

Anjum, Nadia, and Suresh K. O. Rajput. 2021. Forecasting islamic equity indices alpha. *International Journal of Islamic and Middle Eastern Finance and Management* 14: 183–203. [CrossRef]

Aygoren, Hakan, and Emrah Balkan. 2020. The role of efficiency in capital asset pricing: A research on nasdaq tech-nology sector. *Managerial Finance* 46: 1479–93. [CrossRef]

Bank, Matthias, and Franz Insam. 2019. Risk premium contributions of the Fama and French mimicking factors. *Finance Research Letters* 29: 347–56. [CrossRef]

Banz, Rolf W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Eco-Nomics* 9: 3–18. [CrossRef]

Bartram, Söhnke M., Harald Lohre, Peter F. Pope, and Ananthalakshmi Ranganathan. 2021. Navigating the factor zoo around the world: An institutional investor perspective. *Journal of Business Economics* 91: 655–703. [CrossRef]

Bessler, Wolfgang, Georgi Taushanov, and Dominik Wolff. 2021. Factor investing and asset allocation strategies: A comparison of factor versus sector optimization. *Journal of Asset Management* 22: 488–506. [CrossRef]

Boussaidi, Ramzi, and Majid I. AlSaggaf. 2022. Contrarian profits and representativeness heuristic in the MENA stock markets. *Journal of Behavioral and Experimental Economics* 97: 101820. [CrossRef]

Carson, Scott A. 2022. Long-term daily equity returns across sectors of the oil and gas industry, 2000–2019. *Journal of Industry, Competition and Trade* 22: 125–43. [CrossRef]

Carvalho, Gabriel A., Hudson F. Amaral, Juliano L. Pinheiro, and Laíse F. Correia. 2022. Pricing of liquidity risk: New evidence from the Latin American emerging stock markets. *Emerging Markets Finance and Trade* 58: 398–416. [CrossRef]

Chai, Daniel, Mardy Chiah, and Angel Zhong. 2019. Choosing factors: Australian evidence. *Pacific Basin Finance Journal* 58: 101223. [CrossRef]

Cox, Shaun, and James Britten. 2019. The Fama-French five-factor model: Evidence from the Johannesburg stock exchange. *Investment Analysts Journal* 48: 240–61. [CrossRef]

Dichtl, Hubert, Wolfgang Drobetz, and Viktoria S. Wendt. 2021. How to build a factor portfolio: Does the allocation strategy matter? *European Financial Management* 27: 20–58. [CrossRef]

Dirkx, Philipp, and Franziska J. Peter. 2020. The Fama-French five-factor model plus momentum: Evidence for the German market. *Schmalenbach Business Review* 72: 661–84. [CrossRef]

Doraisami, Anita. 2018. The Timor Leste petroleum fund, veterans and white elephants: Fostering intergenerational equity? *Resources Policy* 58: 250–56. [CrossRef]

Ekaputra, Irwan A., and Bambang Sutrisno. 2020. Empirical tests of the Fama-French five-factor model in Indonesia and Singapore. *Afro-Asian Journal of Finance and Accounting* 10: 85–111. [CrossRef]

Fama, Eugene F., and James D. MacBeth. 1973. Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy* 81: 607–36. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 1992. The cross-section of expected stock returns. *The Journal of Finance* 47: 427–65. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33: 3–56. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 2012. Size, value, and momentum in international stock returns. *Journal of Financial Economics* 105: 457–72. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116: 1–22. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 2017. International tests of a five-factor asset pricing model. *Journal of Financial Economics* 123: 441–63. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 2018. Choosing factors. *Journal of Financial Economics* 128: 234–52. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 2020. Comparing cross-section and time-series factor models. *Review of Financial Studies* 33: 1891–926. [CrossRef]

Fang, Tong, Zhi Su, and Libo Yin. 2021. Does the green inspiration effect matter for stock returns? Evidence from the Chinese stock market. *Empirical Economics* 60: 2155–76. [CrossRef]

Hasler, Michael, and Charles Martineau. 2022. Equity return predictability with the ICAPM. *SSRN Electronic Journal*, 1–37. Available online: https://papers.ssrn.com/abstract=3368264 (accessed on 3 November 2023).

Hens, Thorsten, and Fatemeh Naebi. 2021. Behavioural heterogeneity in the capital asset pricing model with an application to the low-beta anomaly. *Applied Economics Letters* 28: 501–7. [CrossRef]

Horváth, Dominik, and Yung-Lin Wang. 2021. The examination of Fama-French model during the COVID-19. *Finance Research Letters* 41: 101848. [CrossRef]

Hu, Grace X., Can Chen, Yuan Shao, and Jiang Wang. 2019. Fama–French in China: Size and value factors in Chinese stock returns. *International Review of Finance* 19: 3–44. [CrossRef]

Huang, Tzu-Lun. 2019. Is the Fama and French five-factor model robust in the Chinese stock market? *Asia Pacific Management Review* 24: 278–89. [CrossRef]

Hundal, Shab, Anne Eskola, and Doan Tuan. 2019. Risk–return relationship in the Finnish stock market in the light of capital asset pricing model (CAPM). *Journal of Transnational Management* 24: 305–22. [CrossRef]

Hung, Phan T. M., Tran T. T. Dai, Phan N. B. Quynh, Le D. Toan, and Vo H. D. Trinh. 2019. The relationship between risk and return—An empirical evidence from real estate stocks listed in Vietnam. *Asian Economic and Financial Review* 9: 1211–26. [CrossRef]

Jareño, Francisco, María de la O. González, and Alba M. Escolástico. 2020. Extension of the Fama and French model: A study of the largest European financial institutions. *International Economics* 164: 115–39. [CrossRef]

de Jesus, Carlos, Gizelle D. Willows, and Alison M. Olivier. 2020. The influence of the market on inflation, not the other way around. *Investment Analysts Journal* 49: 79–91. [CrossRef]

Ji, Ziyang, Victor Chang, Hao Lan, Ching H. R. Hsu, and Raul Valverde. 2020. Empirical research on the Fama-French three-factor model and a sentiment-related four-factor model in the Chinese blockchain industry. *Sustainability* 12: 5170. [CrossRef]

John, Samuel, Elissaios Papyrakis, and Luca Tasciotti. 2020. Is there a resource curse in Timor-Leste? A critical re-view of recent evidence. *Development Studies Research* 7: 141–52. [CrossRef]

Kaya, Emine. 2021. Relative performances of asset pricing models for BIST 100 index. *Revista Espanola de Financiacion y Contabilidad* 50: 280–301. [CrossRef]

Kazmi, Madiha, Umara Noreen, Imran A. Jadoon, and Attayah Shafique. 2021. Downside beta and downside gam-ma: In search for a better capital asset pricing model. *Risks* 9: 223. [CrossRef]

Kostin, Konstantin B., Philippe Runge, and Michel Charifzadeh. 2022. An analysis and comparison of multi-factor asset pricing model performance during pandemic situations in developed and emerging markets. *Mathematics* 10: 142. [CrossRef]

Kubota, Keiichi, and Hitoshi Takehara. 2018. Does the Fama and French five-factor model work well in Japan? *International Review of Finance* 18: 137–46. [CrossRef]

Kynigakis, Iason, and Ekaterini Panopoulou. 2022. Does model complexity add value to asset allocation? Evidence from machine learning forecasting models. *Journal of Applied Econometrics* 37: 603–39. [CrossRef]

Lalwani, Vaibhav, and Madhumita Chakraborty. 2020. Multi-factor asset pricing models in emerging and developed markets. *Managerial Finance* 46: 360–80. [CrossRef]

Levy, Moshe. 2022. An inter-temporal CAPM based on first order stochastic dominance. *European Journal of Operational Research* 298: 734–39. [CrossRef]

Lintner, John. 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The Review of Economics and Statistics* 47: 13. [CrossRef]

Liu, Hao, Hao Zhang, Ya C. Gao, and Xu D. Chen. 2022. Firm age and beta: Evidence from China. *International Review of Economics and Finance* 77: 244–61. [CrossRef]

Lone, Umer M., Mushtaq A. Darzi, and Khalid U. Islam. 2021. Macroeconomic variables and stock market performance: A PMG/ARDL approach for BRICS economies. *Macroeconomics and Finance in Emerging Market Economies* 16: 300–25. [CrossRef]

Lopes, Helder. 2021. Timor-Leste to Graduate from LDC Category and Beyond: Through Structural Transformation and Economic Diversification. ESCAP Working Paper Series. pp. 1–51. Available online: https://repository.unescap.org/handle/20.500.12870/3689 (accessed on 1 November 2023).

López-García, María Nieves, J. E. Trinidad-Segovia, M. A. Sánchez-Granero, and Igor Pouchkarev. 2021. Extending the Fama and French model with a long term memory factor. *European Journal of Operational Research* 291: 421–26. [CrossRef]

Maneemaroj, Panutat, Ravi Lonkani, and Chanon Chingchayanurak. 2021. Appropriate expected return and the relationship with risk. *Global Business Review* 22: 865–78. [CrossRef]

Markowitz, Harry. 1952. Portfolio Selection. *Journal of Finance* 7: 77–91.

Mohanty, Subhransu S. 2019. Does one model fit all in global equity markets? some insight into market factor based strategies in enhancing alpha. *International Journal of Finance and Economics* 24: 1170–92. [CrossRef]

Mosoeu, Selebogo, and Odongo Kodongo. 2020. The Fama-French five-factor model and emerging market equity returns. *Quarterly Review of Economics and Finance* 85: 55–76. [CrossRef]

Mossin, Jan. 1966. Equilibrium in a capital asset market. *Econometrica* 34: 768–83. Available online: http://www.jstor.org/stable/1910098 (accessed on 2 April 2022).

Nazaire, Gregory, Maria Pacurar, and Oumar Sy. 2020. Betas versus characteristics: A practical perspective. *European Financial Management* 26: 1385–413. [CrossRef]

Nichol, Eoghan, and Michael Dowling. 2014. Profitability and investment factors for UK asset pricing models. *Economics Letters* 125: 364–66. [CrossRef]

Pandey, Asheesh, and Rajni Joshi. 2021. Examining asset pricing anomalies: Evidence from Europe. *Business Perspectives and Research* 10: 362–78. [CrossRef]

Richey, Greg. 2017. Fewer reasons to sin: A five-factor investigation of vice stock returns. *Managerial Finance* 43: 1016–33. [CrossRef]

Rojo-Suárez, Javier, Ana B. Alonso-Conde, and Ricardo Ferrero-Pozo. 2022. Liquidity, time-varying betas and anom-alies: Is the high trading activity enhancing the validity of the CAPM in the UK equity market? *International Journal of Finance and Economics* 27: 45–60. [CrossRef]

Roy, Rahul. 2021. A six-factor asset pricing model: The Japanese evidence. *Financial Planning Review* 4: 1–18. [CrossRef]

Ryan, Nina, Xinfeng Ruan, Jin E. Zhang, and Jing A. Zhang. 2021. Choosing factors for the Vietnamese stock market. *Journal of Risk and Financial Management* 14: 96. [CrossRef]

Sahabuddin, Mohammad, Md A. Islam, Mosab I. Tabash, Suhaib Anagreh, Rozina Akter, and Md M. Rahman. 2022. Co-movement, portfolio diversification, investors' behavior and psychology: Evidence from developed and emerging countries' stock markets. *Journal of Risk and Financial Management* 15: 319. [CrossRef]

Saiti, Buerhan, Yusuf Ma, Ruslan Nagayev, and İbrahim G. Yumusak. 2020. The diversification benefit of Islamic investment to Chinese conventional equity investors: Evidence from the multivariate GARCH analysis. *International Journal of Islamic and Middle Eastern Finance and Management* 13: 1–23. [CrossRef]

Scheiner, Charles. 2021. Timor-Leste economic survey: The end of petroleum income. *Asia and the Pacific Policy Studies* 8: 253–79. [CrossRef]

Sehrawat, Neeraj, Amit Kumar, Narander K. Nigam, Kirtivardhan Singh, and Khushi Goyal. 2020. Test of capital market integration using Fama-French three-factor model: Empirical evidence from India. *Investment Management and Financial Innovations* 17: 113–27. [CrossRef]

Shaikh, Salman A., Mohd A. Ismail, Abdul G. Ismail, Shahida Shahimi, and Muhammad H. Muhammad. 2019. Cross section of stock returns on Shari'ah-compliant stocks: Evidence from Pakistan. *International Journal of Islamic and Middle Eastern Finance and Management* 12: 282–302. [CrossRef]

Sharma, Prateek, and Vipul. 2018. Improving portfolio diversification: Identifying the right baskets for putting your eggs. *Managerial and Decision Economics* 39: 698–711. [CrossRef]

Sharpe, William F. 1964. A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19: 425–42.

Silva, Sabrina E., Carolina M. S. Roma, and Robert A. Iquiapaza. 2020. Portfolio turnover and performance of equity investment funds in Brazil. *Revista Contabilidade & Finanças* 31: 332–47. [CrossRef]

Taib, Asmâa A., and Safae Benfeddoul. 2023. The empirical explanatory power of CAPM and the Fama and French Three-Five Factor Models in the Moroccan Stock Exchange. *International Journal of Financial Studies* 11: 47. [CrossRef]

Taussig, Roi D. 2022. New evidence on practical implications of the CAPM: In memory of Simon Benninga. *Journal of Corporate Accounting and Finance* 33: 72–77. [CrossRef]

Timor-Leste Ministry of Finance. 2019. Timor-Leste Petroleum Fund—Annual Report 2019. Available online: https://www.bancocentral.tl/uploads/documentos/documento_1600057279_4497.pdf (accessed on 20 February 2022).

Tiwari, Aviral K., Ibrahim D. Raheem, Seref Bozoklu, and Shawkat Hammoudeh. 2022. The oil price-macroeconomic fundamentals nexus for emerging market economies: Evidence from a wavelet analysis. *International Journal of Finance and Economics* 27: 1569–90. [CrossRef]

Wang, Ching-Ping, Hung H. Huang, and Jin S. Hu. 2017. Reverse-engineering and real options–adjusted CAPM in the Taiwan stock market. *Emerging Markets Finance and Trade* 53: 670–87. [CrossRef]

Yamaka, Woraphon, and Rungrapee Phadkantha. 2021. A convex combination approach for artificial neural net-work of interval data. *Soft Computing* 25: 7839–51. [CrossRef]

Yunus, Nafeesa. 2020. Time-varying linkages among gold, stocks, bonds and real estate. *Quarterly Review of Economics and Finance* 77: 165–85. [CrossRef]

Zaimovic, Azra, Adna Omanovic, and Almira Arnaut-Berilo. 2021. How many stocks are sufficient for equity portfolio diversification? A Review of the Literature. *Journal of Risk and Financial Management* 14: 551. [CrossRef]

Zaremba, Adam. 2020. Small-minus-big predicts betting-against-beta: Implications for international equity alloca-tion and market timing. *Investment Analysts Journal* 49: 322–41. [CrossRef]

Zaremba, Adam, Anna Czapkiewicz, Jan J. Szczygielski, and Vitaly Kaganov. 2019. An application of factor pricing models to the Polish stock market. *Emerging Markets Finance and Trade* 55: 2039–56. [CrossRef]

Zhang, Tianyang, and Sergio H. Lence. 2022. Liquidity and asset pricing: Evidence from the Chinese Stock Markets. *North American Journal of Economics and Finance* 59: 101557. [CrossRef]

Article

Pricing Path-Dependent Options under Stochastic Volatility via Mellin Transform

Jiling Cao, Xi Li and Wenjun Zhang *

Department of Mathematical Sciences, School of Engineering, Computer and Mathematical Sciences,
Auckland University of Technology, Private Bag 92006, Auckland 1142, New Zealand; jiling.cao@aut.ac.nz (J.C.);
xi.li@aut.ac.nz (X.L.)
* Correspondence: wenjun.zhang@aut.ac.nz

Abstract: In this paper, we derive closed-form formulas of first-order approximation for down-and-out barrier and floating strike lookback put option prices under a stochastic volatility model using an asymptotic approach. To find the explicit closed-form formulas for the zero-order term and the first-order correction term, we use Mellin transform. We also conduct a sensitivity analysis on these formulas, and compare the option prices calculated by them with those generated by Monte-Carlo simulation.

Keywords: asymptotic approximation; barrier; down-and-out; floating strike; lookback; Mellin transform; stochastic volatility

MSC: 91G20; 41A60; 44A99; 91G60

1. Introduction

A standard option gives its owner the right to buy (or sell) some underlying asset in the future for a fixed price. Call options confer the right to buy the asset, while put options confer the right to sell the asset. Path-dependent options represent extensions of this concept. For example, a lookback call option confers the right to buy an asset at its minimum price over some time period. A barrier option resembles a standard option except that the payoff also depends on whether or not the asset price crosses a certain barrier level during the option's life. Lookback and barrier options are two of the most popular types of path-dependent options

Following the lead set by Black and Scholes (1973) and assuming that the underlying asset price follows a geometric Brownian motion with constant volatility, Merton (1973) derived a closed-form pricing formula for down-and-out call options. Reiner and Rubinstein (1991) extended Merton's results to other types of barrier options. Goldman et al. (1979) and Conze and Vishwanathan (1991) provided closed-form pricing formulas for lookback options. For a good summary for research on path-dependent options under the Black–Scholes framework, refer to Clewlow et al. (1994). As we know, the assumption that an asset price process follows a geometric Brownian motion with constant volatility does not capture the empirical observations, due to the volatility smile effect. So, it is desirable to overcome this drawback. There are different ways of extending the Black–Scholes model to incorporate the "smile" feature: one way is to consider "local volatility", and the other is to consider "stochastic volatility".

One popular local volatility model is the constant elasticity of variance (CEV) model introduced by Cox (1975, 1996), where a closed-form pricing formula for European call options was presented. Davydov and Linetsky (2001) derived solutions for barrier and lookback option prices under the CEV process in closed form and demonstrated that barrier and lookback option prices and hedge ratios under the CEV process can deviate dramatically from the lognormal values. In Boyle and Tian (1999), the pricing of certain

path-dependent options was re-examined when the underlying asset follows the CEV diffusion process, by approximating the CEV process using a trinomial method.

Heston (1993) assumes that volatility reverts to a long-term mean at a specified rate. Bates (1996) builds upon the Heston model by introducing a jump component for asset prices, which is represented as a compound Poisson process with normally distributed jumps. In a further refinement of the Heston model, the jumps are characterized by infinite activity jumps generated by a tempered stable process, as demonstrated in Zaevski et al. (2014). Despite these advancements, the pricing challenges associated with path-dependent options in the context of stochastic volatility persist, as no analytical solutions are available for these models.

Chiarella et al. (2012) considered the problem of numerically evaluating barrier option prices when the underlying dynamics are driven by the Heston stochastic volatility model and developed a method of lines approach to evaluate the price as well as the delta and gamma of the option. Park and Kim (2013) investigated a semi-analytic pricing method for lookback options in a general stochastic volatility framework. The resultant formula is well connected to the Black–Scholes price that is the first term of a series expansion, which makes computing the option prices relatively efficient. Furthermore, a convergence condition for the expansion was provided with an error bound. Leung (2013) and Wirtu et al. (2017) derived an analytic pricing formula for floating strike lookback options under the Heston model by means of the homotopy analysis method. The price is given by an infinite series whose value can be determined once an initial term is given well.

In addition, Kato et al. (2013) derived a new semi-closed-form approximation formula for pricing an up-and-out barrier option under a certain type of stochastic volatility model, including an SABR model. In a more recent paper by Funahashi and Higuchi (2018), a unified approximation scheme was proposed for a single-barrier option under local volatility models, stochastic volatility models, and their combinations. The basic idea of their approximation is to mimic a target underlying an asset process using a polynomial of the Wiener process. They then translated the problem of solving the first hit probability of the asset price into the problem of solving that of a Wiener process whose distribution of the passage time is known. Finally, utilizing Girsanov's theorem and the reflection principle, they showed that single-barrier option prices can be approximated in a closed form.

The main contribution of this paper is to derive new closed-form approximation formulas for pricing down-and-out put barrier options and floating strike lookback put options under a certain type of stochastic volatility model, which is similar to the one in Cao et al. (2023); Kato et al. (2013); Kim et al. (2023). To achieve our goal, we apply the asymptotic approach discussed in Fouque et al. (2011) and Mellin transform. Mellin transform techniques were used by Panini and Srivastav (2004) to derive integral equation representations for the price of European and American basket put options. Similarly, Yoon (2014) applied Mellin transform to derive a closed-form solution of the option price with respect to a European call option and a European put option with the Hull–White stochastic interest rate. Moreover, Kim and Yoon (2018) derived a closed-form formula of a second-order approximation for a European corrected option price under a stochastic elasticity of variance (SEV) model.

The rest of the paper is organized as follows. Section 2 discusses the model framework and the features of down-and-out and floating strike lookback put options. In Section 3, we provide detailed discussions on an asymptotic approach, which is used to derive approximations to the risk-netural values of these types of options. In Section 4, we apply Mellin transform to derive a closed-form formula of the first-order approximation for down-and-out barrier put options. In Section 5, we apply Mellin transform to derive a closed-form formula of the first-order approximation for floating strike lookback put options. Section 6 presents a sensitivity and comparison analysis and demonstrates that the results given by these closed-form formulas match well with those generated by Monte-Carlo simulation. Section 7 gives a brief summary. Details on Mellin transform and the derivation of the closed-form formulas in Sections 4 and 5 are provided in Appendices A and B, respectively.

2. Basic Model Set-Up and Path-Dependent Options

2.1. Stochastic Volatility Model

Let $\{S_t : t \geq 0\}$ denote the price process of a risky asset on some filtered probability space $(\Omega, \mathscr{F}, (\mathscr{F}_t)_{t\geq 0}, \mathbb{P})$, where $\mathbb{P}$ is the physical probability measure In this paper, we assume that $\{S_t : t \geq 0\}$ evolves according to the following system of stochastic differential equations:

$$
\begin{aligned}
dS_t &= \mu S_t dt + f(Y_t) S_t dW_t^s, \\
dY_t &= \alpha(m - Y_t)dt + \beta\left(\rho dW_t^s + \sqrt{1-\rho^2} dW_t^y\right),
\end{aligned}
\tag{1}
$$

where μ, $\alpha > 0$, $\beta > 0$, and m are constants and f is a function having positive values and specifying the dependence on the hidden process $\{Y_t : t \geq 0\}$. The processes $\{W_t^s : t \geq 0\}$ and $\left\{W_t^y : t \geq 0\right\}$ are independent standard Brownian motions. The constant correlation coefficient ρ with $-1 < \rho < 1$ captures the leverage effect. Here, μ is the drift rate. The mean-reversion process $\{Y_t : t \geq 0\}$ given in Equation (1) is characterized by its typical time to return back to the mean level m of its long-run distribution. The parameter α determines the speed of mean-reversion, and β controls the volatility of $\{Y_t : t \geq 0\}$. In the sequel, we shall refer to the above system as the stochastic volatility (SV) model. In Sections 2 and 3, we will not specify the concrete form of f, but assume that f is bounded and smooth enough, e.g., $f \in C_0^2(\mathbb{R})$. Furthermore, f has to satisfy a sufficient growth condition in order to avoid bad behavior such as the non-existence of moments of $\{S_t : t \geq 0\}$. For numerical results in Section 6, we choose f to take a special form, as used in Fouque et al. (2000, 2011) and Cao et al. (2021).

We apply the well-known Girsanov theorem to change the physical measure $\mathbb{P}$ to a risk-neutral martingale measure $\mathbb{Q}$ by letting

$$
dW_t^{s*} = \frac{\mu - r}{f(Y_t)} dt + dW_t^s y \quad \text{and} \quad dW_t^{y*} = \xi(Y_t)dt + dW_t^y,
$$

where $\xi(Y_t)$ represents the premium of volatility risk. Then, the model equations under the measure $\mathbb{Q}$ can be written as

$$
\begin{aligned}
dS_t &= rS_t dt + f(Y_t) S_t dW_t^{s*}, \\
dY_t &= \left[\alpha(m - Y_t) - \beta\left(\rho \frac{\mu - r}{f(Y_t)} + \xi(Y_t)\sqrt{1-\rho^2}\right)\right] dt \\
&\quad + \beta\left(\rho dW_t^{s*} + \sqrt{1-\rho^2} dW_t^{y*}\right).
\end{aligned}
\tag{2}
$$

Note that $\{W_t^{s*} : t \geq 0\}$ and $\left\{W_t^{y*} : t \geq 0\right\}$ are independent standard Brownian motions under $\mathbb{Q}$. As an Ornstein–Uhlenbeck (OU) process, $\{Y_t : t \geq 0\}$ in Equation (1) has an invariant distribution, which is normal with mean m and variance $\beta^2/2\alpha$. Thus, we can expect that if mean reversion is very fast, i.e., α goes to infinity, the process $\{S_t : t \geq 0\}$ should be close to a geometric Brownian motion. This means that if mean reversion is extremely fast, then the model of Black and Scholes would become a good approximation. In reality, however, it may not be the case. For fast but not extremely fast mean-reversion, the Black–Scholes model needs to be corrected to account for the random characteristics of the volatility of a risky asset. For this purpose, we introduce another small parameter ϵ defined by $\epsilon = 1/\alpha$, as performed by Fouque et al. (2000). For notational convenience, we put $\nu = \beta/\sqrt{2\alpha}$. With the help of these notations, the model equations under $\mathbb{Q}$ are re-written as

$$dS_t = rS_t dt + f(Y_t)S_t dW_t^{s*},$$

$$dY_t = \left[\frac{1}{\epsilon}(m - Y_t) - \frac{\sqrt{2}\nu}{\sqrt{\epsilon}}\Lambda(Y_t)\right]dt + \frac{\sqrt{2}\nu}{\sqrt{\epsilon}}\left(\rho dW_t^{s*} + \sqrt{1-\rho^2}dW_t^{y*}\right),$$

where $\Lambda(\cdot)$, defined by

$$\Lambda(y) := \rho\frac{\mu - r}{f(y)} + \xi(y)\sqrt{1-\rho^2},$$

is the combined market price of risk.

2.2. Path-Dependent Options

Let $V(T)$ denote the payoff of a put option on the risky asset at its expiration T. Then, its risk-neutral price at time $t \in [0, T]$ under our SV model is given by

$$P(t, s, y) = \mathbb{E}^{\mathbb{Q}}\left(e^{-r(T-t)}V(T)|\, S_t = s, Y_t = y\right).$$

Note that $V(T)$ depends on the type of options. In this paper, we consider two types of path-dependent options: down-and-out put options and floating strike lookback put options. For notational convenience, we put $U_t := \min_{0 \le u \le t} S_u$ and $Z_t := \max_{0 \le u \le t} S_u$. The payoff of a down-and-out put option is given by

$$DOP(T) := \max\{K - S_T, 0\} \times \mathbb{1}_{U_T > B},$$

where K is the strike price, B is the barrier level satisfying $0 < B < K$, and $\mathbb{1}_{U_T > B}$ is the indicator function. For a floating strike lookback put option, its payoff has the form of $LP_{float}(T) := Z_T - S_T$. Applying Itô's lemma, we can obtain a partial differential equation (PDE) for $P(t, s, y)$ as follows:

$$
\begin{aligned}
0 =\ & \frac{\partial P}{\partial t} + \frac{1}{2}s^2 f^2(y)\frac{\partial^2 P}{\partial s^2} + r\left(s\frac{\partial P}{\partial s} - P\right) + \frac{\sqrt{2}\rho\nu s}{\sqrt{\epsilon}}f(y)\frac{\partial^2 P}{\partial s \partial y} \\
& + \frac{\nu^2}{\epsilon}\frac{\partial^2 P}{\partial y^2} + \left(\frac{1}{\epsilon}(m - y) - \frac{\sqrt{2}\nu}{\sqrt{\epsilon}}\Lambda(y)\right)\frac{\partial P}{\partial y}.
\end{aligned}
\tag{3}
$$

The boundary conditions for Equation (3) vary depending on the type of options. For example, the boundary conditions for Equation (3) when $V(T) = DOP(T)$ are

$$
\begin{cases}
P(T, s, y) = \max\{K - s, 0\}, & s > B, \\
P(t, B, y) = 0, & 0 \le t \le T.
\end{cases}
$$

When $V(T) = LP_{float}(T)$, the boundary conditions become the following:

$$
\begin{cases}
\dfrac{\partial P}{\partial z}(t, z, y, z) = 0, & 0 \le t \le T, z > 0, \\
P(T, s, y, z) = z - s, & 0 \le s \le z.
\end{cases}
$$

Note that in this case, P is a function of t, s, y, and z (here, $Z_t = z$).

Remark: Since the Mellin transform of the payoff function of a call option is not defined, this paper primarily concentrates on evaluating put options. However, as outlined in Buchen (2001), the pricing of call options can be directly derived from put options through the put–call parity relationship.

3. Asymptotic Expansions

In this section, we apply an asymptotic expansion approach to establish partial differential equations, which will be used to derive an approximate solution to Equation (3) and thus find an approximated value of a put option.

We begin with re-organizing Equation (3) in terms of the orders of ϵ as follows:

$$\frac{1}{\epsilon}\mathcal{L}_0 P + \frac{1}{\sqrt{\epsilon}}\mathcal{L}_1 P + \mathcal{L}_2 P = 0, \tag{4}$$

where the operators $\mathcal{L}_0$, $\mathcal{L}_1$ and $\mathcal{L}_2$ are defined by

$$\mathcal{L}_0 := (m - y)\frac{\partial}{\partial y} + v^2 \frac{\partial^2}{\partial y^2},$$

$$\mathcal{L}_1 := \sqrt{2}\rho v s f(y)\frac{\partial^2}{\partial s \partial y} - \sqrt{2}v\Lambda(y)\frac{\partial}{\partial y}, \text{ and}$$

$$\mathcal{L}_2 := \frac{\partial}{\partial t} + \frac{1}{2}s^2 f^2(y)\frac{\partial^2}{\partial s^2} + r\left(s\frac{\partial}{\partial s} - \cdot\right).$$

In order to obtain an efficient approximate solution to P, as that in Fouque et al. (2011), we apply the following asymptotic expansion of P:

$$P = P_0 + \sqrt{\epsilon}P_1 + \epsilon P_2 + \epsilon\sqrt{\epsilon}P_3 + \ldots, \tag{5}$$

where P_0, P_1, ... are functions corresponding to varying orders of ϵ. Substituting P in Equation (5) into Equation (4) and re-organizing the terms, we obtain

$$\begin{aligned}
0 =& \frac{1}{\epsilon}\mathcal{L}_0 P_0 + \frac{1}{\sqrt{\epsilon}}(\mathcal{L}_1 P_0 + \mathcal{L}_0 P_1) + (\mathcal{L}_0 P_2 + \mathcal{L}_1 P_1 + \mathcal{L}_2 P_0) \\
&+ \sqrt{\epsilon}(\mathcal{L}_0 P_3 + \mathcal{L}_1 P_2 + \mathcal{L}_2 P_1) + \ldots.
\end{aligned} \tag{6}$$

Our aim is to find P_0 and P_1.

Firstly, from the $O(1/\epsilon)$-order term in Equation (6), we obtain $\mathcal{L}_0 P_0 = 0$. If we assume that P_0 does not grow as fast as $e^{y^2/2}$, as was assumed in Choi et al. (2013), we can show that P_0 is independent of y. Secondly, from the $O(1/\sqrt{\epsilon})$-order term in Equation (6), we obtain $\mathcal{L}_1 P_0 + \mathcal{L}_0 P_1 = 0$. Since P_0 is independent of y, then $\mathcal{L}_1 P_0 = 0$. It follows that $\mathcal{L}_0 P_1 = 0$. Again, if we assume that P_1 does not grow as fast as $e^{y^2/2}$, then we can deduce that P_1 is also independent of y.

Next, from the $O(1)$-order term in Equation (6), we obtain

$$\mathcal{L}_0 P_2 + \mathcal{L}_1 P_1 + \mathcal{L}_2 P_0 = 0.$$

Since P_1 is independent of y, we have $\mathcal{L}_1 P_1 = 0$, which implies that

$$\mathcal{L}_0 P_2 + \mathcal{L}_2 P_0 = 0. \tag{7}$$

Seeing Equation (7) as a Poisson equation for P_2 in y, in order for it to have a solution, it is required to satisfy the centering condition

$$\langle\mathcal{L}_2 P_0\rangle = \langle\mathcal{L}_2\rangle P_0 = 0, \tag{8}$$

which is equivalent to

$$\frac{\partial P_0}{\partial t} + rs\frac{\partial P_0}{\partial s} + \frac{1}{2}s^2\langle f^2\rangle\frac{\partial^2 P_0}{\partial s^2} - rP_0 = 0. \tag{9}$$

This is an equation for us to determine the P_0 term. Here, $\langle\cdot\rangle$ denotes the expectation with respect to the invariant distribution of the process $\{Y_t : t \geq 0\}$, i.e.,

$$\langle h\rangle = \int_{-\infty}^{+\infty} h(y)\Phi(y)dy, \quad \text{where} \quad \Phi(y) = \frac{1}{\sqrt{2\pi v^2}}e^{-\frac{(y-m)^2}{2v^2}}.$$

Note that a small ϵ value corresponds to fast-mean reverting. In this case, Y_t approaches a constant and $\langle f^2\rangle$ can be regarded as constant variance, and then Equation (9)

is the Black–Scholes PDE. Thus, for small ϵ, P_0 represents the put option price under the Black–Scholes model.

Following Equation (8), we have

$$\mathcal{L}_2 P_0 = \mathcal{L}_2 P_0 - \langle \mathcal{L}_2 \rangle P_0 = \frac{1}{2}\left(f^2 - \langle f^2 \rangle\right)s^2\frac{\partial^2 P_0}{\partial s^2},$$

which, together with Equation (7), implies

$$\mathcal{L}_0 P_2 = -\frac{1}{2}\left(f^2 - \langle f^2 \rangle\right)s^2\frac{\partial^2 P_0}{\partial s^2}. \tag{10}$$

The solution to Equation (10) can be expressed as

$$P_2 = -\frac{1}{2}(\phi + c)s^2\frac{\partial^2 P_0}{\partial s^2}, \tag{11}$$

where ϕ is a function of y which only satisfies the equation $\mathcal{L}_0\phi = f^2 - \langle f^2 \rangle$, and c is a function of other variables except y.

To derive an equation for P_1, we consider the $O(\sqrt{\epsilon})$-term in Equation (6) and obtain

$$\mathcal{L}_0 P_3 + \mathcal{L}_1 P_2 + \mathcal{L}_2 P_1 = 0.$$

This equation can be regarded as a Poisson equation for P_3 in y, and in order for it to have a solution, the following centering condition must be satisfied:

$$\langle \mathcal{L}_1 P_2 + \mathcal{L}_2 P_1 \rangle = 0. \tag{12}$$

After we substitute P_2 in Equation (11) into Equation (12) and make simplifications, we obtain

$$\frac{\partial P_1}{\partial t} + \frac{1}{2}\langle f^2 \rangle s^2\frac{\partial^2 P_1}{\partial s^2} + rs\frac{\partial P_1}{\partial s} - rP_1 = c_1 s^3\frac{\partial^3 P_0}{\partial s^3} + c_2 s^2\frac{\partial^2 P_0}{\partial s^2}, \tag{13}$$

where

$$c_1 := \frac{\sqrt{2}}{2}\langle f\phi' \rangle\rho\nu \quad \text{and} \quad c_2 := \frac{\sqrt{2}}{2}\left(2\rho\langle f\phi' \rangle - \langle \Lambda\phi' \rangle\right)\nu. \tag{14}$$

This is an equation for us to determine the first correction term P_1.

We summarize the previous formal analysis as the following theorem.

Theorem 1. *Under the SV model governed by Equation (1), an approximation of the risk-neutral value P of a path-dependent put option is given by*

$$P = P_0 + \sqrt{\epsilon}P_1 + o(\sqrt{\epsilon}), \tag{15}$$

for small ϵ, where P_0 and P_1 are determined by Equations (9) and (13) with corresponding boundary conditions, respectively, such that P_0 is the put option price under the Black–Scholes model with constant effective volatility $\sqrt{\langle f^2 \rangle}$ and P_1 is the first-order correction term.

Finally, as mentioned in Section 2, boundary conditions for Equations (8) and (13) depend on the types of options that we consider. We describe the corresponding boundary conditions and solve these equations in the next two sections.

4. Determining P_0 and P_1 for Down-and-Out Put Options

In this section, we use Mellin transform to derive analytical expressions of the P_0 and P_1 terms for down-and-out put options

4.1. P_0 Term for Down-and-Out Put Options

In order to use Mellin transform to calculate the P_0 term for down-and-out put options, noting that P_0 is independent of y under our assumption, we first follow the method in Buchen (2001) and use the boundary condition,

$$P(T, s, y) = \max\{K - s, 0\}, \quad \text{for} \quad s > B,$$

to set up the boundary condition of P_0 for $s \geq 0$ as follows:

$$P_0(T, s) := (K - s)\mathbb{1}_{B<s<K} - \left(\frac{B}{s}\right)^{k_1-1}\left(K - \frac{B^2}{s}\right)\mathbb{1}_{\frac{B^2}{K}<s<B'} \tag{16}$$

where $k_1 = 2r/\langle f^2 \rangle$. Now, we apply Mellin transform to Equation (9) to convert this PDE into the following ODE:

$$\frac{d\hat{P}_0}{dt} + \left(\frac{1}{2}\langle f^2 \rangle (w^2 + w) - rw - r\right)\hat{P}_0 = 0. \tag{17}$$

The solution to Equation (17) is given by

$$\hat{P}_0(t, w) = \hat{\theta}(w)e^{\frac{1}{2}\langle f^2 \rangle \left(w^2 + (1-k_1)w - k_1\right)(T-t)}, \tag{18}$$

where $\hat{\theta}$ is a function of w, determined by the boundary condition (16).

Next, we take inverse Mellin transform of Equation (18) and obtain

$$P_0(t, s) = P_0(T, s) * \mathcal{M}^{-1} e^{\lambda(w+\eta)^2 + \delta},$$

where

$$\lambda = \frac{1}{2}\langle f^2 \rangle (T - t), \ \eta = \frac{1 - k_1}{2}, \ \delta = -\lambda\eta^2 - r(T - t)$$

and the operation $*$ means the convolution. Applying Table A1 in Appendix A and the boundary condition given in Equation (16), we have

$$
\begin{aligned}
P_0(t, s) &= P_0(T, s) * \left(\frac{e^\delta s^\eta}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}(\ln s)^2}\right) \\
&= \int_B^K (K - u)e^\delta \left(\frac{s}{u}\right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{s}{u}\right)\right)^2}\right)\frac{du}{u} - \\
&\quad \int_{\frac{B^2}{K}}^B \left(\frac{B}{u}\right)^{k_1-1}\left(K - \frac{B^2}{u}\right)e^\delta \left(\frac{s}{u}\right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{s}{u}\right)\right)^2}\right)\frac{du}{u}.
\end{aligned} \tag{19}
$$

After some careful calculation, for down-and-out put options, we derive a closed-form expression of the P_0 term as follows:

$$
\begin{aligned}
P_0(t, s) &= Ke^{-r(T-t)}\left(\Phi\left(-\Delta_-\left(\frac{s}{K}\right)\right) - \Phi\left(-\Delta_-\left(\frac{s}{B}\right)\right)\right) - \\
&\quad s\left(\Phi\left(-\Delta_+\left(\frac{s}{K}\right)\right) - \Phi\left(-\Delta_+\left(\frac{s}{B}\right)\right)\right) - \\
&\quad Ke^{-r(T-t)}\left(\frac{B}{s}\right)^{k_1-1}\left[\Phi\left(\Delta_-\left(\frac{B}{s}\right)\right) - \Phi\left(\Delta_-\left(\frac{B^2}{sK}\right)\right)\right] + \\
&\quad B\left(\frac{B}{s}\right)^{k_1}\left[\Phi\left(\Delta_+\left(\frac{B}{s}\right)\right) - \Phi\left(\Delta_+\left(\frac{B^2}{sK}\right)\right)\right],
\end{aligned} \tag{20}
$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution and

$$\Delta_\pm(x) = \frac{1}{\sqrt{\langle f^2 \rangle(T-t)}}\left[\ln(x) + \left(r \pm \frac{1}{2}\langle f^2 \rangle\right)(T-t)\right].$$

Note that P_0 given in Equation (20) is precisely the same as the price of a down-and-out put option given in the literature, e.g., Hull (2015, chp. 26, p. 606) or Haug (2006, chp. 4), if we let $\sigma^2 = \langle f^2 \rangle$. For details of the derivation of formula (20), we refer the reader to Appendix B.

4.2. P_1 Term for Down-and-Out Put Options

For down-and-out put options, the boundary conditions for P_1 are

$$
\begin{cases}
P_1(T,s) & = & 0, & for \ s \geq B, \\
P_1(t,B) & = & 0, & for \ 0 < t < T.
\end{cases}
$$

We again follow the method in Buchen (2001) and extend the boundary conditions $P_1(T,s) = 0$, for $s \geq B$ as $P_1(T,s) = 0$ for all $s \geq 0$.

Next, we apply Mellin transform to Equation (13) to obtain

$$
\frac{d\hat{P}_1}{dt} + \left(\frac{1}{2} \langle f^2 \rangle \left(w^2 + w \right) - rw - r \right) \hat{P}_1 = (-c_1 w(w+1)(w+2) + c_2 w(w+1)) \hat{P}_0.
$$

Solving this equation, we obtain

$$
\hat{P}_1(t,w) = \left[c_1(T-t)w^3 - (c_2 - 3c_1)(T-t)w^2 - (c_2 - 2c_1)(T-t)w \right] \hat{P}_0(t,w).
$$

Finally, applying inverse Mellin transform, we obtain an explicit closed-form expression of P_1 as follows:

$$
\begin{aligned}
P_1(t,s) & = & \mathcal{M}^{-1}\left(\hat{P}_1(t,w) \right) \\
& = & c_1(T-t)\left(-s\frac{d}{ds}P_0(t,s) - 3s^2\frac{d^2}{ds^2}P_0(t,s) - s^3\frac{d^3}{ds^3}P_0(t,s) \right) \\
& & -(c_2 - 3c_1)(T-t)\left(s\frac{d}{ds}P_0(t,s) + s^2\frac{d^2}{ds^2}P_0(t,s) \right) \\
& & -(c_2 - 2c_1)(T-t)\left(-s\frac{d}{ds}P_0(t,s) \right),
\end{aligned}
\tag{21}
$$

where P_0 is given in the previous section and c_1 and c_2 are given in Equation (14).

We summarize the above analysis and calculation on down-and-out put options in the following theorem.

Theorem 2. *Under the SV model governed by Equation* (1), *an approximation of the risk-neutral value P of a down-and-out barrier put option is given by*

$$
P = P_0 + \sqrt{\epsilon}P_1 + o(\sqrt{\epsilon}),
\tag{22}
$$

where P_0 and P_1 are given by Equations (20) *and* (21), *respectively.*

5. Determining P_0 and P_1 for Lookback Put Options

In this section, we use Mellin transform to derive analytical expressions of the P_0 and P_1 terms for floating strike lookback put options

5.1. P_0 Term for Lookback Put Options

For lookback floating strike put options, the boundary conditions of P_0 are

$$
\begin{cases}
\dfrac{\partial P_0}{\partial z}(t,z,z) & = & 0, \\
\dfrac{\partial P_0}{\partial z}(T,s,z) & = & 1, & for \ 0 < s < z.
\end{cases}
$$

Similar to the case of down-and-out put options, we extend the second boundary condition to $0 < s < \infty$ as follows:

$$\frac{\partial P_0}{\partial z}(T,s,z) := \mathbb{1}_{s<z} - \left(\frac{z}{s}\right)^{k_1-1} \cdot \mathbb{1}_{z<s}, \quad \text{for} \quad 0 < s < \infty.$$

Then, by integrating each side of the last equation, we can obtain

$$P_0(T,s,z) = \int_s^z -\left(\frac{\xi}{s}\right)^{k_1-1} d\xi = -\frac{1}{k_1}\left(\frac{z}{s}\right)^{k_1}s + \frac{1}{k_1}s \tag{23}$$

for $s > z$. For convenience, we let $u = s/z$ and $Q_0 = P_0/z$. With these notations, Equation (9) becomes

$$\frac{\partial Q_0}{\partial t} + \frac{1}{2}u^2\langle f^2\rangle\frac{\partial^2 Q_0}{\partial u^2} + ru\frac{\partial Q_0}{\partial u} - rQ_0 = 0, \tag{24}$$

with boundary conditions

$$Q_0(T,u) = -\frac{1}{k_1}u^{1-k_1} + \frac{1}{k_1}u, \quad \text{for} \quad u > 1, \tag{25}$$

and $Q_0(T,u) = 1$, for $0 < u < 1$.

Note that except the boundary conditions, Equation (24) is identical to Equation (9). Applying Mellin transform in the same way as that for the case of down-and-out put options, we can derive the solution to Equation (24) as follows:

$$Q_0(t,u) = \hat{\theta}(w) * \mathcal{M}^{-1}e^{\lambda(w+\eta)^2+\delta}.$$

Again, applying Table A1 and P_0 given in Equation (16), we have

$$\begin{aligned}
Q_0(t,u) &= Q_0(T,u) * e^{\delta}z^{\eta}\left(\frac{1}{2\sqrt{\pi}}\lambda^{-\frac{1}{2}}e^{-\frac{1}{4\lambda}(\ln z)^2}\right) \\
&= \int_0^1 (1-\xi)e^{\delta}\left(\frac{u}{\xi}\right)^{\eta}\left(\frac{1}{2\sqrt{\pi}}\lambda^{-\frac{1}{2}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{u}{\xi}\right)\right)^2}\right)\frac{d\xi}{\xi} + \\
&\quad \int_1^{\infty}\left(\frac{-1}{k_1}\xi^{1-k_1} + \frac{\xi}{k_1}\right)e^{\delta}\left(\frac{u}{\xi}\right)^{\eta}\left(\frac{1}{2\sqrt{\pi}}\lambda^{-\frac{1}{2}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{u}{\xi}\right)\right)^2}\right)\frac{d\xi}{\xi}.
\end{aligned} \tag{26}$$

After calculating integrals, for floating strike lookback put options, we derive a closed-form expression of the P_0 term as follows:

$$\begin{aligned}
P_0(t,s,z) &= ze^{-r(T-t)}\Phi\left(-\Delta_-\left(\frac{s}{z}\right)\right) - s\Phi\left(-\Delta_+\left(\frac{s}{z}\right)\right) \\
&\quad -\frac{z}{k_1}\left(\frac{s}{z}\right)^{1-k_1}e^{-r(T-t)}\Phi\left(-\Delta_-\left(\frac{z}{s}\right)\right) + \frac{s}{k_1}\Phi\left(\Delta_+\left(\frac{s}{z}\right)\right),
\end{aligned} \tag{27}$$

where $\Phi(\cdot)$ is the CDF of the standard normal distribution. Note that P_0 given in Equation (27) is precisely the same as the price of a floating strike put option given in the literature, e.g., Hull (2015, chp. 26, p. 608) or Haug (2006, chp. 4), if we let $\sigma^2 := \langle f^2\rangle$. Details of the derivation of this formula can be found in Appendix B.

5.2. P_1 Term for Lookback Put Options

For floating strike lookback put options, the boundary conditions for P_1 are

$$\begin{cases}
P_1(T,s,z) &= 0, \quad \text{for } 0 < s < z, \\
\frac{\partial P_1}{\partial z}(t,z,z) &= 0, \quad \text{for } 0 < t < T \text{ and } z > 0.
\end{cases}$$

Just like that for the P_0-term for floating strike lookback put options, we let $u = s/z$ and $Q_1 = P_1/z$. With these notation changes, Equation (13) is converted to the following:

$$\frac{\partial Q_1}{\partial t} + \frac{1}{2}\langle f^2 \rangle u^2 \frac{\partial^2 Q_1}{\partial u^2} + ru\frac{\partial Q_1}{\partial u} - rQ_1 = c_1 u^3 \frac{\partial^3 Q_0}{\partial u^3} + c_2 u^2 \frac{\partial^2 Q_0}{\partial u^2} \tag{28}$$

with $Q_1(T, u) = 0$ for $0 < u < 1$.

Note that Equation (28) is essentially the same as Equation (13), except the notational difference. So, we have

$$
\begin{aligned}
Q_1(t, u) \;=\; & c_1(T - t)\left(-u\frac{d}{du}Q_0(t, u) - 3u^2\frac{d^2}{du^2}Q_0(t, u) - u^3\frac{d^3}{du^3}Q_0(t, u)\right) \\
& -(c_2 - 3c_1)(T - t)\left(u\frac{d}{du}Q_0(t, u) + u^2\frac{d^2}{dz^2}Q_0(t, u)\right) \\
& -(c_2 - 2c_1)(T - t)\left(-u\frac{d}{du}Q_0(t, u)\right),
\end{aligned}
\tag{29}
$$

where Q_0 is given previously. Consequently, we have

$$
\begin{aligned}
P_1(t, s, z) \;=\; & c_1(T - t)\left(-s\frac{d}{ds}P_0(t, s, z) - 3s^2\frac{d^2}{ds^2}P_0(t, s, z) - s^3\frac{d^3}{ds^3}P_0(t, s, z)\right) \\
& -(c_2 - 3c_1)(T - t)\left(s\frac{d}{ds}P_0(t, s, z) + s^2\frac{d^2}{ds^2}P_0(t, s, z)\right) \\
& -(c_2 - 2c_1)(T - t)\left(-s\frac{d}{ds}P_0(t, s, z)\right),
\end{aligned}
\tag{30}
$$

where c_1 and c_2 are the same as those defined previously.

We summarize the above analysis and calculation on floating strike lookback put options in the following theorem.

Theorem 3. *Under the SV model governed by Equation* (1), *an approximation of the risk-neutral value P of a floating strike lookback put option is given by*

$$P = P_0 + \sqrt{\epsilon}P_1 + o(\sqrt{\epsilon}), \tag{31}$$

where P_0 and P_1 are given by Equations (27) *and* (30), *respectively.*

6. Numerical Results and Sensitivity Analysis

In this section, we conduct a numerical study to investigate the sensitivity of the first-order correction term P_1 and our approximation results $P_0 + \sqrt{\epsilon}P_1$ with respect to the initial value of underlying asset. This means that we set $t = 0$ throughout this section. We also compare the results given by our closed form formulas with those generated by the Monte-Carlo simulation.

First of all, as conducted by Fouque et al. (2000, 2011) and Cao et al. (2021), we choose f to take the following form:

$$f(y) = 0.35\left(\tan^{-1}(y) + \frac{\pi}{2}\right)/\pi + 0.05.$$

Secondly, the values of other parameters used in this section are given in Table 1 whenever they are required to be fixed.

Table 1. The role and numerical value of parameters.

Parameter	Role	Value
r	risk-free interest rate	0.035
B	barrier level	1500
K	put option strike price	2700
T	maturity time	1
c_1	as defined in Section 3	-0.004
c_2	as defined in Section 3	-0.018

Here, we do not choose precise values of β and ρ, and particular forms of $\xi(y)$ (in Section 2) and $\phi(y)$ (in Section 3) to calculate the above values of c_1 and c_2. Instead, c_1 and c_2 are calibrated from the term structure of the implied volatility surface as described in the book of Fouque et al. (2000). Specifically, the implied volatility I^ϵ of a European vallina call option with fast mean-reverting stochastic process can be approximated by the following formula:

$$I^\epsilon = a\frac{\ln(\frac{K}{s})}{T-t} + b + o(\sqrt{\epsilon})$$

with

$$a = -\frac{c_1}{\langle f^2 \rangle^{3/2}} \quad \text{and} \quad b = \sqrt{\langle f^2 \rangle} + \frac{c_1}{\langle f^2 \rangle^{3/2}}\left(r + \frac{3}{2}\langle f^2 \rangle\right) - \frac{c_2}{\sqrt{\langle f^2 \rangle}}.$$

The parameters a and b are estimated as the slope and intercept of the regression fit of the observed implied volatilities as a linear function of logmoneyness-to-maturity-ratio $\ln(K/s)/(T-t)$. From the calibrated values a and b on the observed implied volatility surface, the parameters c_1 and c_2 are obtained as

$$c_1 = -a\sigma\langle f^2 \rangle^{3/2} \quad \text{and} \quad c_2 = \sqrt{\langle f^2 \rangle}((\sqrt{\langle f^2 \rangle} - b) - a(r + \frac{3}{2}\langle f^2 \rangle)).$$

Thirdly, note that when $t = 0$, $s = z$. Hence, in this case, the formula for P_0 given by Equation (27) is simplified.

Figure 1a shows how the $\sqrt{\epsilon}P_1$-term for a down-and-out barrier put option changes with respect to a variation in ϵ values As we can see, for fixed ϵ, when s increases, P_1 decreases first, and then increases after it hits its trough. When ϵ becomes smaller (equivalently, the mean-reverting speed becomes larger), $\sqrt{\epsilon}P_1$ approaches to a zero. Figure 1b shows how the value of $P_0 + \sqrt{\epsilon}P_1$ for a down-and-out put option varies with respect to the change in ϵ values. As we can see, when the value of ϵ changes from 0.01 to 0.0001, the value of $P_0 + \sqrt{\epsilon}P_1$ does not vary much. In fact, the values of $P_0 + \sqrt{\epsilon}P_1$ match well with the result of Monte-Carlo simulation in all cases. Furthermore, in all cases, the value of $P_0 + \sqrt{\epsilon}P_1$ declines as s increases.
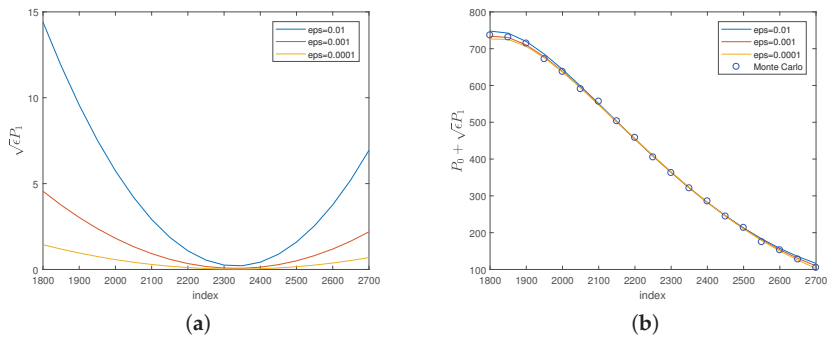
Figure 1. Plots of $\sqrt{\epsilon}P_1$ and $P_0 + \sqrt{\epsilon}P_1$ with different values of ϵ against the initial value of the underlying asset, for the down-and-out put option.

Figure 2a shows how the $\sqrt{\epsilon}P_1$-term for a floating strike lookback put changes with respect to a variation in ϵ values. In a similar pattern, for a fixed ϵ-value, when s increases, P_1 decreases first and then increases after it hits its trough. Similar to the case of down-and-out put options, when ϵ becomes smaller (equivalently, the mean-reverting speed becomes larger), $\sqrt{\epsilon}P_1$ approaches to zero. Figure 2b shows how the value of $P_0 + \sqrt{\epsilon}P_1$ for a floating strike put varies with respect to the change in ϵ values. When the value of ϵ changes from 0.01 to 0.001, the value of $P_0 + \sqrt{\epsilon}P_1$ varies. But when the value of ϵ changes from 0.001 to 0.0001, the value of $P_0 + \sqrt{\epsilon}P_1$ does not vary much. The values of $P_0 + \sqrt{\epsilon}P_1$ match well

with the result of Monte-Carlo simulation when $\epsilon = 0.001$ or 0.0001. Furthermore, in all cases, the value of $P_0 + \sqrt{\epsilon}P_1$ increases as s increases.
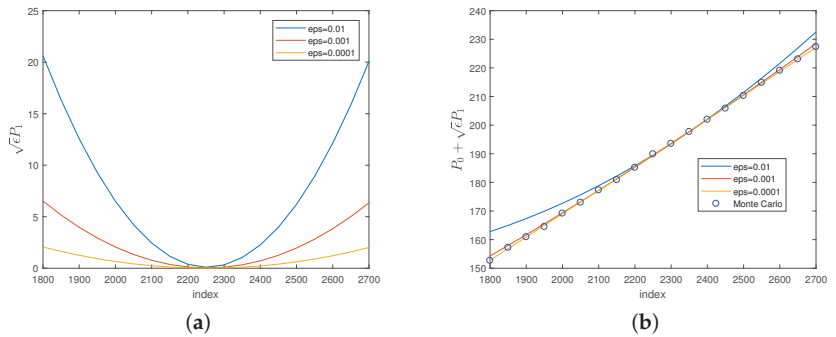
Figure 2. Plots of $\sqrt{\epsilon}P_1$ and $P_0 + \sqrt{\epsilon}P_1$ with different values of ϵ, against the initial value of the underlying asset, for floating strike put options.

Figure 3a illustrates the variation in the value of $P_0 + \sqrt{\epsilon}P_1$ for a down-and-out put option in response to changes in ρ values. As ρ shifts from -0.6 to -0.4, there is a slight decrease in the value of $P_0 + \sqrt{\epsilon}P_1$. Additionally, in all scenarios, the value of $P_0 + \sqrt{\epsilon}P_1$ shows an upward trend as s increases.
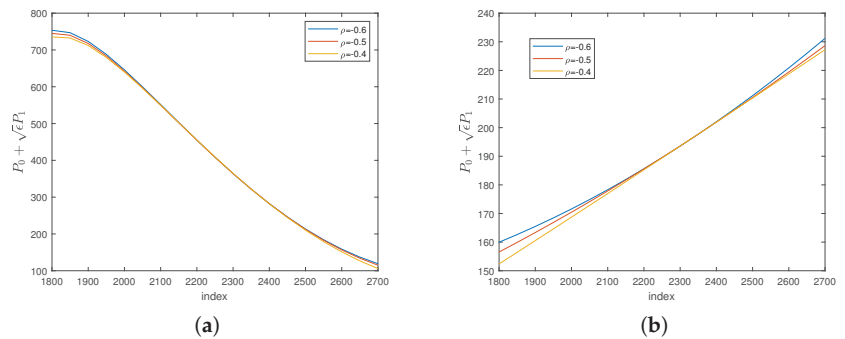
Figure 3. Plots $P_0 + \sqrt{\epsilon}P_1$ with different values of ρ, against the initial value of the underlying asset, for down-and-out put option and floating strike put option.

Figure 3b depicts the change in the value of $P_0 + \sqrt{\epsilon}P_1$ for a floating strike put concerning variations in ρ values. Similar to the previous case, a shift in ρ from -0.6 to -0.4 results in a minor decline in the value of $P_0 + \sqrt{\epsilon}P_1$. Moreover, across all instances, an increase in s is associated with a rise in the value of $P_0 + \sqrt{\epsilon}P_1$.

7. Concluding Remarks

This article establishes explicit closed-form solutions for first order approximations of down-and-out barrier and floating strike lookback put option prices under a stochastic volatility model by means of Mellin transform. The zero-order terms in the solutions for the prices of both types of put options coincide with those in Hull (2015) or Haug (2006) under the classical Back–Scholes model. Our numerical analysis shows that the results given by those explicit closed-form solutions match well with those generated by the Monte-Carlo simulation. This confirms the accuracy of the approximation. Furthermore, we also discussed the sensitivity of the first-order error terms and the approximation with respect to the underlying asset price and the mean-reverting speed of the OU-process which governs the volatility.

This model formula can be employed by financial professionals for the swift and precise pricing of barrier and lookback options. This is demonstrated by the efficiency of our formula in comparison to the conventional Monte-Carlo method. Our pricing formula offers an effective means of assessing barrier and lookback options. Looking ahead, we may extend our methodology to evaluate other path-dependent options in future works, including, but not limited to, Asian options, Russian options, and more.

Appendix A. Mellin Transform

The Mellin transform is an integral transform that may be regarded as the multiplicative version of the two-sided Laplace transform. It is often used in the theory of asymptotic expansions. For a locally Lebesgue integrable function $h : \mathbb{R}^+ \to \mathbb{R}$, the Mellin transform denoted by $\mathcal{M}h$ or $\hat{h}$ is given by

$$\hat{h}(w) = (\mathcal{M}h)(w) := \int_0^{+\infty} s^{w-1} h(s)\, ds, \quad w \in \mathbb{C},$$

and if $a < \mathrm{Re}(w) < b$ and c such that $a < c < b$ exists, the inverse of the Mellin transform is expressed by

$$h(s) = \left(\mathcal{M}^{-1}\hat{h}\right)(s) = \frac{1}{2\pi i} \int_{c-i\infty}^{c+i\infty} s^{-w} \hat{h}(w)\, dw.$$

In this paper, we use the following properties of Mellin transform.

Table A1. List of properties of Mellin transform used in this paper.

Function	Mellin Tansform
h	$\hat{h}$
sh'	$-w\hat{h}$
$s^2 h''$	$w(w+1)\hat{h}$
$s^3 h^{(3)}$	$-w(w+1)(w+2)\hat{h}$
$\frac{e^\delta s^\eta}{2\sqrt{\lambda\pi}} e^{-\frac{1}{4\lambda}(\ln s)^2}$	$e^{\lambda(w+\eta)^2+\delta}$
$sh' + s^2 h''$	$w^2\hat{h}$
$-sh' - 3s^2 h'' - s^3 h^{(3)}$	$w^3\hat{h}$

Here, λ, η, and δ are not related to w or s, and h', h'', and $h^{(3)}$ are the first-order, second-order, and third-order derivatives of h, respectively.

Appendix B. Derivation of Formulas (20) and (27)

In this appendix, we give detailed derivation of the Formulas (20) and (27).

Appendix B.1. Derivation of Formula (20)

From Equation (19), we know that

$$P_0(t,s) = \int_B^K (K-u)e^\delta \left(\frac{s}{u}\right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{s}{u}\right)\right)^2}\right)\frac{du}{u} -$$

$$\int_{\frac{B^2}{K}}^B \left(\frac{B}{u}\right)^{k_1-1}\left(K-\frac{B^2}{u}\right)e^\delta \left(\frac{s}{u}\right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{s}{u}\right)\right)^2}\right)\frac{du}{u}.$$

By letting $v = \ln u$, we convert the first integral to

$$\int_{\ln B}^{\ln K}(K-e^v)s^\eta e^\delta e^{-\eta v}\left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}(\ln s - v)^2}\right)dv$$

$$= \frac{s^\eta e^\delta}{2\sqrt{\lambda\pi}}\left(\int_{\ln B}^{\ln K}Ke^{-\frac{1}{4\lambda}(v^2-2v\ln s+(\ln s)^2+4\lambda\eta v)}dv\right.$$

$$\left.-\int_{\ln B}^{\ln K}e^{-\frac{1}{4\lambda}(v^2-2v\ln s+(\ln s)^2+4\lambda(\eta-1)v)}dv\right)$$

$$= \frac{s^\eta e^\delta}{2\sqrt{\lambda\pi}}\left(\int_{\ln B}^{\ln K}Ke^{-\frac{1}{4\lambda}(v-\ln s+2\lambda\eta)^2+\lambda\eta^2-\eta\ln s}dv\right.$$

$$\left.-\int_{\ln B}^{\ln K}e^{-\frac{1}{4\lambda}[v-\ln s+2\lambda(\eta-1)]^2+\lambda(\eta-1)^2-(\eta-1)\ln s}dv\right),$$

we further apply the following changes in variables:

$$x' := \frac{v-\ln s+2\lambda\eta}{\sqrt{2\lambda}} \quad \text{and} \quad x'' := \frac{v-\ln s+2\lambda(\eta-1)}{\sqrt{2\lambda}}$$

to obtain

$$\int_{\ln B}^{\ln K}(K-e^v)s^\eta e^\delta e^{-\eta v}\left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}(\ln s-v)^2}\right)dv$$

$$= \frac{e^\delta}{\sqrt{2\pi}}\left(Ke^{\lambda\eta^2}\int_{\frac{\ln(\frac{B}{s})+2\lambda\eta}{\sqrt{2\lambda}}}^{\frac{\ln(\frac{K}{s})+2\lambda\eta}{\sqrt{2\lambda}}}e^{-\frac{x'^2}{2}}dx'-se^{\lambda(\eta-1)^2}\int_{\frac{\ln(\frac{B}{s})+2\lambda(\eta-1)}{\sqrt{2\lambda}}}^{\frac{\ln(\frac{K}{s})+2\lambda(\eta-1)}{\sqrt{2\lambda}}}e^{-\frac{x''^2}{2}}dx''\right)$$

$$= Ke^{\delta+\lambda\eta^2}\left[\Phi\left(\frac{\ln(\frac{K}{s})+2\lambda\eta}{\sqrt{2\lambda}}\right)-\Phi\left(\frac{\ln(\frac{B}{s})+2\lambda\eta}{\sqrt{2\lambda}}\right)\right]$$

$$-se^{\delta+\lambda(\eta-1)^2}\left[\Phi\left(\frac{\ln(\frac{K}{s})+2\lambda(\eta-1)}{\sqrt{2\lambda}}\right)-\Phi\left(\frac{\ln(\frac{B}{s})+2\lambda(\eta-1)}{\sqrt{2\lambda}}\right)\right].$$

Now, if we plug into δ, η, and λ into the above formula, we derive

$$\int_{\ln B}^{\ln K}(K-e^v)s^\eta e^\delta e^{-\eta v}\left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}(\ln s-v)^2}\right)dv$$

$$= Ke^{-r(T-t)}\left[\Phi\left(-\Delta_-\left(\frac{s}{K}\right)\right)-\Phi\left(-\Delta_-\left(\frac{s}{B}\right)\right)\right]$$

$$-s\left[\Phi\left(-\Delta_+\left(\frac{s}{K}\right)\right)-\Phi\left(-\Delta_+\left(\frac{s}{B}\right)\right)\right].$$

Similarly, we can evaluate the second integral

$$\int_{\frac{B^2}{K}}^B \left(\frac{B}{u}\right)^{k_1-1}\left(K-\frac{B^2}{u}\right)e^\delta \left(\frac{s}{u}\right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{s}{u}\right)\right)^2}\right)\frac{du}{u}$$

to obtain

$$
Ke^{-r(T-t)} \left(\frac{B}{s}\right)^{k_1-1} \left[\Phi\left(\Delta_-\left(\frac{B}{s}\right)\right) - \Phi\left(\Delta_-\left(\frac{B^2}{sK}\right)\right)\right]
$$

$$
-B\left(\frac{B}{s}\right)^{k_1}\left[\Phi\left(\Delta_+\left(\frac{B}{s}\right)\right) - \Phi\left(\Delta_+\left(\frac{B^2}{sK}\right)\right)\right].
$$

Putting these two integrals together yields Formula (20).

Appendix B.2. Derivation of Formulas (27)

From Equation (26), we have

$$
Q_0(t,u) = \int_0^1 (1-\xi)e^\delta \left(\frac{u}{\xi}\right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{u}{\xi}\right)\right)^2}\right)\frac{d\xi}{\xi} +
$$

$$
\int_1^\infty \left(-\frac{1}{k_1}\xi^{1-k_1} + \frac{\xi}{k_1}\right)e^\delta \left(\frac{u}{\xi}\right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{u}{\xi}\right)\right)^2}\right)\frac{d\xi}{\xi}.
$$

We let $v = \ln\xi$. For the first integral, we have

$$
\int_0^1 (1-\xi)e^\delta \left(\frac{u}{\xi}\right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{u}{\xi}\right)\right)^2}\right)\frac{du}{u}
$$

$$
= \int_{-\infty}^0 u^\eta(1-e^v)e^{\delta-v\eta}\left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}(\ln u - v)^2}\right)dv
$$

$$
= \frac{u^\eta e^\delta}{2\sqrt{\lambda\pi}}\left(\int_{-\infty}^0 e^{-\frac{1}{4\lambda}\left(v^2 - 2v\ln u + (\ln u)^2 + 4\lambda\eta v\right)}dv\right.
$$

$$
\left. - \int_{-\infty}^0 e^{-\frac{1}{4\lambda}\left(v^2 - 2v\ln u + (\ln u)^2 + 4\lambda(\eta-1)v\right)}dv\right)
$$

$$
= \frac{u^\eta e^\delta}{2\sqrt{\lambda\pi}}\left(\int_{-\infty}^0 e^{-\frac{1}{4\lambda}(v - \ln u + 2\lambda\eta)^2 + \lambda\eta^2 - \eta\ln u}dv\right.
$$

$$
\left. - \int_{-\infty}^0 e^{-\frac{1}{4\lambda}(v - \ln u + 2\lambda(\eta-1))^2 + \lambda(\eta-1)^2 - (\eta-1)\ln u}dv\right).
$$

Next, we let

$$
v' := \frac{v - \ln u + 2\lambda\eta}{\sqrt{2\lambda}} \quad \text{and} \quad v'' := \frac{v - \ln u + 2\lambda(\eta-1)}{\sqrt{2\lambda}}.
$$

Then, we have

$$
\int_0^1 (1-\xi)e^\delta \left(\frac{u}{\xi}\right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}}e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{u}{\xi}\right)\right)^2}\right)\frac{du}{u}
$$

$$
= \frac{e^\delta}{\sqrt{2\pi}}\left(\int_{-\infty}^{\frac{-\ln u + 2\lambda\eta}{\sqrt{2\lambda}}} e^{-\frac{v'^2}{2} + \lambda\eta^2}dv' - u\int_{-\infty}^{\frac{-\ln u + 2\lambda(\eta-1)}{\sqrt{2\lambda}}} e^{-\frac{v''^2}{2} + \lambda(\eta-1)^2}dv''\right)
$$

$$
= e^{\delta+\lambda\eta^2}\Phi\left(\frac{-\ln u + 2\lambda\eta}{\sqrt{2\lambda}}\right) - ue^{\delta+\lambda(\eta-1)^2}\Phi\left(\frac{-\ln u + 2\lambda(\eta-1)}{\sqrt{2\lambda}}\right)
$$

$$
= e^{-r(T-t)}\Phi\left(-\Delta_-\left(\frac{s}{z}\right)\right) - \left(\frac{s}{z}\right)\Phi\left(-\Delta_+\left(\frac{s}{z}\right)\right).
$$

For the second integral, we have

$$\int_1^\infty \left(-\frac{1}{k_1}\xi^{1-k_1} + \frac{\xi}{k_1} \right) e^\delta \left(\frac{u}{\xi} \right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}} e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{u}{\xi} \right) \right)^2} \right) \frac{d\xi}{\xi}$$

$$= \int_0^\infty \left(-\frac{1}{k_1}e^{(1-k_1)v} + \frac{1}{k_1}e^v \right) e^\delta u^\eta e^{-v\eta} \left(\frac{1}{2\sqrt{\lambda\pi}} e^{-\frac{1}{4\lambda}(\ln u - v)^2} \right) dv$$

$$= \frac{e^\delta u^\eta}{2k_1\sqrt{\lambda\pi}} \int_0^\infty \left(-e^{\eta v - \frac{1}{4\lambda}(\ln u - v)^2} + e^{v(1-\eta) - \frac{1}{4\lambda}(\ln u - v)^2} \right) dv$$

$$= \frac{e^\delta u^\eta}{2k_1\sqrt{\lambda\pi}} \left(\int_0^\infty -e^{-\frac{1}{4\lambda}(v - \ln u - 2\lambda\eta)^2 + \lambda\eta^2 + \eta \ln u} dv \right.$$

$$\left. + \int_0^\infty e^{-\frac{1}{4\lambda}(v - \ln u - 2\lambda(1-\eta))^2 + \lambda(1-\eta)^2 + (1-\eta)\ln u} dv \right),$$

where we use the fact that $k_1 - 1 + \eta = -\eta$. Furthermore, we introduce a new variable

$$v''' := \frac{v - \ln u - 2\lambda\eta}{\sqrt{2\lambda}}.$$

Then, we have

$$\int_1^\infty \left(-\frac{1}{k_1}\xi^{1-k_1} + \frac{\xi}{k_1} \right) e^\delta \left(\frac{u}{\xi} \right)^\eta \left(\frac{1}{2\sqrt{\lambda\pi}} e^{-\frac{1}{4\lambda}\left(\ln\left(\frac{u}{\xi} \right) \right)^2} \right) \frac{d\xi}{\xi}$$

$$= \frac{e^\delta u^\eta}{k_1\sqrt{2\pi}} \left(\int_{\frac{-\ln u - 2\lambda\eta}{\sqrt{2\lambda}}}^\infty -e^{-\frac{v'''^2}{2}} e^{\lambda\eta^2 + \eta \ln u} dv''' \right.$$

$$\left. + \int_{\frac{-\ln u + 2\lambda(\eta-1)}{\sqrt{2\lambda}}}^\infty e^{-\frac{v''^2}{2}} e^{\lambda(\eta-1)^2 + (1-\eta)\ln u} dv'' \right)$$

$$= -\frac{1}{k_1}e^{\delta + \lambda\eta^2} u^{1-k_1} \Phi\left(\frac{\ln u + 2\lambda\eta}{\sqrt{2\lambda}} \right) + \frac{1}{k_1}ue^{\delta + \lambda(\eta-1)^2} \Phi\left(\frac{\ln u + 2\lambda(1-\eta)}{\sqrt{2\lambda}} \right)$$

$$= -\frac{1}{k_1}\left(\frac{s}{z} \right)^{1-k_1} e^{-r(T-t)} \Phi\left(-\Delta_-\left(\frac{z}{s} \right) \right) + \frac{1}{k_1}\left(\frac{s}{z} \right) \Phi\left(\Delta_+\left(\frac{s}{z} \right) \right).$$

Putting these two integrals together and using the fact that $P_0 = zQ_0$, we can obtain our Formula (27).

References

Bates, David S. 1996. Jumps and stochastic volatility: Exchange rate processes implicit in Deutsche market options. *The Review of Financial Studies* 9: 69–107. [CrossRef]

Black, Fischer, and Myron Scholes. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy* 3: 637–54. [CrossRef]

Boyle, Phelim P., and Yisong S. Tian. 1999. Pricing lookback and barrier options under the CEV process. *Journal of Financial and Quantitative Analysis* 34: 241–64. [CrossRef]

Buchen, Peter. 2001. Image options and the road to barriers. *Risk Magazine* 14: 127–30.

Cao, Jiling, Jeong-Hoon Kim, and Wenjun Zhang. 2021. Pricing variance swaps under hybrid CEV and stochastic volatility. *Journal of Computational and Applied Mathematics* 386: 113220. [CrossRef]

Cao, Jiling, Jeong-Hoon Kim, and Wenjun Zhang. 2023. Valuation of barrier and lookback options under hybrid CEV and stochastic volatility. *Mathematics and Computers in Simulation* 208: 660–76. [CrossRef]

Chiarella, Carl, Boda Kang, and Gunter H. Meyer. 2012. The evaluation of barrier option prices under stochastic volatility. *Computers & Mathematics with Applications* 64: 2034–48.

Choi, Sun-Yong, Jean-Pierre Fouque, and Jeong-Hoon Kim. 2013. Option pricing under hybrid stochastic and local volatility. *Quantitative Finance* 13: 1157–65. [CrossRef]

Clewlow, Les, Javier Llanos, and Chris Strickland. 1994. *Pricing Exotic Options in a Black-Scholes World*. Coventry: Financial Operations Research Centre, University of Warwick, vol. 54, 32p.

Conze, Antoine, and R. Vishwanathan. 1991. Path-dependent options: The case of lookback options. *The Journal of Finance* 46: 1893–907. [CrossRef]

Cox, John. 1975. *Notes on Option Pricing I: Constant Elasticity of Variance Diffusions*. Working Paper. Stanford: Stanford University.

Cox, John. 1996. The constant elasticity of variance option pricing model. *Journal of Portfolio Management* 5: 15–17. [CrossRef]

Davydov, Dmitry, and Vadim Linetsky. 2001. Pricing and hedging path-dependent options under the CEV process. *Management Science* 47: 881–1027. [CrossRef]

Fouque, Jean-Pierre, George Papanicolaou, and Ronnie Sircar. 2000. *Derivatives in Financial Markets with Stochastic Volatility*. Cambridge: Cambridge University Press.

Fouque, Jean-Pierre, George Papanicolaou, Ronnie Sircar, and Knut Sølna. 2011. *Multiscale Stochastic Volatility for Equity, Interest Rate, and Credit Derivatives*. Cambridge: Cambridge University Press.

Funahashi, Hideharu, and Tomohide Higuchi. 2018. An analytical approximation for single barrier options under stochastic volatility models. *Annals of Operations Research* 266: 129–57. [CrossRef]

Goldman, M. Barry, Howard B. Sosin, and Marry Ann Gatto. 1979. Path dependent options: "Buy at the low, sell at the high". *The Journal of Finance* 34: 1111–27.

Haug, Espen. 2006. *The Complete Guide to Option Pricing Formulas*, 2nd ed. New York: McGraw-Hill.

Heston, Steven. 1993. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *The Review of Financial Studies* 6: 327–43. [CrossRef]

Hull, John C. 2015. *Options, Futures and Other Derivatives*, 9th ed. London: Pearson.

Kato, Takashi, Akihiko Takahashi, and Toshihiro Yamada. 2013. An asymptotic expansion formula for up-and-out barrier option price under stochastic volatility model. *Japan Society for Industrial and Applied Mathematics Letters* 5: 17–20. [CrossRef]

Kim, Hyun-Gyoon, Jiling Cao, Jeong-Hoon Kim, and Wenjun Zhang. 2023. A Mellin transform approach to pricing barrier options under stochastic elasticity of variance. *Applied Stochastic Models in Business and Industry* 39: 160–76. [CrossRef]

Kim, So-Yeun, and Ji-Hun Yoon. 2018. An approximated European option price under stochastic elasticity of variance using Mellin transforms. *East Asian Mathematical Journal* 34: 239–48.

Leung, Kwai Sun. 2013. An analytic pricing formula for lookback options under stochastic volatility. *Applied Mathematics Letters* 26: 145–49. [CrossRef]

Merton, Robert C. 1973. The theory of rational option pricing. *Bell Journal of Economics and Management Science* 1: 141–83. [CrossRef]

Panini, Radha, and Ram P. Srivastav. 2004. Option pricing with Mellin transforms. *Mathematical and Computer Modelling* 40: 43–56. [CrossRef]

Park, Sang-Hyeon, and Jeong-Hoon Kim. 2013. A semi-analytic pricing formula for lookback options under a general stochastic volatility model. *Statistics & Probability Letters* 83: 2537–43.

Reiner, Eric, and M. Rubinstein. 1991. Breaking down the barriers. *Risk* 4: 28–35.

Wirtu, Teferi Dereje, Philip Ngare, and Ananda Kube. 2017. Pricing floating strike lookback put option under Heston stochastic volatility. *Global Journal of Mathematical Sciences: Theory and Practical* 9: 427–39.

Yoon, Ji-Hun. 2014. Mellin transform method for European option pricing with Hull-White stochastic interest rate. *Journal of Applied Mathematics* 2014: 759562. [CrossRef]

Zaevski, Tsvetelin S., Young Shin Kim, and Frank J. Fabozzi. 2014. Option pricing under stochastic volatility and tempered stable Lévy jumps. *International Review of Financial Analysis* 31: 101–8. [CrossRef]

Article

Relations among Bitcoin Futures, Bitcoin Spot, Investor Attention, and Sentiment

Arun Narayanasamy [1,*], Humnath Panta [2] and Rohit Agarwal [3]

[1] Department of Finance, University of Northern Iowa, Cedar Falls, IA 50614, USA
[2] School of Business, Cal Poly Humboldt, Arcata, CA 95519, USA; humnath.panta@humboldt.edu
[3] Department of Economics, Finance and Accounting, University of South Carolina Upstate, Spartanburg, SC 29303, USA; rohita@uscupstate.edu
[*] Correspondence: arun.narayanasamy@uni.edu

Abstract: This research investigates the function of price discovery between the Bitcoin futures and the spot markets while also analyzing the impact of investor sentiment and attention on these markets. This study utilizes various statistical models to examine the short-term and long-term relations between these variables, including the bivariate Granger causality model, the ARDL and NARDL models, and the Johansen cointegration procedure with a vector error correction mechanism. The results suggest that there is no statistical evidence of price discovery between the Bitcoin spot price and futures, and the term structure of the Bitcoin futures neither enriches nor impairs this lead lag relation. However, the study finds robust evidence of a long-run cointegrating relation between the two markets and the presence of asymmetry in them. Moreover, this research indicates that investor sentiment exhibits a lead lag relation with both the Bitcoin futures and the spot markets, while investor attention only leads to the Bitcoin spot market, without showing any lead lag relation with the Bitcoin futures. These findings highlight the crucial role of investor behavior in affecting both Bitcoin futures and spot prices.

Keywords: bitcoin; bitcoin futures; investor attention; sentiment; causality; cointegration

JEL Classification: G10; G12; G13; G14; G40

1. Introduction

A recent survey conducted on the most widely held financial assets reported that cryptocurrency is the second most widely held financial asset[1], especially among women. Among the various cryptocurrencies, Bitcoin has been recognized as the best-performing asset in recent times[2]. Introduced in 2008, Bitcoin is a fascinating addition to the financial markets. Among the 9420 cryptocurrencies traded around the world, Bitcoin is generally considered the leading cryptocurrency in terms of market share[3]. In addition to being a dominant asset, Bitcoin has received much attention, not only from regulators and the media but also from academic researchers and investment participants in the financial markets. In fact, December 2017 was the first time that the Bitcoin futures were introduced, owing to the popularity that Bitcoin has enjoyed since its launch. The introduction of Bitcoin futures has facilitated hedging risks related to the underlying Bitcoin spot or other traded cryptocurrencies. It is important to note that although the Bitcoin market share has decreased over time, from more than 90% in 2010 to about 47.13% in 2023, it remains one of the most important digital currencies in the cryptocurrency market.

Several factors substantiate the uniqueness of Bitcoin among all the listed cryptocurrencies[4] and the Bitcoin futures[5]. As the investigation into the regulation of cryptocurrencies has gained significance, so has the research into cryptocurrencies, specifically the Bitcoin spot price and Bitcoin futures. Most of the existing research has focused on the determinants of price discovery in the Bitcoin markets (Alexander et al. 2020; Alexander and Heck

2020; Entrop et al. 2020), the trading activity in Bitcoin (Wustenfeld and Geldner 2022; Scharnowski 2021; Dyhrberg et al. 2018), the short-term and long-term determinants of the value of Bitcoin (Dubey 2022; Mai et al. 2018; Li and Wang 2017), the market efficiency of Bitcoin (Kochling et al. 2019; Urquhart 2016), the impact of the launch of the Bitcoin futures on the Bitcoin spot (Akanksha et al. 2021; Kim et al. 2020; Liu et al. 2020), the diversification benefits and the linkage of Bitcoin with other financial assets (Wang et al. 2022; Baur and Dimpfl 2021; Qarni and Gulzar 2021; Guesmi et al. 2019), the price discovery leadership between the Bitcoin futures and spot markets (Akyildirim et al. 2020; Baur and Dimpfl 2019; Corbet et al. 2018; Kapar and Olmo 2019), the hedging properties of the Bitcoin futures (Sebastiao and Godinho 2020; Chan et al. 2019), the large contract size associated with the Bitcoin futures and its impact on trading activity (Park 2022; Akyildirim et al. 2021), the process of price discovery among the various Bitcoin trading exchanges (Pagnottoni and Dimpfl 2019), the illegal activities related to Bitcoin (Foley et al. 2019), the impact of investor attention on Bitcoin (Smales 2022; Choi 2021; Lin 2021), and the impact of investor sentiment on Bitcoin (Koutmos 2023; Mokni et al. 2022; Naeem et al. 2021; Guegan and Renault 2021).

Despite the considerable research on the Bitcoin futures and spot prices, it is important to highlight that various challenges and limitations persist, reinforcing the necessity for further investigation into the specific relations concerning these assets. The limitations of past research include conflicting results regarding the price discovery function between the Bitcoin futures and spot price, restrictions concerning the long-run relation between the Bitcoin futures and the spot, limitations associated with the connection between investor attention and the pricing of the Bitcoin futures and spot, limitations associated with the relation between investor sentiment and the pricing of the Bitcoin futures and spot, limitations in understanding the nature of participants in the Bitcoin market, limitations in understanding the activities of market participants for the varying levels of investor attention, and constraints related to the term structure of the Bitcoin futures and its impact on the underlying spot prices.

This unique study addresses the limitations mentioned above and investigates the relations between the Bitcoin futures, spot, crypto sentiment, and Bitcoin attention, to reconcile the conflicting evidence in the existing literature. Our findings contribute to the existing Bitcoin pricing literature in many ways. Firstly, we find no statistical evidence of price discovery between the Bitcoin spot and futures market, and the term structure of the Bitcoin futures does not play a significant role in this relation. The reason for the lack of price discovery functionality could be the increase in informational efficiency for these two assets over time. The amount of trading activity and the number of market participants in both these assets have increased significantly over time, contributing to efficiency. Secondly, we discover a lead lag relation between Bitcoin sentiment and the futures and spot price, suggesting that changes in market participants' opinions and perceptions simultaneously affect trading activity in both markets. The reason for this significant relation is likely the activities that market participants conduct upon the fundamental change in sentiment. While a portion of the participants would engage with the futures to speculate or hedge their risks, another portion of the participants would engage in the actual underlying spot asset, driving the significance of this relation. Thirdly, investor attention leads the Bitcoin spot market but does not exhibit a lead lag relation with the Bitcoin futures. The reason could be the fact that the participants in the spot asset market are the ones who react to changes in the level of attention, affecting the changes in the spot price, and the futures market participants are relatively immune to changes in attention. Fourthly, we find robust statistical evidence of a long-term cointegrating relation between the Bitcoin spot price and futures, with the findings consistent across different cointegration procedures. In essence, this implies that one series could be expressed as another owing to the common characteristics between them. Finally, the speed of adjustment towards equilibrium in the long-term cointegrating relation is stronger and more significant when using the nonlinear autoregressive distributed lag procedure (NARDL) than the ARDL or Johansen procedures.

This additionally adds validity to the robustness of our significant long-term relations between the series.

The subsequent sections of this research are organized as follows. Section 2 presents an extensive examination of the pertinent literature. Section 3 discusses the methods or research design. Section 4 of this study describes the data. Section 5 presents the empirical evidence as results, and Section 6 finally concludes the paper.

2. Literature Review

The Bitcoin futures were introduced by both the Chicago Board Options Exchange (CBOE) and a week later by the Chicago Mercantile Exchange (CME) in December 2017. These Bitcoin futures were primarily created to present two major functionalities, i.e., price discovery and hedging in the underlying spot Bitcoin market, and past research has highlighted the efficacy of the Bitcoin futures as a hedging tool (Sebastiao and Godinho 2020; Alexander et al. 2020; Nekhili 2020; Kochling et al. 2019). As much as some empirical studies have shown that the Bitcoin futures are an effective instrument for hedging, there is also evidence of the limitation of the hedging properties of these futures, warranting further investigation into their dynamics (Hung et al. 2021; Alexander and Heck 2020; Hattori and Ishida 2020). Likewise, one of the most important functionalities provided by the futures contracts towards the underlying spot asset activity is price discovery (Silber 1981).

In general, the prices of these futures contracts mirror the expectations of investors in the corresponding asset market for the near future. This anticipation should be factored into the values of the underlying spot assets, contributing to a process of price discovery. The evidence of price discovery from past research is mixed, confirming the need to investigate the relations further. Although some shared factors influence both the Bitcoin spot prices and futures, there is evidence suggesting that the futures have a relatively greater impact on the price discovery process (Akyildirim et al. 2020; Kapar and Olmo 2019). Furthermore, there is alternative evidence, using Hasbrouck's (1995) information share methodology, indicating that the Bitcoin spot price leads and exerts an influence on the future prices with regard to price discovery (Baur and Dimpfl 2019; Corbet et al. 2018). These inconsistent findings could be a result of the relatively small data sample that is used in these studies, as well as a sharp decline in the market conditions for Bitcoin[6], further highlighting the need to analyze these relations with larger sample periods. There is also evidence that as the Bitcoin futures contracts become shorter and shorter, the accuracy of the futures contract in aiding the price discovery of the underlying spot market increases (Matsui and Gudgeon 2020). This highlights the need not just to reinvestigate the lead lag relation between the Bitcoin futures and spot but also to use the near-term and next-term futures contract to identify the magnitude of this linkage. The preceding discussion suggests that, in the literature, there is no prevailing consensus regarding the price discovery process in the Bitcoin market.

In addition to the extant literature looking at the short-run dynamics between the Bitcoin futures and spot, several studies look at the long-run relations between these assets (Lee and Rhee 2022; Wu et al. 2021; Hung et al. 2021; Hu et al. 2020; Kapar and Olmo 2019; Cheah et al. 2018). Some past research evidence shows a long-run relation between the Bitcoin futures and spot and suggests that the futures dominate the spot assets using a fractionally cointegrated framework (Wu et al. 2021; Cheah et al. 2018). Although the fractional cointegration framework allows the underlying series to take fractional integration values and analyzes the long-run equilibrium relations, it has no allowance for deterministic trends, and the convergence to equilibrium does not necessarily have an optimal rate. Additionally, the memory parameter is unknown in the fractionally cointegrated system. These limitations support the estimation of the long-run dynamics using a standard cointegrated framework. Some studies have looked at these long-run relations assuming a time-varying cointegrating coefficient (Lee and Rhee 2022; Hu et al. 2020), but the challenge with these findings is the oversight of the common factors that remain stable and that affect both the Bitcoin futures and spot simultaneously over time.

Notably, the challenge with these studies substantiates the need to revisit the analysis of long-run relations using a standard cointegration framework, that allows for asymmetry in addition to relaxing the assumptions of the order of integration of the underlying series.

Intrinsically, research has been conducted to analyze the aspects that provide Bitcoin with its value or the factors behind its price fluctuations. It has been distinctly observed that the Bitcoin price reflects several more factors than the standard interactions of demand, supply, and fundamental news (Griffin and Shams 2020; Eross et al. 2019; Panagiotidis et al. 2019). To further understand the impact of non-standard factors on the pricing of the Bitcoin futures and spot, some studies have looked at the role of social interactions in Bitcoin prices (Lyocsa et al. 2020; Gronwald 2019; Geuder et al. 2019; Garcia et al. 2014). While these studies use different socio-economic signals, like the volume of information searched about Bitcoin, the prices in online exchanges, the volume of word-of-mouth communication in online media, and user base growth, they show evidence that highlights the impact of online searches, word of mouth, and an expanding user base on Bitcoin pricing and not the pricing of the Bitcoin futures. The reason for these findings is that as the media publishes articles about price increases in Bitcoin, it fosters and acts as a stimulus for search activities among investors, thereby impacting trading activity and the pricing of the underlying asset. These findings are equally robust for negative news or unfavorable attention as well and can result in significant price declines (Chevapatrakul and Mascia 2019). In essence, while a Google search of a traditional currency will not impact its value or volume, it can, in all likelihood, drive the prices of Bitcoin (Aalborg et al. 2019). Owing to the extensive spectrum of available socio-economic signals and the limited few that have been studied in past research, it is necessary to further analyze the impact of these signals on both the Bitcoin futures and the spot assets using alternate measures.

Further, in looking at the factors that give Bitcoin its value, a small number of studies have looked at the impact of sentiment or public opinion on the price movements of Bitcoin (Entrop et al. 2020; Rognone et al. 2020; Dastgir et al. 2019; Aalborg et al. 2019; Karalevicius et al. 2018). While some results are in support of the fact that sentiment does affect Bitcoin pricing (Rognone et al. 2020; Karalevicius et al. 2018), there is also evidence that sentiment does not play any role in impacting the prices of Bitcoin (Dastgir et al. 2019; Entrop et al. 2020), and, most importantly, none of these past studies look at the relation of sentiment with the Bitcoin futures. This paper addresses this limitation by looking at the impact on both the future as well as the spot. An indication of bi-directional causality between the Bitcoin attention variable measured by Google Trends search queries and the Bitcoin asset returns softens any explanatory linkage (Dastgir et al. 2019; Fry 2018) and limits itself to looking at only the spot asset. Although these results are conflicting and limited, the impact that social media plays in the acceptance of Bitcoin as an asset and the ensuing trading activity are broadly analyzed, with results indicating that bullish posts predict positive returns and bearish posts predict negative returns (Chen et al. 2020; Mai et al. 2018). The aforementioned discussion of sentiment and its impact on Bitcoin pricing implies that there is no conformity in the literature on the linkage between sentiment and Bitcoin pricing, supporting the need to investigate this relation further.

In summary, after examining numerous relevant literature works pertaining to the Bitcoin futures, Bitcoin spot, investor attention, and sentiment, along with identifying their incongruities, this study attempts to answer several questions that include the following: Do the Bitcoin futures and the Bitcoin spot have any long-run relation? Conditional on the existence of a long-run relation, is there any asymmetry in this long-run relation? Do the Bitcoin futures and the Bitcoin spot have any short-run causal relation? Do the Bitcoin futures and Bitcoin attention have any short-run causal relation? Do the Bitcoin futures and Bitcoin sentiment have any short-run causal relation? Conditional on the existence of a short-run causal relation between the Bitcoin futures and spot, do the shorter-term Bitcoin futures (near-term versus next-term) have greater statistical significance in the short-run causal relation with the Bitcoin spot?

3. Methodology

We are interested in evaluating whether the movements in the Bitcoin futures, the Bitcoin spot prices, crypto sentiment, and investor attention have a short-term causal or long-term cointegrated relation. The interest in the relations among these variables arises from the limitations in the extant literature identified in the previous sections. We accomplish this analysis by following four distinct steps. First, we investigate the stationarity of each series used in the study employing multiple tests (Dickey and Fuller 1979, 1981; Cheng et al. 2021; Kwiatkowski et al. 1992 (KPSS); Ng and Perron 2001). After checking for stationarity, second, we evaluate the short-term causal relation using bivariate Granger causality tests (Granger 1969, 1980, 1988). Third, after determining the optimal lag lengths, we evaluate the long-term cointegrated relation using several methods. We use the Johansen cointegration method (Johansen 1988, 1991, 1995) as well as the autoregressive distributed lag (ARDL) method (Pesaran and Shin 1999; Pesaran et al. 1996, 2001) and the nonlinear autoregressive distributed lag (NARDL) method (Demir et al. 2021; Mhadhbi et al. 2021; Dutta et al. 2019; Shin et al. 2014). Fourth, and finally, we evaluate the error correction model. We provide detailed explanations of each method used in this study below.

3.1. Stationarity

Stationarity is considered as invariance under a time shift, and any time series that is stationary is treated to have properties that are not conditional on the time in which they are observed. In general, a stationary time series will have no discernible pattern in the long term. Although the features of a constant mean and variance are not particularly imperative in estimating the parameters in econometric models, they can significantly impact model selection, since these features are essential for the calculation of reliable test statistics. Hence, it becomes essential to first test whether the relevant variables have the problem of a unit root and determine the orders of integration for each of the series used in the study (Enders 2014; Chan 2010) before we can determine the statistical specification of the model and conduct either a causality or cointegration test. We first conduct the stationarity tests using the standard augmented Dickey–Fuller (ADF) method (Cheng et al. 2021). The augmented Dickey–Fuller (ADF) test is a standard test for stationarity and is estimated using the following general equation.

$$Y_t = \alpha + \beta T + \delta Y_{t-1} + \sum_{i=1}^{m-1} \theta_i \Delta Y_{t-i} + \varepsilon_t \tag{1}$$

where $\delta = \sum_{i-1}^{m} \rho_{i-1}$ and $\theta_i = -\sum_{k=i+1}^{m} \rho_k$.

The dependent variable Y_t is lagged to represent higher-order autoregressive processes and to eliminate serial correlation. In Equation (1), a time trend is also included. This is done to test the presence of a deterministic trend. It is important to note that we consider and include a variable for a subsequent Johansen cointegration analysis if and only if it meets the criterion of being non-stationary and integrated of at least order $I(1)$. Correspondingly, we consider and include a variable for the ARDL cointegration test only if it meets the criterion of not being integrated of higher order, like $I(2)$ or greater. The null hypothesis of the ADF test (a unit root exists, $\delta = 0$) is rejected when the test statistic that is computed is greater than the critical value. Failing to reject it will imply that the series is non-stationary, and we must differentiate it until it becomes stationary. To check the robustness of our findings and to verify the consistency of the stationary properties, we also conduct the KPSS test (Kwiatkowski et al. 1992). Contrary to the ADF test, the KPSS test evaluates for a null hypothesis of no unit root or stationarity and an alternative hypothesis of the existence of a unit root or non-stationarity. The KPSS test can be depicted by Equation (2) below.

$$KPSS = \hat{\eta} = T^{-2} \left(\sum_{t=1}^{T} S_t^2 \right) / \hat{\sigma}_N^2 \tag{2}$$

where $\hat{\eta}$ denotes the respective KPSS statistic for the testing of stationarity around the mean and $\hat{\sigma}_N^2$ is a consistent estimator of the long-term variance of residuals. Past research has

shown evidence that both the ADF and the KPSS tests experience some size problems as well as finite sample power problems (Sephton 2008). In order to address this issue, we also run stationarity tests using the Ng and Perron test, which uses the detrending of the generalized least squares method (Ng and Perron 2001; Phillips and Perron 1988). The Ng and Perron test statistics are depicted in Equation (3) below.

$$\left.\begin{aligned} \text{MZ}_a &= \left(T^{-1}y_T^d - \hat{\sigma}_N^2\right)\left(2T^{-2}\textstyle\sum_{t=1}^{T} y_{t-1}^d\right)^{-1} \\ MSB &= \left(T^{-2}\textstyle\sum_{t=1}^{T} y_{t-1}^d / \hat{\sigma}_N^2\right)^{1/2} \\ \text{MZ}_t &= \text{MZ}_a \times MSB \end{aligned}\right\} \tag{3}$$

where y_t^d represents the GLS detrended data and $\hat{\sigma}_N^2$ represents a consistent estimator for the persistent variance of residuals over the long term. The null hypothesis of the Ng and Perron test is that the unit root exists, and the alternate hypothesis is that the series is stationary. If the computed statistics for MZ_a and MZ_t are less than the critical values in absolute terms, we reject the null hypothesis. The Ng and Perron test is also used to complement the results of both the ADF and KPSS tests.

3.2. Granger Causality

After analyzing the stationarity properties of the series, we look at the short-term causal lead lag relations using the bivariate Granger causality tests (Granger 1969, 1980, 2001). In essence, the method analyzes whether a particular time series is a factor of another series by reducing the forecast error of the overall model. "Causality" here does not necessarily mean a cause-and-effect relation between the variables but rather the "precedence" of one variable over the other in time series data. In this way, the use of time series information facilitates an understanding of the direction of causality. The bivariate linear Granger causality used in this study can be shown by a generic two-equation model as below.

$$Y_t = A + \textstyle\sum_{i=1}^{p} B_i Y_{t-i} + \sum_{j=1}^{p} C_j X_{t-j} + U_t \tag{4}$$

$$X_t = D + \textstyle\sum_{k=1}^{p} E_k Y_{t-k} + \sum_{l=1}^{p} F_l X_{t-l} + V_t \tag{5}$$

where all X_t and Y_t are stationary variables, the optimal lag length in the system is shown by p, and U_t and V_t are the random errors. To test whether X_t Granger causes Y_t, we need to determine whether any lags of X_t are statistically significant in Equation (4). We do this using an F-test for linear restrictions. Empirically, in testing the null hypotheses shown below, we are testing for the presence of a linear causal relation between Y_t and X_t.

$$\begin{aligned} H1: C_1 &= C_2 = \cdots\cdots\cdots = C_p = 0, \text{ and} \\ H2: E_1 &= E_2 = \cdots\cdots\cdots = E_p = 0 \end{aligned} \tag{6}$$

By testing the hypotheses stated in Equation (6), we have several possibilities for causal relations between Y_t and X_t, which include either a lack of causal relation between the variables or a unidirectional or bidirectional relation between the variables.

In addition, it is important to emphasize that the results of the bivariate Granger causality tests can be affected by the choice of the lag lengths in Equations (4) and (5) (Thornton and Batten 1985; Guilkey and Salemi 1982). Specifically, if we use more lags than the true order, the power of the test will be affected. In addition, if we use fewer lags than the true order, the estimates from the regression will be biased. Moreover, the residuals from the regression will be serially correlated. Therefore, we adopt Hsiao's approach (Hsiao 1982, 1981, 1979) to select lag lengths that minimize the Akaike Information Criterion (AIC) and the prediction error (Akaike 1969a, 1969b, 1974, 1981). We use the unrestricted vector autoregression (VAR) procedure to examine each series for its optimal lag length, using the pth order VAR model as expressed by Enders (Enders 2014; Chan 2010).

3.3. Johansen Cointegration

The cointegration test is generally used to analyze whether there is a long-term relation between two or more time series. If a cointegrating relation is absent, the two variables can move arbitrarily through time and away from each other. In this paper, we analyze the existence of cointegration among variables specified using three methods. The first method is explained in this sub-section, and the other two methods are explained in the following sub-section. The Johansen cointegration procedure (Johansen 1995; Engle and Granger 1987) is a general dynamic systems technique that allows for more than one cointegrating relation. The variables are parametrized in terms of the lagged levels of the system variables in addition to the lagged first differences under this approach. Consider a VAR model of order *m* as shown below in Equation (7).

$$Y_i = \mu + A_1 Y_{i-1} + \cdots + A_m Y_{i-m} + \varepsilon_i \tag{7}$$

where Y_i is an n by 1 vector of variables that are first-order integrated and are denoted as *I(1)*, μ is the first moment of the series, $A_1 \ldots A_m$ are the coefficient matrices for each lag, and ε_i is the noise term with a mean of zero. If the vectors are cointegrated, we can form a vector error correction model (VECM). Then, Equation (7) can be modified, as shown below in Equation (8).

$$\Delta Y_i = \mu + \Gamma Y_{i-1} + \sum_{j=1}^{m-1} \Gamma_j \Delta Y_{i-1} + \varepsilon_i \tag{8}$$

where $\Gamma = \sum_{j=1}^{m} A_j - I$ and $\Gamma_j = -\sum_{k=j+1}^{m} A_k$.

In Equation (8), $\Delta Y_i := Y_i - Y_{i-1}$ is the differencing operator, Γ is the coefficient matrix for the first lag, and Γ_j are the matrices for each differenced lag.

The Johansen test successively assesses whether the rank (r) is equal to zero or one, continuing up to (r) being equal to $n-1$, where n is the number of time series variables used to conduct the test. The null hypothesis is that there is no cointegrating relation. When the rank is greater than zero, it indicates the existence of some cointegrating relation between the series examined. Eventually, r is the number of cointegrating relations. Thus, A is the parameter in the vector error correction model (VECM) that acts as an adjustment parameter, which must have a negative sign and statistical significance, and β is the cointegration vector of each column of the Johansen model.

Specifically, for the Johansen procedure, this cointegration test is conducted in two main forms: trace tests (λ_{trace}) and maximum eigenvalue tests (λ_{max}). These are the primary tests used in canonical corrections to help to determine the number of cointegrating vectors (CIVs) among the series of interest.

It is also important to note that, similar to the way in which the Granger causality tests were affected by the choice of the optimal lag length, the Johansen cointegration test is also affected by the lag length choice. Additionally, while the λ_{trace} value tests the null hypothesis that the number of cointegrating vectors is $\leq r$ against an alternative hypothesis, the λ_{max} value tests the null hypothesis that the number of cointegrating vectors is = r against an alternate hypothesis of r + 1. The statistics of this procedure are defined by Equations (9) and (10) shown below.

$$\lambda_{max}(r, r+1) = -T \ln\left(1 - \hat{\lambda}_{r+1}\right) \tag{9}$$

$$\lambda_{trace}(r) = -T \sum_{i=r+1}^{n} \ln\left(1 - \hat{\lambda}_i\right) \tag{10}$$

3.4. ARDL and NARDL Cointegration

We have previously examined a cointegration approach (Johansen 1995, 1991; Johansen and Juselius 1990; Engle and Granger 1987) to assess the long-term connection between the variables of interest. One of the major limitations of this approach is that the series has to be integrated into at least order one, *I*(1). To overcome this problem, this study employs

the widely recognized ARDL model (Pesaran et al. 2001). The ARDL model offers the flexibility that the set of variables can be integrated in a different order as it is the most general dynamic unrestricted model in the economic literature (Sari et al. 2008; Ghatak and Siddiki 2001). Additionally, a dynamic error correction model can be derived from ARDL by using a linear transformation, which will essentially enhance the speed of adjustment towards equilibrium (Banerjee et al. 1993). A standard linear ARDL(p, q) cointegration model with two time series Y_t and X_t has the form shown in Equation (11).

$$\Delta Y_t = A_0 + \rho\, Y_{t-1} + \theta\, X_{t-1} + \gamma\, Z_t + \sum_{j=1}^{p-1} A_j \Delta\, Y_{t-j} + \sum_{j=0}^{q-1} \pi_j \Delta\, X_{t-j} + E_t \qquad (11)$$

where Z_t is a vector of deterministic regressors and E_t is an iid stochastic process. According to the null hypothesis, the two series are not cointegrated, implying that the coefficients of the lagged levels of the two variables in Equation (10) are jointly zero ($\rho = \theta = 0$). The hypothesis of this model can be tested using a modified F-test or a t-test, as shown in the prior literature (Pesaran et al. 2001). It is important to note that the combination of stochastic regressors in the ARDL model is linear and signifies symmetric adjustments in both the long and short run. We can further extend this methodology to include nonlinearities (Demir et al. 2021; Mhadhbi et al. 2021; Dutta et al. 2019; Shin et al. 2014). The nonlinear ARDL(p,q) model is depicted below in Equation (12).

$$Y_t = \sum_{j=1}^{p} \varnothing_j Y_{t-j} + \sum_{j=0}^{q} \left(\theta_j^{+'} X_{t-j}^{+} + \theta_j^{-'} X_{t-j}^{-} \right) + E_t \qquad (12)$$

This NARDL model is capable of explaining the asymmetry in the long-run relation, and the model's hypothesis can be evaluated through the use of the bounds testing procedure similar to the ARDL model (Demir et al. 2021; Pesaran et al. 1996, 2001; Mhadhbi et al. 2021; Shin et al. 2014).

3.5. Error Correction Model

Based on the outcomes of the Johansen, ARDL, and NARDL cointegration tests, the establishment of a cointegrating relation between variables Yt and Xt implies the existence of a long-term equilibrium relation between them. However, to evaluate the short-run relations and properties of the cointegrated series, we use the error correction model (ECM) technique. In brief, the ECM, consistent with the long-run cointegrating relation, represents how Y_t and X_t behave in the short term. This ECM is shown to contain important information on both the long-term and short-term properties of the model with disequilibrium. For example, if two variables are integrated in order one, $I(1)$, and there is a linear combination between them that is integrated with order zero, $I(0)$, then we will have an error correction term (ECT) that is statistically significant and has a negative sign, indicating the speed of adjustment towards equilibrium. The general equation for the error correction model can be represented as shown below in Equations (13) and (14).

$$\Delta Y_t = M + N_i \Delta Y_{t-i} + O_j \Delta X_{t-i} + \varphi\, ECT_{y,t-i} + U_{y,t} \qquad (13)$$

$$\Delta X_t = P + Q_i \Delta Y_{t-i} + R_j \Delta X_{t-i} + \varphi\, ECT_{x,\, t-i} + V_{x,t} \qquad (14)$$

where, most importantly, M, N, O, P, Q, and R are the coefficients of the above models; φ is the coefficient of adjustment towards equilibrium in the long term; U and V are random error terms; and ECT denotes the deviations from the long-term equilibrium between the two lagged series. The ECM, in a way, captures an element of the speed of adjustment at which a dependent variable returns to equilibrium. Moreover, in comparing the ECT of the Johansen procedure versus the ARDL versus the NARDL approach, the ECT of the NARDL model should have the most enhanced speed of adjustment towards equilibrium among the three models considered. This should be followed by the ARDL model and the Johansen model (Demir et al. 2021; Mhadhbi et al. 2021; Nkoro and Uko 2016; Shin et al. 2014; Banerjee et al. 1993).

4. Data and Summary Statistics

4.1. Data

To evaluate and analyze the relations between the variables used in this study, three separate types of data with daily frequency were gathered: (1) Bitcoin futures and spot data, (2) investor attention data, and (3) investor sentiment data. The Bitcoin futures and spot data were obtained from Bloomberg. The important attention data were gathered from Google Trends (Google Search Volume Index—GSVI), which allows us to compare the relative popularity of search terms for specific time periods and regions. The unique Bitcoin sentiment data, which were based on a multifactorial crypto market sentiment analysis, were collected from the alternative.me website. The time period for which the data were gathered for the different variables used in this study spanned from 1 February 2018 to 8 September 2022. Figure 1 shows a graph of the Bitcoin spot, the two rolling Bitcoin futures, and the sentiment and attention variables through time, and Table 1 describes all the variables used in this study.
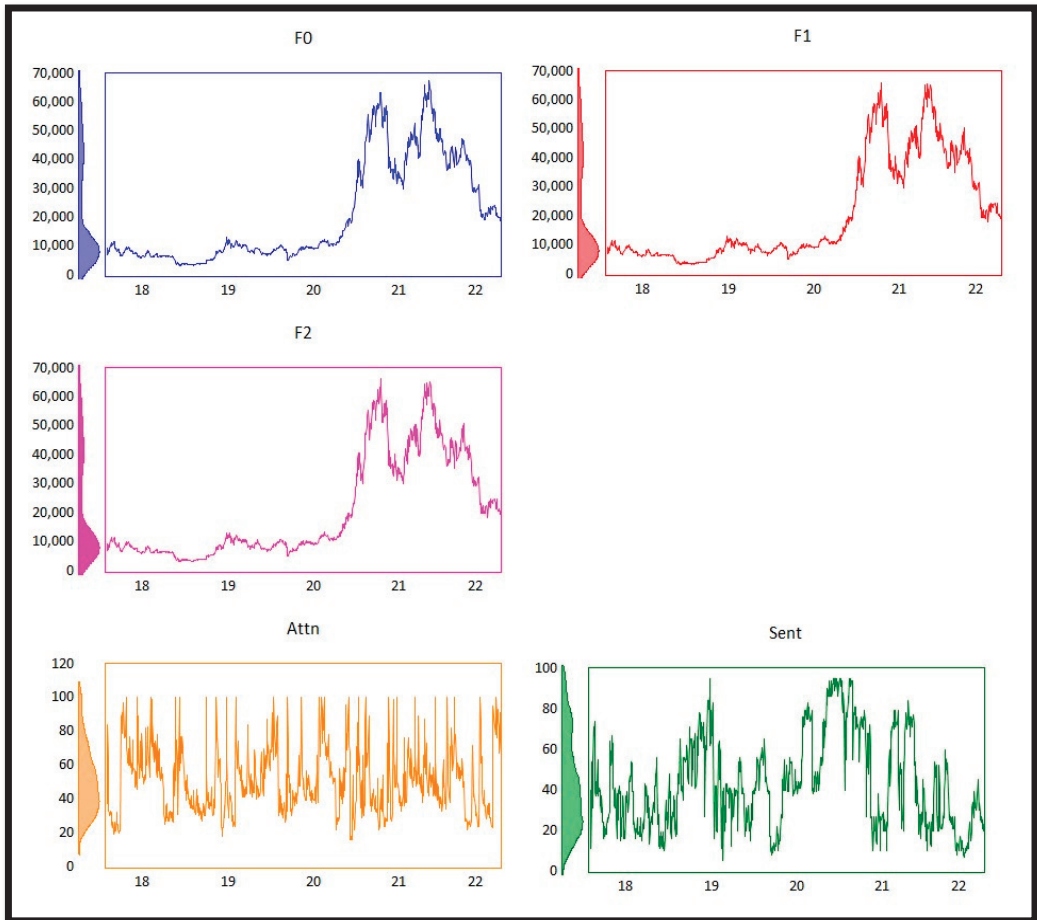
Figure 1. Bitcoin spot, futures, attention, and sentiment over time.

Table 1. Description of the variables used in this paper.

Variable Name Indicator	Variable Symbol
Bitcoin Spot	F0
Bitcoin—Near-Term Futures	F1
Bitcoin—Next-Term Futures	F2
Bitcoin Sentiment—Fear and Greed Index	Sent
Bitcoin Attention—Google Search Volume Index	Attn

This table provides the descriptions of the different variables used in this research paper.

4.2. Bitcoin Futures and Spot

Bitcoin is a decentralized cryptocurrency and technically the crypto industry's first asset, was launched in 2009, following the white paper by Satoshi Nakamoto in 2008. There is a maximum supply of 21 million bitcoins, and their trading has gained significance over time, causing them to be the largest and most popular cryptocurrency by market capitalization (Howell et al. 2020; Hashemi Joo et al. 2020). The active trading of Bitcoin and the fervent interest of investors has given rise to both regulated and unregulated Bitcoin derivative markets. Bitcoin futures contracts were first offered and traded on 10 December 2017 by the Chicago Board Options Exchange (CBOE). Similarly, they were offered and first traded on 17 December 2017 by the Chicago Mercantile Exchange (CME). This study only includes the listings from the CME, gathered from Bloomberg, since the volumes traded on the CME are significantly larger than those for the CBOE and the listings are regulated by the Commodity Futures Trading Commission (CFTC)[7]. All CME Bitcoin futures contracts generally expire on the last Friday of the month and are cash-settled rather than taking the delivery of actual Bitcoin[8].

We organize two Bitcoin futures price series. This is accomplished by grouping the two adjacent futures contracts based on time to maturity. The groups "nearest-term" and "next-term" represent the futures contracts that have the shortest and the second shortest time to maturity at a given date for the Bitcoin spot. The two futures price series are gathered, each with rolling contracts[9], and are recorded in terms of another currency[10]. It is important to note that this price varies across exchanges, mainly due to different fee policies and cash-out methods. The implication of this is that the source that we gather our data from matters. Since exchange-based data providers present their own quotes, whereas extrinsic data providers like Bloomberg or Coinbase compute their own indexes, usually a weighted average of all prices across major exchanges, we use the data from Bloomberg, consistent with past research (Baur and Dimpfl 2021; Janson and Karoubi 2021; Hattori and Ishida 2020, 2021; Cermak 2017).

4.3. Investor Attention

In examining several investor attention measures, such as turnover, extreme returns, news, etc., past research has shown evidence that the Google Search Volume Index (GSVI) acts as a direct and unambiguous measure of retail investor attention; it leads all other investor attention measures and an increase in GSVI results in an increase in stock liquidity due to the positive price pressure on the underlying asset from retail investors (Ding and Hou 2015; Da et al. 2011; Mondria et al. 2010). The GSVI compares the relative popularity of search words relative to the entire volume of searches on Google (Woloszko 2020; Stephens-Davidowitz and Varian 2015), and an increasing GSVI measure does not imply that there are more searches currently than in the past. It only means that a larger share of the searches on Google are dedicated to the specific search word; hence, its use as a proxy for information demand should be performed with care. Nevertheless, since the number of users using a Google search is a random sample of the total internet users, this measure lends itself to valid interpretation (Aslanidis et al. 2022; Tong et al. 2022). It is also important to note that while searching for words, the GSVI includes all searches that contain the word of interest, and it does not differentiate when additional inconsequential words are included in the word of interest. In order to increase the relevance of the attention measure, we

augment the main search word, "Bitcoin", with individual additional search words such as "Futures", "Price", "Spot", and "Trading", along with a plus operator, and we take the average of all the resulting measures for a particular date. These additional words and the plus operator aid in mitigating the problem of a resultant generic search, as well as the tacit challenge of smaller search volumes (Heyman et al. 2019; Yung and Nafar 2017; Han et al. 2017).

4.4. Investor Sentiment

The Fear and Greed Index (FGI) for Bitcoin, which is used as a measure of investor sentiment in this study, is gathered from a unique and comprehensive data source[11] that, in essence, tries to capture the emotional state of the cryptocurrency market. The FGI is constructed based on the expectation that investors are likely to become greedy (fearful) when the market price for assets increases (decreases), inducing an overreaction. Extreme fear can result when Bitcoin prices are far below their intrinsic value, and too much greed can result when prices are far above what they should be worth (Gunay et al. 2022; Mokni et al. 2022). It is important to note that the FGI takes values that range between 0 and 100. The range of values 0 (100) indicates the behavior or emotional state of investors, i.e., "Extreme Fear (Extreme Greed)". This multifactor FGI is constructed based on six important and distinctive factors.

The FGI uses volatility (25%) as the first factor because an unanticipated volatility increase is typically construed as an indicator of fear in investors. The ratio of market momentum to volume (25%) is used as the second factor because high buying volumes in a positive market are typically considered as greedy actions by investors. Social media (15%) is used as the third factor because an unusually high interaction rate on social media in terms of the count of posts and various hashtags is typically interpreted as an increase in public interest or corresponding greedy behavior by investors. Although the fourth factor of surveys (15%) or people polls is currently paused, they were included as a measure of direct perception by investors in the Bitcoin market for the time period for which we gather data.

Dominance (10%) is used as the fifth factor in constructing the FGI because the market capitalization share of Bitcoin with reference to all the available cryptocurrencies, in essence, measures the substitution of investment capital and fear among investors. Bitcoin dominance is generally understood to be caused by the reduction of investments in speculative alt-coin. Over time, Bitcoin has become the safe haven of cryptocurrency. Finally, the search trends (10%) are used as the sixth and last factor in constructing the FGI. In essence, the FGI uses both endogenous and exogenous factors in its construction (Gunay et al. 2022; Mokni et al. 2022; Guler 2021). To the best of our knowledge, this paper is the first to use the FGI to evaluate the linkages between investor sentiment and the Bitcoin futures and spot.

4.5. Descriptive Statistics, Returns, and Correlations

Table 2 exhibits the descriptive statistics for all variables across the entire study period. The average Bitcoin spot price (20,385.18) is quite similar to the average of both the near-term futures (20,413.64) and next-term futures (20,493.34). Both the Bitcoin futures (F1 and F2) are marginally higher than the spot (F0) in average price levels. Although the median values have a similar relation in comparing the futures and the spot, essentially, they are only 50% of the average values. The average Bitcoin sentiment (43.31) and its respective median (40) indicate that the sentiment on average is classified as "Fear"[12]. Similarly, the mean Bitcoin attention (50.90) is very close to the median (48). All the variables used in this study show positive skewness, indicating longer right tails, and the Jarque–Bera test results show evidence that all the variables are not normally distributed. The kurtosis numbers for all the variables were greater than +2, indicating a peaked distribution.

Table 2. Descriptive statistics in levels.

Entire Period Data Statistics (1 February 2018 to 8 September 2022)					
	F0	F1	F2	Sent	Attn
Mean	20,385.18	20,413.64	20,493.34	43.31	50.90
Median	10,224.14	10,362.27	10,495.62	40.00	48.00
Maximum	67,566.83	66,149.11	66,379.90	95.00	100.00
Minimum	3242.49	3138.02	3110.48	5.00	16.00
Std. Dev.	17,720.81	17,673.61	17,651.18	22.47	18.98
Skewness	0.98	0.98	0.97	0.51	0.62
Kurtosis	2.50	2.49	2.47	2.29	2.81
Jarque–Bera	195.32	193.61	191.08	74.11	74.01
Probability	0.00	0.00	0.00	0.00	0.00
Sum	2.33×10^7	2.33×10^7	2.34×10^7	4.95×10^4	5.81×10^4
Sum Sq. Dev.	3.58×10^{11}	3.56×10^{11}	3.55×10^{11}	5.76×10^5	4.11×10^5
Observations	1142	1142	1142	1142	1142

This table shows the descriptive statistics of the entire period's data in levels. The summary statistics reported include mean, median, max, min, standard deviation, skewness, kurtosis, Jarque–Bera, probability, sum, sum square deviation, and number of observations. The variables are defined in Table 1.

Table 3 illustrates the descriptive statistics for returns or changes in the different series. Panel A gives the summary statistics, and Panel B gives the counts of each series within the specified percentage bins. The average daily return of the Bitcoin spot (0.17%) in Panel A is almost 29% lower than the average daily Bitcoin futures return (0.22%). Moreover, the longer-term rolling futures have a lower median value than the shorter-term rolling futures. It should be noted that the Bitcoin futures (F1 and F2) offer higher returns on average compared to the Bitcoin spot (F0), but their standard deviations are higher as well. Regarding the count distribution of the returns shown in Panel B, most of the returns for the Bitcoin spot and futures fall within the range of −5% to +5% (approximately 75% of the returns data)[13]. In looking at the direction of the returns and the count distribution of the returns, it is clear that both the Bitcoin futures and the spot prices move in the same direction, indicating the least benefit for hedging. The sentiment and attention variables indicate that investors pay more attention to the Bitcoin asset when overall sentiment can be classified as "Fear" or "Extreme Fear".

Table 3. Summary statistics.

Panel A: Entire Period Data Statistics					
	F0	F1	F2	Sent	Attn
Mean	0.17%	0.22%	0.22%	3.18%	2.62%
Median	0.12%	0.15%	0.11%	0.00%	−1.89%
Maximum	18.75%	23.06%	23.10%	560.00%	316.67%
Minimum	−37.17%	−37.02%	−36.81%	−75.56%	−73.53%
Std. Dev.	4.56%	5.53%	5.55%	30.81%	27.55%
Skewness	−44.91%	−7.39%	−7.58%	650.00%	443.45%
Kurtosis	619.55%	341.53%	334.84%	9881.63%	3585.56%
Count	1141	1141	1141	1141	1141
Panel B: Count Distribution of Returns					
	F0	F1	F2	Sent	Attn
<−10%	24	42	43	242	269
−10 to −5%	81	102	100	104	188
−5 to −2%	176	183	186	116	103
−2 to −0.5%	177	165	162	41	39
−0.5 to 0.5%	180	132	131	114	81
0.5 to 2%	203	161	162	33	16
2 to 5%	179	194	193	87	98
5 to 10%	92	105	110	124	92
>10%	29	57	54	280	255

This table presents the summary statistics of the entire period's data in returns, i.e., from 2 February 2018 to 8 September 2022. Panel A shows the descriptive statistics of each variable used in this study. Panel B shows the count distribution of the returns for each variable. The variables are defined in Table 1.

Table 4 presents the Spearman rank-order correlation statistics. As is evident, the correlations between the near-term and next-term Bitcoin futures and spot are almost close to +1, reinforcing the inference that we made when looking at the count distribution of returns, namely that these securities offer the least benefit for diversification. Interestingly, while the sentiment variable is also positively correlated with all the Bitcoin variables (F0, F1, and F2), the attention variable is negatively correlated with the Bitcoin variables. The correlations of the attention variable are marginally lower than zero, indicating that if the attention to Bitcoin among investors increases, a likely reason is fear sentiments, which would plausibly exert negative price pressure on both the Bitcoin spot and futures.

Table 4. Correlation statistics.

Variables	F0	F1	F2	Sent	Attention
F0	1				
F1	0.9993 **	1			
F2	0.9991 **	0.9999 **	1		
Sent	0.2543 **	0.2530 **	0.2521 **	1	
Attn	−0.0531	−0.0547	−0.0552	0.0124	1

This table shows the Spearman rank correlation coefficients among the bitcoin spot, near-term bitcoin future, next-term bitcoin future, bitcoin sentiment, and bitcoin investor attention. The variables are defined in Table 1. ** indicates significance at the 5% level.

5. Empirical Results

To facilitate the investigation of both the short-term and long-term relations between the Bitcoin futures, spot, sentiment, and attention variables, we need to empirically evaluate the stationarity of each series and identify the order of integration. Since both the bivariate Granger causality tests used to analyze the short-term relations and the Johansen test used to evaluate the long-term relations are sensitive to the choice of lag length, we need to examine the lag structure resulting from an unrestricted VAR model for each time series and determine the optimal lag lengths. The long-term relations can be evaluated using the ARDL and NARDL tests, which eliminate the limitation on the order of integration imposed by the Johansen procedure. The NARDL model specifically allows for an asymmetric long-run relation between the variables. The following sections analyze the results of each of these tests, and, in the end, we evaluate an error correction model for the Johansen, ARDL, and NARDL procedures.

5.1. Stationarity and Optimal Lag Length

Examining the stationarity properties of all the variables is critical in performing a cointegration analysis. The ADF (Dickey and Fuller 1979, 1981), KPSS (Kwiatkowski et al. 1992), and Ng and Perron (Ng and Perron 2001) tests are used to evaluate the unit root. Panels A and B in Table 5 present the results of the ADF and the KPSS and NG and Perron test results, respectively. Regardless of whether we include only an intercept or both an intercept and a trend when ADF tests are performed on the level data of the various series, we do not reject the null hypothesis of non-stationarity for three out of the five variables under examination. The Bitcoin futures (F1 and F2) and the Bitcoin spot (F0) variables show evidence of a unit root and must be first differenced. The null hypothesis is rejected for the sentiment and attention variables, and both these variables are stationary in level form. Considering the three series (F0, F1, and F2), which have unit roots in the level form, the Panel A results of the ADF test on the first differences show evidence to reject the null hypothesis of non-stationarity. Thus, these three series become stationary after the first differencing and are integrated into order one, *I*(1).

Table 5. Unit root test statistics.

Panel A:						
	F0	F1	F2	Sent	Attn	Critical Values
ADF Tests on Levels						
With Intercept	−1.2503	−1.3115	−1.3068	−4.9547	−11.2451	−2.8639 **
p-values	0.6543	0.6261	0.6283	0.0000 **	0.0000 **	
With Intercept and Trend	−1.3029	−1.4342	−1.4402	−4.9368	−11.2390	−3.4138 **
p-values	0.8865	0.8506	0.8488	0.0003 **	0.0000 **	
ADF Tests on First Difference						
With Intercept	−22.0040	−37.7488	−38.4890	−41.5384	−30.1591	−2.8639 **
p-values	0.0000 **	0.0000 **	0.0000 **	0.0000 **	0.0000 **	
With Intercept and Trend	−22.0020	−37.7393	−38.4790	−41.5246	−30.1469	−3.4138 **
p-values	0.0000 **	0.0000 **	0.0000 **	0.0000 **	0.0000 **	
Panel B:						
	F0	F1	F2	Sent	Attn	Critical Values
KPSS Tests on Level						
With Intercept and Trend	0.3711	0.3689	0.3680	0.0336	0.0392	0.1460 **
KPSS Tests on First Difference						
With Intercept and Trend	0.1419	0.1328	0.1286	0.0178	0.0327	0.1460 **
Ng and Perron Tests on Level (With Intercept and Trend)						
MZa	−4.2466	−4.8360	−4.8306	−44.0376	−194.3780	−17.3000 **
MZt	−1.3597	−1.4647	−1.4658	−4.6610	−9.8294	−2.9100 **
Ng and Perron Tests on First Difference (With Intercept and Trend)						
MZa	−458.9110	−564.7240	−563.1830	−3.3718	−2.5451	−17.3000 **
MZt	−15.1475	−16.8036	−16.7806	−1.2553	−1.12117	−2.9100 **

This table shows the unit root statistics for each variable used in the study using levels and first differences. Panel A presents the unit root test results using the augmented Dickey–Fuller tests. Panel B shows the unit root tests results using the KPSS and the Ng and Perron tests. The variables are defined in Table 1. ** indicates significance at the 5% level.

In contrast to the ADF test, which posits the null hypothesis of non-stationarity, the KPSS test sets the null hypothesis as a stationary series and the alternative hypothesis as non-stationarity. Panel B in Table 5 presents results obtained for level data, which show evidence to reject the null hypothesis for three of the variables (F0, F1, and F2) and fail to reject the null hypothesis for the sentiment and attention variables. We fail to reject the null hypothesis for the F0, F1, and F2 series when the KPSS tests are applied to the first differenced data. The results of the KPSS tests are consistent with the ADF tests. Since these stationarity tests (ADF and KPSS) are known to suffer potentially from severe finite sample power and size problems (De Jong et al. 2007; Keblowski and Welfe 2004; DeJong et al. 1992), we further augment the stationarity results by conducting the Ng and Perron test. This test provides good power and reliable size properties to reconfirm the results of the ADF and KPSS tests. In Panel B of Table 5, we present the Ng and Perron test results. Based on the results, we reject the null hypothesis of stationarity on the level for variables F0, F1, and F2. After the first differencing, all three non-stationary series become stationary, and these results are consistent with the previous ADF and KPSS tests.

In conjunction with the stationarity tests, it is critical to examine the lag structure of the unrestricted VAR model for each series in determining the optimal lag length. Table 6 shows the results of the lag length analysis. Following past research, the lag length is chosen using the Akaike Information Criterion (AIC) (Hatemi-J and Hacker 2008; Akaike 1974, 1981). The results in Table 6 show that the AIC criterion indicates that the model must have optimally eight lags for the Bitcoin futures (F1 and F2) and spot (F0) variables. Although the lag lengths for the sentiment and attention variables are depicted, they cannot

be used in the cointegration analysis since the variables are already stationary and not integrated of at least order one, $I(1)$. The eight lags that are determined will be used in the long-run relation analysis for the respective variables.

Table 6. VAR—Lag length analysis.

	F0	F1	F2	Sent	Attn
LR	7	8	8	2	3
FPE	8	8	8	2	3
AIC	8 **	8 **	8 **	2 **	3 **
SC	1	2	2	2	1
HQ	2	2	2	2	3

This table shows the VAR—optimal lag length analysis. These results are used to conduct the cointegration analysis. The optimal lag length for each variable is determined to perform the cointegration analysis. Since sentiment and attention are not integrated of at least order one, they are not included in the cointegration analysis. The table shows the optimal number of lags as per different criteria, i.e., LR, FPE, AIC, SC, and HQ. The optimal lag length is selected based on the AIC. The variables are defined in Table 1. ** indicates significance at the 5% level.

5.2. Granger Causality

We next analyze the short-term relations between the different variables using the bivariate Granger causality test (Granger 1969, 1980). "Causality", in this case, simply implies that the past values of one variable can be used to predict the future values of another variable and is tested using a standard F-test. The results of the Granger causality tests are shown in Table 7.

Table 7. Granger causality results.

Null Hypothesis	F-Statistic	Probability	Conclusion
D(F1) does not Granger cause D(F0)	3.0219	0.0023 **	Bidirectional Causality
D(F0) does not Granger cause D(F1)	21.5939	0.0000 **	
D(F2) does not Granger cause D(F0)	2.3975	0.0145 **	Bidirectional Causality
D(F0) does not Granger cause D(F2)	21.7286	0.0000 **	
D(F2) does not Granger cause D(F1)	3.1640	0.0015 **	Bidirectional Causality
D(F1) does not Granger cause D(F2)	3.7429	0.0002 **	
SENT does not Granger cause D(F0)	0.6479	0.7376	Unidirectional Causality
D(F0) does not Granger cause SENT	41.2038	0.0000 **	
SENT does not Granger cause D(F1)	1.5790	0.1266	Unidirectional Causality
D(F1) does not Granger cause SENT	26.7861	0.0000 **	
SENT does not Granger cause D(F2)	1.5716	0.1288	Unidirectional Causality
D(F2) does not Granger cause SENT	26.8062	0.0000 **	
ATTN does not Granger cause D(F0)	2.1761	0.0269 **	Unidirectional Causality
D(F0) does not Granger cause ATTN	1.6223	0.1140	
ATTN does not Granger cause D(F1)	2.6111	0.0078 **	Bidirectional Causality
D(F1) does not Granger cause ATTN	1.9518	0.0493 **	
ATTN does not Granger cause D(F2)	2.6118	0.0078 **	Bidirectional Causality
D(F2) does not Granger cause ATTN	2.0662	0.0363 **	
ATTN does not Granger cause SENT	1.2223	0.2821	No Causality
SENT does not Granger cause ATTN	0.8206	0.5844	

This table shows the bivariate Granger causality test results. The variables are defined in Table 1. In the null hypothesis, when the sentence starts with D() of a particular variable name, it implies that this series had to be first differenced to make it stationary. ** indicates significance at the 5% level.

The pairwise Granger causality tests between the Bitcoin spot (F0) and near-term (F1) and next-term (F2) Bitcoin futures show evidence of bidirectional causality. This indicates that the market between the futures and spot for Bitcoin is efficient, and there is no price

discovery function between these assets. The informational content that affects the prices of both the futures and spot for Bitcoin is effectively reflected in the prices simultaneously. In contrast, all three variables (F0, F1, and F2) have unidirectional causality with the Bitcoin sentiment variable. The lagged values of the Bitcoin spot as well as the Bitcoin futures show evidence of predicting the future values of Bitcoin sentiment. There is a clear lead lag relation between the price movements of the Bitcoin spot and futures and the Bitcoin sentiment variable. The evidence supports the short-term dependence of the sentiment variable on the price movements of the spot and futures variables.

Evaluating the results of the retail investor Bitcoin attention variable, there is a unidirectional causal relation between the attention and Bitcoin spot variable and a bidirectional causal relation between the attention and Bitcoin futures variables. This evidence implies that retail investors who pay attention to the Bitcoin asset engage in transactions of the asset in the spot market and effectively impact the price of the Bitcoin spot. Interestingly, however, the information content in the same attention variable does not translate into activity in the Bitcoin futures asset, and the relation between them is efficient. Finally, the attention and sentiment variables have no causal relation with each other. Overall, these findings are significant, and this study is the first to collectively highlight these short-term relations.

5.3. Johansen, ARDL, and NARDL Cointegration

This study employs three cointegration techniques to analyze the long-term relation between the Bitcoin spot (F0) and the Bitcoin futures (F1 and F2). First, we employ the Johansen cointegration test developed by Johansen (Johansen 1988, 1991). Second, we employ the ARDL bounds cointegration test (Sari et al. 2008; Pesaran et al. 2001). Third, we employ the NARDL bounds cointegration test (Demir et al. 2021; Mhadhbi et al. 2021; Shin et al. 2014). As part of the Johansen test, we use both the trace test statistic and the maximum eigenvalue statistic to test the hypothesis of cointegration. After confirming previously that all three series (F0, F1, and F2) are integrated of the same order, $I(1)$, and identifying the optimal lag length of eight lags, the multivariable Johansen test is conducted on these three variables and the results are shown in Table 8.

Table 8. Johansen cointegration results.

Series	Trace Test Statistics				
	No. of CE(s)	**Eigenvalue**	**Statistic**	**Critical Value**	**Probability**
F0 F1 F2	None	0.0817	116.3167	29.7971	0.0000 **
	At most 1	0.0157	19.7959	15.4947	0.0105 **
	At most 2	0.0016	1.8515	3.8415	0.1736
Series	**Max Test Statistics**				
	No. of CE(s)	**Eigenvalue**	**Statistic**	**Critical Value**	**Probability**
F0 F1 F2	None	0.0817	96.5208	21.1316	0.0000 **
	At most 1	0.0157	17.9444	14.2646	0.0125 **
	At most 2	0.0016	1.8515	3.8415	0.1736

This table reveals the Johansen cointegration results. The variables are defined in Table 1. ** indicates significance at the 5% level.

The results in Table 8 indicate that between the Bitcoin spot (F0) and the two futures, near-term (F1) and next-term (F2), there is evidence of at least one cointegrating relation. The number of cointegrating relations is tested sequentially, starting at zero and incrementing to one and two. The null hypotheses of having zero and one cointegrating vectors are both rejected at the 5% significance level. We fail to reject the last sequential analysis with two cointegrating vectors after analyzing both the trace and the maximum eigenvalue test statistics. This implies that, at most, there is one long-run relation between the variables analyzed.

Following the results of the Johansen procedure, we relax the requirement of the series being integrated of at least order one, *I(1)*, and conduct the ARDL test. The results of the ARDL test are shown in Table 9. Panel A presents the statistics for the bounds test, and Panel B shows the long-run coefficients. As part of the ARDL bounds cointegration test, we use the F-test and the t-test statistics to test the hypothesis of cointegration. Since the absolute values of both the F-test and the t-test are greater than the upper bound *I(1)* statistic values, we reject the null hypothesis and determine that these variables do have a long-run cointegrating relation among them. Following the results of the ARDL procedure, we relax the requirement of a symmetric relation between the Bitcoin futures and spot and conduct the NARDL test, which allows for asymmetry. The results of the NARDL test are shown in Panels C and D in Table 9. Panel C presents the statistics for the bounds test, and Panel D shows the long-run coefficients. As part of the NARDL bounds cointegration test, we use the F-test to evaluate the results. Since the absolute values of the F-test are greater than the upper bound *I(1)* statistic values, we reject the null hypothesis of no cointegrating relations and conclude that these variables do have a long-run cointegrating relation among them. Moreover, the responses to both lagged positive and negative changes are significant, with the response to lagged positive changes being stronger than the response to lagged negative changes. This is captured in the long-run coefficients in Panel D of Table 9. In summary, the findings of the Johansen, ARDL, and NARDL procedures are consistent, and there is evidence of at least one long-run relation between the Bitcoin futures and spot variables and the existence of asymmetry in this relation.

Table 9. ARDL and NARDL cointegration tests.

Panel A: ARDL Cointegration Bounds Test			
Test Statistic	Value	Lower Bound—I(0)	Upper Bound—I(1)
F-statistic	12.9757 **	2.7200	3.8300
t-statistic	−6.2156 **	−1.9500	−3.0200

Panel B: ARDL Long-Run Coefficients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
F1	1.8757 **	0.1659	11.3065	0.0000
F2	−0.8745 **	0.1656	−5.2796	0.0000

Panel C: NARDL Cointegration Bounds Test			
Test Statistic	Value	Lower Bound—I(0)	Upper Bound—I(1)
F-statistic	20.8186 **	2.5600	3.4900

Panel D: NARDL Long-Run Coefficients				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
F1_POS	1.9532 **	0.1265	15.4352	0.0000
F1_NEG	1.7802 **	0.1434	12.4101	0.0000
F2_POS	−0.9563 **	0.1274	−7.5037	0.0000
F2_NEG	−0.7847 **	0.1440	−5.4486	0.0000

This table presents both the ARDL and NARDL cointegration results. Panel A shows the bounds test results for the ARDL model and Panel B shows the long-run coefficients from the ARDL model. Panel C shows the bounds test results for the NARDL model and Panel D shows the long-run coefficients from the NARDL model. The variables are defined in Table 1. ** indicates significance at the 5% level.

5.4. Error Correction Model

Consistent with the findings of our cointegration analysis and identifying the existence of at least one long-run cointegrating relation between the Bitcoin futures and spot variables and some asymmetry, the short-run dynamics need to be analyzed using the error correction model (ECM) instead of the VAR model. The general equation of the error correction model is shown in Equations (12) and (13) of Section 3. To establish a long-run equilibrium between the series analyzed, the coefficient, or the speed of adjustment term from the

model, must satisfy two conditions. First, the coefficient should be negative; second, it should be statistically significant.

The results of the ECM are presented in Table 10. The dependent variable used in the ECM is the Bitcoin spot (F0), and the independent variables are the Bitcoin futures (F1 and F2 or F1_POS, F1_NEG, F2_POS, and F2_NEG) for the Johansen, ARDL, and NARDL procedures. Using the optimal lag length of eight and the ECM from the Johansen procedure, the coefficient in Table 10 satisfies both the conditions of being negative and statistically significant. This implies that the variables share a common trend, which describes the long-run relation between them. To augment these findings and by including an automatic selection of lag lengths and the ARDL or the NARDL procedure, the respective coefficients show evidence to satisfy both requirements, and the results are consistent with the error correction coefficient from the Johansen procedure. These results not only reinforce the existence of a long-run relation but also show evidence of asymmetry in this relation. In summary, the relation between the Bitcoin futures and spot undergoes an adjustment process through time towards long-run equilibrium, and the NARDL model has the fastest speed of adjustment towards long-run equilibrium.

Table 10. ECM estimation.

Method	Dependent Variable	Independent Variables	Error Correction Term (Speed of Adjustment)			
			Coefficient	Standard Error	t-Statistic	Probability
Johansen Test	F0	F1 F2	−0.2498 **	0.0849	−2.9430	0.0033
ARDL Test	F0	F1 F2	−0.4434 **	0.0710	−6.2448	0.0000
NARDL Test	F0	F1_POS F1_NEG F2_POS F2_NEG	−0.6407 **	0.0572	−11.2018	0.0000

This table shows the results of the error correction model (ECM) results using the Johansen test and the ARDL tests. The dependent variables are the spot Bitcoin. The independent variables in all specifications are the two Bitcoin futures series. The variables are defined in Table 1. ** indicates significance at the 5% level.

6. Conclusions

The main objective of this research is to investigate the causal relations in the short term and the cointegrated relations in the long term among the Bitcoin spot price and futures, investor sentiment, and investor attention. Previous research on the cryptocurrency market has produced mixed results concerning the price discovery function between the Bitcoin assets. It has also been limited in scope in analyzing the long-term relation between the Bitcoin futures and spot and obscure in examining the impact of investor sentiment and attention on Bitcoin assets. This study aims to provide conclusive evidence of the price discovery function between the spot price and futures and investigate the effect of investor sentiment and attention on both the Bitcoin futures and spot assets. To achieve these objectives, the study employs the bivariate Granger causality methodology to investigate the short-term relations among the variables of interest. Furthermore, to investigate the long-term relations among the variables of interest, the study employs the ARDL, NARDL, and Johansen cointegration procedures with an error correction mechanism.

While the measure of investor attention (GSVI) and crypto sentiment (FGI) used in this study enables us to capture the interlinkages with Bitcoin assets, we wish to highlight the limitations in the scope and the existence of alternate measures (Ali et al. 2022). The purpose of this study is not to look at all the available measures of sentiment and investor attention; we leave this to future research. Moreover, from a methodology perspective, there are alternate methods like the cross-quantilogram, dynamic conditional correlation (DCC), wavelet multiscale decomposition, quantile-on-quantile methods, etc., which could be employed to look at the interlinkages, and we leave this also to future research. Importantly, we also do not answer two important questions as part of this study. This is related to identifying the nature of the participants in the Bitcoin market and the levels of attention at which they become active participants in the asset market, thereby influencing the prices. While there is some recent work on these questions (Ülkü et al. 2023), we highlight the need

to examine this relation as part of future work. Overall, our focus is only on Bitcoin assets, and, in the future, we will extend this to include all other crypto assets.

In summary, this research contributes to the existing Bitcoin pricing literature by providing the first empirical examination of the interlinkages between the Bitcoin futures, Bitcoin spot, investor sentiment, and attention as a collective. The study has five main findings. First, there is no statistical evidence of price discovery between the Bitcoin futures and the spot market, and the term structure of the Bitcoin futures does not play a significant role in this relation. Second, there is statistical evidence of a lead lag relation between Bitcoin sentiment and both the Bitcoin futures and spot prices. This suggests that changes in market participants' opinions and perceptions lead to trading activity in both markets, affecting the prices of both assets. Third, investor attention statistically leads the Bitcoin spot market but does not exhibit any lead lag relation with the Bitcoin futures. Fourth, the study finds statistical evidence of a long-run cointegrating relation between the Bitcoin futures and spot prices, and the results are robust to the type of cointegration procedure used. Finally, the speed of adjustment towards equilibrium in the long-run cointegration relation is stronger and more significant when using the NARDL and ARDL procedures than the Johansen test, despite the robustness of all testing procedures.

The findings of this study have practical implications for retail and institutional investors, portfolio managers, regulators, and institutions such as the SEC, CBOE, and CME in understanding the interplay of these assets and market forces for optimal trading decisions and market regulation. In conclusion, this study contributes to the finance literature by establishing connections between the Bitcoin spot, futures, investor sentiment, and investor attention in the cryptocurrency markets. The consideration of investor sentiment and attention allows for a better understanding of the drivers of the spot and futures market and the identification of both long-term and short-term and lead lag relations, ultimately leading to a deeper comprehension of investor and asset behavior in the financial markets. Lastly, these findings can aid retail investors, institutional investors, portfolio managers, and regulators in making informed trading decisions, considering the price discovery between the Bitcoin futures and the spot market, as well as the impact of investor sentiment and attention on Bitcoin prices, thereby enhancing the efficiency of cryptocurrency markets.

Author Contributions: Conceptualization, A.N. and H.P.; Methodology, A.N.; Software, A.N.; Validation, A.N., H.P. and R.A.; Formal analysis, A.N.; Investigation, A.N. and H.P.; Resources, A.N., R.A. and H.P.; Data curation, R.A.; Writing—original draft preparation, A.N. and H.P.; Writing—review and editing, H.P. and R.A.; Visualization, A.N., H.P. and R.A.; Supervision, H.P.; Project administration, A.N. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

[1] https://markets.businessinsider.com/news/currencies/cryptocurrency-women-investing-asset-holdings-cash-markets-bitcoin-etoro-survey-2023-1 (accessed on 22 June 2023).

[2] https://beincrypto.com/bitcoin-the-best-performing-asset-class-in-q1-2023/ and https://www.forbes.com/advisor/investing/cryptocurrency/top-10-cryptocurrencies/ (accessed on 22 June 2023).

[3] https://www.slickcharts.com/currency and https://coinmarketcap.com/ (accessed on 22 June 2023).

[4] https://argoblockchain.com/articles/5-features-that-make-bitcoin-a-unique-asset-class (accessed on 22 June 2023).

[5] https://www.fxstreet.com/cryptocurrencies/resources/brokers-what-are-bitcoin-futures (accessed on 22 June 2023).

[6] The sample period used by Akyildirim et al. (2020) is from 12 December 2017 to 26 February 2018. The sample period used by Kapar and Olmo (2019) is from from 18 December 2017 to 16 May 2018. The sample period used by Baur and Dimpfl (2019) is

from 12 December 2017 to 18 October 2018. Finally, the sample period used by Corbet et al. (2018) spans from 26 September 2017 to 22 February 2018.

7   The CBOE announced in March 2019 that it was reviewing its approach to Bitcoin derivatives and would stop listing the Bitcoin futures contracts. In June 2019, the CBOE stopped adding new futures, so the trading of CBOE Bitcoin futures has ceased.

8   More information about the CME Bitcoin futures can be found on their website at https://www.cmegroup.com/education/bitcoin/cme-bitcoinfutures-frequently-asked-questions.html (accessed on 22 June 2023).

9   The nearest-term futures contain prices from the futures contract with the nearest maturity. When this current contract expires, it rolls into the futures contract with the next nearest maturity. This also occurs with the next-term or the second month maturity contracts.

10   In this study, we use the US Dollar (USD).

11   More details on this index can be found on the website https://alternative.me (accessed on 22 June 2023).

12   Although the counts of the individual classifications of "Extreme Greed", "Greed", "Neutral", "Fear", and "Extreme Fear" are not shown year after year, the counts were significantly more towards the "Fear" and "Extreme Fear" classification, year after year.

13   Almost 50% of the returns for the Bitcoin futures and spot were positive and 50% were negative.

References

Aalborg, Halvor Aarhus, Peter Molnár, and Jon Erik de Vries. 2019. What can explain the price, volatility and trading volume of Bitcoin? *Finance Research Letters* 29: 255–65. [CrossRef]

Akaike, Hirotugu. 1969a. Fitting Autoregressions for Prediction. *Annals of the Institute of Statistical Mathematics* XXI: 243–47.

Akaike, Hirotugu. 1969b. Statistical Predictor Identification. *Annals of the Institute of Statistical Mathematics* XXI: 201–17.

Akaike, Hirotugu. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–23. [CrossRef]

Akaike, Hirotugu. 1981. Likelihood of a model and information criteria. *Journal of Econometrics* 16: 3–14. [CrossRef]

Akanksha, Jalan, Roman Matkovskyy, and Andrew Urquhart. 2021. What effect did the introduction of Bitcoin futures have on the Bitcoin spot market? *The European Journal of Finance* 27: 1251–81.

Akyildirim, Erdinc, Oguzhan Cepni, Shaen Corbet, and Gazi Salah Uddin. 2021. Forecasting mid-price movement of bitcoin futures using machine learning. *Annals of Operations Research*, 1–32. [CrossRef]

Akyildirim, Erdinc, Shaen Corbet, Paraskevi Katsiampa, Neil Kellard, and Ahmet Sensoy. 2020. The development of Bitcoin futures: Exploring the interactions between cryptocurrency derivatives. *Finance Research Letters* 34: 101234. [CrossRef]

Alexander, Carol, and Daniel F. Heck. 2020. Price discovery in Bitcoin: The impact of unregulated markets. *Journal of Financial Stability* 50: 100776. [CrossRef]

Alexander, Carol, Jaehyuk Choi, Heungju Park, and Sungbin Sohn. 2020. BitMEX bitcoin derivatives: Price discovery, informational efficiency, and hedging effectiveness. *Journal of Futures Markets* 40: 23–43. [CrossRef]

Ali, Fahad, Elie Bouri, Nader Naifar, Syed Jawad Hussain Shahzad, and Mohammad AlAhmad. 2022. An examination of whether gold-backed Islamic cryptocurrencies are safe havens for international Islamic equity markets. *Research in International Business and Finance* 63: 101768. [CrossRef]

Aslanidis, Nektarios, Aurelio F. Bariviera, and Óscar G. López. 2022. The link between cryptocurrencies and Google Trends attention. *Finance Research Letters* 47: 102654. [CrossRef]

Banerjee, Anindya, Juan J. Dolado, John W. Galbraith, and David Hendry. 1993. *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Baur, Dirk G., and Thomas Dimpfl. 2019. Price Discovery in Bitcoin Spot or Futures? *Journal of Futures Markets* 39: 803–17. [CrossRef]

Baur, Dirk G., and Thomas Dimpfl. 2021. The volatility of Bitcoin and its role as a medium of exchange and a store of value. *Empirical Economics* 61: 2663–83. [CrossRef] [PubMed]

Cermak, Vavrinec. 2017. Can Bitcoin Become a Viable Alternative to Fiat Currencies? An Empirical Analysis of Bitcoin's Volatility Based on a GARCH Model. Business Economics, ERN: International Finance. Available online: https://creativematter.skidmore.edu/cgi/viewcontent.cgi?article=1067&context=econ_studt_schol (accessed on 22 June 2023).

Chan, Ngai Hang. 2010. *Time Series Applications to Finance with R and S-Plus*. Hoboken: John Wiley&Sons, Inc.

Chan, Wing Hong, Minh Le, and Yan Wendy Wu. 2019. Holding Bitcoin longer: The dynamic hedging abilities of Bitcoin. *The Quarterly Review of Economics and Finance* 71: 107–13. [CrossRef]

Cheah, Eng-Tuck, Tapas Mishra, Mamata Parhi, and Zhuang Zhang. 2018. Long Memory Interdependency and Inefficiency in Bitcoin Markets. *Economics Letters* 167: 18–25. [CrossRef]

Chen, Conghui, Lanlan Liu, and Ningru Zhao. 2020. Fear sentiment, uncertainty, and bitcoin price dynamics: The case of COVID-19. *Emerging Market Finance and Trade* 56: 2298–309. [CrossRef]

Cheng, Tingting, Junli Liu, Wenying Yao, and Albert Bo Zhao. 2021. The impact of COVID-19 pandemic on the volatility connectedness network of global stock market. *Pacific Basin Finance Journal* 71: 101678. [CrossRef]

Chevapatrakul, Thanaset, and Danilo V. Mascia. 2019. Detecting overreaction in the Bitcoin market: A quantile autoregression approach. *Finance Research Letters* 30: 371–77. [CrossRef]

Choi, Hyungeun. 2021. Investor attention and bitcoin liquidity: Evidence from bitcoin tweets. *Finance Research Letters* 39: 101555.

Corbet, Shaen, Brian Lucey, Maurice Peat, and Samuel Vigne. 2018. Bitcoin Futures—What Use are They? *Economics Letters* 172: 23–27. [CrossRef]

Da, Zhi, Joseph Engelberg, and Pengjie Gao. 2011. In Search of Attention. *The Journal of Finance* 66: 1461–99. [CrossRef]

Dastgir, Shabbir, Ender Demir, Gareth Downing, Giray Gozgor, and Chi Keung Marco Lau. 2019. The causal relationship between Bitcoin attention and Bitcoin returns: Evidence from the Copula-based Granger causality test. *Finance Research Letters* 28: 160–64. [CrossRef]

De Jong, Robert M., Christine Amsler, and Peter Schmidt. 2007. A robust version of the KPSS test based on indicators. *Journal of Econometrics* 137: 311–33. [CrossRef]

DeJong, David N., John C. Nankervis, N. Eugene Savin, and Charles H. Whiteman. 1992. The power problems of unit root test in time series with autoregressive errors. *Journal of Econometrics* 53: 323–43. [CrossRef]

Demir, Ender, Serdar Simonyan, Conrado-Diego García-Gómez, and Chi Keung Marco Lau. 2021. The asymmetric effect of bitcoin on altcoins: Evidence from the nonlinear autoregressive distributed lag (NARDL) model. *Finance Research Letters* 40: 101754. [CrossRef]

Dickey, David A., and Wayne A. Fuller. 1979. Distribution of estimators for time series regressions with unit root. *Journal of the American Statistical Association* 74: 427–31.

Dickey, David A., and Wayne A. Fuller. 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49: 1057–72. [CrossRef]

Ding, Rong, and Wenxuan Hou. 2015. Retail investor attention and stock liquidity. *Journal of International Financial Markets, Institutions and Money* 37: 12–26. [CrossRef]

Dubey, Priti. 2022. Short-run and long-run determinants of bitcoin returns: Transnational evidence. *Review of Behavioral Finance* 14: 533–44.

Dutta, Anupam, Elie Bouri, and David Roubaud. 2019. Nonlinear relationships amongst the implied volatilities of crude oil and precious metals. *Resources Policy* 61: 473–78. [CrossRef]

Dyhrberg, Anne H., Sean Foley, and Jiri Svec. 2018. How investible is Bitcoin? Analyzing the liquidity and transaction costs of Bitcoin markets. *Economics Letters* 171: 140–43. [CrossRef]

Enders, Walter. 2014. *Applied Econometric Time Series*, 4th ed. New York: John Wiley.

Engle, Robert F., and Clive W. J. Granger. 1987. Cointegration and Error Correction: Representation, Estimation and Testing. *Econometrica* 55: 251–76. [CrossRef]

Entrop, Oliver, Bart Frijns, and Marco Seruset. 2020. The determinants of price discovery on bitcoin markets. *Journal of Futures Markets* 40: 816–37. [CrossRef]

Eross, Andrea, Frank McGroarty, Andrew Urquhart, and Simon Wolfe. 2019. The intraday dynamics of bitcoin. *Research in International Business and Finance* 49: 71–81. [CrossRef]

Foley, Sean, Jonathan R. Karlsen, and Tālis J. Putniņš. 2019. Sex, drugs, and Bitcoin: How much illegal activity is financed through cryptocurrencies? *Review of Financial Studies* 32: 1798–853. [CrossRef]

Fry, John. 2018. Booms, busts and heavy-tails: The story of Bitcoin and cryptocurrency markets? *Economic Letters* 171: 225–29. [CrossRef]

Garcia, David, Claudio J. Tessone, Pavlin Mavrodiev, and Nicolas Perony. 2014. The digital traces of bubbles: Feedback cycles between socio-economic signals in the Bitcoin economy. *Journal of the Roayal Society Interface* 11: 20140623. [CrossRef] [PubMed]

Geuder, Julian, Harald Kinateder, and Niklas F. Wagner. 2019. Cryptocurrencies as financial bubbles: The case of Bitcoin. *Finance Research Letters* 31. Available online: https://www.sciencedirect.com/science/article/abs/pii/S1544612318306846 (accessed on 22 June 2023).

Ghatak, Subrata, and Jalal U. Siddiki. 2001. The use of the ARDL approach in estimating virtual exchange rates in India. *Journal of Applied Statistics* 28: 573–83. [CrossRef]

Granger, Clive W. J. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* 37: 424–38. [CrossRef]

Granger, Clive W. J. 1980. Testing for Causality: A Personal Viewpoint. *Journal of Economic Dynamics and Control* II: 329–52.

Granger, Clive W. J. 1988. Some recent developments in the concept of causality. *Journal of Econometrics* 39: 199–211. [CrossRef]

Granger, Clive W. J. 2001. *Essays in Econometrics: The Collected Papers of Clive W.J. Granger*. Cambridge: Cambridge University Press.

Griffin, John M., and Amin Shams. 2020. Is Bitcoin Really Untethered? *Journal of Finance. American Finance Association* 75: 1913–64. [CrossRef]

Gronwald, Marc. 2019. Is Bitcoin a Commodity? On price jumps, demand shocks, and certainty of supply. *Journal of International Money and Finance* 97: 86–92. [CrossRef]

Guegan, Dominique, and Thomas Renault. 2021. Does investor sentiment on social media provide robust information for Bitcoin returns predictability? *Finance Research Letters* 38: 101494. [CrossRef]

Guesmi, Khaled, Samir Saadi, Ilyes Abid, and Zied Ftiti. 2019. Portfolio diversification with virtual currency: Evidence from bitcoin. *International Review of Financial Analysis* 63: 431–37. [CrossRef]

Guilkey, David K., and Michael K. Salemi. 1982. Small Sample Properties of Three Tests for Granger-Causal Ordering in a Bivariate Stochastic System. *Review of Economics and Statistics* LXIV: 668–80.

Guler, Derya. 2021. The Impact of Investor Sentiment on Bitcoin Returns and Conditional Volatilities during the Era of COVID-19. *Journal of Behavioral Finance* 24: 276–89. [CrossRef]

Gunay, Samet, Shahnawaz Muhammed, and Duc Khuong Nguyen. 2022. Identifying the Role of Investor Sentiment Proxies in NFT Market: Comparison of Google Trend, Fear-Greed Index and VIX. Available online: https://www.researchgate.net/publication/360860266_IDENTIFYING_the_ROLE_of_INVESTOR_SENTIMENT_PROXIES_in_NFT_MARKET_COMPARISON_of_GOOGLE_TREND_FEAR-GREED_INDEX_and_VIX (accessed on 22 June 2023). [CrossRef]

Han, Liyan, Qiuna Lv, and Libo Yin. 2017. Can investor attention predict oil prices? *Energy Economics* 66: 547–58. [CrossRef]

Hasbrouck, Joel. 1995. One Security, Many Markets: Determining the Contributions to Price Discovery. *Journal of Finance* 50: 1175–99. [CrossRef]

Hashemi Joo, Mohammad, Yuka Nishikawa, and Krishnan Dandapani. 2020. Cryptocurrency, a successful application of blockchain technology. *Managerial Finance* 46: 715–33. [CrossRef]

Hatemi-J, Abdulnasser, and R. Scott Hacker. 2008. Optimal lag-length choice in stable and unstable VAR models under situations of homoscedasticity and ARCH. *Journal of Applied Statistics* 35: 601–15.

Hattori, Takahiro, and Ryo Ishida. 2020. Did the introduction of Bitcoin futures crash the Bitcoin market at the end of 2017? *The North American Journal of Economics and Finance* 56: 101322. [CrossRef]

Hattori, Takahiro, and Ryo Ishida. 2021. The relationship between arbitrage in futures and spot markets and Bitcoin price movements: Evidence from the Bitcoin markets. *Journal of Futures Markets* 41: 105–14. [CrossRef]

Heyman, Dries, Michiel Lescrauwaet, and Hannes Stieperaere. 2019. Investor attention and short-term return reversals. *Finance Research Letters* 29: 1–6. [CrossRef]

Howell, Sabrina T., Marina Niessner, and David Yermack. 2020. Initial Coin Offerings: Financing Growth with Cryptocurrency Token Sales. *The Review of Financial Studies* 33: 3925–74. [CrossRef]

Hsiao, Cheng. 1979. Autoregressive Modelling of Canadian Money and Income Data. *Journal of the American Statistical Association* LXXIV: 553–60.

Hsiao, Cheng. 1981. Autoregressive Modelling and Money-Income Causality Detection. *Journal of Monetary Economics* VII: 85–106.

Hsiao, Cheng. 1982. Autoregressive Modelling and Causal Ordering of Economic Variables. *Journal of Economic Dynamics and Control* IV: 243–59.

Hu, Yang, Yang Greg Hou, and Les Oxley. 2020. What role do futures markets play in Bitcoin pricing? Causality, cointegration and price discovery from a time-varying perspective? *International Review of Financial Analysis* 72: 101569. [CrossRef]

Hung, Jui-Cheng, Hung-Chun Liu, and J. Jimmy Yang. 2021. Trading activity and price discovery in Bitcoin futures markets. *Journal of Empirical Finance* 62: 107–20. [CrossRef]

Janson, Nathalie, and Bruno Karoubi. 2021. The Bitcoin: To be or not to be a Real Currency? *The Quarterly Review of Economics and Finance* 82: 312–19. [CrossRef]

Johansen, Søren. 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12: 231–54. [CrossRef]

Johansen, Søren. 1991. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59: 1551–80. [CrossRef]

Johansen, Søren. 1995. *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford: Oxford Press.

Johansen, Soren, and Katarina Juselius. 1990. Maximum likelihood estimation and inference on cointegration with applications to the demand for money. *Oxford Bulletin of Economics and Statistics* 52: 169–210. [CrossRef]

Kapar, Burcu, and Jose Olmo. 2019. An Analysis of Price Discovery Between Bitcoin Futures and Spot Markets. *Economics Letters* 174: 62–64. [CrossRef]

Karalevicius, Vytautas, Niels Degrande, and Jochen De Weerdt. 2018. Using sentiment analysis to predict interday Bitcoin price movements. *The Journal of Risk Finance* 19: 56–75. [CrossRef]

Keblowski, Piotr, and Aleksander Welfe. 2004. The ADF–KPSS test of the joint confirmation hypothesis of unit autoregressive root. *Economics Letters* 85: 257–63. [CrossRef]

Kim, Wonse, Junseok Lee, and Kyungwon Kang. 2020. The effects of the introduction of Bitcoin futures on the volatility of Bitcoin returns. *Finance Research Letters* 33: 101204. [CrossRef]

Kochling, Gerrit, Janis Müller, and Peter N. Posch. 2019. Does the introduction of futures improve the efficiency of Bitcoin? *Finance Research Letters* 30: 367–70. [CrossRef]

Koutmos, Dimitrios. 2023. Investor sentiment and bitcoin prices. *Review of Quantitative Finance and Accounting* 60: 1–29. [CrossRef]

Kwiatkowski, Denis, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin. 1992. Testing the null hypothesis of stationary against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics* 54: 159–78. [CrossRef]

Lee, Yong, and Joon Hee Rhee. 2022. A VECM analysis of Bitcoin price using time-varying cointegration approach. *Journal of Derivatives and Quantitative Studies* 30: 197–218. [CrossRef]

Li, Xin, and Chong Alex Wang. 2017. The technology and economic determinants of cryptocurrency exchange rates: The case of Bitcoin. *Decision Support Systems* 95: 49–60. [CrossRef]

Lin, Zih-Ying. 2021. Investor attention and cryptocurrency performance. *Finance Research Letters* 40: 101702. [CrossRef]

Liu, Ruozhou, Shanfeng Wan, Zili Zhang, and Xuejun Zhao. 2020. Is the introduction of futures responsible for the crash of Bitcoin? *Finance Research Letters* 34: 101259. [CrossRef]

Lyocsa, Štefan, Peter Molnár, Tomáš Plíhal, and Mária Širaňová. 2020. Impact of macroeconomic news, regulation and hacking exchange markets on the volatility of bitcoin. *Journal of Economic Dynamics and Control* 119: 103980. [CrossRef]

Mai, Feng, Zhe Shan, Qing Bai, Xin Wang, and Roger H. L. Chiang. 2018. How Does Social Media Impact Bitcoin Value? A Test of the Silent Majority Hypothesis. *Journal of Management Information Systems* 35: 19–52. [CrossRef]

Matsui, Toshiko, and Lewis Gudgeon. 2020. *The Speculative (In)Efficiency of the CME Bitcoin Futures Market*. Springer Proceedings in Business and Economics, Mathematical Research for Blockchain Economy. Cham: Springer, pp. 91–103.

Mhadhbi, Mayssa, Mohamed Imen Gallali, Stéphane Goutte, and Khaled Guesmi. 2021. On the asymmetric relationship between stock market development, energy efficiency and environmental quality: A nonlinear analysis. *International Review of Financial Analysis* 77: 101840. [CrossRef]

Mokni, Khaled, Ahmed Bouteska, and Mohamed Sahbi Nakhli. 2022. Investor sentiment and Bitcoin relationship: A quantile-based analysis. *The North American Journal of Economics and Finance* 60: 101657. [CrossRef]

Mondria, Jordi, Thomas Wu, and Yi Zhang. 2010. The determinants of international investment and attention allocation: Using internet search query data. *Journal of International Economics* 82: 85–95. [CrossRef]

Naeem, Muhammad Abubakr, Imen Mbarki, and Syed Jawad Hussain Shahzad. 2021. Predictive role of online investor sentiment for cryptocurrency market: Evidence from happiness and fears. *International Review of Economics and Finance* 73: 496–514. [CrossRef]

Nekhili, Ramzi. 2020. Are bitcoin futures contracts for hedging or speculation? *Investment Management and Financial Innovations* 17: 1–9. [CrossRef]

Ng, Serena, and Pierre Perron. 2001. Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69: 1519–54. [CrossRef]

Nkoro, Emeka, and Aham Kelvin Uko. 2016. Autoregressive Distributed Lag (ARDL) cointegration technique: Application and interpretation. *Journal of Statistical and Econometric Methods* 5: 63–91.

Pagnottoni, Paolo, and Thomas Dimpfl. 2019. Price discovery on Bitcoin markets. *Digital Finance* 1: 139–61. [CrossRef]

Panagiotidis, Theodore, Thanasis Stengos, and Orestis Vravosinos. 2019. The effects of markets, uncertainty and search intensity on bitcoin returns. *International Review of Financial Analysis*, 63, 220–42.

Park, Beum-Jo. 2022. The COVID-19 pandemic, volatility, and trading behavior in the Bitcoin futures market. *Research in International Business and Finance* 59: 101519. [CrossRef] [PubMed]

Pesaran, M. Hashem, and Yongcheol Shin. 1999. An Autoregressive Distributed Lag Modelling Approach to Cointegration Analysis. In *Chapter 11 in Econometrics and Economic Theory in the 20th Century the Ragnar Frisch Centennial Symposium*. Edited by Strom Steinar. Cambridge: Cambridge University Press, pp. 371–413.

Pesaran, M. Hashem, Yongcheol Shin, and Richard J. Smith. 1996. *Testing for the Existence of a Long run Relationship*. DAE Working paper No. 9622. Cambridge: Department of Applied Economics, University of Cambridge.

Pesaran, M. Hashem, Yongcheol Shin, and Richard J. Smith. 2001. Bounds Testing Approaches to the Analysis of Level Relationships. *Journal of Applied Econometrics* 16: 289–326. [CrossRef]

Phillips, Peter C. B., and Pierre Perron. 1988. Testing for a Unit Root in Time Series Regression. *Bimetrika* 75: 335–46. [CrossRef]

Qarni, Muhammad Owais, and Saiqb Gulzar. 2021. Portfolio diversification benefits of alternative currency investment in Bitcoin and foreign exchange markets. *Financial Innovation* 7: 17. [CrossRef]

Rognone, Lavinia, Stuart Hyde, and S. Sarah Zhang. 2020. News sentiment in the cryptocurrency market: An empirical comparison with Forex. *International Review of Financial Analysis* 69: 101462. [CrossRef]

Sari, Ramazan, Bradley T. Ewing, and Ugur Soytas. 2008. The relationship between disaggregate energy consumption and industrial production in the United States: An ARDL approach. *Energy Economics* 30: 2302–13. [CrossRef]

Scharnowski, Stefan. 2021. Understanding Bitcoin liquidity. *Finance Research Letters* 38: 101477.

Sebastiao, Helder, and Pedro Godinho. 2020. Bitcoin futures: An effective tool for hedging cryptocurrencies. *Finance Research Letters* 33: 101230. [CrossRef]

Sephton, Peter. 2008. On the finite sample size and power of the generalized KPSS test in the presence of level breaks. *Applied Economics Letters* 15: 833–43. [CrossRef]

Shin, Yongcheol, Byungchul Yu, and Matthew Greenwood-Nimmo. 2014. Modelling asymmetric cointegration and dynamic multipliers in a nonlinear ARDL framework. In *Festschrift in Honor of Peter Schmidt*. New York: Springer, pp. 281–314.

Silber, William L. 1981. Innovation, competition, and new contract design in futures markets. *Journal of Futures Market* 1: 123–55. [CrossRef]

Smales, Lee A. 2022. Investor attention in cryptocurrency markets. *International Review of Financial Analysis* 79: 101972. [CrossRef]

Stephens-Davidowitz, Seth, and Hal Varian. 2015. *A Hands-on Guide to Google Data*. Mountain View: Google, Inc.

Thornton, Daniel L., and Dallas S. Batten. 1985. Lag-Length selection and Tests of Granger Causality Between Money and Income. *Journal of Money Credit and Banking* XVII: 164–78.

Tong, Zezheng, John W. Goodell, and Dehua Shen. 2022. Assessing causal relationships between cryptocurrencies and investor attention: New results from transfer entropy methodology. *Finance Research Letters* 50: 103351. [CrossRef]

Ülkü, Numan, Fahad Ali, Saidgozi Saydumarov, and Deniz İkizlerli. 2023. COVID caused a negative bubble. Who profited? Who lost? How stock markets changed? *Pacific-Basin Finance Journal* 79: 102044. [CrossRef]

Urquhart, Andrew. 2016. The inefficiency of Bitcoin. *Economics Letters* 148: 80–82. [CrossRef]

Wang, Panpan, Xiaoxing Liu, and Sixu Wu. 2022. Dynamic Linkage between Bitcoin and Traditional Financial Assets: A Comparative Analysis of Different Time Frequencies. *Entropy* 24: 1565. [CrossRef]

Woloszko, Nicolas. 2020. *Tracking Activity in Real Time with Google Trends*. OECD Economics Department Working Papers, No. 1634. Paris: OECD Publishing.

Wu, Jinghong, Ke Xu, Xinwei Zheng, and Jian Chen. 2021. Fractional cointegration in bitcoin spot and futures markets. *Journal of Futures Markets* 41: 1478–94. [CrossRef]

Wustenfeld, Jan, and Teo Geldner. 2022. Economic uncertainty and national bitcoin trading activity. *The North American Journal of Economics and Finance* 59: 101625.

Yung, Kenneth, and Nadia Nafar. 2017. Investor attention and the expected returns of reits. *International Review of Economics and Finance* 48: 23–439. [CrossRef]

Article

Does Fiscal Consolidation Affect Non-Performing Loans? Global Evidence from Heavily Indebted Countries (HICs)

Habib Ur Rahman [1], Adam Arian [2] and John Sands [2,*]

1 Faculty of Higher Education (Accounting and Finance), Holmes Institute, Gold Coast, QLD 4217, Australia
2 School of Business, Faculty of Business, Education, Law and Arts, University of Southern Queensland, Darling Heights, QLD 4350, Australia
* Correspondence: john.sands@usq.edu.au

Abstract: This study explores fiscal consolidations' impact on non-performing loans (NPLs) in highly indebted countries (HICs) following the global financial crisis (GFC) and subsequent sovereign debt crisis. A dynamic panel data estimator was applied to obtain the unbiased estimator due to NPLs' time persistence. The findings reveal that fiscal consolidation measures increase NPLs since they limit the household and business loan-serving capacity. Extended analysis categorises fiscal consolidation episodes into (1) the fiscal consolidation weak form (FCWE) and (2) the fiscal consolidation strong form (FCSE). The extended analysis results reveal that the FCWE and FCSE improve NPLs by 1.55% and 31.10%, respectively. The weak-to-strong form transition of the fiscal consolidation analysis resulted in improving NPLs by 28.55 percentage points. NPL definition challenges, the potential influence of loan restructuring and regulatory restrictions, and implications for policymakers and financial institutions in managing NPLs in highly indebted economies were explored. Investigating the potentially different effects of both forms of fiscal consolidation (FCWE and FCSE) on NPLs in countries with different definitions of NPLs, including a comparison study between different definitions, was identified as an implication for future research. Finally, future studies should examine how restrictions on IFRS 9 may affect the FCWE and NPL as well as FCSE and NPL associations.

Keywords: non-performing loans; fiscal consolidation; highly indebted countries; global financial crisis; sovereign debt crisis

1. Introduction

The economic recessions that followed the global financial crisis (GFC) of 2007–2009 had varying impacts on credit quality worldwide. Initially, banks heavily exposed to United States residential mortgage-backed securities saw a decline in asset quality. Subsequently, the GFC evolved into a sovereign debt crisis in the Eurozone, further straining the debt-serving capacity of businesses and households, especially in highly indebted countries (HICs)[1] (Louzis et al. 2012; Siakoulis 2017). Notably, Ireland witnessed a significant 24.82 percentage point increase in NPLs from 2003 to 2013, while Ukraine experienced a surge of 51.54 percentage points in NPLs from 2007 to 2017, mirroring the trend seen in other highly indebted economies.

During the sovereign debt crisis, highly indebted economies had to adopt stringent fiscal measures to address their debt burdens. Many of these countries, including Ireland, Spain, Colombia, and Jordan, implemented such measures, increasing the fiscal burden on businesses and households. Businesses face weak demand for their products and services due to the lower disposable income available to households. In these economic conditions, lower free cash flows available to firms and less disposable income for households elevate credit risk for the banking sector. Along these lines, Konstantakis et al. (2016) also mentioned that this fiscal burden negatively affects the debt-servicing capacity, leading to a rise in NPLs. There is a wide range of research on the macroeconomic determinants of

NPLs (Alizadeh Janvisloo and Muhammad (2013), Beck et al. (2013), Dimitrios et al. (2016), Fallanca et al. (2021), Fofack (2005), Makri et al. (2014), Messai and Jouini (2013), Roland et al. (2013), Zheng et al. (2019), Kartal et al. (2021), Konstantakis et al. (2016), Thornton and Di Tommaso (2020), Tanasković and Jandrić (2015), Vithessonthi (2016), Le et al. (2020), Boumparis et al. (2019), and Jiang et al. (2018)). While there is existing research on the macroeconomic determinants of NPLs, evidence is scarce regarding the fiscal determinants, with limited empirical studies, such as that of Siakoulis (2017), focusing solely on the Eurozone and using the cyclical adjusted primary balance as a measure of the fiscal stance. This study aims to fill this gap by providing empirical evidence on the impact of fiscal consolidation on NPLs in highly indebted countries, making significant contributions to the current literature.

NPLs refer to loans where borrowers cannot meet scheduled repayment obligations outlined in the loan agreement. The NPL ratio represents the proportion of defaulted loans to the total gross loans. Defaulted loans are defined as those with overdue interest and principal payments exceeding three months, while total gross loans encompass the entire loan portfolio's value. It is crucial to note that NPLs reflect the gross value of loans recorded on financial statements and not just the specifically overdue amount. However, caution must be exercised when comparing NPL data across countries due to variations in national accounting practices, standards, taxation policies, and supervision frameworks.

Furthermore, NPL data may not fully capture impaired loans resulting from bank-specific loan restructuring. Therefore, the interpretation of NPL data requires careful consideration, as reporting bodies employ different methodologies. Barisitz (2013) conducted a comprehensive study to shed light on the definition of NPLs across diverse economies, identifying two key criteria: loans overdue by more than 90 days or strong evidence indicating significant weaknesses in the loan or borrower's financial position[2].

The discussion on loan restructuring is essential due to regulatory restrictions regarding the transfer of NPLs into performing loans. For example, restructured loans typically need to remain non-performing for a probationary period of one year, followed by an additional two years before their status can be changed. The treatment of restructured loans in historical non-performing loan data remains unclear[3]. Furthermore, a situation may arise where a borrower fulfils payment obligations for one loan but fails to do so for another. This raises concerns regarding the classification of loans as performing, considering the well-defined weakness highlighted in the second key element of Barisitz's (2013) comprehensive definition of NPLs. Such discussions indicate that the restructuring of NPLs can impact the health of a balance sheet. Consequently, global data on NPLs may present a slightly different picture of balance sheet health, while accounting and financial standards, such as the IFRS and GAAP, primarily focus on impaired loans rather than NPLs[4].

NPLs arise from financial sector difficulties (Konstantakis et al. 2016), which are intertwined with macroeconomic conditions. During an economic expansionary phase, NPLs tend to decrease as the debt-servicing capacity of businesses and households improves. However, the financial sector may extend credit to less-creditworthy customers, increasing NPLs during an economic downturn. Recessionary conditions diminish the income streams of consumers and firms, contributing to a rise in NPLs. Furthermore, the financial sector has been found to influence overall economic growth (Carlstrom and Fuerst (1997), De Bock and Demyanets (2012), Fisher (1933), and Kiyotaki and Moore (1997)). Increased levels of NPLs result in higher associated costs, reducing the owner's equity of banks, increasing the credit risk of commercial banks (Jiajia et al. (2023)), and potentially leading to insolvencies and systemic failures. Under such circumstances, the banking system faces challenges in effectively channelling savings into investments and transmitting monetary policy to the real economy.[5] This study primarily focuses on the transmission mechanism through which the economy affects the banking sector, specifically examining the impact of the debt-servicing capacity of individual and corporate borrowers on fiscal stance improvement (Perotti (1996)).

However, identifying the effect of a fiscal stance is complicated by potential endo-geneity issues. The existing literature employs various measures to analyse fiscal policy effects. For instance, Siakoulis (2017) utilised the cyclical adjusted primary balance, which represents the budget balance under normal economic activity levels. A positive change in this indicator indicates an increased tax burden, adversely affecting the debt-servicing capacity of businesses and households. Another approach employed in the literature (Alesina and Ardagna (2010), Barrios et al. (2010), Mirdala (2013), Perotti (1996), and Rahman (2018)) is the use of strict fiscal measures to identify fiscal effects. This approach focuses on fiscal consolidation episodes aimed at reducing the burden of sovereign debt. Two measurement approaches are commonly used for fiscal consolidation episodes: the "cold shower" approach and gradual consolidation. The cold shower approach identifies a fiscal consolidation episode when the cyclical adjusted primary balance improves by more than 1.5% of the GDP per year. Gradual consolidation, on the other hand, refers to a situation where the cyclical adjusted primary balance does not deteriorate by more than 0.5% of the GDP per year. Building upon the work of Rahman (2018), a strong episode of fiscal consolidation is defined as a period in which the cyclical adjusted primary balance improves by 1.5% of the GDP per year or two consecutive years with an improvement of at least 1% of the GDP per year. To the best of our knowledge, this study represents the first attempt to analyse the impact of both weak and strong forms of fiscal consolidation on NPLs.

The existing literature (Beck et al. (2013), Dimitrios et al. (2016), Fallanca et al. (2021), Kjosevski and Petkovski (2021), Louzis et al. (2012), Makri et al. (2014), Messai and Jouini (2013), Roland et al. (2013), Siakoulis (2017), Vogiazas and Nikolaidou (2011), and Zheng et al. (2019)) has identified a wide range of determinants of the NPLs. Following this strand of research, this study uses economic growth and unemployment to capture the effect of the economic cycle. The existing empirical literature shows that the debt-servicing capacity of businesses and households has a negative relationship with economic growth and a positive relationship with unemployment. Inflation also affects the debt-serving capacity of businesses and households through different channels. The first channel increases the debt-serving capacity of businesses and households. For instance, the higher level of inflation reduces the real value of an outstanding loan, making debt serving much easier. Conversely, the second channel deteriorates the debt-serving capacity of these agents. For instance, the higher level of inflation reduces the real income of borrowers, which deteriorates the borrowers' capacity to repay the loans.

Furthermore, the monetary policy announcements affect the NPLs in a variable loan rate environment. In particular, the monetary policy actions to reduce the level of inflation are highly likely to reduce the debt-serving capacity of borrowers, since the lenders adjust their rates to maintain the real returns. In other words, they increase their interest rates in response to the increasing policy rates. These rate adjustments deteriorate the loan-paying capacity of borrowers. Therefore, the impact of inflation on the NPLs can be positive or negative. This study also uses private debt as a control variable, and it is expected that private credit is positively associated with NPLs, since the increase in private credit restricts the capacity of businesses and households to refinance their debt. To summarise, this study uses economic growth, unemployment, inflation, and domestic credit for the private sector as the control variables to separate the effect of fiscal measures from the general macroeconomic factors.

Our empirical investigation reveals that fiscal consolidation measures increase the NPLs in highly indebted countries since these measures limit the loan-serving capacity of households and businesses. The existing literature used the positive change in the CAPB to analyse its impact on NPLs. We called this positive change in the CAPB the weak form of fiscal consolidation. Our findings reveal that the weak form of fiscal consolidation improves NPLs by 1.55%. In other words, any positive change in the cyclical adjusted primary balance improves NPLs by one and a half percent. To the best of our knowledge, the impact of the strong form of fiscal consolidation on NPLs has yet to be analysed. Therefore, this study

contributes to the existing empirical literature by adding evidence on the strong form of fiscal consolidation. The strong form of fiscal consolidation is defined as a period where the cyclical adjusted primary balance improves by 1.5 percent of the GDP per year or a period of two consecutive years where the cyclical adjusted primary balance is improved by at least 1 percent of the GDP per year. Our extended analysis indicates that the strong form of fiscal consolidation improves the NPLs by 30.10%. The most striking observation to emerge from the comparison between the weak form and the strong form of fiscal consolidation is that moving from the weak form to the strong form improves the NPLs by 28.55 percentage points. The results of this study are consistent with the theoretical expectation.

The rest of this paper is organised as follows. Section 2 presents the relevant literature on the topic. Section 3 presents the data, empirical model, and estimation strategy, which is further categorised into (1) The Data Set and HICS Classification, (2) Empirical Model and Estimation Strategy, and (3) Methodological Notes on Fiscal Consolidation Episodes. Section 4 presents the results and discussion. This study is concluded in Section 5.

2. Literature Review

The empirical literature on the association between macroeconomic conditions and credit quality is extensive (Dimitrios et al. (2016), Fallanca et al. (2021), Kjosevski and Petkovski (2021), Tanasković and Jandrić (2015), and Zheng et al. (2019)) and focuses primarily on the loan-serving capacity of households and businesses. For instance, several studies have reported a positive relationship between unemployment and NPLs, since a higher level of unemployment lowers the loan-serving capacity of households and businesses. Most of the studies on NPLs are country-specific and conducted after the financial and debt crises. For instance, Vogiazas and Nikolaidou (2011) investigated the determinants of the NPLs in Romania's banking and financial system. Their empirical investigation revealed that unemployment, inflation, external debt, and M2 are the critical determinants of credit quality in the Romanian financial system.

Apart from the macroeconomic factors, bank-specific factors also affect the quality of loans in any banking sector. Furthermore, it might be relevant to note that these bank-specific and macroeconomic effects vary between different loan categories. Along these lines, Louzis et al. (2012) reported that the macroeconomic factors, including the GDP, unemployment, public debt, and interest rates, explain the NPLs of all categories. They further reported that management quality is also one of the key determinants of NPLs. The weighted average loan rate was also reported as a determinant (Greenidge and Grosvenor (2010)). One strand of research used the panel of banks to investigate the determinants of the NPLs (Alizadeh Janvisloo and Muhammad (2013), Messai and Jouini (2013), Quagliariello (2007), and Salas and Saurina (2002)).

Another strand of research used cross-country data to identify the specific determinants of credit quality. We follow this approach since cross-country analysis incorporates country-specific variations in the trends of NPLs. This country-specific variation has a couple of sources, including accounting standards. Econometrically, cross-country analysis provides more robust results than time series analysis. Using an unbalanced panel of 75 countries, Roland et al. (2013) reported that share prices, real GDP growth, lending interest rates, and exchange rates are critical determinants of NPLs. However, they reported some variation in the exchange rate and share price effects. The countries with pegged and managed exchange rates had a higher impact of exchange rates on the NPLs. Working in a similar vein, Makri et al. (2014) used the unbalanced panel data from 2002 to 2008 of 14 Eurozone economies and reported that the state of the economy is significantly linked with the loan quality. In particular, they reported that the NPLs of the previous year, GDP, unemployment, and public debt are the key determinants of NPLs. Similarly, Fofack (2005) used the unbalanced panel data of 16 sub-Saharan African countries from 1993 to 2003 and reported the strong causality between NPLs and (1) economic growth, (2) interest rates, and (3) real exchange rates. Comparatively recently, some panel studies were also published on the determinants of NPLs. For instance, Kjosevski and Petkovski (2021) investigated

the macroeconomic and bank-specific determinants of NPLs from the Baltic states using panel data from 21 commercial banks. By applying the fixed-effect Generalised Method of Moments difference and system, they reported that GDP growth, inflation, public debt, and unemployment are key determinants of NPLs. However, they also reported the bank-specific determinants, including return on assets, total assets ratio, return on equity, and the growth of gross loans.

To the best of our knowledge, we could not find any paper on the linkage between fiscal policy and NPLs other than that of Siakoulis (2017). Using a global dataset from 31 countries, Siakoulis reported that fiscal pressure, as measured by the changes in the cyclical adjusted primary balance, determines the NPLs.

3. Research Design

3.1. The Data Set and HICS Classification

The data set used in this study is a balanced panel which consists of NPLs, fiscal consolidation episodes, and a set of control variables. The control variables consist of economic growth, unemployment, inflation, and domestic credit to the private sector. The complete definitions, acronyms, and indicator codes are given in Table 1. Fiscal consolidation and NPLs are the key variables of interest in this study. A separate section is included on the detailed calculation of the weak and strong episodes of the fiscal consolidation episodes.

Table 1. Definitions, acronyms, and indicator codes of the variables.

Names of the Variables	Acronyms	Indicator Codes
Dependent Variable		
Bank NPLs to gross loans (%)	BNPL	GFDD.SI.02
Independent Variable		
Weak episode of fiscal consolidation	FCWE	Author's calculation
Strong episode of fiscal consolidation	FCSE	Author's calculation
Control Variables		
GDP growth (annual %)	GDPG	NY.GDP.MKTP.KD.ZG
Unemployment, total (% of the total labour force) (national estimate)	UNEM	SL.UEM.TOTL.NE.ZS
Inflation, GDP deflator (annual %)	INFL	NY.GDP.DEFL.KD.ZG
Domestic credit to the private sector (% of GDP)	LCPD	FS.AST.PRVT.GD.ZS
Variable to calculate HICS counties		
Central government debt, total (% of GDP)	CGTD	GC.DOD.TOTL.GD.ZS

Note: See the section Methodological Notes on Fiscal Consolidation Episodes for the detailed calculation of the weak and strong episodes of the fiscal consolidation.

The highly indebted countries (HICs) were selected based on the entire stock of the direct long-term contractual obligations of the government to others. The central government debt (percent of the GDP) is the most suitable measure for this entire stock. These data calculate the average debt of all the countries available in the World Development Indicators (World Bank 2022). For this purpose, we used the latest data on the central government debt (percent of the GDP) for the last 10 years. Then, the countries were sorted in descending order, and the first 35 countries were selected as the HICs countries. These global data enabled us to investigate the common patterns in the NPLs.

3.2. Empirical Model and Estimation Strategy

Following the recent literature (Louzis et al. (2012) and Rahman et al. (2020)), this study applied the dynamic panel data estimator to arrive at the unbiased estimator due to the time persistence in NPLs. Equation (1) presents the dynamic panel data specification to analyse the impact of fiscal consolidation on NPLs:

$$NPL_{i,t} = \beta NPL_{i,t-1} + \gamma(L)x_{i,t} + \delta_i + \varepsilon_{i,t;} \ |\beta| < 1, \ i = 1,\dots,N, \ t = 1,\dots,T. \quad (1)$$

The subscripts i and t represent the cross-sectional and time dimensional of the panel dataset, respectively. $NPL_{i,t}$ represents the NPLs as the dependent variable, and $\gamma(L)$ denotes a lag polynomial vector, while $x_{i,t}$ is the $kX1$ vector of independent variables other than $y_{i,t-1}$. Here, δ_i and $\varepsilon_{i,t}$ represent the unobserved individual effect and the error term, respectively.

The existing literature (see Siakoulis (2017)) provides evidence of time persistency in NPLs. Therefore, Equation (1) is consistently estimated using the Generalised Method of Moments in a framework proposed by Arellano and Bond (1991). Later, this approach was generalised by Arellano and Bover (1995) and Bulundell and Bond (1998). This estimation approach is based on the first difference transformation and the subsequent elimination of δ_i. In this case, Equation (1) can be written as follows:

$$\Delta NPL_{i,t} = \beta \Delta NPL_{i,t-1} + \gamma(L)\Delta x_{i,t} + \Delta \varepsilon_{i,t} \tag{2}$$

In Equation (2), Δ represents the first difference operator. As mentioned above, $x_{i,t}$ represents the explanatory variables. The explanatory variable set includes the primary and control variables of interest. In particular, this set includes (1) weak episodes of fiscal consolidation (FCWE), (2) strong episodes of fiscal consolidation (FCSE), (3) the GDP growth (GDPG), (4) unemployment (UNEM), (5) inflation (INFL), and (6) domestic credit to the private sector (LCPD). We used the FCWE as the positive change in the cyclical adjusted primary balance. However, the FCSE is a dummy variable which takes a value of one for the case of strong episodes of fiscal consolidation and zero otherwise. Following Rahman (2018), a strong episode of fiscal consolidation is a period where the cyclical adjusted primary balance improves by 1.5 percent of the GDP per year or a period of two consecutive years where the cyclical adjusted primary balance is improved by at least 1 percent of the GDP per year. For further details on these variables, see Table 1. However, detailed notes on the calculation of the fiscal consolidation episodes are provided in Section 3.3 (Methodological Notes on Fiscal Consolidation Episodes). Based on this set of explanatory variables, Equation (2) can be rewritten as

$$\begin{aligned} \Delta NPL_{i,t} = {} & \beta_1 \Delta NPL_{i,t-1} + \gamma_1 \Delta FCWE_{i,t} + \gamma_2 \Delta FCSE_{i,t} + \gamma_3 \Delta GDPG_{i,t} \\ & + \gamma_4 \Delta UNEM_{i,t} + \gamma_5 \Delta INFL_{i,t} + \gamma_6 \Delta LCPD_{i,t} + \Delta \varepsilon_{i,t} \end{aligned} \tag{3}$$

It might be relevant to note that the error term $[\Delta \varepsilon_{i,t}]$ in Equation (2) is, by definition, correlated with the lagged dependent variable $\Delta NPL_{i,t-1}$. This correlation imposes a bias in the estimation process of the model. One of the possible ways to arrive at the unbiased coefficient is using the higher-order lags of the dependent variables as instruments. For instance, it is expected in Equation (2) that the second lag of NPLs $\Delta NPL_{i,t-2}$ is correlated with its first lag $\Delta NPL_{i,t-1}$ and uncorrelated with the error term $[\Delta \varepsilon_{i,t}]$ for the third period and above. Therefore, $\Delta NPL_{i,t-2}$ can be used as an instrument in Equation (2). This discussion reveals that lags of orders of two and above satisfy the moment condition of $E = 0$ for $t = 3, \ldots, T$ and $s \geq 2$. Another source of biasedness originates from (1) the possible endogeneity of the explanatory variables and (2) the correlation with the error term $[\Delta \varepsilon_{i,t}]$. The explanatory variables should be strictly exogenous. For the case of strictly exogenous variables, the historical and future values are uncorrelated with the error term $[\Delta \varepsilon_{i,t}]$. In other words, the strict exogenous variables satisfy the moment condition of $E = 0$ for $t = 3, \ldots, T$ and all the values of s. However, the restrictive assumption of the strict exogenous is no longer valid in reverse causality. Following Cameron and Trivedi (2010), the current lagged values of the explanatory variables are the only valid instruments for the weak or predetermined explanatory variables implying the moment condition of $E = 0$ for $t = 3, \ldots, T$ and $s \geq 2$ (see Gholami et al. (2023), Gholami et al. (2022) and Siakoulis (2017)). Based on these orthogonality restrictions, the estimates of the GMM are consistent. Furthermore, we apply the Sargan specification test for the null hypothesis, and therefore the instruments must be valid for that moment's conditions. This test is asymptotically distributed as a chi-square. The reported J-statistics are simply the Sargan

statistics. Furthermore, we apply the Arellano–Bond test for zero autocorrelation in the first-difference error. Consequently, the null hypothesis of no autocorrelation is expected to be rejected at order one and not at the higher orders.

3.3. Methodological Notes on Fiscal Consolidation Episodes

Identification of the fiscal effect on banks' NPLs is difficult due to potential endogeneity (Siakoulis 2017). Therefore, the budgetary impact should be recorded when the economy is at a normal activity level. Along these lines, Siakoulis (2017) used the cyclical adjusted primary balance to analyse the effect of fiscal policy on NPLs. Theoretically, the cyclical adjusted primary balance removes the endogenous components of spending and revenues. In particular, the cyclical adjusted primary balance reveals the fiscal position after removing the cyclical and automatic movements. Considering these theoretical aspects, the positive change in the cyclical adjusted primary balance is considered a weak fiscal consolidation episode. However, this study extends the empirical literature by incorporating the second type of fiscal consolidation episode. Following the existing empirical literature (Alesina and Ardagna (2010), Alesina and Perotti (1995), Mirdala (2013), and Rahman (2018)), we incorporate the strong episodes of fiscal consolidation in Equation (3).

Fiscal consolidation is the improved fiscal stance to reduce the burden of sovereign debt. Following Mirdala (2013), this is accomplished through a set of fiscal arrangements on the side of the government budget's revenue and expenditures. The existing literature has provided several approaches to measure the episodes of fiscal consolidation (Alesina and Ardagna (2010), Barrios et al. (2010), and Mirdala (2013)). The most common approach to measuring fiscal consolidation episodes was given by Mirdala (2013), which is a revised version of the approach proposed by Barrios et al. (2010). Mirdala (2013) used two approaches, including (1) the cold shower approach and (2) gradual consolidation. According to the cold shower approach, the episode of fiscal consolidation is when the cyclical adjusted primary balance improves by more than 1.5 percent of the GDP per year. However, gradual consolidation is when the cyclical adjusted primary balance will not deteriorate by more than half a percent of the GDP per year. Following Rahman (2018), a strong episode of fiscal consolidation is a period where the cyclical adjusted primary balance improves by 1.5 percent of the GDP per year or a period of two consecutive years where the cyclical adjusted primary balance is improved by at least 1 percent of the GDP per year.

Cyclical Adjusted Primary Balance

There are different approaches to calculating the cyclical adjusted primary balance. However, Mirdala (2013) revealed that the main algorithm follows the same procedures"

1.   The first step is to estimate the potential GDP.
2.   The second step is to determine the responses of the key revenues and expenditures to a fluctuation in the cyclical GDP.
3.   The third step is to adjust these cyclical components calculated in the second step from the revenue and expenditures.

The existing literature has different approaches to estimating the cyclical components, as mentioned in the second step above. One approach is to estimate the income elasticities of the main budgetary variables, including revenue and expenditures (Altar et al. (2010) and Bouthevillain et al. (2001)). However, most empirical studies apply the Hodrick and Prescott (HP) filter (Hodrick and Prescott (1997)) to calculate the cyclical components. Following this strand of research, we apply the HP filter to calculate the cyclical components for the fiscal variables. Mirdala (2013) revealed that the cyclical adjusted primary balance is calculated by subtracting the cyclical components from the primary government balance. This can be written as follows:

$$CAPB_t = PB_t - B_t^c = PB_t - \sum_{i=1}^{n} B_{t,i}^c \tag{4}$$

where $CAPB, PB,$ and B^c represent the cyclical adjusted primary balance, primary balance, and cyclical components, respectively. Here, the primary balance is calculated by subtracting the interest payable from the actual government budget balance. This can be represented as follows:

$$PB_t = B_t - E^I \tag{5}$$

$B_{t,i}^c$ in Equation (4) represents a cyclical component of each budget category, including revenue and expenditure. As mentioned above, we apply the HP filter to calculate the cyclical components. Being a two-sided linear filter, the HP filter minimises the variance in y around s and computes the smoothed series s of y (Hodrick and Prescott (1997)). This computation is subject to a penalty that constrains the second difference of s. And the HP filter selects s to minimise. This can be represented as follows:

$$\sum_{t=1}^{T} (y_t - s_t)^2 + \lambda \sum_{t=2}^{T-1} ((s_{t+1} - s_t) - (s_t - s_{t-1}))^2 \tag{6}$$

In Equation (6), the smoothness of variance is controlled by λ. And as $\lambda = \infty$, s approaches a linear trend. EViews 12 is used to apply this HP filter and estimate the cyclical components.

4. Results and Discussion

Table 2 presents the descriptive statistics of the variables selected for this study. Critical analysis of Table 2 reveals that 6.03% of the total bank loans were NPLs. The highest level of NPLs was observed in Ukraine, Indonesia, and Uruguay. It might be relevant to note that the highest level of NPLs was observed in Ukraine from 2017 to 2019.[6] What is interesting about the data in this table is that it reveals useful insights about fiscal consolidation. The descriptive statistics of the weak episodes of fiscal consolidation reveal that the cyclical adjusted primary balance was reduced by 0.41% of the GDP in the top 35 HICs from 2000 to 2020. The main source of variation in the cyclical adjusted primary balance was within the economies. The construction of strong episodes of fiscal consolidation is complex. However, 12.93 percent of the budgetary efforts could be considered strong commitment of the government to improving the fiscal stance.

Table 2. Descriptive statistics.

Variable		Mean	Std. Dev.	Min	Max	Observations
BNPL	Overall	6.0354	6.9033	0.2000	54.5413	N = 624
	Between		4.7869	0.9927	22.8881	n = 34
	Within		5.1332	−13.8527	37.6886	T = 18.3529
FCWE	Overall	−0.4096	2.9920	−16.1102	15.0072	N = 686
	Between		1.6505	−4.9682	2.0327	n = 35
	Within		2.5048	−17.0294	14.0880	T-bar = 19.6
FCSE	Overall	0.1293	0.3357	0.0000	1.0000	N = 735
	Between		0.0935	0.0000	0.3333	n = 35
	Within		0.3228	−0.2041	1.0816	T = 21
GDPG	Overall	3.0991	3.6954	−18.9795	25.1763	N = 735
	Between		1.3947	−0.2410	5.8194	n = 35
	Within		3.4299	−15.6395	23.2489	T = 21
UNEM	Overall	7.8553	6.1962	0.0000	33.2900	N = 734
	Between		5.4026	1.1276	27.2562	n = 35
	Within		3.1570	−9.7195	21.3905	T = 20.9714
INFL	Overall	5.6538	10.0784	−5.9922	185.2908	N = 735
	Between		6.1587	−0.4652	33.1890	n = 35
	Within		8.0422	−19.1933	157.7556	T = 21
LCPD	Overall	4.2207	0.6416	2.8236	5.7189	N = 643
	Between		0.6005	3.3519	5.2135	n = 35
	Within		0.2348	3.2692	5.1246	T-bar = 18.3714

Note: BNPL, FCWE, FCSE, GDPG, UMEN, INFL, and LCDP indicate the bank NPLs to gross loans, weak episodes of fiscal consolidation, strong episodes of fiscal consolidation, gross domestic product growth, unemployment, inflation, and the natural log of domestic credit to the private sector, respectively.

The correlation analysis results in Table 3 revealed that there was no evidence of mul-
ticollinearity since the independent variables were not correlated. Following the empirical
strategy, the models were estimated using the GMM after testing the appropriateness of
the estimation techniques.[7] Along these lines, the existing empirical literature reveals that
stationarity of the panel data should be ensured for appropriate estimation of the dynamic
panel model (Buck et al. (2008) and Chang et al. (2011)).

Table 3. Correlation analysis.

		FCSE	FCWE	GDPG	UNEM	INFL	LCPD
FCSE	*Correlation*	*1.0000*					
	t-Statistics	-----					
	Probability	-----					
FCWE	*Correlation*	*0.0901 ***	*1.0000*				
	t-Statistics	2.2569	-----				
	Probability	0.0244	-----				
GDPG	*Correlation*	*−0.1234 ***	*0.3286 ****	*1.0000*			
	t-Statistics	−3.1044	8.6831	-----			
	Probability	0.0020	0.0000	-----			
UNEM	*Correlation*	*0.1325 ****	*0.0236*	*−0.1511 ****	*1.0000*		
	t-Statistics	3.3364	0.5882	−3.8141	-----		
	Probability	0.0009	0.5566	0.0002	-----		
INFL	*Correlation*	*0.1234 ****	*0.1193 ****	*0.0753 ***	*0.0287*	*1.0000*	
	t-Statistics	3.1030	2.9980	1.8837	0.7157	-----	
	Probability	0.0020	0.0028	0.0601	0.4745	-----	
LCPD	*Correlation*	*0.0358*	*−0.2521 ****	*−0.2146 ****	*0.0323*	*−0.3703 ****	*1.0000*
	t-Statistics	0.8932	−6.5033	−5.4832	0.8076	−9.9501	-----
	Probability	0.3721	0.0000	0.0000	0.4196	0.0000	-----

Note: BNPL, FCWE, FCSE, GDPG, UMEN, INFL, and LCDP indicate the bank NPLs to gross loans, weak episodes
of fiscal consolidation, strong episodes of fiscal consolidation, gross domestic product growth, unemployment,
inflation, and the natural log of domestic credit to the private sector, respectively. ** < 0.05; *** < 0.01.

Stationarity testing is particularly important when T is less than N, as demonstrated
in our dataset.[8] There was another reason to test the stationarity before applying the GMM,
as the first-difference GMM only takes care of the first order of integration. Therefore,
stationarity testing should be applied to ensure that none of the series is integrated into
an order of two. For this purpose, this study applied three cross-sectionally independent
panel unit root tests, including (1) Levin, Lin, and Chu, (2) the ADF-Fisher Chi-square, and
(3) the PP-Fisher Chi-square.

Table 4 presents the results of these tests at the level and the first difference. A critical
analysis of this table reveals that all the series were stationary, at least at the first difference.
These results ensure that it is econometrically appropriate to estimate Equation (3). Table 5
presents the results of the impact of fiscal consolidation on the NPLs of the HICs. Columns
1 and 2 present the coefficients of the test statistics of Equation (2), which were estimated
using the panel GMM with the first difference transformations. This model is named Model
1. Columns 3–8 present the coefficients of the test statistics of Equation (2) estimated using
pooled (Model 2), random (Model 3), and fixed effects (Model 4). The estimates of Models
2–4 were part of the robustness analysis. This study aims to analyse the impact of fiscal
consolidation on NPLs in highly indebted economies. The NPLs are the ratio of defaulting
loans to total gross loans[9]. The dataset from the wide scope of economics enabled us to
investigate the common patterns in the NPLs.

Table 4. Panel unit root tests.

	At Level			At First Difference		
	LLC	**ADF-F**	**PP-F**	**LLC**	**ADF-F**	**PP-F**
BNPL	−18.0683	610.5140	463.3610	−5.8995	202.3380	542.7890
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
FCWE	−7.4605	164.7540	169.5170	−19.7814	450.2270	451.2990
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
FCSE	−8.7310	117.9560	120.9550	−30.7682	400.6090	296.5490
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GDPG	−8.5565	195.2070	190.8560	−28.4972	621.2740	615.6850
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
UNEM	−4.1440	154.6120	149.4780	−17.6138	363.9070	694.3550
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
INFL	−10.2159	243.4020	225.0290	−36.5655	726.6920	710.1670
	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
LCPD	−1.5656	84.8266	85.9664	−2.9989	135.6620	225.6220
	0.0587	0.1094	0.0944	0.0014	0.0000	0.0000

Note: BNPL, FCWE, FCSE, GDPG, UMEN, INFL, and LCDP indicate the bank NPLs to gross loans, weak episodes of fiscal consolidation, strong episodes of fiscal consolidation, gross domestic product growth, unemployment, inflation, and the natural log of domestic credit to the private sector, respectively.

Table 5. Impact of fiscal consolidation on the NPLs of the HICs.

	Model 1 (GMM)		Model 2 (Pooled)		Model 3 (Random)		Model 4 (Fixed)	
	Coefficient (1)	t-Stat (2)	Coefficient (3)	t-Stat (4)	Coefficient (5)	t-Stat (6)	Coefficient (7)	t-Stat (8)
BNPL (-1)	0.8386 ***	173.0265						
FCWE	0.0155 **	1.8658	0.0220 ***	0.2291	0.1097 **	1.1784	0.1213 **	1.2527
FCSE	0.3010 ***	4.5585	3.4370	4.2808	1.7588	2.7890	1.4868	2.3277
GDPG	−0.3097 ***	−41.4851	−0.2696 ***	−2.9762	−0.2155 **	−2.7372	−0.1856 **	−2.2930
UNEM	0.1464 ***	29.9918	0.1673 ***	4.2974	0.3856 ***	5.7880	0.4944 ***	6.1327
INFL	0.0547 ***	23.6857	0.0016	0.0331	0.0163	0.4649	0.0175	0.4964
LCPD	−0.1006 **	−2.8072	−2.3737 ***	−5.8425	−2.5276 ***	−3.5083	−2.4397 **	−2.7017
C			14.3889 ***	7.8559	13.6066 ***	4.1882	12.0663 ***	3.0342
J-statistic		29.4914						
Prob. (J-statistic)		0.3375						
Instrument rank		34.0000						
Arellano–Bond Serial Correlation Test								
AR (1)								
M-Statistic	−0.4684 ***							
Prob.		0.0000						
AR (2)								
M-Statistic	−0.1908							
Prob.		0.8487						
Lagrange Multiplier Tests for Random Effects								
Breusch–Pagan			486.5730 ***					
Prob.				0.0000				
Correlated Random Effects—Hausman Test								
Chi-Sq. Statistic					11.8314 **			
Prob.						0.0658		

Note: BNPL, FCWE, FCSE, GDPG, UMEN, INFL, and LCDP indicate the bank NPLs to gross loans, weak episodes of fiscal consolidation, strong episodes of fiscal consolidation, gross domestic product growth, unemployment, inflation, and the natural log of domestic credit to the private sector, respectively. For further details on these variables, see Table 1. ** < 0.05; *** < 0.01.

The first row of Table 5 reveals that the coefficient of the lagged dependent variable was positive and statistically significant at a five percent significance level. These results indicate that the NPLs increase in the current period if these loans increased in the previous period.[10] It is highly likely that the fiscal consolidation measures increased the NPLs since

these measures limit the loan-serving capacity of households and businesses. The results of this study are consistent with the theoretical expectation. The first two columns of Table 5 show that fiscal consolidation improved the NPLs in highly indebted countries. These results are consistent with those of Siakoulis (2017). It might be relevant to note that Siakoulis (2017) used the cyclical adjusted primary balance to measure the fiscal policy effects. He specifically used the positive change in the cyclical adjusted primary balance to measure the effect of austere fiscal policy. Conceptually, the first measure of fiscal consolidation is similar to the approach used by Siakoulis (2017). This study called it the weak form of fiscal consolidation, and the results of Table 5 reveal that the weak form of fiscal consolidation improved the NPLs by 1.55%. In other words, any positive change in the cyclical adjusted primary balance improved the NPLs by one and a half percent. To the best of our knowledge, no one has analysed the impact of the strength of fiscal consolidation on NPLs. Therefore, this study contributes to the existing empirical literature by adding evidence on the strong form of fiscal consolidation.

The coefficient of FCSE (the strong form of fiscal consolidation) was positive and statistically significant, revealing that the period of fiscal consolidation improved the NPLs by 30.10%. The most striking observation to emerge from the comparison between the weak form and the strong form of fiscal consolidation is that moving from the weak to the strong form improved the NPLs by 28.55 percentage points.[11] However, these results should be interpreted carefully since there are significant differences in the definitions of the NPLs across jurisdictions, despite some recent efforts in the form of IFRS 9.

The next row of Table 5 presents the results of the economic growth. The coefficient of economic growth was negative and statistically significant. These results reveal the strong dependence of the debt-serving capacity of businesses and households on economic growth. These results are consistent with those of Louzis et al. (2012) and Siakoulis (2017). Similar results were observed for the case of domestic credit to the private sector. The next control variable of this study was unemployment, and its coefficients were positive and significant. This finding broadly supports the work of other studies in this area (Louzis et al. (2012) and Siakoulis (2017)). A possible explanation might be that businesses can predict their debt-serving capacity and take steps to cut their costs. One of the possible methods of cost-cutting is reducing the labour force. However, reducing the labour force cannot always avoid debt-serving problems. Another possible explanation is that the higher level of unemployment deteriorates the debt-serving capacity of business and households.

Table 5 further reveals that the coefficient of inflation was also positive and statistically significant. These results indicate that the higher level of inflation affected the borrower's debt-serving capacity through various channels. Along these lines, Siakoulis (2017) further revealed that the impact of inflation can be positive or negative on NPLs. For instance, the higher level of inflation reduces the real value of an outstanding loan, making debt serving much easier. Conversely, the higher level of inflation reduces the real income of borrowers, which deteriorates the borrowers' capacity to repay the loans.

Furthermore, the monetary policy announcements affect NPLs in a variable loan rate environment. In particular, the monetary policy actions to reduce the level of inflation are highly likely to reduce the debt-serving capacity of borrowers, since the lenders adjust their rates to maintain real returns. In other words, they increase their interest rates in response to the increasing policy rates. These rate adjustments deteriorate the loan-paying capacity of borrowers. The inflation coefficients (see Table 5) were positive and statistically significant. The data of NPLs were not strictly comparable across countries due to a couple of differences in the national accounting, accounting standards, taxation, and supervision regimes.

5. Robustness Tests

We tested the robustness of our findings using the alternative models as elaborated in Section 4 above. In particular, we used three alternative models, including (1) the pooled model (see Model 2 in Table 5), (2) the random model (see Model 3 in Table 5), and (3) the

fixed effects model (see Model 4 in Table 5). A comparison of the coefficients of the GMM model with the coefficients of our three alternative models revealed that our estimated coefficients maintained their orders of magnitude and statistical significance in most of the cases. In particular, our main findings were robust to different models.

6. Conclusions

This study investigated the impact of fiscal consolidation on NPLs in highly indebted countries. Identification of the fiscal effect on banks' NPLs is difficult due to potential endogeneity. Consequently, the budgetary impact should be recorded when the economy is at a normal activity level. Along these lines, the existing literature applied the cyclical adjusted primary balance to measure the impact of fiscal consolidation on NPLs. We extended this literature by measuring the episodes of fiscal consolidation. For this purpose, we categorised the fiscal consolidation episodes into two types, including (1) the weak form of fiscal consolidation and (2) the strong form of fiscal consolidation (see Section 3.3 (Methodological Notes on Fiscal Consolidation Episodes)). Considering the theoretical aspects, the positive change in the cyclical adjusted primary balance was considered a weak fiscal consolidation episode. A strong episode of fiscal consolidation was when the cyclical adjusted primary balance improved by 1.5 percent of the GDP per year or two consecutive years where the cyclical adjusted primary balance was improved by at least 1 percent of the GDP per year. We used the World Development Indicators (World Bank 2022) to arrive at 35 highly indebted countries to analyse the impact of fiscal consolidation on NPLs.

We applied the dynamic panel data estimator to arrive at the unbiased estimator due to the time persistence in the NPLs. We applied the GMM and some alternative estimation techniques for empirical investigation, including the pooled, random, and fixed-effect models. These empirical investigations revealed that fiscal consolidation improved the NPLs in highly indebted countries. Our results suggest that both forms of fiscal consolidation (FCWE and FCSE) improved the NPLs by 1.55% and 31.10%, respectively. Our analysis also revealed that the weak-to-strong form transition of fiscal consolidation improved the NPLs by 28.55 percentage points. Policymakers should consider that the weak form of fiscal consolidation has a very low impact on NPLs, and such fiscal steps are safe for the banking sector.

Conversely, the strong forms of fiscal consolidation had strong detrimental effects on the banking sector's balance sheets. However, these results should be interpreted carefully, since the definitions of NPLs vary across countries. A comparatively recent addition to the international financial reporting standard (IFRS 9 available at IFRS-IFRS 9 Financial Instruments) puts some restrictions on financial institutions and banks to assess the credit losses on loans and recognise these loans based on the forward-looking approach. Despite these guidelines, there is a need for a universally accepted criterion on the classification of loans, since loans are the most sizeable assets of the statement of financial position of banks.

Future research should investigate the impact of both forms of fiscal consolidation (FCWE and FCSE) on NPLs in countries with similar or different definitions of NPLs or include a comparison study between different definitions. Additionally, the effect of the restriction of IFRS 9 on the FCWE and FCSE association with NPLs is another future research implication that recent reporting impositions may influence. Furthermore, some threshold of the central government debt (percent of the GDP) can be used to extend the panel of HICs. Later, a comparative study can be conducted by including the less-indebted countries.

Author Contributions: Conceptualization, A.A. and H.U.R.; methodology, H.U.R.; software, A.A. and H.U.R.; validation, H.U.R. and J.S.; formal analysis, A.A. and H.U.R.; investigation, J.S.; resources, J.S.; data curation, H.U.R., A.A. and J.S.; writing—original draft preparation, H.U.R.; writing—review and editing, A.A. and J.S.; visualization, A.A. and J.S.; supervision, J.S.; project administration, H.UR. and A.A.; funding acquisition, J.S. All authors have read and agreed to the published version of the manuscript.

Notes

1. See Section 3.1 (The Data Set and HICS Classification) for detailed notes on the classification of HICS.
2. For detailed discussion on the secondary elements, see Barisitz (2013).
3. For further discussion, see ITS 227/2015, as discussed in Siakoulis (2017). Also, see Chang (2006).
4. For the relevant discussion, see IFRS 9.
5. Interested readers can see Kankpang et al. (2023) for further discussion on the impact of NPLs on profitability of banks. Also, see Muchiri and Omwenga (2023) for further discussion on the impact of provision of NPLs on the financial performance of commercial banks in Kenya.
6. The descriptive statistics do not cover the latest crises since the latest available data values are from 2020. However, the descriptive analysis revealed some insights from 2017's data.
7. For further discussion, see Nelson and Plosser (1982) and Rahman and Ali (2022).
8. For further discussion, see Buck et al. (2008) and Chang et al. (2011).
9. The loans are considered defaulting loans if the payments of interest and principles are overdue by more than three months, and the total gross loans are the total value of the loan portfolio. Furthermore, it might be relevant to note that the NPLs are the gross value of the loans recorded on the statement of financial position instead of the amount that is overdue.
10. For further discussion on the economic interpretation of a lagged dependent variable, see Louzis et al. (2012) and Sorge and Virolainen (2006).
11. For the relevant discussion, also see Gavin and Hausmann (1996).

References

Alesina, Alberto, and Roberto Perotti. 1995. Fiscal expansions and adjustments in OECD countries. *Economic Policy* 10: 205–48. [CrossRef]

Alesina, Alberto, and Silvia Ardagna. 2010. Large changes in fiscal policy: Taxes versus spending. *Tax Policy and the Economy* 24: 35–68. [CrossRef]

Alizadeh Janvisloo, Mohammadreza, and Junaina Muhammad. 2013. Non-performing loans sensitivity to macro variables: Panel evidence from Malaysian commercial banks. *American Journal of Economics* 3: 16–21.

Altar, Moisa, Ciprian Necula, and Gabriel Bobeica. 2010. Estimating the cyclically adjusted budget balance for the Romanian economy. A robust approach. *Romanian Journal of Economic Forecasting* 13: 79–99.

Arellano, Manuel, and Olympia Bover. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* 68: 29–51. [CrossRef]

Arellano, Manuel, and Stephen Bond. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies* 58: 277–97. [CrossRef]

Barisitz, Stephan. 2013. Non-performing loans in Western Europe—A selective comparison of countries and national definitions. *Focus on European Economic Integration Q* 1: 28–47.

Barrios, Salvador, Sven Langedijk, and Lucio R. Pench. 2010. EU Fiscal Consolidation after the Financial Crisis Lessons from Past Experiences. Bank of Italy Occasional Paper. Available online: https://ssrn.com/abstract=1985234 (accessed on 20 April 2023).

Beck, Roland, Petr Jakubik, and Anamaria Piloiu. 2013. Non-performing loans: What matters in addition to the economic cycle?

Boumparis, Periklis, Costas Milas, and Theodore Panagiotidis. 2019. Non-performing loans and sovereign credit ratings. *International Review of Financial Analysis* 64: 301–14. [CrossRef]

Bouthevillain, Carine, Philippine Cour-Thimann, Gerrit Van den Dool, Pablo Hernández De Cos, Geert Langenus, Matthias F. Mohr, Sandro Momigliano, and Mika Tujula. 2001. *Cyclically Adjusted Budget Balances: An Alternative Approach*. ECB Working Paper No. 77. Available online: https://ssrn.com/abstract=356222 (accessed on 20 April 2023).

Buck, Trevor, Xiaohui Liu, and Rodion Skovoroda. 2008. Top executive pay and firm performance in China. *Journal of International Business Studies* 39: 833–850. [CrossRef]

Bulundell, Richard, and Stephen Bond. 1998. Initial conditions and Moment Conditions in Dynamic Panel Data Models. *Journal of Econometrics* 87: 115–143. [CrossRef]

Cameron, Adrian Colin, and Pravin K. Trivedi. 2010. *Microeconometrics Using Stata*. College Station: Stata Press, A Stata Press Publication, StataCorp LLC, vol. 2.

Carlstrom, Charles T., and Timothy S. Fuerst. 1997. Agency costs, net worth, and business fluctuations: A computable general equilibrium analysis. *The American Economic Review* 87: 893–910.

Chang, Hsin-Chen, Bwo-Nung Huang, and Chin Wei Yang. 2011. Military expenditure and economic growth across different groups: A dynamic panel Granger-causality approach. *Economic Modelling* 28: 2416–23. [CrossRef]

Chang, Yoonhee Tina. 2006. *Role of Non-Performing Loans (NPLs) and Capital Adequacy in Banking Structure and Competition*. University of Bath School of Management Working Paper No. 2006, Bath. 16, CCP Working Paper No. 06-15, 1-33. Available online: https://ssrn.com/abstract=938475 (accessed on 20 April 2023).

De Bock, Mr Reinout, and Mr Alexander Demyanets. 2012. *Bank Asset Quality in Emerging Markets: Determinants and Spillovers*. Washington: International Monetary Fund.

Dimitrios, Anastasiou, Louri Helen, and Tsionas Mike. 2016. Determinants of non-performing loans: Evidence from Euro-area countries. *Finance Research Letters* 18: 116–19. [CrossRef]

Fallanca, Mariagrazia, Antonio Fabio Forgione, and Edoardo Otranto. 2021. Do the Determinants of Non-Performing Loans Have a Different Effect over Time? A Conditional Correlation Approach. *Journal of Risk and Financial Management* 14: 21. [CrossRef]

Fisher, Irving. 1933. The debt-deflation theory of great depressions. *Econometrica: Journal of the Econometric Society* 1: 337–57. [CrossRef]

Fofack, Hippolyte. 2005. *Non-Performing Loans in Sub-Saharan Africa: Causal Analysis and Macroeconomic Implications*. World Bank Policy Research Working Paper No. 3769. Washington: The World Bank. Available online: https://ssrn.com/abstract=849405 (accessed on 20 April 2011).

Gavin, Michael, and Ricardo Hausmann. 1996. The Roots of Banking Crises: The Macroeconomic Context. IDB Working Paper No. 262. January. Available online: https://ssrn.com/abstract=1815948 (accessed on 20 April 2011).

Gholami, Amir, John Sands, and Habib Ur Rahman. 2022. Environmental, Social and Governance disclosure and value generation: Is the financial industry different? *Sustainability* 14: 2647. [CrossRef]

Gholami, Amir, John Sands, and Syed Shams. 2023. Corporates' sustainability disclosures impact on cost of capital and idiosyncratic risk. *Meditari Accountancy Research* 31: 861–86. [CrossRef]

Greenidge, Kevin, and Tiffany Grosvenor. 2010. Forecasting non-performing loans in barbados. *Journal of Business, Finance & Economics in Emerging Economies* 5: 79–108.

Hodrick, Robert J., and Edward C. Prescott. 1997. Postwar US business cycles: An empirical investigation. *Journal of Money, Credit, and Banking* 29: 1–16. [CrossRef]

Jiajia, Liu, Guo Kun, Tang Fangcheng, Wang Yahan, and Wang Shouyang. 2023. The effect of the disposal of non-performing loans on interbank liquidity risk in China: A cash flow network-based analysis. *The Quarterly Review of Economics and Finance* 89: 105–19. [CrossRef]

Jiang, Chunxia, Angelos Kanas, and Philip Molyneux. 2018. Public policy and financial stability: The impact of PCA and TARP on U.S. bank non-performing loans. *International Journal of Finance & Economics* 23: 376–92. [CrossRef]

Kankpang, Alphonsus Kechi, Suleiman Gbenga Lawal, and Ashishie P. Uklala. 2023. Credit risk and profitability of deposit money banks in nigeria. *Nigerian Journal of Management Sciences* 24: 177–87.

Kartal, Mustafa Tevfik, Derviş Kirikkaleli, and Fatih Ayhan. 2021. Nexus between non-performing loans and economic growth in emerging countries: Evidence from Turkey with wavelet coherence approach. *International Journal of Finance & Economics* 28: 1250–60. [CrossRef]

Kiyotaki, Nobuhiro, and John Moore. 1997. Credit cycles. *Journal of Political Economy* 105: 211–48. [CrossRef]

Kjosevski, Jordan, and Mihail Petkovski. 2021. Macroeconomic and bank-specific determinants of non-performing loans: The case of baltic states. *Empirica* 48: 1009–28. [CrossRef]

Konstantakis, Konstantinos N., Panayotis G. Michaelides, and Angelos T. Vouldis. 2016. Non performing loans (NPLs) in a crisis economy: Long-run equilibrium analysis with a real time VEC model for Greece (2001–2015). *Physica A: Statistical Mechanics and Its Applications* 451: 149–61. [CrossRef]

Le, Chau, Aleksandar Šević, Panayiotis G Tzeremes, and Trong Ngo. 2020. Bank efficiency in Vietnam: Do scale expansion strategies and non-performing loans matter? *International Journal of Finance & Economics* 27: 822–43. [CrossRef]

Louzis, Dimitrios P., Angelos T. Vouldis, and Vasilios L. Metaxas. 2012. Macroeconomic and bank-specific determinants of non-performing loans in Greece: A comparative study of mortgage, business and consumer loan portfolios. *Journal of Banking & Finance* 36: 1012–27.

Makri, Vasiliki, Athanasios Tsagkanos, and Athanasios Bellas. 2014. Determinants of non-performing loans: The case of Eurozone. *Panoeconomicus* 61: 193–206. [CrossRef]

Messai, Ahlem Selma, and Fathi Jouini. 2013. Micro and macro determinants of non-performing loans. *International Journal of Economics and Financial Issues* 3: 852–60.

Mirdala, Rajmund. 2013. Lessons learned from tax versus expenditure based fiscal consolidation in the European transition economies. *Journal of Applied Economic Sciences* 8: 73–98.

Muchiri, Naomi Wangari, and Jane Queen Omwenga. 2023. Liquidity capacity and financial performance of commercial banks in kenya. *Reviewed Journal International of Business Management* 4: 165–85.

Nelson, Charles R., and Charles R. Plosser. 1982. Trends and random walks in macroeconmic time series: Some evidence and implications. *Journal of Monetary Economics* 10: 139–62. [CrossRef]

Perotti, Roberto. 1996. Fiscal consolidation in Europe: Composition matters. *The American Economic Review* 86: 105–10.

Quagliariello, Mario. 2007. Banks' riskiness over the business cycle: A panel analysis on Italian intermediaries. *Applied Financial Economics* 17: 119–38. [CrossRef]

Rahman, Habib Ur. 2018. *Vulnerability to Crisis, Fiscal Consolidation and Banking Sector Stability: Evidence from Selected Economies*. Robina: Bond University.

Rahman, Habib Ur, and Asif Ali. 2022. Revisiting the role of audit and compensation 'committees' characteristics in the financial performance of the non-financial sector through the lens of the difference generalised method of moments. *Cogent Business & Management* 9: 2085365.

Rahman, Habib Ur, Muhammad Waqas Yousaf, and Nageena Tabassum. 2020. Bank-specific and macroeconomic determinants of profitability: A revisit of Pakistani banking sector under dynamic panel data approach. *International Journal of Financial Studies* 8: 42. [CrossRef]

Roland, Beck, Jakubik Petr, and Piloiu Anamaria. 2013. *Non-Performing Loans What Matters in Addition to the Economic Cycle*. European Central Bank: Working Paper Series 1515. Frankfurt: European Central Bank.

Salas, Vicente, and Jesus Saurina. 2002. Credit risk in two institutional regimes: Spanish commercial and savings banks. *Journal of Financial Services Research* 22: 203–24. [CrossRef]

Siakoulis, Vasileios. 2017. Fiscal Policy Effects on Non-Performing Loan Formation. Bank of Greece Working Paper No. 224. May 1. Available online: https://ssrn.com/abstract=4192683 (accessed on 18 August 2022).

Sorge, Marco, and Kimmo Virolainen. 2006. A comparative analysis of macro stress-testing methodologies with application to Finland. *Journal of Financial Stability* 2: 113–51. [CrossRef]

Tanasković, Svetozar, and Maja Jandrić. 2015. Macroeconomic and institutional determinants of non-performing loans. *Journal of Central Banking Theory and Practice* 4: 47–62. [CrossRef]

Thornton, John, and Caterina Di Tommaso. 2020. The effect of non-performing loans on credit expansion: Do capital and profitability matter? Evidence from European banks. *International Journal of Finance & Economics* 26: 4822–39. [CrossRef]

Vithessonthi, Chaiporn. 2016. Deflation, bank credit growth, and non-performing loans: Evidence from Japan. *International Review of Financial Analysis* 45: 295–305. [CrossRef]

Vogiazas, Sofoklis D., and Eftychia Nikolaidou. 2011. Investigating the determinants of non-performing loans in the Romanian banking system: An empirical study with reference to the Greek crisis. *Economics Research International* 2011: 214689. [CrossRef]

World Bank. 2022. World Development Indicators. Available online: https://datbank.worldbank.org/source/world-development-indicators (accessed on 28 June 2023).

Zheng, Changjun, Probir Kumar Bhowmik, and Niluthpaul Sarker. 2019. Industry-specific and macroeconomic determinants of non-performing loans: A comparative analysis of ARDL and VECM. *Sustainability* 12: 325. [CrossRef]

Article

Spatial Multivariate GARCH Models and Financial Spillovers

Rosella Giacometti [1], Gabriele Torri [1,2,*], Kamonchai Rujirarangsan [3] and Michela Cameletti [3]

[1] Department of Management, University of Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy; rosella.giacometti@unibg.it

[2] Department of Finance, VŠB-TU Ostrava, Sokolská třída 33, 701 21 Ostrava, Czech Republic

[3] Department of Economics, University of Bergamo, Via dei Caniana 2, 24127 Bergamo, Italy; michela.cameletti@unibg.it (M.C.)

[*] Correspondence: gabriele.torri@unibg.it

Abstract: We estimate the risk spillover among European banks from equity log-return data via Conditional Value at Risk (CoVaR). The joint dynamic of returns is modeled with a spatial DCC-GARCH which allows the conditional variance of log-returns of each bank to depend on past volatility shocks to other banks and their past squared returns in a parsimonious way. The backtesting of the resulting risk measures provides evidence that (i) the multivariate GARCH model with Student's t distribution is more accurate than both the standard multivariate Gaussian model and the Filtered Historical Simulation (FHS), and (ii) the introduction of a spatial component improves the assessment of risk profiles and the market risk spillovers.

Keywords: spatial multivariate GARCH; spatial weights; CoVaR

JEL Classification: C31; C32; G21; G32

1. Introduction

The interconnectedness of risk between banks is an increasingly hot topic. In the last decades, several countries have simultaneously faced severe economic conditions with spillover effects of risk across the EU. Due to the direct and indirect links among the banks, the stand-alone measurement of a Value-at-Risk (VaR) of each bank cannot provide a comprehensive representation of the risk (Adrian and Brunnermeier 2014; Billio et al. 2012; Rahman 2014).

Recently, multivariate GARCH models have been playing a crucial role in estimating risk interconnectedness. The constant conditional correlation GARCH model (CCC-GARCH) proposed by Bollerslev (1990) is computationally less complex than other multivariate models (see, among others, Bollerslev et al. 1988; Diebold and Nerlove 1989; Engle et al. 1990). However, it does not capture the dynamic interactions between the volatilities. The BEKK model proposed by Engle and Kroner (1995) (the acronym stands for Baba, Engle, Kraft, and Kroner) allows for the dependence of conditional variances and covariance of one variable on the lagged values of another variable, so that spillovers in variances can be modeled. However, it is highly computationally intensive due to the large number of parameters. A more parsimonious model is the Dynamic Conditional Correlation GARCH (DCC-GARCH), introduced by Engle (2002), that introduces an autoregressive process for the conditional correlation matrix, allowing us to model its dynamics with only two parameters in addition to the ones of the CCC-GARCH model.

An approach for modeling explicitly the volatility interactions is the introduction of a spatial component that accounts for the effect of direct bilateral exposures, closeness, or similarities between different financial institutions. Borovkova and Lopuhaa (2012) adopted a spatial GARCH approach to handle the spillover effects where the spatial weights are obtained from the GDP data and alternatively from the market capitalization of the US and

European countries' stock market and embedded in the Extended CCC-GARCH model (E-CCC), see Jeantheau (1998). As a result, they better capture the high kurtosis of squared returns. Keiler and Eder (2013) studied the systematic risk that integrates the interaction between the micro and macro stress situations as spatial econometrics parameters. Analogously, Chen (2017) showed that when the spatial weights are derived from credit rating downgrades, the multivariate spatial BEKK model can capture the spillover effects among the southern European stock index: Portugal, Italy, Ireland, Greece, and Spain (PIIGS). Zhang et al. (2018) applied the multivariate GARCH with a dynamic panel of spatial weight matrices based on the GDP. The work studies the countries' interconnectedness of returns and uses the estimated parameters to forecast the portfolio risk of six stock indices.

Our contribution is to introduce a dynamic conditional correlation GARCH (DCC-GARCH) model with a spatial component based on the credit exposure similarity among banks derived from the EU-wide stress test data. We add the spatial components into a DCC-GARCH model to investigate whether we can better capture the spillover effects thanks to this additional information and alternative distributional assumptions of the DCC-GARCH model. We discuss and implement the estimation of the model using both Gaussian and Student's t distributional assumptions. In particular, we estimate the individual risk via Value at Risk (VaR) and spillover risk via CoVaR. The results of the different models are compared, both amongst themselves and with the one obtained thanks to Filtered Historical Simulations (FHS). Finally, the results are backtested (Abad et al. 2014; Caporin 2008; Christoffersen and Pelletier 2004; Kupiec 1995) to evaluate the accuracy of the different VaR and CoVaR estimates, showing the superiority of the spatial DCC-GARCH model compared to the other models.

Finally, we point out that the applications of spatial DCC-GARCH models not only allow us to estimate CoVaR, but also permit investors to estimate more accurately the distribution of future returns of a set of assets to develop optimal portfolio strategies. The proposed framework has relevant applications for financial regulators interested in accurately measuring risk spillovers and systemic risk in financial systems, but it also caters to risk managers who want to measure the risk related to the interconnectedness among institutions.

The remainder of this paper is organized as follows. In Section 2, the spatial DCC-GARCH model and the estimation procedure are discussed, and the methodology for the financial application is presented. Section 3 presents the data and the empirical results, and in Section 4 we present our conclusions and discuss the results in relation to the literature.

2. Matherials and Methods

2.1. Modelling and Inference

A common feature of financial time series is the presence of volatility clustering (see, e.g., Cont 2001).[1] Common tools used to address such features are Generalized Auto-Regressive Conditional Heteroscedasticity (GARCH) models (Bollerslev 1986), which generalize the ARCH models introduced by Engle (1982). Let r_t be the return discrete-time process with zero mean. The standardized disturbances ε_t are independent and identically distributed (i.i.d.) with zero mean, $E(\varepsilon_t \mid \varepsilon_{t-1}, \ldots) = 0$, and unit variance, $Var(\varepsilon_t \mid \varepsilon_{t-1}, \ldots) = 1$. Then, the GARCH($p, q$) process for return r_t is defined as

$$r_t = \sqrt{h_t}\varepsilon_t, \quad t = 1, \ldots, T \tag{1}$$

and

$$h_t = \omega + \sum_{k=1}^{q} \alpha_k r_{t-k}^2 + \sum_{k=1}^{p} \beta_k h_{t-k}, \tag{2}$$

where h_t is the conditional variance, $\omega > 0$, $\alpha_k \geq 0$ and $\beta_k \geq 0 \ \forall k$.

When studying spillover risk, it is natural to look for multivariate extensions of the GARCH model to characterize the joint evolution of stock returns. Before presenting the

multivariate model it is useful to define the following quantities of interest. Assuming a market with N assets, then at time $t = 1, \ldots, T$ we have:

- r_t is the vector of assets' returns at time t,
- H_t is the conditional covariance matrix,
- $h_t = diag(H_t)$ is the vector of the univariate conditional variances,
- D_t is a squared matrix with the conditional standard deviations h_t on the main diagonal and zero otherwise.
- R_t is the positive definite conditional correlation matrix,
- Q_t is the conditional covariance matrix of the standardized residuals,
- $\bar{Q}$ is the unconditional covariance matrix of the standardized residuals,

Full generalizations of a univariate model, such as the VEC GARCH model (Bollerslev et al. 1988; Ling and McAleer 2003) or the BEKK model, (Engle and Kroner 1995) have been extensively discussed in the literature. Using matrix notation, it is possible to characterize a multivariate GARCH as follows:

$$r_t = H_t^{1/2} \varepsilon_t, \tag{3}$$

and

$$H_t = A_0 + \sum_{k=1}^{q} A_k r_{t-k} r_{t-k}' + \sum_{k=1}^{p} B_k H_{t-k}, \tag{4}$$

where H_t, A_k, B_k are $N \times N$ matrices and ε_t is an $\mathbb{R}^N$ valued i.i.d. sequence of random variables with zero-mean and unit-variances (see Engle and Kroner (1995) for the restrictions required to ensure stationarity and positive semi-definiteness of the conditional covariance matrix).

Multivariate GARCH models have the drawback of having a large number of parameters, making the estimation complex and computationally challenging, hence these models are suitable only if the dimensionality N is very small. A solution to the dimensionality problem is to pose further restrictions on the multivariate process. A common restricted specification is the Constant Conditional Correlation model (CCC) proposed by Bollerslev (1990) that assumes that the conditional covariance matrix is constant over time, requiring focusing solely on the estimation of conditional variances. According to the CCC-GARCH model, Equation (1) is given by (3) and Equation (2) can be written as follows:

$$h_t = \omega + \sum_{k=1}^{q} \alpha_k \odot r_{t-k}^2 + \sum_{k=1}^{p} \beta_k \odot h_{t-k}, \tag{5}$$

where ω is the $N \times 1$ dimensional vector of unconditional variances with $\omega \in \mathbb{R}^+$, α_k and β_k are the $N \times 1$ dimensional vector of ARCH and GARCH parameters of order q and p with $\alpha_{k,i} \in \mathbb{R}_0^+$, $\beta_{k,i} \in \mathbb{R}_0^+$, and $\odot$ is the Hadamard product.

The CCC-GARCH model assumes that the conditional covariance matrix, H_t, can be factorized as

$$H_t = D_t R D_t, \tag{6}$$

where the correlation matrix is assumed to be constant throughout time ($R_t = R, \ \forall t$) and the conditional standard deviation matrix D_t is a diagonal matrix given by

$$D_t = diag\left(\sqrt{h_t}\right). \tag{7}$$

The generic element of conditional covariance matrix H_t is constructed as

$$[H_t]_{ij} = \sqrt{h_{it}} \rho_{ij} \sqrt{h_{jt}}, \quad i \neq j; \ i, j = 1, \ldots, N, \tag{8}$$

where $\rho_{ij} = [R]_{ij}$ is the constant conditional correlation coefficient between the ith and jth variables.

The multivariate GARCH model with a dynamic conditional correlation structure (DCC), introduced by Engle (2002), improves the dynamic relationship, assuming a time-varying correlation matrix as follows

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t. \tag{9}$$

The dynamic correlation model allows $\boldsymbol{R}_t$ to be time-varying, and its dynamics are modeled assuming a GARCH(1,1) process for the covariance of the standardized residuals. Hence $\boldsymbol{R}_t$ is decomposed into

$$\boldsymbol{R}_t = diag(\boldsymbol{Q}_t^{-1}) \boldsymbol{Q}_t diag(\boldsymbol{Q}_t^{-1}), \tag{10}$$

where

$$\boldsymbol{Q}_t = \bar{\boldsymbol{Q}}(1 - \gamma - \delta) + \gamma(\boldsymbol{\epsilon}_{t-1}\boldsymbol{\epsilon}'_{t-1}) + \delta\boldsymbol{Q}_{t-1}, \tag{11}$$

where γ and δ are ARCH parameters and GARCH parameters of the DCC model, respectively. By following the GARCH model from Equation (2), the generic element of the time-varying conditional covariance matrix of the standardized residuals $[\boldsymbol{Q}_t]_{ij} = q_{ij,t}$ can be expressed as

$$q_{ij,t} = \bar{q}_{ij}(1 - \gamma - \delta) + \gamma(\epsilon_{i,t-1}\epsilon_{j,t-1}) + \delta q_{ij,t-1}, \tag{12}$$

where $\bar{q}_{ij} = [\bar{\boldsymbol{Q}}]_{ij}$. The process is mean-reverting as long as $0 < \delta < 1$ and $\gamma + \delta < 1$. In the particular case of $\gamma + \delta = 1$, the process will follow the exponential smoother matrix of the standard residuals, as described in Engle (2002). Finally, the generic conditional correlation

$$\rho_{ij,t} = \frac{q_{ij,t}}{\sqrt{q_{ii,t}q_{jj,t}}}, \tag{13}$$

can be written into matrix form as in Equation (10). Substituting the conditional correlation matrix into Equation (9), the DCC is given by

$$\boldsymbol{H}_t = \boldsymbol{D}_t \boldsymbol{R}_t \boldsymbol{D}_t = \boldsymbol{D}_t diag(\boldsymbol{Q}_t^{-1}) \boldsymbol{Q}_t diag(\boldsymbol{Q}_t^{-1}) \boldsymbol{D}_t. \tag{14}$$

Restricted GARCH models beyond CCC and DCC-GARCH have been discussed by Caporin (2008) and Billio et al. (2021), who introduce spatial matrices within BEKK models for measuring risk spillover. In these approaches, the interaction components of the model are based on spatial weight matrices provided exogenously (for instance on the basis of geographical distances among assets, or some similarity metrics). These models allow easier and more accurate estimation by effectively imposing restrictions on the parameter space. An alternative approach to improve the estimation of multivariate GARCH models is to introduce sparsity in the parameter estimates by using an L_1 penalization, as suggested by Dhaene et al. (2022).

2.1.1. Spatial DCC-GARCH

In this work, we introduce a spatial extension of the DCC-GARCH model. The model is based on the approach of Borovkova and Lopuhaa (2012). In particular, to enrich the DCC-GARCH model with a spatial component we introduce a spatial matrix $\boldsymbol{W}$ into the vector of the conditional variances $\boldsymbol{h}_t$. The resulting conditional variance is expressed as

$$\boldsymbol{h}_t = \boldsymbol{A}_0 + \sum_{k=1}^{q} (\boldsymbol{A}_{1,k} + \boldsymbol{A}_{2,k}\boldsymbol{W})\boldsymbol{r}^2_{t-k} + \sum_{k=1}^{p} (\boldsymbol{B}_{1,k} + \boldsymbol{B}_{2,k}\boldsymbol{W})\boldsymbol{h}_{t-k}, \tag{15}$$

where $\boldsymbol{A}_0 = (a_{0,1}, \ldots, a_{0,N})'$, $\boldsymbol{A}_{1,k}$, $\boldsymbol{A}_{2,k}$, $\boldsymbol{B}_{1,k}$, and $\boldsymbol{B}_{2,k}$ are diagonal matrices. The term $\boldsymbol{W} = [\boldsymbol{W}_{ij}]$ is the weight matrix ($i, j = 1, \ldots, N$) with $\sum_{j=1}^{N} w_{ij} = 1$ and $w_{ii} = 0$ $\forall i$, given by

$$
\mathbf{W} \quad = \quad \begin{bmatrix} 0 & w_{12} & \cdots & w_{1N} \\ w_{21} & 0 & \cdots & w_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N1} & w_{N2} & \cdots & 0 \end{bmatrix}.
$$

The i-th element of $\mathbf{h}_t$ becomes

$$
h_{t,i} = a_{0,i} + a_{1,i} r_{t-1,i}^2 + a_{2,i} X_{t-1,i} + b_{1,i} h_{t-1,i} + b_{2,i} Y_{t-1,i}, \tag{16}
$$

where $X_{t-1,i} = \sum_{j=1}^{N} w_{ij} r_{t-1,j}^2$ and $Y_{t-1,i} = \sum_{j=1}^{N} w_{ij} h_{t-1,j}$. The introduction of the spatial component results in two exogenous spatial variables in the conditional variance equation and two additional parameters $a_{2,i}$ and $b_{2,i}$, which measure the influence of the aggregated lagged variances and squared returns of all the other assets. These two new variables measure the aggregated spillover effects. To complete the Spatial DCC-GARCH model, we then estimate the conditional correlation matrix following the two-step procedure described in Engle and Sheppard (2001), see Section 2.1.2.

The condition for the weak stationarity of the spatial GARCH model follows from the corresponding stationarity condition for E-CCC models, derived by Jeantheau (1998) and Conrad and Karanasos (2010) for E-CCC models. The positivity conditions on all GARCH coefficients are not necessary for the positivity of variance and in many empirical cases, these may be too restrictive, ruling out possible negative volatility feedback. One author Conrad and Karanasos (2010) studied the E-CCC models and stated necessary and sufficient conditions (in terms of the process parameters) for the positivity of variance; these conditions are summarized in Theorem 1 of their paper. It can be seen easily that our spatial DCC-GARCH(1,1) model is equivalent to the E-DCC model of order one, with the particular form of the parameter matrices $A = A_1 + A_2 W$ and $B = B_1 + B_2 W$. So for the conditional variances to be positive, the conditions (C1)–(C3) of Theorem 1 of Conrad and Karanasos (2010) must apply. The proposed spatial DCC-GARCH(1,1) model is weakly stationary if the modulus of the largest eigenvalue of the matrix $A_1 + B_1 + (A_2 + B_2)W$ is less than 1. In that case, the unconditional variances are given by $A_0(I - (A_1 + B_1) - (A_2 + B_2)W)^{-1}$. More specifically the unconditional variance of the ith bank is given by

$$
\sigma_i^2 = \frac{a_{0,i} + (a_{2,i} + b_{2,i}) \sum_{j=1}^{n} w_{ij} \sigma_j^2}{1 - (a_{1,i} + b_{1,i})} \tag{17}
$$

and it is positive for $a_{0,i} > 0$, $a_{2,i} + b_{2,i} > 0$, $(a_{1,i} + b_{1,i}) < 1$.

2.1.2. Estimation of the Multivariate Spatial GARCH(1,1) Model

We follow a two-step procedure for the DCC-GARCH estimation, as described in Engle and Sheppard (2001) and Engle (2002). The first step is devoted to the estimation of (16) where the exogenous variable $Y_{t,i}$ is not observable since it is a function of the conditional variance of the other assets. Hence, following Borovkova and Lopuhaa (2012), we start by estimating the standard univariate GARCH(1,1) models without the external regressors to obtain the initial parameters $\left(a_{0,i}^0, a_{1,i}^0, b_{1,i}^0\right)$ and the estimated variances $\left(h_{1,i}^0, \ldots, h_{T,i}^0\right)$. Then we use an iterative procedure in which we alternate the following two steps:

- Compute the exogenous variables $(Y_{t-1,i})$ given the weights (w_{ij}) and the initially estimated variances $\left(h_{1,i}^0, \ldots, h_{T,i}^0\right)$;
- Estimate the complete set of parameters $\left(a_{0,i}^1, a_{1,i}^1, b_{1,i}^1, a_{2,i}^1, b_{2,i}^1\right)$ and the new estimated variances $\left(h_{1,i}^1, \ldots, h_{T,i}^1\right)$ according to Equation (16).

We iterate this procedure until the percentage variation of the estimate is less than a small threshold. For more details please refer to Borovkova and Lopuhaa (2012).

In the second step, as in Engle (2002), we maximize the quasi log-likelihood that, when the standardized error ϵ_t follows a multivariate Gaussian distribution is

$$
\begin{aligned}
\log\big(L(\boldsymbol{\theta_2}|\hat{\boldsymbol{\theta}}_1; \boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)\big) &= -\tfrac{1}{2}\sum_{t=1}^{T}\Big(N\log(2\pi) + 2\log|\boldsymbol{D}_t| + \log(|\boldsymbol{R}_t|) + \boldsymbol{r}_t'\boldsymbol{D}_t^{-1}\boldsymbol{R}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{r}_t\Big) \\
&= -\tfrac{1}{2}\sum_{t=1}^{T}\Big(N\log(2\pi) + 2\log|\boldsymbol{D}_t| + \log(|\boldsymbol{R}_t|) + \boldsymbol{\epsilon}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\epsilon}_t\Big),
\end{aligned}
\tag{18}
$$

where $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$ is the set of parameters of the multivariate distribution, with subsets $\boldsymbol{\theta}_1 = (\boldsymbol{A}_0, \boldsymbol{A}_1, \boldsymbol{A}_2, \boldsymbol{B}_1, \boldsymbol{B}_2)$ being the spatial GARCH parameters estimated in the first step, and $\boldsymbol{\theta}_2 = (\gamma, \delta)$ the parameters of the time-varying conditional correlation that are estimated in the second step.

Excluding $\boldsymbol{D}_t$ and other additive and multiplicative constants, we maximize the following function:

$$
-\sum_{t=1}^{T}\Big(\log(|\boldsymbol{R}_t|) + \boldsymbol{\epsilon}_t'\boldsymbol{R}_t^{-1}\boldsymbol{\epsilon}_t\Big).
\tag{19}
$$

The quasi-log-likelihood function under the Student's t distribution is

$$
\begin{aligned}
\log\big(L(\boldsymbol{\theta_2}|\hat{\boldsymbol{\theta}}_1; \boldsymbol{r}_1, \ldots, \boldsymbol{r}_T)\big) &= \sum_{t=1}^{T}\Big(\log\big(\Gamma\big(\tfrac{\nu+N}{2}\big)\big) - \log\big(\Gamma\big(\tfrac{\nu}{2}\big)\big) - \tfrac{N}{2}\log(\pi(\nu-2))\Big) + \\
&\quad -\tfrac{1}{2}\log(|\boldsymbol{D}_t\boldsymbol{R}_t\boldsymbol{D}_t|) - \tfrac{\nu+N}{2}\log\Big(1 + \tfrac{\boldsymbol{r}_t^T\boldsymbol{D}_t^{-1}\boldsymbol{R}_t^{-1}\boldsymbol{D}_t^{-1}\boldsymbol{r}_t}{\nu-2}\Big),
\end{aligned}
\tag{20}
$$

where ν is the degrees of freedom, $\boldsymbol{\theta_2} = (\gamma, \delta, \nu)$ is the set of parameters estimated in the second step, and $\Gamma(\cdot)$ is the Gamma function. The estimation of the model is implemented in R using the packages `rugarch` and `rmgarch` for the estimation of univariate GARCH models in the first step, and the DCC-GARCH model in the second step, respectively.

Concerning the complexity of estimation, we see that the spatial models add two parameters $(a_{i,2}, b_{i,2})$ for each asset. Hence the number of additional parameters scales linearly with the size of the dataset considered. We also see that Student's t model has one extra parameter compared to the Gaussian model, and that in the limit for $\nu \to \infty$, the former converges to the latter. Moreover, the spatial model nests the non-spatial DCC-GARCH models, where the coefficients of the spatial components are restricted to zero. The Spatial DCC-GARCH model can therefore be considered parsimonious in terms of the number of parameters, especially compared to VEC GARCH or the BEKK model. One drawback of the proposed model is that it requires the exogenous identification of a spatial matrix.

2.1.3. Spatial Weight Matrix

To estimate the spatial DCC-GARCH described in Section 2.1.1, we need to specify the weight matrix $\boldsymbol{W}$ which incorporates the spatial structure defined a priori. The most intuitive way to compute the weights is to consider the geographical distance between the issuers' market cities. However, according to Borovkova and Lopuhaa (2012), the obtained weights are not economically meaningful, and as an alternative, they consider a different set of information and compute distance in terms of GDP and market capitalization. In our work, we investigate whether the banks' similarity of the structure of credit exposure provides some benefit in catching risk spillover effects. Hence we propose to consider the cosine similarity between exogenous information relative to the credit exposure of each bank derived from the EU-wide stress test under the European Banking Authority (EBA). The higher the cosine similarity the stronger the closeness of banks's credit exposure. Suppose two attribute vectors of length L, $U_{i,L} = (u_{i,1}, u_{i,2}, \ldots, u_{i,L})$ and

$U_{j,L} = (u_{j,1}, u_{j,2}, \dots, u_{j,L})$ which describe the credit exposure information of bank i and j with $i, j = 1, \dots, N$. We define the cosine similarity as follows:

$$C_{ij} = \frac{\sum_{l=1}^{L} u_{i,l} \cdot u_{j,l}}{\sqrt{\sum_{l=1}^{L} u_{i,l}^2 \cdot \sum_{l=1}^{L} u_{j,l}^2}}, \quad i, j = 1, \dots, N, \quad i \neq j. \tag{21}$$

We set $C_{ii} = 0 \ \forall_i$ and we normalize the rows of C by dividing each element by the sum of the row. Doing so, we obtain the matrix W that is the spatial weight matrix used in Equation (15).

2.2. Financial Application: CoVaR

Financial institutions use VaR to measure the standalone risk. However, the measurement of individual risk is not able to explain the linkages between other financial institutions and the financial system. Systemic risk is the possibility that an event at the institutional level could trigger severe instability or collapse of an entire industry or economy. The work Adrian and Brunnermeier (2014) introduces CoVaR to help regulators to measure risk spillovers.

The Value at Risk (VaR) at level $q \in (0, 1)$ of a random variable r with cumulative distribution function $F_r(.)$ is defined as

$$VaR_q(r) = -\inf \left\{ x \in \mathbb{R} : F_r(x) \geq q \right\},$$

where $100(1 - q)\%$ denotes the confidence level of the VaR.[2] Restricting our analysis to continuous probability distribution functions, VaR can be implicitly defined as the q-quantile of the probability distribution function

$$VaR_q(r) = -F_r^{-1}(q).$$

The Conditional Value-at-Risk (CoVaR) (see Adrian and Brunnermeier 2014), denoted by $CoVaR_q^{S|\mathbb{C}(r_i)}$, is implicitly defined by the q-quantile for a continuous probability distribution function of the financial system S conditional on some event related to $\mathbb{C}(r_i)$, where r_i is the return of institution i such that

$$\Pr\left(r_S \leq -CoVaR_q^{S|\mathbb{C}(r_i)} | \mathbb{C}(r_i) \right) = q.$$

The CoVaR can capture the contribution of systemic risk by conditioning the VaR to a stressed situation. It captures the spillover of risk between a particular institution and the financial system, and it is commonly used to assess the systemic risk of a bank in a financial system. Inspired by this idea, we concentrate our attention on a CoVaR pairwise analysis between institutions in order to quantify the spillover between couples of banks.[3]

The conditioning event $\mathbb{C}(r_i)$ in the original paper by Adrian and Brunnermeier (2014) is defined as the return of the conditioning asset i being equal to its negative VaR, that is $\mathbb{C}(r_i) := (r_i | r_i = -VaR(r_i))$. In this work, we follow the alternative approach of Girardi and Ergün (2013) that considers as a conditioning event the return r_i being smaller or equal than the following quantity: $\mathbb{C}(r_i) := (r_i | r_i \leq -VaR(r_i))$. This formulation allows us to consider more severe distress events and improves the consistency of the measure with respect to the dependence parameter, allowing for backtesting. Following Girardi and Ergün (2013), the redefined $CoVaR_q^{j|i}$ is obtained solving

$$\Pr\left(r_j \leq -CoVaR_q^{j|i}, r_i \leq -VaR_q^i \right) = q^2. \tag{22}$$

Let $f(r_j, r_i)$ be the bivariate probability distribution function of future returns, estimated using the DCC-GARCH model with either Gaussian or Student's t innovations, $CoVaR_q^{j|i}$ is implicitly defined as the quantity that solves

$$\int_{-\infty}^{CoVaR_q^{x|y}} \int_{-\infty}^{VaR_q^y} f(x,y) \, dy \, dx = q^2. \tag{23}$$

We compute the integral (23) on a grid of 100 values for $CoVaR_q^{j|i}$ to find the approximated solution under the different distributional assumptions.[4]

2.2.1. CoVaR Based on Filtered Historical Simulations (FHS)

In order to compare our result with a model-free approach, we consider the Filtered Historical Simulations (FHS).

FHS is a well-known tool for multivariate forecasting and simulation of time series that avoids the need for distributional assumptions for the returns' joint dynamic, relying instead on past realizations. The main novelty of this approach compared to historical simulation is to rescale the innovation by the volatility that prevails on a specific day, allowing therefore to reflect the current market conditions (Barone-Adesi et al. 2002; Giannopoulos and Tunaru 2005; Gurrola-Perez and Murphy 2015). To provide a distribution-free benchmark model for the analysis, we compute the VaR and CoVaR via FHS. Consider a time window of length T and let r_t be the series of historical returns with $t \in [1, T]$. The volatility weighted returns series can be computed as follows: $z_t = r_t \times \hat{\sigma}_{T+1} / \hat{\sigma}_t$, where $\hat{\sigma}_t$ is the volatility estimated with an Exponentially Weighted Moving Average procedure (EWMA) with decay factor $\lambda = 0.9$ at time t and $\hat{\sigma}_{T+1}$ is the one-day-ahead estimate of volatility at the end of the estimation period. In practice, implementing FHS for the estimation of VaR and CoVaR requires the following steps:

- compute the residual (or *devol*) time series, dividing the returns by EWMA estimated volatility $\hat{\sigma}_t$. This allows us to sample from approximately serially independent and identically distributed data;
- compute the estimated empirical distribution of $\hat{r}_{T+1}$ (*revol*), multiplying the devol time series z_t by the latest estimate of volatility $\hat{\sigma}_{T+1}$ and assigning to each of the possible outcome a weight $1/T$,
- estimate $VaR_{q,T+1}^i$ and $CoVaR_{q,T+1}^{j|i}$ by computing the empirical quantile of $\hat{r}_{i,T+1}$ and $\hat{r}_{j,T+1} | \left(\hat{r}_{i,T+1} < -VaR_{q,T+1}^i \right)$, respectively.

The FHS approach has the advantage of being non-parametric, although it has the drawback of requiring a large number of observations to accurately estimate risk, especially for the CoVaR. For this reason, it is not suitable for small values of q. For instance, with $q = 0.01$ the expected number of exceedances of the CoVaR for an estimation window of 10,000 daily observations (approx 40 years) is 1, while for $q = 0.05$ it is 25.

2.2.2. Backtesting VaR and CoVaR

In order to test the goodness of our VaR and CoVaR estimates we estimate a time series of length τ of one-day-ahead estimates, each computed on an estimation window of $T = 1000$ daily observations. We consider tests based on the number of violations and specifically unconditional and conditional coverage tests (Christoffersen and Pelletier 2004; Kupiec 1995), as well as tests based on asymmetric loss functions for the VaR and CoVaR (Caporin 2008). The model that provides estimates of VaR and CoVaR with the correct number and distribution of exceedances and/or lower loss function values will be considered the more accurate.

2.2.3. Tests Based on the Number of Violations

In order to determine the accuracy of the proposed model, we consider two tests based on the number of violations.

Denote by

- r_t^i the ex-post realized returns of institution i with $t = 1, \ldots, \tau$;
- $VaR_{q,t}^i$ the ex-ante Value-at-Risk forecasts at $t-1$ for time t, where q is the expected coverage;
- I_t^i a sequence of violation for a given interval of the Value-at-Risk forecast:

$$I_t^i = \begin{cases} 1, & \text{if } r_t^i \le -VaR_{q,t}^i \\ 0, & \text{if } r_t^i > -VaR_{q,t}^i \end{cases}. \tag{24}$$

The first test is the Kupiec test or unconditional coverage (UC) test (Kupiec 1995). The null hypothesis that the observed failure rate p is equal to the failure rate, suggested by the confidence level of VaR, q, is tested. Thus, the null hypothesis assumes that the observed violation rate is equal to the expected violation rate. If the null hypothesis is rejected, the model is considered inaccurate at the 95% confidence level.

The conditional coverage (CC) test proposed by Christoffersen and Pelletier (2004) indicates that the number of violations must be independently distributed along the testing period where the dependence can be described as a first-order Markov sequence with a transition probability matrix given by

$$\Pi = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix},$$

where π_{01} is the probability that, conditional on today being a non-violation, the next period is a violation, and π_{11} is the probability that, conditional on today being a violation, the next period is a violation. The hypothesis to test for the conditional coverage property is $H_0 : \pi_{01} = \pi_{11}$ which assesses the independence of failures on consecutive time periods.

Girardi and Ergün (2013) proposed the backtesting of $CoVaR_{q,t}^{j|i}$ via a straightforward application of the standard Kupiec and Christoffersen tests considering the violations (i.e., $r_t^j \le -CoVaR_{q,t}^{j|i}$) for those time periods in which $r_t^i \le -VaR_{q,t}^i$. Having that in mind we compute a second hit sequence, $I_t^{j|i}$, on the sub-sample in which $r_t^i \le -VaR_{qt}^i$ as follows:

$$I_t^{j|i} = \begin{cases} 1, & \text{if } r_t^j \le -CoVaR_{q,t}^{j|i} \\ 0, & \text{if } r_t^j > -CoVaR_{q,t}^{j|i} \end{cases}, \tag{25}$$

where the number of observations of the second hit sequence is equal to the number of violations of the first hit sequence. Hence for the tests on CoVaR, the sequence of violation $I_t^{j|i}$ can be used instead of I_t^i.

2.2.4. Backtesting Based on Loss Functions

The backtesting based on the confidence level of VaR estimates shows the accuracy of an individual model; however, the comparison between the different models can be limited. To overcome the drawback, Lopez (1999) proposed backtesting based on a loss function. The method focuses on the magnitude of the failure when the violation occurs. Thus, the VaR estimates under the loss function can provide the model's performance as a numerical score. The value of the loss function at time t can be given as

$$l_t^i = \begin{cases} g\left(r_t, VaR_{q,t}^i\right) & \text{if } r_t \le -VaR_{q,t}^i \\ h\left(r_t, VaR_{q,t}^i\right) & \text{if } r_t > -VaR_q^{i,t} \end{cases},$$

where $g(\cdot)$ and $h(\cdot)$ are the loss functions applied to exceedances and values within the VaR, respectively. Finally $L_i = \sum_{t=1}^{T} l_t^i$ is defined as the total loss. The best model can be identified by the lowest total loss. Other works by Abad et al. (2014), Caporin (2008), and Cesarone and Colucci (2016) show several alternative specifications for the loss functions $g(\cdot)$ and $h(\cdot)$, defined from the regulator and investor's point of view. In the regulator's view, we consider the size of the loss only if the violation occurs:

$$h\left(r_t, VaR_q^i\right) = 0 \ \text{ if } \ r_t > -VaR_q^i.$$

On the contrary, the investor is interested in both sides, as an overestimation of VaR may trigger limitations from the risk management, or lead to higher capital requirements imposed by the regulator. In particular we consider the functions

$$g(r_{t,i}, VaR_{q,t}^i) = |r_t + VaR_{q,t}^i|$$

and

$$h(r_{t,i}, VaR_{q,t}^i) = \frac{q}{1-q} g(r_{t,i}, VaR_{q,t}^i).$$

We underline that the resulting loss function l_t^i is strictly related to the Koenker loss function used for the estimation of quantile regression, defined as

$$l(X, \xi, q) = (1-q)(X - \xi)_+ + q(X - \xi)_-$$

where $(\cdot)_+ = \min(X, 0)$ and $(\cdot)_- = \min(-X, 0)$. In case of independent and identically distributed returns the minimization $\arg \min_\xi l(X, \xi, \alpha)$ is the value at risk. For further details we refer to Koenker and Bassett (1978), Rockafellar and Uryasev (2013), and Giacometti et al. (2021).

We extend the backtesting procedure to the case of the CoVaR as before, estimating the measure $l_t^{j|i}$ on a sub-sample in which $r_t^i \leq -VaR_q^{i,t}$.

3. Results

3.1. Data

We consider ten years of weekly data from seven representative banks in Italy, France, Germany, the United Kingdom, the Netherlands, Spain, and Belgium. The data span from 20 September 2010 to 18 September 2020, including 2566 daily equity log-returns. The data are downloaded from Refinitiv Eikon. We perform a rolling analysis with an estimation window of $T = 1000$ daily observations (approximately 4 years), forecasting one-day-ahead VaR and CoVaR, for a total of $\tau = 1566$ out-of-sample daily observations. We use the same windows for both the DCC-GARCH models and the FHS estimation.

Table 1 reports the descriptive statistics and tests the output of the log returns for the out-of-sample period. We see that all the banks in the sample with the exception of KBC Group had negative average returns. The series have typically negative skewness and excess kurtosis, as expected from equity time series. The results of the Engle ARCH test (Engle 1982) indicate that the null hypothesis of homoscedastic returns is rejected, suggesting the need for GARCH models. The autocorrelograms and partial autocorrelograms of the returns, not reported for brevity, do not highlight relevant serial correlation structure, while the autocorrelogram of squared residuals (also omitted for brevity) show significant and persistent autocorrelations, confirming the heteroscedasticity of the data. Next, we consider the correlation between the banks. Figure 1 shows that correlations are positive and high. Figure 2 studies the evolution of correlations, computed using 6-month rolling windows. The right panel represents the dynamics of the 21 pair-wise correlations over time, while the left panel shows the Frobenius norm of the correlation matrices to provide a synthetic representation. We see that the Frobenius norm changes over time, suggesting that a CCC-GARCH model is not appropriate for the dataset. On the contrary,

the time-varying correlation matrix is consistent with the assumptions of a DCC-GARCH, and the high variability in the individual correlations leaves space for spatial models, which could better characterize the multivariate stochastic process.

Table 1. Descriptive statistics of daily equity log-returns. The table reports statistics on the univariate distributions (mean, standard deviation, skewness, and kurtosis), and the p-values of the Engle ARCH test (null hypothesis: the process is homoscedastic). We use the following abbreviations for each bank: Intesa Sanpaolo S.p.A.–Turin, Italy (ISP), Crédit Agricole Group–Montrouge, France (ACA), Deutsche Bank AG–Frankfurt am Main, Germany (DB), Barclays Plc–London, United Kingdom (BCS), ING Groep NV–Amsterdam, Netherlands (ING), Banco de Sabadell S.A.–Alicante, Spain (SAB), KBC Group NV–Bruxelles, Belgium (KBC).

Bank	Mean	StDev	Skewness	Kurtosis	ARCH Test (p-Value)
ISP	-1×10^{-4}	0.0256	-0.8101	11.6198	$<2.2 \times 10^{-16}$
ACA	-1×10^{-4}	0.0253	-0.3810	11.2140	$<2.2 \times 10^{-16}$
DB	-6×10^{-4}	0.0244	0.1203	8.1958	$<2.2 \times 10^{-16}$
BCS	-4×10^{-4}	0.0235	-0.6568	12.9321	$<2.2 \times 10^{-16}$
ING	-1×10^{-4}	0.0236	-0.4705	11.4810	$<2.2 \times 10^{-16}$
SAB	-8×10^{-4}	0.0248	-0.3435	10.9513	$<2.2 \times 10^{-16}$
KBC	-1×10^{-4}	0.0254	-0.3002	10.0297	$<2.2 \times 10^{-16}$
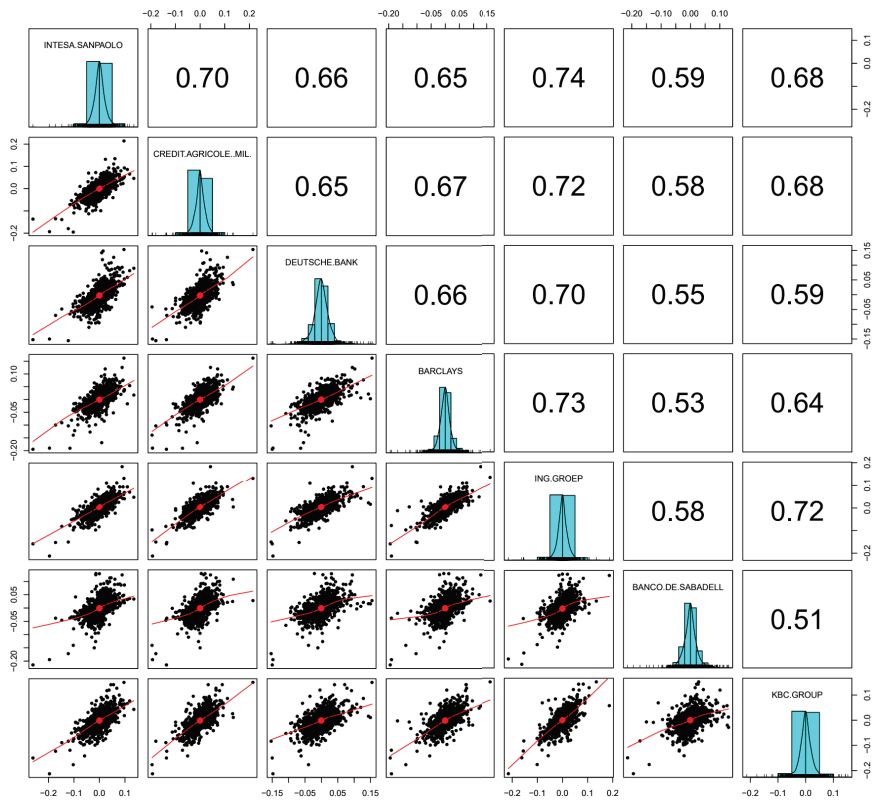
Figure 1. Correlations, marginal distributions, and bivariate scatterplots with LOESS local regression lines of equity daily log-returns.
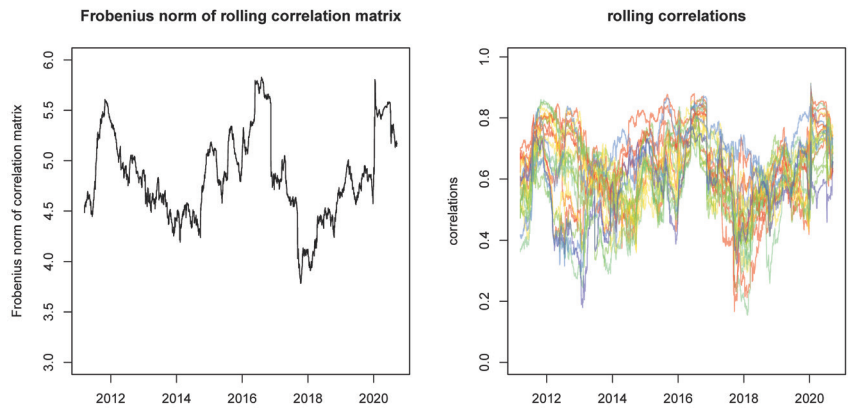
Figure 2. Frobenius norm of rolling correlation matrix (**left** panel), and 21 pair-wise rolling correlations (**right** panel). Rolling correlations are computed using a 6-month window of daily observations.

Spatial Weight Data

For the construction of the spatial matrix, we analyze the data from the EU-wide stress test under the European Banking Authority (EBA). The EBA stress test aims to evaluate financial institutions' resilience to adverse market conditions. It also provides the overall assessment of systematic risk in the European banking system. In the EU-wide stress test analysis report, we consider the base scenarios for each bank and the relative credit exposure information: exposure values, risk exposure amounts, stock of provision, and leverage ratio under the internal ratings-based (IRB) approach or Standardized approach (STA) referred to credit exposure in specific asset classes,[5] as presented on the EBA's website (EBA 2021). We compute for each bank the vector of percentage exposure in each class with respect to the total exposure and the similarity between couples of vectors for the different banks, as in Equation (21). This indicator provides a broad view of the similarity between the banks' credit structures and exposures.

We then rescale the values such that each row sums to 1, as shown in Table 2. The spatial matrix is based on the EU-wide stress test of 2018 and is kept fixed for the entire analysis. To ascertain the matrix weight's consistency over time, we perform the inequality test for couples of matrices by Jennrich (1970) on matrices computed in different years. In particular, we compare the normalized cosine similarity matrix weight from the EU-wide stress test of 2014 vs. 2016, 2016 vs. 2018, and 2018 vs. 2014. We do not reject the null hypotheses at a 1% significance level, suggesting that the spatial components of the EU-wide stress test do not change significantly over time.

Table 2. Normalized cosine similarity matrix.

Bank	ISP	ACA	DB	BCS	ING	SAB	KBC
ISP	0	0.1482	0.1763	0.1221	0.1875	0.1873	0.1786
ACA	0.1453	0	0.1807	0.1745	0.1480	0.1562	0.1953
DB	0.1689	0.1766	0	0.1347	0.1306	0.2140	0.1753
BCS	0.1388	0.2025	0.1600	0	0.1668	0.1520	0.1799
ING	0.2130	0.1715	0.1549	0.1667	0	0.1596	0.1343
SAB	0.1834	0.1560	0.2188	0.1308	0.1375	0	0.1734
KBC	0.1761	0.1964	0.1804	0.1559	0.1165	0.1746	0

3.2. Empirical Results

We estimate the DCC-GARCH(1,1) models using the equity log returns, considering four specifications that differ in terms of distribution and inclusion of the spatial component:

Gaussian DCC (GaussDCC), spatial Gaussian DCC (GaussSpDCC), Student's t DCC (tDCC), and spatial Student's t (tSpDCC).[6] The procedure is numerically stable, and only for a small percentage of the estimation window, do the DCC-GARCH models fail to converge. In such cases, we carry over the result from the previous estimation window.

Table 3 reports the Akaike, Bayesian, Shibata, and Hannan-Quinn information criteria, averaged across the 1566 estimation windows. Information criteria allow us to assess the quality of the model in relation to the data, controlling for both the quality of the fit and the number of parameters. Therefore, we use them to assess whether the inclusion of a distribution with more parameters (Student's t, compared to the Gaussian) and the use of the spatial component actually improve the quality of the fit or, on the contrary, the added complexity of the model affects negatively the estimation. We see that the models based on the Student's t distribution have lower information measures according to all the measures considered (denoting a better model for the data), and that the introduction of the spatial component improves the performance of the model.

Table 3. Akaike, Bayesian, Shibata, and Hannan-Quinn information criteria. The table reports the average across estimation windows. The smallest value for each measure is highlighted in bold.

	Akaike	**Bayes**	**Shibata**	**Hannan-Quinn**
GaussDCC	−39.18	−38.96	−39.18	−39.09
GaussSpDCC	−39.34	−39.06	−39.35	−39.23
tDCC	−39.96	−39.71	−39.98	−39.87
tSpDCC	**−40.04**	**−39.71**	**−40.04**	**−39.91**

To further illustrate the characteristics of the model, Table 4 shows the average value of the coefficients computed on the 1566 rolling windows (for brevity we report the averages of the parameters related to each bank). We notice that a relatively large part of the spatial parameters ($a_{2,i}, b_{2,i}$) are statistically significant at the 95% significance level, suggesting that the spatial components have some explanatory power, motivating their introduction in the model. We also see that the DCC parameters γ and δ are almost always statistically significant (confirming the presence of time-varying correlations). The ν parameter is also significant and has values close to 6, suggesting that the innovations have fat-tails and that the Student's t model is more suitable than the Gaussian one (that is the limit of the Student's t model with $\nu \to \infty$) Table A1 in the Appendix A reports the coefficients of the tSpDCC model for the first estimation window as an example.

Table 4. The table reports the average values of the coefficients of the different specifications across the estimation windows of the multivariate GARCH. The percentages of estimation windows in which the coefficients are significantly different from 0 at the 95% confidence level are reported in italics. For brevity we report the values of $a_{0,i}$, $a_{1,i}$, $a_{2,i}$, $b_{1,i}$, $b_{2,i}$ averages across the seven banks considered in the analysis.

	a_0	a_1	b_1	a_2	b_2	γ	δ	ν
GaussDCC	0	0.065	0.907	-	-	0.016	0.885	-
	19%	*57%*	*97%*	*-*	*-*	*74%*	*97%*	*-*
GaussSpDCC	0	0.059	0.722	0.062	0.109	0.024	0.858	-
	37%	*29%*	*81%*	*31%*	*12%*	*93%*	*99%*	*-*
tDCC	0	0.066	0.893	-	-	0.014	0.915	5.90
	27%	*69%*	*96%*	*-*	*-*	*95%*	*99%*	*100%*
tSpDCC	0	0.048	0.720	0.051	0.115	0.015	0.924	6.15
	27%	*32%*	*74%*	*25%*	*11%*	*92%*	*100%*	*100%*

3.2.1. VaR and CoVaR

In this section, we use the spatial DCC-GARCH model to compute pair-wise CoVaRs and study risk spillover in the European banking system. We report in Table 5 the estimates of the average $VaR_{5\%}^{i}$ (diagonal elements) and $CoVaR_{5\%}^{j|i}$ (off-diagonal elements) for the four DCC-GARCH models and for the Filtered Historical Simulations (FHS) estimates. The one-day-ahead forecast of $VaR_{5\%}^{i}$ is computed using the conditional variance estimate and the parametric distribution of the model. The corresponding $CoVaR_{5\%}$ is computed numerically, according to (23) using the time-varying covariance matrices.

Table 5. Estimated $VaR_{5\%}^{i}$ (diagonal elements) and $CoVaR_{5\%}^{j|i}$ (off-diagonal elements) for the spatial and non-spatial DCC GARCH models with Student's t and Gaussian innovations, and for FHS. The reported values are the average of the out-of-sample estimates computed across 1566 daily rolling windows.

Bank	ISP	ACA	DB	BCS	ING	SAB	KBC
			$VaR_{5\%}^{i}$ and $CoVaR_{5\%}^{j|i}$—FHS				
ISP	0.034	0.065	0.076	0.060	0.064	0.083	0.068
ACA	0.064	0.033	0.081	0.057	0.064	0.078	0.064
DB	0.066	0.065	0.041	0.072	0.062	0.077	0.061
BCS	0.065	0.069	0.090	0.034	0.059	0.072	0.063
ING	0.065	0.071	0.076	0.061	0.032	0.078	0.066
SAB	0.067	0.066	0.077	0.058	0.065	0.042	0.063
KBC	0.065	0.067	0.076	0.061	0.066	0.079	0.028
			$VaR_{5\%}^{i}$ and $CoVaR_{5\%}^{j|i}$—GaussDCC				
ISP	0.035	0.055	0.063	0.050	0.052	0.063	0.048
ACA	0.056	0.034	0.062	0.050	0.051	0.062	0.048
DB	0.056	0.055	0.039	0.051	0.051	0.062	0.047
BCS	0.055	0.054	0.063	0.032	0.051	0.061	0.047
ING	0.057	0.056	0.064	0.051	0.031	0.063	0.050
SAB	0.055	0.054	0.061	0.049	0.050	0.040	0.047
KBC	0.056	0.055	0.062	0.050	0.052	0.062	0.030
			$VaR_{5\%}^{i}$ and $CoVaR_{5\%}^{j|i}$—GaussSpDCC				
ISP	0.035	0.057	0.064	0.050	0.052	0.063	0.048
ACA	0.056	0.035	0.064	0.050	0.052	0.062	0.048
DB	0.056	0.056	0.040	0.051	0.052	0.061	0.047
BCS	0.055	0.056	0.064	0.032	0.051	0.060	0.047
ING	0.058	0.058	0.065	0.051	0.032	0.063	0.050
SAB	0.056	0.056	0.062	0.049	0.051	0.039	0.047
KBC	0.057	0.057	0.063	0.050	0.053	0.061	0.030
			$VaR_{5\%}^{i}$ and $CoVaR_{5\%}^{j|i}$—tDCC				
ISP	0.033	0.088	0.100	0.078	0.083	0.104	0.078
ACA	0.087	0.033	0.099	0.079	0.083	0.102	0.077
DB	0.087	0.087	0.038	0.081	0.083	0.101	0.075
BCS	0.085	0.085	0.101	0.031	0.082	0.100	0.075
ING	0.089	0.089	0.102	0.082	0.031	0.105	0.078
SAB	0.087	0.085	0.097	0.077	0.082	0.040	0.075
KBC	0.088	0.088	0.099	0.079	0.083	0.102	0.029
			$VaR_{5\%}^{i}$ and $CoVaR_{5\%}^{j|i}$—tSpDCC				
ISP	0.033	0.087	0.098	0.075	0.081	0.100	0.075
ACA	0.085	0.034	0.097	0.076	0.081	0.098	0.075
DB	0.085	0.086	0.038	0.077	0.081	0.097	0.073
BCS	0.083	0.085	0.098	0.030	0.081	0.096	0.073
ING	0.087	0.088	0.100	0.078	0.031	0.100	0.076
SAB	0.085	0.084	0.095	0.074	0.080	0.039	0.073
KBC	0.086	0.087	0.097	0.076	0.081	0.098	0.029

We observe that, as expected, the CoVaR is always greater in absolute values than the VaR figures on the diagonal. Furthermore, we see that the estimates of the Value at Risk are similar across the five considered models. On the contrary, the estimates of CoVaR are significantly larger for the DCC-GARCH models based on the Student's t distribution,

suggesting that the Gaussian model may potentially underestimate the risk of joint distress and risk spillover, as expected by the stylized fact of fat tails and high tail correlations in financial time series (see, e.g., Cont 2001). Finally, we see that the estimates of CoVaR of the tSpDCC model are slightly smaller than the tDCC model. The FHS estimates yield similar results to the other models in terms of VaR, and lie in the middle between the Gaussian and Student's t models in terms of CoVaR. Figure 3 shows the out-of-sample equity log returns and the estimate of the Value at Risk. We see that the dynamics are similar for all the models, and that the two models with a spatial component share some similar dynamics in specific time periods.
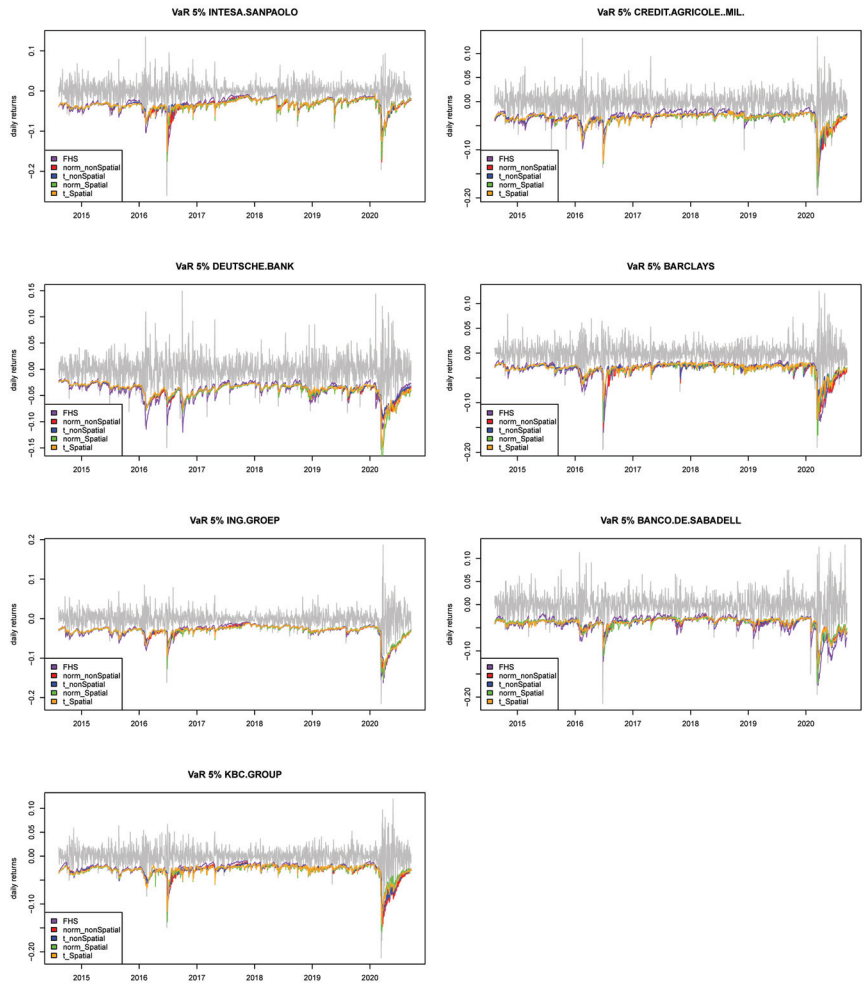
Figure 3. Out-of-sample equity log-returns of the seven banks considered in the study and VaR estimates.

Figure 4 shows the $VaR_{5\%}^{j}$ of each bank j together with the average $CoVaR_{5\%}^{j|i}$ for $i \neq j$ (for brevity we report only the results estimated using the tSpDCC model). We see that for all the banks, the CoVaR is always higher than the VaR. Focusing on the dynamics, we see two main spikes in the series that affected all banks: one in mid-2016 (corresponding to the Brexit Referendum), and one when the COVID-19 crisis started in March 2020. The former

shock was short-lived, and risk measures returned to normal levels quickly, while the Covid crisis had more long-lasting effects, with CoVaR decreasing slowly in the following months (although with differences across banks, for instance, Intesa San Paolo risk levels returned to a normal level quicker than ING Group). In other periods, the dynamics of CoVaR are diversified across banks, with some institutions (in particular Intesa San Paolo) characterized by several spikes (likely related to idiosyncratic or regional shocks) while other institutions such as Credit Agricole or Barclays characterized by a more stable risk profile.
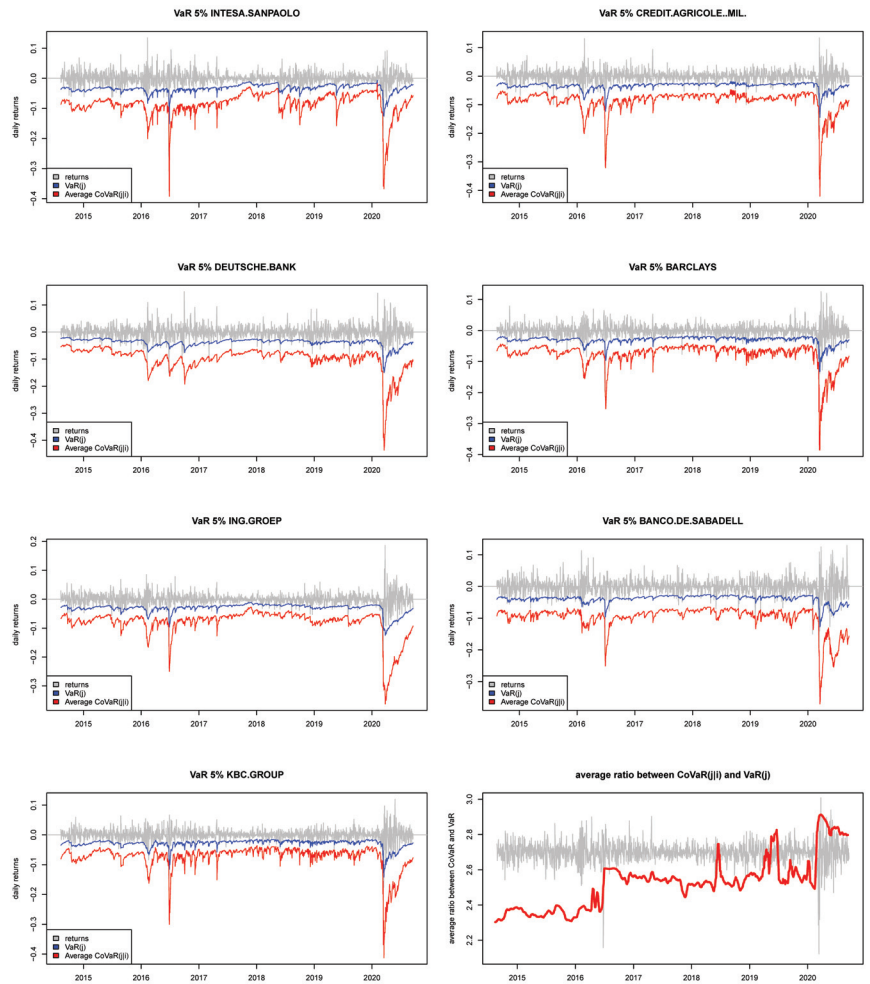
Figure 4. Out-of-sample equity log-returns, VaR estimated using the tSpDCC model (blue lines), and the average of the CoVaRs for each bank (red lines). The last panel reports the average ratio between the CoVaR and the VaR (moving average over 10 days), and the (rescaled) average log-return of the seven banks for reference.

The last panel of Figure 4 shows the average ratio between $CoVaR_{5\%}^{j|i}$ and $VaR_{5\%}^{j}$ for each couple of institutions, and the rescaled average log-returns of the seven banks included as reference. We see that the shocks associated with Brexit and Covid had the effect of increasing the ratio, thus increasing the risk spillover in the system. The effect

is persistent, as the ratio remained higher for the periods after the shocks, despite the levels of risks reduced more quickly after the event, suggesting the presence of long-lasting spillover effects.

3.2.2. Backtesting Results

We now present two backtesting analyses in order to assess the quality of the estimation of VaR and CoVaR based on different models. First, we report the results for the backtesting of $VaR_{0.05}^i$ with the Kupiec (unconditional) and Christoffersen (conditional) coverage tests. Table 6 provides the p-value of the unconditional coverage (UC) and conditional coverage (CC) tests. We see that, regarding the UC test, in most cases, the null hypothesis of correct exceedances is not rejected at the 95% confidence level, meaning that the models identify correctly the expected number of exceedances. The spatial component, according to this statistic does not provide benefits, leading typically to lower p-values and higher rejection rates. Concerning the CC test, the null hypothesis of correct and independent exceedances is rejected in a higher number of cases, indicating the presence of some residual clustering of the exceedances. Summing up, according to the CC tests, the introduction of the spatial component improves the estimation of Value at Risk, leading to higher p-values and lower rejection rates. The FHS estimation has good performance in terms of the UC test, while it is the worst one considered in terms of the CC test (the null hypothesis is rejected for all banks except for Banco de Sabadell).

Table 6. p-Values of the unconditional (Panel A) and conditional (Panel B) for $VaR_{5\%}^i$. The null hypotheses are "Correct Exceedances" (unconditional test) and "correct & independent exceedances" (conditional test). p-values lower than 5% are highlighted in bold.

Panel A: Unconditional Coverage Test for $VaR_{5\%}^i$, p-Value.					
Bank	**FHS**	**GaussDCC**	**GaussSpDCC**	**tDCC**	**tSpDCC**
ISP	0.700	0.113	0.067	0.459	0.113
ACA	0.756	0.143	**0.004**	0.534	**0.037**
DB	0.534	0.670	0.589	0.270	0.122
BCS	0.670	0.789	0.327	0.756	0.589
ING	0.513	0.534	0.390	0.880	0.789
SAB	0.379	0.935	0.589	0.844	0.756
KBC	0.224	0.180	0.272	0.615	0.700
Panel B: Conditional Coverage Test for $VaR_{5\%}^i$, p-Value.					
Bank	**FHS**	**GaussDCC**	**GaussSpDCC**	**tDCC**	**tSpDCC**
ISP	**0.001**	**0.002**	0.125	**0.017**	0.119
ACA	**0.033**	**0.002**	**0.017**	**0.007**	0.069
DB	**0.022**	**0.013**	**0.042**	**0.008**	**0.009**
BCS	**0.038**	0.255	0.366	0.200	0.413
ING	**0.001**	**0.002**	**0.012**	**0.001**	**0.013**
SAB	0.455	0.997	0.655	0.882	0.774
KBC	**0.023**	0.095	0.156	0.079	**0.035**

Table 7 reports a summary of the UC and CC tests for the CoVaR. Panels A and B show the average number of exceedances, the average p-value across all the combinations of assets, and the percentage of p-values that are smaller than 5% (i.e., the null hypothesis is rejected at the 95% confidence level) for the UC test and CC test, respectively.[7] We see that for the GaussDCC and GaussSpDCC based on the Gaussian distribution, the null hypotheses for both the UC and CC tests are rejected for all the bilateral CoVaRs. On the contrary, for the Student's t models the null hypotheses are never rejected for both the UC and CC tests. Finally, we observe that the spatial model tSpDCC performances are aligned with those of tDCC. The results of the FHS are slightly better than the Gaussian DCC-GARCH models but are worse than the Student's t DCC-GARCH models. The results, although they do not highlight relevant differences between the spatial and non-spatial models, do confirm that the models based on the Gaussian distribution are not suitable to properly measure spillover risk, while Student's t models have much better results when compared to FHS estimation of CoVaR.

Table 7. Average exceedances, *p*-values, and percentage of *p*-values smaller than 0.05 of the unconditional (Panel A) and conditional (Panel B) for $CoVaR_{5\%}^{j|i}$. The null hypotheses are "Correct Exceedances" (unconditional test) and "correct & independent exceedances" (conditional test). Averages and percentages are computed across all the couples of assets. For $q = 5\%$ the expected number of exceedances is 3.9.

Panel A: Unconditional Coverage Test for $CoVaR_{5\%}^{j|i}$.					
	FHS	**GaussDCC**	**GaussSpDCC**	**tDCC**	**tSpDCC**
Average exceedances	12.0	18.9	17.1	4.4	5.0
Average *p*-value	0.047	0.000	0.000	0.639	0.472
(% *p*-values < 0.05)	(88.1%)	(100%)	(100%)	(0%)	(0%)
Panel B: Conditional Coverage Test for $CoVaR_{5\%}^{j|i}$.					
	FHS	**GaussDCC**	**GaussSpDCC**	**tDCC**	**tSpDCC**
Average *p*-value	0.061	0.000	0.000	0.446	0.508
(% *p*-values < 0.05)	(81.0%)	(100%)	(100%)	(0%)	(0%)

Finally, we assess the quality of the estimation of VaR and CoVaR using the loss function methodology described in Section 2.2.4. Table 8 reports the value of the loss function for the investor and the regulator (lower values are better). We see that, consistently with the results for CC and UC tests, the introduction of the spatial component has a positive effect for both the models based on Gaussian and Student's t distribution. Contrary to expectations, the Student's t models do not outperform the Gaussian models according to this metric, showing similar or slightly worse performance. The similarity between Gaussian and Student's t distribution may be due to the fact that we are considering a low confidence level (95% for $q = 0.05$): for a higher confidence level (e.g., 99%, $q = 0.01$) the shape of the tails may matter more, and the Student's t model may perform better. In this analysis, we do not consider higher confidence levels as the backtesting of the CoVaR would become impossible.

Table 8. Backtesting based on loss function for $VaR_{5\%}^i$. Panel A reports the investor's point of view and Panel B reports the regulator's point of view (see Section 2.2.4). For each line the best value is highlighted in bold.

Panel A: Backtesting Based on Loss Function. $VaR_{5\%}^i$ (Investor's Point of View).					
Bank	**FHS**	**GaussDCC**	**GaussSpDCC**	**tDCC**	**tSpDCC**
ISP	4.11	4.06	3.97	4.08	**3.96**
ACA	4.21	4.24	**4.14**	4.24	4.18
DB	**4.70**	4.74	4.76	4.71	4.77
BCS	3.95	3.87	**3.75**	3.91	3.80
ING	3.74	3.74	**3.70**	3.77	3.72
SAB	4.92	4.85	**4.83**	4.89	4.87
KBC	3.54	3.61	3.48	3.55	**3.47**
mean	4.17	4.16	**4.09**	4.16	4.11
Panel B: Backtesting Based on Loss Function. $VaR_{5\%}^i$ (Regulator's Point of View).					
Bank	**FHS**	**GaussDCC**	**GaussSpDCC**	**tDCC**	**tSpDCC**
ISP	1.24	1.15	**1.05**	1.29	1.16
ACA	1.44	1.35	**1.17**	1.44	1.33
DB	**1.34**	1.47	1.44	1.56	1.59
BCS	1.15	1.21	**1.11**	1.35	1.28
ING	1.06	1.11	**1.05**	1.19	1.12
SAB	**1.45**	1.58	1.60	1.63	1.64
KBC	1.14	1.09	**0.97**	1.10	1.02
mean	1.26	1.28	**1.20**	1.37	1.31

Looking at Table 9, we see the average value of the loss functions for the CoVaR estimations. The results highlight once again a beneficial effect of the spatial component, that leads to improvements regardless of the distribution and the loss function used. We

also see that the average loss is clearly smaller for the Student's t models, highlighting their ability to better estimate the spillover risk compared to the Gaussian models. The results for the FHS show that such a model is not as accurate as tSpDCC and tDCC, although it performs better than Gaussian DCC-GARCH models.

Overall, the results suggest that the introduction of a spatial component improves the performance of the DCC-GARCH model, both in terms of the information criteria, and in terms of the out-of-sample estimation of CoVaR, as confirmed by the backtesting. The results also confirm the better fit of the Student's t models compared to the Gaussian models.

Table 9. Backtesting based on loss function for $CoVaR_{5\%}^{j|i}$. The null hypotheses are "Correct Exceedances" (unconditional test) and "correct & independent exceedances" (conditional test). Panel A reports the investor's point of view and Panel B reports the regulator's point of view (see Section 2.2.4). For each line, the best value is highlighted in bold.

Panel A: Backtesting Based on Loss Function. $CoVaR_{5\%}^{j|i}$ (Investor's Point of View).					
	FHS	**GaussDCC**	**GaussSpDCC**	**tDCC**	**tSpDCC**
average loss	0.392	0.465	0.414	0.393	**0.371**

Panel B: Backtesting Based on Loss Function. $CoVaR_{5\%}^{j|i}$ (Regulator's Point of View).					
	FHS	**GaussDCC**	**GaussSpDCC**	**tDCC**	**tSpDCC**
average loss	0.233	0.359	0.307	0.153	**0.142**

4. Discussion and Conclusions

As the spillover effects of risk become a problem in interconnected banking systems, this study introduces a spatial DCC-GARCH(1,1) to provide a more accurate measurement of joint tail risk in a parsimonious way. The model aims to improve the standard DCC-GARCH model (Engle 2002), without introducing the estimation complexity typical of VEC GARCH and BEKK model (Bollerslev et al. 1988; Engle and Kroner 1995; Ling and McAleer 2003). After discussing the estimation of the model, we perform an empirical analysis on the equity log returns of seven large European banks, using a matrix that reflects the similarity in credit structure and exposures. We compare four multivariate GARCH specifications (with and without spatial components, and with two alternative distributions for the innovations). The comparison of information criteria suggests that the proposed spatial model with Student's t innovation provides a better fit compared to the alternatives.

A common approach to measuring the spillover risk is the usage of pairwise CoVaR Adrian and Brunnermeier (2014), which measures the tail risk of an institution conditional to the distress of another institution. We estimate pairwise CoVaRs using our spatial DCC-GARCH(1,1) model. Compared to other GARCH-based estimation procedures for CoVaR (see, e.g., Girardi and Ergün 2013), our framework allows us to consider a time-varying correlation matrix thanks to the DCC component, and a network dimension thanks to the spatial component. Indeed, as highlighted in the literature, the network component of risk is more and more relevant (see, e.g., Billio et al. 2012; Diebold and Yılmaz 2014), and spatial GARCH models allow us to include it while maintaining the estimation feasible.

To test the reliability of the VaR and pairwise CoVaR, we first examine their accuracy via the UC and CC tests considering different GARCH specifications and a non-parametric FHS approach. The results show that the Student's t spatial DCC-GARCH(1,1) model (tSpDCC) provides the lowest rejection rate for $CoVaR_{5\%}$ compared to other models. Second, we investigate the models via backtesting based on loss functions. The analysis confirms that the Student's t spatial DCC GARCH(1,1) model outperforms the other DCC-GARCH specifications, as well as the filtered historical simulations model in terms of the estimation of $CoVaR_{5\%}$. Overall, from a methodological perspective, we conclude that the multivariate GARCH model with the Student's t and the spatial component obtained thanks to the proposed similarity matrix can improve the assessment of credit risk profiles and the credit risk market's spillover.

Concerning the economic analysis and interpretation of the empirical results, the spillover analysis shows that the dynamics of CoVaR were diversified across European banks, and that in the out-of-sample period (2014–2020) there were two main shocks common to all the institutions: the Brexit Referendum (mid-2016) and the COVID-19 Crisis (started in the first half of 2020). Both periods were associated with spikes in VaR and CoVaR, and persistent increases in the ratio of CoVaR over VaR (denoting therefore a long-lasting increase in the interconnectedness and risk spillovers). The persistent increase in CoVaRs after Brexit is consistent with Li (2020), which studies the behavior of European stock markets in a multivariate time-varying setting, finding that market co-volatility continues to be substantial and persists after Brexit despite the fact that the market adjusted quickly to the shock. Similarly, the increase of CoVaR during the Covid Crisis confirms results from the literature that show how spillover and interconnectedness increased in the first part of the Covid period. In particular, Aslam et al. (2021) studied twelve European markets using the methodology from Diebold and Yılmaz (2014) on high-frequency data and found more stable spillovers in the Covid period compared to the previous, and Foglia et al. (2022) show an increase of volatility connectedness during the Covid period across 30 major Eurozone banks.

Finally, we point out that the proposed framework not only has relevant applications for financial regulators, but it is also relevant for the asset management industry. Indeed, the proposed model allows us to estimate the joint distribution of future returns, providing asset managers with reliable inputs for optimal portfolio strategies (Meucci 2005), and risk managers with data useful to measure the risk of investment funds and to conduct stress-testing analyses based on hypothetical scenarios (see, e.g., Alexander and Sheedy 2008; Koliai 2016).

Future works may further extend the current work by testing alternative spatial matrices and applying the model to other markets and larger datasets.

Data Availability Statement: Restrictions apply to the availability of these data. Data were obtained from the financial data provider Refinitiv Eikon and are available with the permission of Refinitiv.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A

Table A1. Estimated parameters of Student's t spatial DCC-GARCH(1,1) model (September 2010–September 2014).

$$
\begin{bmatrix} h_{1,t} \\ h_{2,t} \\ h_{3,t} \\ h_{4,t} \\ h_{5,t} \\ h_{6,t} \\ h_{7,t} \end{bmatrix}
=
\begin{bmatrix}
\mathbf{1.39\times10^{-5}}\;(0.000) \\
\mathbf{8.31\times10^{-6}}\;(0.000) \\
2.22\times10^{-16}\;(1.000) \\
4.37\times10^{-6}\;(0.460) \\
4.92\times10^{-6}\;(0.847) \\
3.49\times10^{-5}\;(0.523) \\
6.07\times10^{-6}\;(0.341)
\end{bmatrix}
$$

$$
+\;
\operatorname{diag}\!\begin{bmatrix}
\mathbf{0.033}\,(0.008) \\ 0.019\,(0.110) \\ \mathbf{0.032}\,(0.045) \\ 0.015\,(0.545) \\ 0.039\,(0.829) \\ 0.078\,(0.250) \\ \mathbf{-0.047}\,(0.015)
\end{bmatrix}
\;+\;
\operatorname{diag}\!\begin{bmatrix}
\mathbf{0.067}\,(0.026) \\ \mathbf{0.087}\,(0.000) \\ 0.009\,(0.500) \\ \mathbf{0.061}\,(0.006) \\ 0.041\,(0.565) \\ -0.013\,(0.534) \\ 0.035\,(0.241)
\end{bmatrix}
\mathbf{W}
\;\underbrace{\begin{bmatrix} r^2_{1,t-1} \\ r^2_{2,t-1} \\ r^2_{3,t-1} \\ r^2_{4,t-1} \\ r^2_{5,t-1} \\ r^2_{6,t-1} \\ r^2_{7,t-1} \end{bmatrix}}
$$

$$
+\;
\operatorname{diag}\!\begin{bmatrix}
\mathbf{0.936}\,(0.000) \\ \mathbf{0.951}\,(0.000) \\ \mathbf{0.927}\,(0.000) \\ \mathbf{0.945}\,(0.000) \\ \mathbf{0.927}\,(0.000) \\ \mathbf{0.871}\,(0.000) \\ \mathbf{0.918}\,(0.000)
\end{bmatrix}
\;+\;
\operatorname{diag}\!\begin{bmatrix}
\mathbf{-0.061}\,(0.000) \\ 0.016\,(0.258) \\ -0.031\,(0.326) \\ -0.016\,(0.951) \\ 0.017\,(0.494) \\ 0 \\ -0.006\,(0.924)
\end{bmatrix}
\mathbf{W}
\;\begin{bmatrix} h_{1,t-1} \\ h_{2,t-1} \\ h_{3,t-1} \\ h_{4,t-1} \\ h_{5,t-1} \\ h_{6,t-1} \\ h_{7,t-1} \end{bmatrix}
$$

$$
\begin{bmatrix}
Q_{1,1,t} & \cdots & Q_{1,7,t} \\
Q_{2,1,t} & \cdots & Q_{2,7,t} \\
\vdots & \ddots & \vdots \\
Q_{7,1,t} & \cdots & Q_{7,7,t}
\end{bmatrix}
=
\left(1 - \underset{(0.000)}{\mathbf{0.008}} - \underset{(0.000)}{\mathbf{0.976}} \right) \bar{Q}
+ \underset{(0.000)}{\mathbf{0.008}}
\begin{bmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \\ \epsilon_{3,t-1} \\ \epsilon_{4,t-1} \\ \epsilon_{5,t-1} \\ \epsilon_{6,t-1} \\ \epsilon_{7,t-1} \end{bmatrix}
\begin{bmatrix} \epsilon_{1,t-1} \\ \epsilon_{2,t-1} \\ \epsilon_{3,t-1} \\ \epsilon_{4,t-1} \\ \epsilon_{5,t-1} \\ \epsilon_{6,t-1} \\ \epsilon_{7,t-1} \end{bmatrix}'
+ \underset{(0.000)}{\mathbf{0.976}}
\begin{bmatrix}
Q_{1,1,t-1} & \cdots & Q_{1,7,t-1} \\
Q_{2,1,t-1} & \cdots & Q_{2,7,t-1} \\
\vdots & \ddots & \vdots \\
Q_{7,1,t-1} & \cdots & Q_{7,7,t-1}
\end{bmatrix}
$$

$$
v = \underset{(0.000)}{\mathbf{7.181}}
$$

Parameters of the Student-t Spatial DCC-GARCH(1,1) model estimated on the first rolling window (September 2010–September 2014). *p*-value in brackets, values significant at 95% confidence level in bold.

Notes

[1]   On the contrary the return are not typically serially correlated, consistently with the efficient market hypothesis Fama (1970).

[2]   We use the convention that a higher value of *r* is preferable to a lower value, as in the cases when *r* represents returns or wealth. Other works consider a random variable *x* such that lower values of *x* are preferable, as in the case of losses (e.g., Artzner et al. 1999). The signs in the definitions then need to be changed accordingly.

[3]   The construction of a network of bilateral CoVaR is inspired by Adrian and Brunnermeier (2014). An alternative approach to extend CoVaR to a network dimension is proposed by Torri et al. (2021) and uses penalized multivariate quantile regression.

[4]   Under both Student t and Gaussian distribution it can be shown that for positive correlation $CoVaR_q^{x|y}$ is always greater than VaR_q^x. Hence we consider a grid of values ranging from VaR_q^x to infinity, such that the corresponding quantiles are equally spaced. In case of null or negative correlations, we use 0 as the lower value of the grid.

[5]   Central governments Institutions, Corporates, Retail, Equity, Securitization, and Other non-credit obligation assets.

[6]   In preliminary analyses we tested different orders of GARCH and we also considered an ARMA component, finding no relevant differences. We omit them for brevity.

[7]   We do not report the complete results for all the bivariate CoVaR estimations for brevity. The results are available upon request.

References

Abad, Pilar, Sonia Benito, and Carmen López. 2014. A comprehensive review of value at risk methodologies. *The Spanish Review of Financial Economics* 12: 15–32. [CrossRef]

Adrian, Tobias, and Markus Brunnermeier. 2014. *CoVaR.* Staff Reports 348. New York: Federal Reserve Bank of New York. [CrossRef]

Alexander, Carol, and Elizabeth Sheedy. 2008. Developing a stress testing framework based on market risk models. *Journal of Banking & Finance* 32: 2220–36.

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath. 1999. Coherent measures of risk. *Mathematical Finance* 9: 203–28. [CrossRef]

Aslam, Faheem, Paulo Ferreira, Khurrum Shahzad Mughal, and Beenish Bashir. 2021. Intraday volatility spillovers among european financial markets during COVID-19. *International Journal of Financial Studies* 9: 5. [CrossRef]

Barone-Adesi, Giovanni, Kostas Giannopoulos, and Les Vosper. 2002. Backtesting derivative portfolios with filtered historical simulation (fhs). *European Financial Management* 8: 31–58. [CrossRef]

Billio, Monica, Massimiliano Caporin, Lorenzo Frattarolo, and Loriana Pelizzon. 2021. Networks in risk spillovers: A multivariate garch perspective. *Econometrics and Statistics, in press.* [CrossRef]

Billio, Monica, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon. 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* 104: 535–59. [CrossRef]

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]

Bollerslev, Tim. 1990. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model. *The Review of Economics and Statistics* 72: 498–505. [CrossRef]

Bollerslev, Tim, Robert F. Engle, and Jeffrey M. Wooldridge. 1988. A capital asset pricing model with time-varying covariances. *Journal of Political Economy* 96: 116–31. [CrossRef]

Borovkova, Svetlana, and Rik Lopuhaa. 2012. Spatial GARCH: A spatial approach to multivariate volatility modeling. *SSRN Electronic Journal.* [CrossRef]

Caporin, Massimiliano. 2008. Evaluating value-at-risk measures in the presence of long memory conditional volatility. *The Journal of Risk* 10: 79–110. [CrossRef]

Cesarone, Francesco, and Stefano Colucci. 2016. A quick tool to forecast value-at-risk using implied and realized volatilities. *Journal of Risk Model Validation* 10: 71–101. [CrossRef]

Chen, Xiurong. 2017. Impact effects and spatial volatility spillover effects of sovereign credit rating downgrades–empirical analysis of multivariate spatial-bekk-garch model based on symbolic transfer entropy. *Boletín Técnico* 55: 614–23.

Christoffersen, Peter, and Denis Pelletier. 2004. Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics* 2: 84–108. [CrossRef]

Conrad, Christian, and Menelaos Karanasos. 2010. Negative volatility spillovers in the unrestricted eccc-garch model. *Econometric Theory* 26: 838–62. [CrossRef]

Cont, Rama. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1: 223. [CrossRef]

Dhaene, Geert, Piet Sercu, and Jianbin Wu. 2022. Volatility spillovers: A sparse multivariate garch approach with an application to commodity markets. *Journal of Futures Markets* 42: 868–87. [CrossRef]

Diebold, Francis X., and Marc Nerlove. 1989. The dynamics of exchange rate volatility: A multivariate latent factor arch model. *Journal of Applied Econometrics* 4: 1–21. [CrossRef]

Diebold, Francis X., and Kamil Yılmaz. 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics* 182: 119–34. [CrossRef]

EBA. 2021. *2021 EU-Wide Stress Test—Methodological Note*. Technical Report. La Défense: European Banking Authority. Available online: https://www.eba.europa.eu/sites/default/documents/files/document_library/RiskAnalysisandData/EU-wideStressTesting/2021/LaunchoftheST/962559/2021EU-widestresstest-MethodologicalNote.pdf (accessed on 3 September 2023).

Engle, Robert. 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* 20: 339–50.

Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50: 987–1007. [CrossRef]

Engle, Robert F., and Kenneth F. Kroner. 1995. Multivariate simultaneous generalized arch. *Econometric Theory* 11: 122–50. [CrossRef]

Engle, Robert F., and Kevin Sheppard. 2001. *Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate Garch*. Working Paper 8554. Cambridge: National Bureau of Economic Research. [CrossRef]

Engle, Robert F., Victor K. Ng, and Michael Rothschild. 1990. Asset pricing with a factor-arch covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics* 45: 213–37. [CrossRef]

Fama, Eugene F. 1970. Efficient capital markets: A review of theory and empirical work. *The Journal of Finance* 25: 383–417. [CrossRef]

Foglia, Matteo, Abdelhamid Addi, and Eliana Angelini. 2022. The eurozone banking sector in the time of COVID-19: Measuring volatility connectedness. *Global Finance Journal* 51: 100677. [CrossRef]

Giacometti, Rosella, Gabriele Torri, and Sandra Paterlini. 2021. Tail risks in large portfolio selection: Penalized quantile and expectile minimum deviation models. *Quantitative Finance* 21: 243–61. [CrossRef]

Giannopoulos, Kostas, and Radu Tunaru. 2005. Coherent risk measures under filtered historical simulation. *Journal of Banking & Finance* 29: 979–96.

Girardi, Giulio, and A. Tolga Ergün. 2013. Systemic risk measurement: Multivariate garch estimation of CoVaR. *Journal of Banking & Finance* 37: 3169–80. [CrossRef]

Gurrola-Perez, Pedro, and David Murphy. 2015. *Filtered Historical Simulation Value-at-Risk Models and Their Competitors*. London: Bank of England.

Jeantheau, Thierry. 1998. Strong consistency of estimators for multivariate arch models. *Econometric Theory* 14: 70–86. [CrossRef]

Jennrich, Robert I. 1970. An asymptotic χ^2 test for the equality of two correlation matrices. *Journal of the American Statistical Association* 65: 904–12.

Keiler, Sebastian, and Armin Eder. 2013. *CDS Spreads and Systemic Risk: A Spatial Econometric Approach*. Discussion Papers 01/2013. Frankfurt: Deutsche Bundesbank. [CrossRef]

Koenker, Roger, and Gilbert Bassett, Jr. 1978. Regression quantiles. *Econometrica: Journal of the Econometric Society* 46: 33–50. [CrossRef]

Koliai, Lyes. 2016. Extreme risk modeling: An evt–pair-copulas approach for financial stress tests. *Journal of Banking & Finance* 70: 1–22.

Kupiec, Paul H. 1995. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* 3: 73–84. [CrossRef]

Li, Hong. 2020. Volatility spillovers across european stock markets under the uncertainty of brexit. *Economic Modelling* 84: 1–12. [CrossRef]

Ling, Shiqing, and Michael McAleer. 2003. Asymptotic theory for a vector arma-garch model. *Econometric Theory* 19: 280–310. [CrossRef]

Lopez, Jose Antonio. 1999. *Methods for Evaluating Value-at-Risk Estimates*. Economic Policy Review. San Francisco: Federal Reserve Bank of San Francisco, pp. 3–17. [CrossRef]

Meucci, Attilio. 2005. *Risk and Asset Allocation*. Berlin: Springer, vol. 1.

Rahman, Dima. 2014. Are banking systems increasingly fragile? Investigating financial institutions? CDS returns extreme co-movements. *Quantitative Finance* 14: 805–30. [CrossRef]

Rockafellar, R. Tyrrell, and Stan Uryasev. 2013. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science* 18: 33–53. [CrossRef]

Torri, Gabriele, Rosella Giacometti, and Tomáš Tichỳ. 2021. Network tail risk estimation in the european banking system. *Journal of Economic Dynamics and Control* 127: 104125. [CrossRef]

Zhang, Wei-Guo, Guo-Li Mo, Fang Liu, and Yong-Jun Liu. 2018. Value-at-risk forecasts by dynamic spatial panel GJR-GARCH model for international stock indices portfolio. *Soft Computing* 22: 5279–97. [CrossRef]

Article

Tensors Associated with Mean Quadratic Differences Explaining the Riskiness of Portfolios of Financial Assets

Pierpaolo Angelini [1,*] and Fabrizio Maturo [2]

[1] Dipartimento di Scienze Statistiche, Università La Sapienza, 00185 Rome, Italy
[2] Faculty of Economics, Universitas Mercatorum, 00186 Rome, Italy; fabrizio.maturo@unimercatorum.it
[*] Correspondence: pier.angelini@uniroma1.it

Abstract: Bound choices such as portfolio choices are studied in an aggregate fashion using an extension of the notion of barycenter of masses. This paper answers the question of whether such an extension is a natural fashion of studying bound choices or not. Given *n* risky assets, the question of why it is appropriate to treat only two risky assets at a time inside the budget set of the decision-maker is handled in this paper. Two risky assets are two goods. They are two marginal goods. The question of why they always give rise to a joint good inside the budget set of the decision-maker is addressed by this research work. A single risky asset is viewed as a double one using four nonparametric joint distributions of probability. The variability of a joint distribution of probability always depends on the state of information and knowledge associated with a given decision-maker. For this reason, two variability tensors are defined to identify the riskiness of the same risky asset. A multilinear version of the Sharpe ratio is shown. It is based on tensors. After computing the expected return on an *n*-risky asset portfolio, its riskiness is obtained using mean quadratic differences developed through tensors.

1. Introduction

In this paper, bound choices such as portfolio choices are studied without adding new axiomatic constructions or using known ones. Previous studies tend to add or use such formalistically abstract constructions (Echenique 2020; Chambers et al. 2017; Nishimura et al. 2017). Such constructions are exact (Halevy et al. 2018). Nevertheless, in our opinion, they are empty, and this characteristic is perhaps inevitable. Conversely, we propose an approach of an operational nature based on metric measures (choices being made by a given decision-maker and expressed by means of specific measures put forward by Corrado Gini are dealt with by Wang et al. (2018)). The advantages of this approach are essentially two. First, such measures indicating rational choices can be introduced without a problem. This is because they comply with any reasonable axiomatic construction (Cassese et al. 2020). Second, such measures are in accordance with one of the fundamental needs of science, which must work with notions of ascertained validity in a pragmatic sense. In our opinion, science must not take combinations of axioms as indefectible concepts, but it must be based on actual experiences, which are at least conceptually possible. Such experiences are subjected to a measure. A remarkable point of this research work is the following. Such a point is connected with how a measure can be obtained. In our opinion, bound choices must be studied under conditions of uncertainty and riskiness (Angelini and Maturo 2021b). They are real and unavoidable conditions (Chudjakow and Riedel 2013; Machina 1987). It follows that we focus on the notion of probability and its properties. This notion is intrinsically subjective (Pfanzagl 1967). A theorem enunciating the notion of utility to be a metric measure is shown by us. Hence, prevision (probability) and utility are two metric

measures (that which has been made by Viscusi and Evans (2006) and Abdellaoui et al. (2013) is enlarged by us in this paper). Prevision (probability) and utility are innovatively discovered to be two sides of the same coin.

In this study, probability judgments always depend on the state of information and knowledge associated with a given individual. We focus on a specific interpretation referred to Bayes' theorem. It is essential to our purposes to explain why we focus on the following geometric interpretation referred to Bayes' theorem, where two stages are distinguished. Every bound choice is intrinsically a barycenter of masses subjectively distributed over a finite set of alternatives. In the first stage, all the barycenters of masses are considered. Their number is infinite. All of them give rise to a convex set. It is the budget set of the decision-maker. In the second stage, one of the barycenters is chosen, so a probabilistic but convergent element is associated with a rational choice.

This paper fills a conceptual and mathematical gap existing in the current literature. It is possible to enlarge the notion of rational behavior. It follows that the optimization principle can be enlarged as well. How is this possible? Given two goods denoted by $_1X$ and $_2X$, their possible values meant as pure numbers are expressed by the sets $I(_1X)$ and $I(_2X)$. A given decision-maker chooses $\mathbf{P}(_1X)$ and $\mathbf{P}(_2X)$ inside his or her budget set. This means that a given decision-maker is indifferent to the exchange of $_1X$, identified with $I(_1X)$, for $\mathbf{P}(_1X)$ and of $_2X$, identified with $I(_2X)$, for $\mathbf{P}(_2X)$. Also, this implies that he or she is indifferent to the exchange of X_{12} for $\mathbf{P}(X_{12})$, where $\mathbf{P}(X_{12})$ extends the notion of barycenter of masses. This is because $\mathbf{P}(X_{12})$ is the determinant based on four measures denoted by $\mathbf{P}(_1X_1X)$, $\mathbf{P}(_1X_2X)$, $\mathbf{P}(_2X_1X)$, and $\mathbf{P}(_2X_2X)$. This means that four nonparametric joint distributions of probability are considered. X_{12} is a multiple good of order 2, whose elements are $_1X$ and $_2X$. The possible values for X_{12} coincide with the components of a tensor. $\mathbf{P}(X_{12})$ is a multiple choice associated with a multiple good. What will be said in this paper is more general than one might think at first. This is because the mathematical notion of α-product on which $\mathbf{P}(_1X_1X)$, $\mathbf{P}(_1X_2X)$, $\mathbf{P}(_2X_1X)$, and $\mathbf{P}(_2X_2X)$ are based is discovered. Such a notion uses subjective probabilities intrinsically connected with exchangeable or symmetric events. Such a notion does not only explain bound choices but also can treat multilinear issues of statistical inference. This makes explicit where the results of this paper can be applied. Every bound choice is studied using subjective tools, probability, and utility, inside a subset of a linear space over $\mathbb{R}$. Linear spaces over $\mathbb{R}$ with a different dimension are here handled. A specific element is held fixed: possible and objective alternatives whose number is finite are always summarized. Possible and objective alternatives are real data. They can be viewed as sampling data. This makes explicit how the results of this paper can be applied. It is then possible to find out a strict connection between how bound choices are dealt with within this context and the least-squares model (as an alternative, a connection between economics and mathematics based on differential equations could be developed by examining Oderinu et al. (2023) as well).

1.1. Bound Choices Made by the Decision-Maker under Claimed Conditions of Certainty

We do not study more than two goods at a time inside the budget set of the decision-maker. This is because we use mathematical methods via a quadratic metric. Every bound choice being made by a given decision-maker inside his or her budget set is a measure obtained using a quadratic metric. It is not convenient to use a non-quadratic metric. For instance, in statistics, variance, standard deviation, and the covariance of two variables are indices obtained using a quadratic metric. It is certainly possible to study n goods, with $n > 2$ which is an integer. Nevertheless, whenever we want to obtain a measure, it is not convenient to study more than two goods at a time. Another remarkable issue developed in this paper is the following. Conceptually, the conditions of certainty referred to nonrandom goods[1] have to be understood as intrinsically fictitious. Given two nonrandom goods with downward-sloping demand curves, that which is chosen by a given decision-maker is denoted by (x_1, x_2) (primordial and fundamental aspects about revealed preference theory studying bound choices are dealt with by Samuelson (1948)). In our opinion, the objects

of decision-maker choice are explicitly bilinear and disaggregate measures (an analysis based on the two-good assumption is made by Cherchye et al. (2018)). Hence, there exists a one-to-one correspondence between two-dimensional points of the budget set of the decision-maker and bilinear and disaggregate measures. Each measure is decomposed into two linear measures. Each of them is a one-dimensional point. We establish the following:

Definition 1. *Given two nonrandom goods with downward-sloping demand curves, that which is demanded for each of them under claimed conditions of certainty by a given decision-maker is an average quantity. We write*

$$x_1 = x_1^1 \, p_1^1 + \ldots + x_1^m \, p_1^m \tag{1}$$

and

$$x_2 = x_2^1 \, p_2^1 + \ldots + x_2^m \, p_2^m, \tag{2}$$

where $\{p_1^i\}$ *and* $\{p_2^j\}$ *are two sets of m nonnegative masses. Their sum is always equal to* 1 *with regard to each of them. Each mass of them is always between* 0 *and* 1, *endpoints included. The possible quantities which can be demanded for good* 1 *are expressed by* $\{x_1^1, \ldots, x_1^m\}$, *whereas the possible quantities which can be demanded for good* 2 *are given by* $\{x_2^1, \ldots, x_2^m\}$.

The possible quantities which can be demanded for good 1 and good 2 are pure numbers. They are possible events. Their nature is objective. They are not directly observed, but they are estimated. What is directly observed is given by (x_1, x_2). The possible quantities which can be demanded for good 1 and good 2 are the components of two vectors of E^m, where E^m is an m-dimensional linear space over $\mathbb{R}$ with a Euclidean structure. Given an orthonormal basis of E^m, any vector whatsoever of E^m is always expressed as a linear combination of basis vectors. The real coefficients of this linear combination are its components. One and only one set of components of a vector of E^m uniquely identifies it. In this paper, good 1 and good 2 are jointly considered, so the weighted average of m^2 possible quantities which can be demanded for good 1 and good 2 is also studied. Such quantities are obtained by taking the Cartesian product given by $\{x_1^1, \ldots, x_1^m\} \times \{x_2^1, \ldots, x_2^m\}$ into account. Such quantities are the components of an affine tensor. The notion of event is always subdivisible, so m^2 possible alternatives can be studied. This means that a nonparametric joint distribution of probability is dealt with[2]. Every weighted average of m possible quantities which can be demanded for good 1 and good 2 is always found between the lowest possible quantity and the highest possible one (the rationality of the behaviors associated with decision-makers viewed to be as consumers is dealt with by Varian (1983)). The same is true regarding $m \times m = m^2$ possible quantities (that which is demanded by a given decision-maker being faced with his or her budget constraint is studied by Varian (1982)). All coherent weighted averages of m^2 possible alternatives identify a two-dimensional convex set. It is a continuous subset of $\mathbb{R} \times \mathbb{R}$. It is the budget set of the decision-maker. Two one-dimensional convex sets coinciding with two closed line segments appear as well. They belong to two mutually orthogonal axes of a two-dimensional Cartesian coordinate system. Strictly speaking, we deal with two half-lines, where each of them extends indefinitely from zero toward positive real numbers before being restricted. At the first stage, all coherent weighted averages of m^2 possible alternatives are handled. Their number is infinite. All coherent weighted averages of m possible alternatives for good 1 and good 2 are also dealt with. Their number is infinite. At a second stage, (x_1, x_2) is chosen. This choice depends on further hypotheses of an empirical nature. Boundary points that are found on each axis of a two-dimensional Cartesian coordinate system identify degenerate averages[3]. The budget line identifying the budget set of the decision-maker is a hyperplane embedded in a two-dimensional Cartesian coordinate system. Its negative slope depends on the prices of good 1 and good 2. We write

$$b_1 \, x_1 + b_2 \, x_2 \leq b, \tag{3}$$

where b_1, b_2, and b are positive real numbers, to identify his or her budget constraint. The slope of (3) is given by $-\frac{b_1}{b_2}$. Its horizontal intercept is given by $\frac{b}{b_1}$, whereas its vertical one is given by $\frac{b}{b_2}$. A specific pair of known and objective prices is denoted by (b_1, b_2), whereas the objective amount of money the decision-maker has to spend is expressed by b. Conditions of certainty are fictitious. This is because actual situations are uncertain. In particular, variations in the total amount of money the decision-maker has to spend could happen. Also, risks of external origin determining variations in his or her income could occur as well. This means that if (b_1, b_2, b) represents the decision-maker budget, then b must be assumed of an uncertain nature at the time of choice. The same b can appear even when the state of information and knowledge associated with a given decision-maker is assumed to have become complete later. On the other hand, if there is no ignorance anymore because further information is later acquired, then it is also possible to observe a parallel shift outward or inward of the budget line. Given (x_1, x_2), the weighted average of m^2 possible alternatives is a summarized element of the Fréchet class. According to our approach, the decision-maker also chooses this summarized element in addition to $(x_1, x_2)^4$.

We establish the following:

Definition 2. *The set of all weighted averages of m^2 possible alternatives, with the same given marginal weighted averages of m possible quantities which can be demanded for good 1 and m possible quantities which can be demanded for good 2, constitutes the Fréchet class.*

We note the following:

Remark 1. *The possible quantities which can be demanded for each nonrandom good taken into account are possible points (pure numbers) belonging to sets whose elements are finite in number. By definition, a hyperplane embedded in a two-dimensional Cartesian coordinate system never separates a coherent summary of possible points from their sets. In other terms, the budget line never separates (x_1, x_2) from $\{x_1^1, \ldots, x_1^m\}$, $\{x_2^1, \ldots, x_2^m\}$, and $\{x_1^1, \ldots, x_1^m\} \times \{x_2^1, \ldots, x_2^m\}$. This characterizes the points of the convex set. The budget set of the decision-maker is a convex set.*

1.2. A Random Good: Logical and Probabilistic Aspects

Assets are goods that provide a flow of services over time. A flow of consumption services can be provided by assets. A flow of money that can be used to purchase consumption can also be provided by assets. Financial assets provide a monetary flow. For instance, the flow of services provided by financial assets can be the flow of interest payments. In this paper, we focus on the future return provided by financial assets under conditions of uncertainty and riskiness. This future or expected return must be estimated by a given individual with respect to observed returns in the past. One of the observed returns can be the actual return. Financial assets such as risky assets are studied under conditions of uncertainty and riskiness. They are random goods. A random good is a random quantity[5] viewed as a specification of what will be chosen in each different outcome of a random process. The different outcomes of a random process are different random events. A random good is intrinsically characterized by a nonparametric probability distribution consisting of a list of different outcomes and the probability associated with each outcome (Gilio and Sanfilippo (2014)). The decision-maker chooses a nonparametric probability distribution of obtaining different random events. We establish the following:

Definition 3. *Let $id_{\mathbb{R}}: \mathbb{R} \to \mathbb{R}$ be the identity function on $\mathbb{R}$, where $\mathbb{R}$ is a linear space over itself. Given m incompatible and exhaustive events, a random good denoted by X is the restriction of $id_{\mathbb{R}}$ to $I(X) = \{x^1, x^2, \ldots, x^m\} \subset \mathbb{R}$ such that we write $id_{\mathbb{R}|I(X)}: I(X) \to \mathbb{R}$.*

A random good is nothing but a random variable X on a sample space denoted by Ω. It is a function from Ω into the set $\mathbb{R}$ of real numbers such that the pre-image of any

interval of $\mathbb{R}$ is an event in Ω. Our intervals are: $[x^1, x^1], \ldots, [x^m, x^m]$. The points in Ω are real numbers only. Given an orthonormal basis of E^m, a random good is represented by a vector whose contravariant components coincide with the elements of $I(X) = \Omega$ before transferring them on a one-dimensional straight line, on which an origin, a unit of length, and an orientation are established. We write

$$I(X) = \{x^1, x^2, \ldots, x^m\}, \tag{4}$$

with $x^1 < x^2 < \ldots < x^m$ without loss of generality. It is clear that we have $\inf I(X) = x^1$ and $\sup I(X) = x^m$. A located vector at the origin of E^m is completely established by its endpoint. An ordered m-tuple of real numbers can be either a point of an affine space denoted by $\mathcal{E}^m$ or a vector of E^m. Accordingly, $\mathcal{E}^m$ and E^m are isomorphic. Each event is generically denoted by E_i, $i = 1, \ldots, m$. We write

$$X = x^1 |E_1| + x^2 |E_2| + \ldots + x^m |E_m|, \tag{5}$$

where we have

$$|E_i| = \begin{cases} 1, & \text{if } E_i \text{ is true} \\ 0, & \text{if } E_i \text{ is false} \end{cases} \tag{6}$$

for every $i = 1, \ldots, m$. Regarding a given set of information and knowledge, we consider the finest possible partition of X into elementary events. The nature of this partition is always relative, arbitrary, and temporary. That alternative which will turn out to be verified a posteriori is nothing but a random point contained in $I(X)$ (von Neumann 1936). This point contained in $I(X)$ is a real number. It expresses everything there is to be said whenever uncertainty ceases. Each possible value for X could uniquely be expressed by

$$\{x^1 + a, \ x^2 + a, \ \ldots, \ x^m + a\}, \tag{7}$$

where $a \in \mathbb{R}$ is an arbitrary constant. We consider infinite changes of origin in this way (Angelini and Maturo 2021a).

We deal with ordered m-tuples of real numbers (that which is objectively possible is dealt with by Coletti et al. (2016)). All possible values for X are uncertain, so it makes sense that the decision-maker attributes to each of them a probability. $I(X)$ with the assignment of probabilities is a probability space denoted by $(\Omega, \mathcal{F}, \mathbf{P})$. The set of all possible outcomes is denoted by Ω. This set is embedded in a larger space with a linear structure. We write $\mathcal{F} = \{\emptyset, \Omega\}$ to denote a set of events[6], whereas $\mathbf{P}$ is a function of probability or prevision defined as an expression of the subjective opinion of a given decision-maker[7]. We think of probability as being a mass. It is always a nonnegative and additive function. Its value is equal to 1 on the whole space of the possible values for the random good taken into account. The notion of probability is not undefined within this context (Anscombe and Aumann 1963). It is the degree of belief in the occurrence of a single event attributed by a given decision-maker at a given instant and with a given set of information and knowledge (Schmeidler 1989). Uncertainty about an event depends on the existence of imperfect information and knowledge by the decision-maker (Capotorti et al. 2014). We speak about uncertainty in the simple sense of ignorance (Jurado et al. 2015). Uncertainty consists of two different aspects. Possibility and probability are the two aspects of it. In this paper, they are studied inside linear spaces over $\mathbb{R}$. Possibility and probability are expressed by two vectors of E^m used to obtain $\mathbf{P}(X)$, where $\mathbf{P}(X)$ is viewed to be as a scalar or inner product written in the form

$$\mathbf{P}(X) = x^1 \, p_1 + x^2 \, p_2 + \ldots + x^m \, p_m. \tag{8}$$

We write

$$\mathbf{x} = (x^1, x^2, \ldots, x^m) \tag{9}$$

to denote what is objectively possible, whereas we write

$$\mathbf{p} = (p_1, p_2, \ldots, p_m) \tag{10}$$

to denote what is subjectively probable. Since we have

$$p_1 + p_2 + \ldots + p_m = 1, \tag{11}$$

with $0 \leq p_i \leq 1$, $i = 1, \ldots, m$, all those evaluations such that (11) holds are coherent. Their number is equal to ∞^{m-1}. Given $x^1, x^2, \ldots, x^m$, a random process consists of ∞^{m-1} possible choices of masses such that a weighted average of m values given by $x^1, x^2, \ldots, x^m$ takes place. If X is viewed to be as a vector of E^m, then it is a linear combination of m incompatible and exhaustive events expressed by

$$X = x^1 |E_1| \mathbf{e}_1 + x^2 |E_2| \mathbf{e}_2 + \ldots + x^m |E_m| \mathbf{e}_m, \tag{12}$$

where $\mathcal{B}_m = \{\mathbf{e}_i\}$, $i = 1, \ldots, m$, is an orthonormal basis of E^m. Regarding $\mathcal{B}_m$, we write

$$\mathbf{x} = x^1 \mathbf{e}_1 + x^2 \mathbf{e}_2 + \ldots + x^m \mathbf{e}_m. \tag{13}$$

If the Einstein summation convention is used, then it gives

$$\mathbf{x} = x^i \mathbf{e}_i. \tag{14}$$

1.3. The Objectives of the Paper

All the objectives of this paper are innovative. Bound choices are based on possible alternatives. Every choice is a barycenter of masses distributed over a finite set of possible alternatives. The latter is embedded in a larger and more manageable space. Regarding choices being made under conditions of uncertainty and riskiness, possible alternatives are not estimated, but they are observed. What is chosen by a given decision-maker inside his or her budget set coincides with a coherent summary of a nonparametric joint distribution of mass. This summary is a bilinear measure. It is always decomposed into two linear measures. The budget set of the decision-maker consists of points such that each point of it has two Cartesian coordinates. Each of them is a summary of a nonparametric marginal distribution of mass related to a marginal good. Given the two marginal distributions of mass, all possible joint distributions of mass constitute the Fréchet class. We admit that it is useful to compare a concrete (nonparametric) probability distribution with a model which is not a continuous function such as the density function of a continuous random variable, but it is itself a distribution of mass. The latter is characterized by probabilities that are finitely but not countably additive. All possible joint distributions of mass are summarized by a given decision-maker. He or she chooses one of these summaries according to his or her variable state of information and knowledge. He or she can choose a coherent summary such that there is no linear correlation between good 1 and good 2. He or she could also choose a coherent summary such that there is an inverse or direct linear relationship between good 1 and good 2. Regarding the Fréchet class, two extreme limit cases together with an intermediate case are accordingly taken into account. They are paradigmatic cases. Each of them identifies the above model. If good 1 and good 2 are two risky assets, then it is methodologically possible to validate that the notion of risk is intrinsically subjective. We develop the notion of mean quadratic difference put forward by Corrado Gini. We develop it via a tensorial approach. The variability of a distribution of mass always depends on how the decision-maker estimates all the masses under consideration. It follows that the origin of this variability is not random within this context. It is not standardized because the decision-maker makes explicit, from time to time, the knowledge hypothesis underlying it. The origin of the variability of a distribution of mass is not connected with the theory of measurement errors, where such errors are random. Regarding the Sharpe ratio, after computing the expected return on an n-risky asset portfolio, its riskiness is obtained using

the notion of mean quadratic difference. The decision-maker always maximizes his or her subjective utility connected with average quantities. In this paper, the notion of utility is a metric measure as well. What is said in this paper can be extended. Multilinear relationships between variables are discovered and handled, so an extension of the least-squares model can be made. In economics, it is frequent that there is one-way causation in the sense that given variables influence another variable, but there is no feedback in the opposite direction. This means that a specific variable does not influence other variables. Conversely, all the multilinear indices we propose in this paper allow the studying of relationships between variables in such a way that there is a two-way causation. Hence, it is possible to study variables influencing each other.

In Section 2 of the paper, antisymmetric tensors identifying multilinear indices are handled. In Section 3, risky assets viewed to be as random goods are studied. In Section 4, analytic conditions allowing the studying of a single risky asset as a double one are developed. Section 5 shows a variability tensor. Section 6 shows another variability tensor. In Section 7, a multilinear approach to the Sharpe ratio is dealt with. In Section 8, future perspectives of our research are outlined after discussing the main results contained in the paper.

2. Two Random Goods That Are Jointly Considered: From Disaggregate Choices to Aggregate Ones

2.1. Bound Choices Made by a Given Decision-Maker under Conditions of Uncertainty and Riskiness: Their Decomposition Inside His or Her Budget Set

Two random goods which are jointly considered inside the budget set of the decision-maker can be handled through the same framework characterizing bound choices being made by him or her under claimed conditions of certainty (portfolio choices with transient price impact are studied by Ekren and Muhle-Karbe (2019)). Given two marginal random goods denoted by $_1X$ and $_2X$, the number of the possible values for each of them is first equal to m. $_1X$ and $_2X$ are linearly independent. Hence, we consider two mutually orthogonal axes of a two-dimensional Cartesian coordinate system, on which an origin, the same unit of length, and an orientation are established. The possible values for each random good taken into account are transferred on a one-dimensional straight line. Thus, we do not consider an m-dimensional point, but we deal with m one-dimensional points on a one-dimensional straight line. We pass from E^m to a linear space over $\mathbb{R}$ with its dimension which is equal to 1. There exists a one-to-one correspondence between a one-dimensional linear subspace of E^m and a one-dimensional straight line, on which an origin, a unit of length, and an orientation are chosen. A one-dimensional linear subspace of E^m contains all collinear vectors[8] regarding one of the two vectors belonging to E^m. Its contravariant components coincide with the possible values for a marginal good. Two one-dimensional linear subspaces of E^m are dealt with. These subspaces identify two one-dimensional straight lines, on which an origin, the same unit of length, and an orientation are chosen. They establish the budget set of the decision-maker. They establish an uncountable subset of the direct product of $\mathbb{R}$ and $\mathbb{R}$ denoted by $\mathbb{R} \times \mathbb{R}$. Its dimension is equal to 2. All the m^2 possible values for two random goods which are jointly considered give rise to $_1X _2X$. All the m^2 possible values for $_1X _2X$ identify, together with m^2 probabilities, $\mathbf{P}(_1X _2X)$. If $\mathbf{P}(_1X _2X)$ is bilinear, where $\mathbf{P}$ stands for prevision or mathematical expectation of a joint random good denoted by $_1X _2X$, then $\mathbf{P}(_1X)$ and $\mathbf{P}(_2X)$ are linear. We write $\mathbf{P}(_1X _2X) \equiv (\mathbf{P}(_1X), \mathbf{P}(_2X))$ because we deal with a bilinear measure coinciding with a two-dimensional point. If $\mathbf{P}$ is linear, then $_1X$ must always be a random good with its possible values which are all nonnegative. The same is true by considering the possible values for $_2X$ on the vertical axis. If $\mathbf{P}$ is linear, then it is first additive and convex.

Given two random goods, $\mathbf{P}(_1X)$ and $\mathbf{P}(_2X)$ tell us how much the decision-maker is choosing to demand for one of the two random goods taken into account and how much he or she is choosing to demand for the other. His or her budget set is established by the negative slope of the budget line coinciding with a hyperplane embedded in a

two-dimensional Cartesian coordinate system. His or her budget set is also established by the two mutually orthogonal axes taken into account. In particular, we consider two half-lines. His or her budget set is accordingly a right triangle belonging to the first quadrant of a two-dimensional Cartesian coordinate system. The vertex of the right angle of the triangle taken into account coincides with the point given by $(0,0)$. $_1X$, $_2X$, and $_1X\,_2X$ are first discrete goods. They are studied as continuous goods when and only when all their coherent previsions are taken into account at the first stage. The budget line is an equation of a linear function expressed in an implicit form. The prices of the prevision bundle[9] denoted by $\mathbf{P}(_1X\,_2X) \equiv (\mathbf{P}(_1X), \mathbf{P}(_2X))$ are formally two constants of the straight line expressed in an implicit form such that their ratio gives its slope. The budget constraint of the decision-maker requires that the amount of money spent on the two random goods be no more than the total amount he or she has to spend. The budget constraint derives from

$$c_1\,(_1X) + c_2\,(_2X) \le c. \tag{15}$$

It is written in the form

$$c_1\,\mathbf{P}(_1X) + c_2\,\mathbf{P}(_2X) \le c, \tag{16}$$

where (c_1, c_2) are the objective prices of the two random goods, whereas the objective amount of money the decision-maker has to spend is equal to c. Please note that c_1, c_2, and c are positive real numbers. The slope of the budget line expressed by

$$c_1\,\mathbf{P}(_1X) + c_2\,\mathbf{P}(_2X) = c \tag{17}$$

is given by

$$-\frac{c_1}{c_2}. \tag{18}$$

The budget line can always be drawn. It is possible to establish its horizontal and vertical intercepts every time. This means that we pass from m to $m + 1$ possible alternatives for each marginal random good. Structures open to the adjunction of new entities as new circumstances arise are considered by us. They are linear spaces over $\mathbb{R}$ with a different dimension. Structures open are considered because the notion of event is intrinsically subdivisible. The prices of the two random goods taken into account are determined whenever the budget line is drawn. Three convex sets are established. They are two one-dimensional convex sets and one two-dimensional convex set. The first one-dimensional convex set is found between zero expressed by $(0,0)$ and the horizontal intercept of the budget line given by $\frac{c}{c_1}$. The second one is found between zero, expressed by $(0,0)$, and the vertical intercept of it given by $\frac{c}{c_2}$. The third two-dimensional convex set is given by all the points that are found inside the plane region bounded by the right triangle. Please note that (17) always passes through the point whose coordinates are given by

$$\left(\sup I(_1X), \sup I(_2X)\right). \tag{19}$$

If the budget line changes its negative slope, then the budget set of the decision-maker changes. He or she chooses a point belonging to his or her changed budget set. His or her state of information and knowledge changes. It is clear that (16) is analogous to (3). We pass from E^{m+1} to a linear space over $\mathbb{R}$ with its dimension which is equal to 1. There exists a one-to-one correspondence between a one-dimensional linear subspace of E^{m+1} and a one-dimensional straight line, on which an origin, a unit of length, and an orientation are chosen. Two one-dimensional linear subspaces of E^{m+1} are dealt with. These subspaces identify two one-dimensional straight lines, on which an origin, the same unit of length, and an orientation are chosen. They establish the budget set of the decision-maker.

The decision-maker's choice functions for the two marginal random goods under consideration are expressed by

$$\mathbf{P}(_1X) = \left\{\mathbf{P}(_1X)[(c_1, c_2, c)]\right\}, \tag{20}$$

and

$$\mathbf{P}(_2X) = \big\{\mathbf{P}(_2X)[(c_1, c_2, c)]\big\}, \tag{21}$$

where **P** is additive and convex as a consequence of its coherence. We note the following:

Remark 2. *The decision-maker estimates both marginal masses associated with $_1X$ and $_2X$ and the joint ones associated with $_1X\,_2X$. Marginal masses associated with $_1X$ and $_2X$ give rise to $\mathbf{P}(_1X)$ and $\mathbf{P}(_2X)$. Given $(\mathbf{P}(_1X), \mathbf{P}(_2X))$, a bilinear and disaggregate measure coinciding with $\mathbf{P}(_1X\,_2X)$ is a summarized element of the Fréchet class. Given $(\mathbf{P}(_1X), \mathbf{P}(_2X))$, the decision-maker also chooses a summarized element of the Fréchet class such that $\mathbf{P}(_1X)$ and $\mathbf{P}(_2X)$ never change. This element is obtained using the notion of α-product outside the budget set of the decision-maker.*

A remarkable point of this paper is the following. The decision-maker can choose a coherent summary of a joint distribution of mass identifying a summarized element of the Fréchet class such that there is no linear correlation between random good 1 and random good 2, so they are stochastically independent. Given the same marginal masses, $_1X$ and $_2X$ are stochastically independent if each joint mass in a joint distribution is the product of its corresponding marginal masses. In particular, if $_1X$ and $_2X$ are two risky assets, then the decision-maker is risk neutral. He or she could also choose a coherent summary of a joint distribution of mass such that there is an inverse or direct linear relationship between $_1X$ and $_2X$. This means that the decision-maker is, respectively, risk averse or risk loving. In fact, given the same marginal masses, an aggregation of joint masses such that $_1X$ tends to increase when $_2X$ increases shows a direct linear relationship between $_1X$ and $_2X$. Conversely, given the same marginal masses, an aggregation of joint masses such that $_1X$ tends to decrease when $_2X$ increases shows an inverse linear relationship between $_1X$ and $_2X$. Regarding the Fréchet class, two extreme limit cases together with an intermediate case are dealt with. They are paradigmatic cases.

2.2. Two Jointly Considered Random Goods Depending on the Notion of Ordered Pair and Their α-Product

Two marginal random goods denoted by $_1X$ and $_2X$ always give rise to a joint random good denoted by $_1X\,_2X$. All its possible values are obtained by considering the Cartesian product of the possible values for $_1X$ and $_2X$. The horizontal and vertical intercepts must be added to $I(_1X)$ and $I(_2X)$, respectively. We write $I(_1X) \cup \big\{\frac{c}{c_1}\big\}$ and $I(_2X) \cup \big\{\frac{c}{c_2}\big\}$. The values of $I(_1X) \cup \big\{\frac{c}{c_1}\big\}$ and $I(_2X) \cup \big\{\frac{c}{c_2}\big\}$ coincide with the contravariant components of two $(m+1)$-dimensional vectors uniquely expressed as linear combinations of $m+1$ basis vectors of E^{m+1}. We put $I(_1X) \cup \big\{\frac{c}{c_1}\big\} = I^*(_1X)$ and $I(_2X) \cup \big\{\frac{c}{c_2}\big\} = I^*(_2X)$.

Another remarkable point of this paper is that the notion of ordinal utility is a metric measure (Maturo and Angelini 2023). Prevision (probability) and utility are formally the two sides of the same coin, so it is possible to present the following:

Theorem 1. *Let $_1X$ and $_2X$ be two logically independent random goods. They are jointly considered inside the budget set of the decision-maker. Their possible values are expressed by $I(_1X) \cup \big\{\frac{c}{c_1}\big\}$ and $I(_2X) \cup \big\{\frac{c}{c_2}\big\}$. If each coherent prevision of $_1X\,_2X$ denoted by $\mathbf{P}(_1X\,_2X)$ is decomposed into two linear previsions, then its properties coincide with the ones of well-behaved preferences.*

We prove this theorem in another paper of ours. Since indifference curves cannot cross, given any two prevision bundles belonging to two different indifference curves, this theorem tells us that the decision-maker can rank them as to their distance from $(0,0)$ measured along the 45-degree line. One of the prevision bundles is strictly better than the other if and only if its distance from $(0,0)$ measured along the 45-degree line is greater than the other. A numerical example of this can easily be shown using the Pythagorean theorem. It is possible to write

$$^2d(O, \mathbf{P}) = \sqrt{\sum_{i=1}^{2} \mathbf{P}(_iX)^2} \qquad (22)$$

to denote the distance of $\mathbf{P}$ from $O = (0,0)$, where $\mathbf{P}$ stands for $(\mathbf{P}(_1X), \mathbf{P}(_2X))$. We write

$$\mathbf{P} = \begin{pmatrix} \mathbf{P}(_1X) \\ \mathbf{P}(_2X) \end{pmatrix}. \qquad (23)$$

The bundles for which the decision-maker is indifferent to $(\mathbf{P}(_1X), \mathbf{P}(_2X))$ form the indifference curve. Its slope is negative. It can be imagined by identifying preferences for perfect substitutes without loss of generality. It intersects the 45-degree line in a point only. All other indifference curves intersect the 45-degree line. Each of them intersects the 45-degree line in a point only. Preferences are not directly observable. In our approach, the notion of utility has then an independent meaning other than its being what a given decision-maker maximizes.

We establish the following:

Definition 4. *All the events associated with an ordered pair of random goods are obtained by considering the Cartesian product of the possible values for two logically independent random goods denoted by $_1X$ and $_2X$. Such random goods give rise to a joint random good denoted by $_1X_2X$. The latter is a function written in the form $_1X_2X: I^*(_1X) \times I^*(_2X) \to \mathbb{R}$, where we have $_1X_2X(_{(1)}x^i, _{(2)}x^j) = {}_{(1)}x^i {}_{(2)}x^j$, with $i, j = 1, \ldots, m+1$.*

We are faced with

$$_1X_2X = {}_{(1)}x^1 {}_{(2)}x^1 |_{(1)}E_1||_{(2)}E_1| + \ldots + {}_{(1)}x^{m+1} {}_{(2)}x^{m+1} |_{(1)}E_{m+1}||_{(2)}E_{m+1}|, \qquad (24)$$

where it is possible to write

$$|_{(1)}E_i||_{(2)}E_j| = \begin{cases} 1, & \text{if } _{(1)}E_i \text{ and } _{(2)}E_j \text{ are both true} \\ 0, & \text{otherwise} \end{cases} \qquad (25)$$

for every $i, j = 1, \ldots, m+1$.

Since $_1X$ and $_2X$ are two random goods, where each of them has $m+1$ possible values, two random goods giving rise to $_1X_2X$ are logically independent whenever there exist $[(m+1) \cdot (m+1)]$ possible values for $_1X_2X$ (the notion of measure associated with possible values for a random entity is dealt with by Nunke and Savage (1952)). Given $(_1X, _2X)$, we are faced with two different partitions. Each of them is characterized by $m+1$ incompatible and exhaustive events[10]. The covariant components of an affine tensor of order 2 represent the joint masses of the nonparametric joint distribution of $_1X$ and $_2X$. Their number is overall equal to $(m+1)^2$ (coherent probabilities associated with possible values for random entities are handled by Regazzini (1985)). We say that an ordered pair of random goods denoted by $(_1X, _2X)$ is represented by an ordered triple of geometric entities denoted by

$$\left(_{(1)}\mathbf{x}, _{(2)}\mathbf{x}, p_{ij} \right), \qquad (26)$$

with $(i, j) \in I_{m+1} \times I_{m+1}$, where we write $I_{m+1} = \{1, 2, \ldots, m+1\}$. We consider the notion of α-product between $_{(1)}\mathbf{x}$ and $_{(2)}\mathbf{x}$. It is possible to establish a quadratic metric on E^{m+1} in this way. This notion is a scalar or inner product obtained using the joint masses denoted by p_{ij} of the nonparametric joint distribution of $_1X$ and $_2X$ together with the contravariant components of $_{(1)}\mathbf{x}$ and $_{(2)}\mathbf{x}$. We then write

$$\langle _{(1)}\mathbf{x}, _{(2)}\mathbf{x} \rangle_\alpha = {}_{(1)}x^i {}_{(2)}x^j p_{ij} = {}_{(1)}x^i {}_{(2)}x_i = \mathbf{P}(_1X_2X), \qquad (27)$$

where

$$_{(2)}x^j p_{ij} = {}_{(2)}x_i \tag{28}$$

is a vector homography by means of which we pass from $_{(2)}x^j$ to $_{(2)}x_i$ using p_{ij}. All covariant components of an $(m+1)$-dimensional vector are obtained by means of vector homographies involving p_{ij}. For instance, from the following Table.

Random Good 1 \ Random Good 2	0	4	5	Sum
0	0	0	0	0
2	0	0.1	0.2	0.3
3	0	0.5	0.2	0.7
Sum	0	0.6	0.4	1

It follows that we have $\mathbf{P}(_1X\,_2X) = 11.8$. Given the contravariant components of $_{(2)}\mathbf{x}$ identifying the following column vector

$$\begin{pmatrix} 0 \\ 4 \\ 5 \end{pmatrix},$$

its covariant components are expressed by

$$0 \cdot 0 + 4 \cdot 0 + 5 \cdot 0 = 0,$$

$$0 \cdot 0 + 4 \cdot 0.1 + 5 \cdot 0.2 = 1.4,$$

and

$$0 \cdot 0 + 4 \cdot 0.5 + 5 \cdot 0.2 = 3,$$

so it is possible to write the following result

$$\left\langle \begin{pmatrix} 0 \\ 2 \\ 3 \end{pmatrix}, \begin{pmatrix} 0 \\ 1.4 \\ 3 \end{pmatrix} \right\rangle = \langle _{(1)}\mathbf{x},\, _{(2)}\mathbf{x} \rangle_\alpha = \mathbf{P}(_1X\,_2X) = 11.8.$$

On the other hand, after calculating the covariant components of $_{(1)}\mathbf{x}$ in a similar way, we write

$$\left\langle \begin{pmatrix} 0 \\ 1.7 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 4 \\ 5 \end{pmatrix} \right\rangle = \langle _{(1)}\mathbf{x},\, _{(2)}\mathbf{x} \rangle_\alpha = \mathbf{P}(_1X\,_2X) = 11.8.$$

After transferring the possible values for $_1X$ and $_2X$ on two one-dimensional straight lines, $\mathbf{P}(_1X\,_2X)$ lives inside a subset of a two-dimensional linear space over $\mathbb{R}$. $\mathbf{P}(_1X\,_2X)$ is a measure of a metric nature living inside a subset of a linear space over $\mathbb{R}$ denoted by $\mathbb{R} \times \mathbb{R}$. Please note that $\mathbf{P}(_1X\,_2X)$ is identified with a two-dimensional point. We write

$$(\mathbf{P}(_1X), \mathbf{P}(_2X)) \tag{29}$$

to identify $\mathbf{P}(_1X\,_2X)$ inside the budget set of the decision-maker. This means that $\mathbf{P}(_1X\,_2X)$ is always decomposed into two linear measures. Each of them is identified with a one-dimensional point inside the budget set of the decision-maker. The notion of α-norm is a particular case of the one of α-product. From the following Table.

	Random Good 1	0	2	3	Sum
Random Good 1					
0		0	0	0	0
2		0	0.3	0	0.3
3		0	0	0.7	0.7
Sum		0	0.3	0.7	1

It follows that we write $\|_{(1)}\mathbf{x}\|_{\alpha}^2 = \mathbf{P}(_1X\,_1X) = 7.5$, whereas from the following Table.

	Random Good 2	0	4	5	Sum
Random Good 2					
0		0	0	0	0
4		0	0.6	0	0.6
5		0	0	0.4	0.4
Sum		0	0.6	0.4	1

It follows that we have $\|_{(2)}\mathbf{x}\|_{\alpha}^2 = \mathbf{P}(_2X\,_2X) = 19.6$.

2.3. Two Jointly Considered Random Goods That Are Independent of the Notion of Ordered Pair

Let $_1X$ and $_2X$ be two random goods, where each of them is characterized by $m+1$ possible values. We note the following:

Remark 3. *Given an orthonormal basis of E^{m+1}, the possible values for two separately considered random goods are represented by the contravariant components of two vectors of E^{m+1}. If we are not interested in fusing $_1X$ and $_2X$, then the possible values for two logically independent random goods which are jointly considered could coincide with the contravariant components of an affine tensor of order 2. If we are conversely interested in fusing $_1X$ and $_2X$, then the possible values for a stand-alone and double random good denoted by X_{12} are represented by the contravariant components of an antisymmetric tensor of order 2.*

Since we want to pass from an ordered pair of marginal random goods to two marginal random goods which are jointly considered regardless of the notion of ordered pair, we define a double random good denoted by

$$X_{12} = \{_1X, _2X\}. \tag{30}$$

It is a multiple random good of order 2. The possible values for X_{12} coincide with the contravariant components of an antisymmetric tensor of order 2. After choosing $(m+1)^2$ joint masses connected with $_1X\,_2X$, where we write

$$_1X\,_2X\colon I^*(_1X) \times I^*(_2X) \to \mathbb{R}, \tag{31}$$

it is necessary to consider four nonparametric joint distributions characterizing $_1X\,_1X$, $_1X\,_2X$, $_2X\,_1X$, and $_2X\,_2X$, with

$$_1X\,_1X\colon I^*(_1X) \times I^*(_1X) \to \mathbb{R}, \tag{32}$$

$$_2X\,_2X\colon I^*(_2X) \times I^*(_2X) \to \mathbb{R}, \tag{33}$$

and

$$_2X\,_1X\colon I^*(_2X) \times I^*(_1X) \to \mathbb{R}, \tag{34}$$

to release X_{12} from the notion of ordered pair. Please note that $_1X\,_1X$ and $_2X\,_2X$ give rise to joint distributions such that all off-diagonal joint masses of a two-way table, where the number of rows is equal to the one of columns, coincide with zero. After choosing $(m+1)^2$ joint masses connected with $_1X\,_2X$, the distributions associated with $_1X\,_1X$, $_1X\,_2X$, $_2X\,_1X$, and $_2X\,_2X$ are automatically determined.

The mathematical expectation of $_iX\,_jX$, with $i, j = 1, 2$, is of a bilinear nature. This means that it is separately linear in each marginal random good (the notion of prevision of a random entity is studied by Berti et al. (2001)).

Thus, we present the following:

Theorem 2. *The mathematical expectation of $X_{12} = \{_1X, _2X\}$ denoted by $\mathbf{P}(X_{12})$ coincides with the determinant of a square matrix of order 2. Each element of such a determinant is a real number coinciding with the mean value of $_iX\,_jX$ denoted by $\mathbf{P}(_iX\,_jX)$, where we have $i, j = 1, 2$.*

This theorem is proved by us in another paper of ours.

What is actually demanded for X_{12} by the decision-maker coincides with $\mathbf{P}(X_{12})$. It is a multiple choice associated with a multiple good. It is an aggregate measure that is obtained after observing what the decision-maker actually chooses inside his or her budget set. He or she chooses $(\mathbf{P}(_1X), \mathbf{P}(_2X))$ whenever the prices and income are, respectively, c_1, c_2, and c. He or she also chooses $\mathbf{P}(_1X\,_2X)$, so he or she chooses those joint masses such that an element of the Fréchet class is summarized. A remarkable point of this paper is the following. Since a given decision-maker is indifferent to the exchange of $_1X$ for $\mathbf{P}(_1X)$ and of $_2X$ for $\mathbf{P}(_2X)$, he or she is also indifferent to the exchange of X_{12} for $\mathbf{P}(X_{12})$, where we write

$$\mathbf{P}(X_{12}) = \begin{vmatrix} \mathbf{P}(_1X\,_1X) & \mathbf{P}(_1X\,_2X) \\ \mathbf{P}(_2X\,_1X) & \mathbf{P}(_2X\,_2X) \end{vmatrix}. \tag{35}$$

$\mathbf{P}(X_{12})$ extends the notion of barycenter of masses. In particular, the property of the barycenter known as stable equilibrium is extended. Given $_1X$ and $_2X$ and their average quantities, we consider all deviations from $\mathbf{P}(_1X)$ and $\mathbf{P}(_2X)$ of the possible values for $_1X$ and $_2X$ (Rockafellar et al. 2006).

We then present the following:

Theorem 3. *The variance of $X_{12} = \{_1X, _2X\}$ denoted by $\mathrm{Var}(X_{12})$ coincides with the determinant of a square matrix of order 2. Each element of such a determinant is a real number coinciding with the variance of $_1X$ and $_2X$, and with their covariance.*

This theorem is proved by us in another paper of ours.

We note the following:

Remark 4. *The origin of the variability of X_{12} depends on the variable state of information and knowledge associated with a given decision-maker. This is because all deviations from $\mathbf{P}(_1X)$ and $\mathbf{P}(_2X)$ of the possible values for $_1X$ and $_2X$ depend on his or her variable state of information and knowledge.*

A nonlinear (multilinear) metric is the expression given by

$$\|_{12}d\|_\alpha^2 = \begin{vmatrix} \|_{(1)}\mathbf{d}\|_\alpha^2 & \langle_{(1)}\mathbf{d}, _{(2)}\mathbf{d}\rangle_\alpha \\ \langle_{(2)}\mathbf{d}, _{(1)}\mathbf{d}\rangle_\alpha & \|_{(2)}\mathbf{d}\|_\alpha^2 \end{vmatrix} = \|_{(1)}\mathbf{d}\|_\alpha^2 \|_{(2)}\mathbf{d}\|_\alpha^2 - \left(\langle_{(1)}\mathbf{d}, _{(2)}\mathbf{d}\rangle_\alpha\right)^2. \tag{36}$$

It is the area of a 2-parallelepiped. Its edges are two marginal random goods with their possible values that are subjected to two changes of origin. The strict components of $_{12}d$ are the coordinates of such edges denoted by $_{(1)}\mathbf{d}$ and $_{(2)}\mathbf{d}$. We also write

$$\|_{12}d\|_\alpha^2 = \text{Var}(X_{12}) = \begin{vmatrix} \text{Var}(_1X) & \text{Cov}(_1X, {}_2X) \\ \text{Cov}(_2X, {}_1X) & \text{Var}(_2X) \end{vmatrix}, \tag{37}$$

so the property of the barycenter known as the minimum of the moment of inertia is extended.

3. Random Goods Whose Possible Values Are of a Monetary Nature: Risky Assets

3.1. Risky Assets Studied inside the Budget Set of the Decision-Maker

Let $_1X$ and $_2X$ be two risky assets. In this subsection, we study them inside the budget set of the decision-maker. In the first stage, all coherent expected returns on the portfolio denoted by $X_{12} = \{_1X, {}_2X\}$ consisting of two risky assets are expressed by

$$\frac{c_1}{c_1 + c_2} \mathbf{P}(_1X) + \frac{c_2}{c_1 + c_2} \mathbf{P}(_2X) \le \frac{c}{c_1 + c_2}. \tag{38}$$

Given $_1X$ and $_2X$, where $_1X$ and $_2X$ are the components of X_{12}, whenever we use the principle characterizing a linear and quadratic metric to establish the expected return on a two-risky asset portfolio, we focus on the components of X_{12} only. We focus on $_1X$ and $_2X$ only. The left-hand side of (38) is a weighted average of the two expected returns on the two risky assets taken into account (Markowitz 1952). The two expected returns on the two risky assets taken into account are themselves two weighted averages. A coherent expected return on a joint risky asset denoted by $_1X {}_2X$ is given by $\mathbf{P}(_1X {}_2X)$. A nonparametric joint distribution of mass is summarized by means of $\mathbf{P}(_1X {}_2X)$. The latter is decomposed into $\mathbf{P}(_1X)$ and $\mathbf{P}(_2X)$ inside the budget set of the decision-maker. In the first stage, all coherent expected returns on a joint risky asset give rise to a two-dimensional convex set. The decision-maker divides his or her relative monetary wealth given by

$$\frac{c_1}{c_1 + c_2} \tag{39}$$

and

$$\frac{c_2}{c_1 + c_2} \tag{40}$$

between the two risky assets taken into account, where we observe

$$\frac{c_1}{c_1 + c_2} + \frac{c_2}{c_1 + c_2} = 1. \tag{41}$$

The budget set of the decision-maker established by the budget constraint given by (16) does not change whenever we multiply all objective prices and income by a positive number. The best rational choice being made by him or her from his or her budget set does not change either. His or her best rational choice depends on his or her subjective preferences (Angelini and Maturo 2022a). His or her best rational choice depends on further hypotheses of an empirical nature. Please note that (39) and (40) can be viewed as the prices associated with average quantities chosen by a given decision-maker, whereas $\frac{c}{c_1 + c_2}$ is the amount of money he or she has to spend. Formally, the two prices are constants expressed by real numbers. Their ratio identifies the slope of a hyperplane embedded in a two-dimensional linear space over $\mathbb{R}$. We note the following:

Remark 5. *It is possible to study real data given by time series connected with annual returns referred to marginal risky assets. It is possible to make a coherent prevision about the return associated with each marginal risky asset based on observed data in different stock markets. Each time series is associated with a stock market. Real data given by time series are possible alternatives*

that are summarized. Their nature is intrinsically objective. From the slope of the budget line which can be drawn, it is possible to observe the prices of the two risky assets viewed to be as two marginal random goods. It is also possible to wonder if the decision-maker taken into account maximizes, or does not maximize, his or her subjective utility connected with weighted averages. This is because the notion of ordinal utility is itself a metric measure.

3.2. Risky Assets Studied outside the Budget Set of the Decision-Maker

Given $_1X$ and $_2X$ and their expected returns, we consider all deviations from $\mathbf{P}(_1X)$ and $\mathbf{P}(_2X)$ of the possible values for $_1X$ and $_2X$. We denote them by $_{(1)}\mathbf{d}$ and $_{(2)}\mathbf{d}$, respectively. Please note that $\mathbf{P}(_1X)$ and $\mathbf{P}(_2X)$ are chosen by the decision-maker inside his or her budget set. We are now found outside it. Given

$$\mathbf{y} = \lambda_1 \,_{(1)}\mathbf{d} + \lambda_2 \,_{(2)}\mathbf{d}, \tag{42}$$

with $\lambda_1 = \frac{c_1}{c_1+c_2}$, $\lambda_2 = \frac{c_2}{c_1+c_2} \in \mathbb{R}$, it is possible to obtain

$$\|\mathbf{y}\|_\alpha^2 = (\lambda_1)^2 \|_{(1)}\mathbf{d}\|_\alpha^2 + 2\lambda_1 \lambda_2 \langle_{(1)}\mathbf{d}, \,_{(2)}\mathbf{d}\rangle_\alpha + (\lambda_2)^2 \|_{(2)}\mathbf{d}\|_\alpha^2, \tag{43}$$

with

$$\|_{(1)}\mathbf{d}\|_\alpha^2 = \mathrm{Var}(_1X), \tag{44}$$

$$\|_{(2)}\mathbf{d}\|_\alpha^2 = \mathrm{Var}(_2X), \tag{45}$$

and

$$\langle_{(1)}\mathbf{d}, \,_{(2)}\mathbf{d}\rangle_\alpha = \mathrm{Cov}(_1X, \,_2X). \tag{46}$$

Whenever we use a linear and quadratic metric, we focus on the riskiness of $_1X$ and $_2X$ only. In fact, we consider $\mathrm{Var}(_1X)$, $\mathrm{Var}(_2X)$, and $\mathrm{Cov}(_1X, \,_2X)$. A linear metric is the α-norm of $\mathbf{y}$ given by (43). In particular, it is possible to write

$$\|_{(1)}\mathbf{d} - _{(2)}\mathbf{d}\|_\alpha^2 = \|_{(1)}\mathbf{d}\|_\alpha^2 + \|_{(2)}\mathbf{d}\|_\alpha^2 - 2\langle_{(1)}\mathbf{d}, \,_{(2)}\mathbf{d}\rangle_\alpha. \tag{47}$$

Such an expression shows the notion of α-distance between two marginal risky assets. Their possible values are subjected to two changes of origin.

4. Conditions Allowing the Studying of a Marginal Risky Asset as a Double Risky Asset

Given a marginal risky asset denoted by $_1X$, we want to study it as a double risky asset denoted by X_{12}, where X_{12} intrinsically consists of four joint risky assets. We must study four joint distributions of mass. They must be all summarized[11]. We note that two conditions must be satisfied to represent $_1X$ as X_{12}. First, we write

$$\mathbf{P}(_1X) = \sum_{i_1=1}^{m+1} {}_{(1)}x^{i_1}\, p_{i_1} \tag{48}$$

to denote the expected return on $_1X$. We say that $_1X$ is the component of a double risky asset, where

$$p = p_{i_1 i_2} \tag{49}$$

is an affine tensor of order 2 whose covariant components express all joint masses taken into account (the conditions of coherence are studied by Berti and Rigo (2002)). Such an affine tensor must satisfy the following relationship given by

$$\sum_{i_1=1}^{m+1} {}_{(1)}x^{i_1}\, p_{i_1} = \sum_{i_1, i_2=1}^{m+1} {}_{(1)}x^{i_1}\, p_{i_1 i_2}. \tag{50}$$

We then say that the two sides of (50) are equal if and only if we have

$$\sum_{i_1=1}^{m+1} p_{i_1} = \sum_{i_1, i_2=1}^{m+1} p_{i_1 i_2}. \tag{51}$$

Since we write

$$\sum_{i_1=1}^{m+1} p_{i_1} = \sum_{i_1, i_2=1}^{m+1} p_{i_1 i_2} = 1, \tag{52}$$

it follows that the first condition tells us that $_1X$ and $_1X\,_2X$ are two finite partitions of events such that the sum of their associated masses is equal to 1.

The second condition tells us that $_1X$ and $_1X\,_2X$ must have the same summarized measure which is obtained using **P**. This means that $_1X$ and $_1X\,_2X$ must have the same expected return. We therefore write

$$\sum_{i_1, i_2=1}^{m+1} {}_{(1)}x^{i_1} {}_{(2)}x^{i_2} p_{i_1 i_2} = \sum_{i_1, i_2=1}^{m+1} {}_{(1)}x^{i_1} p_{i_1 i_2}. \tag{53}$$

It follows that the two sides of (53) are equal if and only if we have

$$_{(2)}x^{i_2} = 1, \quad \forall i_2 \in I_{m+1}. \tag{54}$$

Hence, we note the following:

Remark 6. *Let $\mathcal{B}_{m+1} = \{\mathbf{e}_i\}$, $i = 1, \ldots, m+1$, be an orthonormal basis of E^{m+1}. The possible values for the other risky asset such that $_1X$ is studied as X_{12} are the contravariant components, all of them coinciding with 1, of a vector of E^{m+1}. They form the set denoted by*

$$\{1^i\}. \tag{55}$$

Its number of elements is equal to $m+1$. Such components are not vectorially intrinsic because they depend on the basis of E^{m+1} being chosen. If we pass from $\mathcal{B}_{m+1}$ to $\mathcal{B}'_{m+1} = \{\mathbf{e}_{i'}\}$, $i' = 1, \ldots, m+1$, then the contravariant components of such a vector transform like the ones of any other vector of E^{m+1}. We therefore write

$$1^{i'} = a_i^{i'}\, 1^i = \sum_{i=1}^{m+1} a_i^{i'}, \tag{56}$$

where $A = (a_i^{i'})$ is an $(m+1) \times (m+1)$ matrix expressing a change of basis.

Remark 7. *The vector of E^{m+1} whose contravariant components form the set expressed by*

$$\{\phi^1 = 1, \phi^2 = 1, \ldots, \phi^{m+1} = 1\} \tag{57}$$

is denoted by $\boldsymbol{\phi}$.

It is evident that $\boldsymbol{\phi}$ identifies a degenerate risky asset. It has 1 as its unique possible value.

From a Marginal Distribution of Mass to Four Joint Distributions: A Numerical Example

A nonparametric marginal distribution of mass of $_1X$ can be interpreted as a joint distribution of $_1X$ and $_2X = \boldsymbol{\phi}$. For instance, from the following Table.

$\diagdown \; {}_2X = \phi$ ${}_1X$	1	1	1	**Sum**
0	0	0	0	0
2	0	0.3	0	0.3
3	0	0	0.7	0.7
Sum	0	0.3	0.7	1

It follows that we have $\mathbf{P}({}_1X) = \mathbf{P}({}_1X\,{}_2X) = \mathbf{P}({}_2X\,{}_1X) = 2.7$. Since we observe $\mathbf{P}({}_1X\,{}_1X) = 7.5$ and $\mathbf{P}({}_2X\,{}_2X) = 1$, the riskiness of ${}_1X$ can be expressed by

$$\sigma^2_{{}_1X} = \begin{vmatrix} \mathbf{P}({}_1X\,{}_1X) = 7.5 & \mathbf{P}({}_1X\,{}_2X) = 2.7 \\ \mathbf{P}({}_2X\,{}_1X) = 2.7 & \mathbf{P}({}_2X\,{}_2X) = 1 \end{vmatrix} = 0.21.$$

The riskiness of ${}_1X$ is expressed through a known index. It is shown in a more general fashion. In fact, the riskiness of ${}_1X$ is determined as if ${}_1X$ coincides with $X_{12} = \{{}_1X, \boldsymbol{\phi}\}$ (other specific risk measures are handled by Herdegen and Khan (2022)).

5. A Marginal Risky Asset Identified with a Variability Tensor

The possible values for a double risky asset denoted by X_{12} coincide with the strict contravariant components of an antisymmetric tensor of order 2. In general, let ${}_{12}f$ be an antisymmetric tensor of order 2. We write

$$ {}_{12}f^{(i_1 i_2)} = \begin{vmatrix} {}_{(1)}x^{i_1} & {}_{(1)}x^{i_2} \\ {}_{(2)}x^{i_1} & {}_{(2)}x^{i_2} \end{vmatrix} = {}_{(1)}x^{i_1}{}_{(2)}x^{i_2} - {}_{(1)}x^{i_2}{}_{(2)}x^{i_1} \tag{58}$$

to identify the strict contravariant components of it. If ${}_1X$ is viewed as a double risky asset, then the strict contravariant components of an antisymmetric tensor of order 2 identifying ${}_1X$ are given by

$$ {}_{(1)}f^{(i_1 i_2)} = \begin{vmatrix} {}_{(1)}x^{i_1} & {}_{(1)}x^{i_2} \\ \phi^{i_1} = 1 & \phi^{i_2} = 1 \end{vmatrix}. \tag{59}$$

We prove the following:

Theorem 4. *A nonparametric distribution of mass characterizing a marginal risky asset denoted by ${}_1X$ is summarized using the notion of α-norm of an antisymmetric tensor of order 2 denoted by ${}_{(1)}f$. A measure of riskiness of ${}_1X$ is obtained by calculating the α-norm of ${}_{(1)}f$ denoted by $\|{}_{(1)}f\|^2_\alpha$.*

Proof. Since it is possible to write

$$\phi^{i_1} p_{i_1 i_2} = \phi_{i_2} = p_{i_2}, \tag{60}$$

the covariant components of $\boldsymbol{\phi}$ represent the masses associated with the possible values for ${}_1X$ by a given decision-maker (Angelini and Maturo 2020). It follows that we observe

$$\phi^{i_1} \phi_{i_1} = 1, \tag{61}$$

where (61) can also be written in the form expressed by

$$\|\boldsymbol{\phi}\|^2_\alpha = 1. \tag{62}$$

The expected return on ${}_1X$ is vectorially expressed by

$$ {}_{(1)}x^{i_1} \phi_{i_1} = {}_{(1)}x_{i_1} \phi^{i_1} = {}_{(1)}\bar{\mathbf{x}}, \tag{63}$$

where we have

$$_{(1)}\bar{\mathbf{x}} = \begin{pmatrix} _{(1)}\bar{x}^1 = \mathbf{P}(_1X) \\ _{(1)}\bar{x}^2 = \mathbf{P}(_1X) \\ \vdots \\ _{(1)}\bar{x}^{m+1} = \mathbf{P}(_1X) \end{pmatrix}.$$

(64)

The strict covariant components of $_{(1)}f$ are given by

$$_{(1)}f_{(i_1 i_2)} = \begin{vmatrix} _{(1)}x_{i_1} & _{(1)}x_{i_2} \\ \phi_{i_1} & \phi_{i_2} \end{vmatrix}.$$

(65)

They are obtained by considering all feasible decompositions of two expected returns on the two elements of X_{12} (a geometric approach connected with more general random entities is shown by Pompilj (1957)). We consider different vector homographies to obtain all covariant components of the two vectors denoted by $_{(1)}\mathbf{x}$ and $\boldsymbol{\phi}$ identifying the two elements of X_{12}. We compute the mean quadratic difference of $_1X$ by taking two different requirements into account (variability measures put forward by Corrado Gini are handled by Berkhouch et al. (2018)). First, the α-norm of an antisymmetric tensor of order 2 is always calculated by considering its strict components. Second, the notion of mean quadratic difference of $_1X$ requires that all possible differences be considered (Furman et al. 2017). This means that the non-strict components of $_{(1)}f$ are even taken into account. We then write

$$^2\Delta^2(_1X) = \|_{(1)}f\|^2_\alpha = {}_{(1)}f^{(i_1 i_2)} {}_{(1)}f_{(i_1 i_2)} = \frac{1}{2} {}_{(1)}f^{i_1 i_2} {}_{(1)}f_{i_1 i_2},$$

(66)

where we have

$$\frac{1}{2} = \frac{1}{2!}.$$

(67)

Please note that (67) appears whenever we do not consider the strict components of an antisymmetric tensor of order 2. By taking (59) and (65) into account, we obtain

$$\frac{1}{2} {}_{(1)}f^{i_1 i_2} {}_{(1)}f_{i_1 i_2} = \frac{1}{2} \begin{vmatrix} _{(1)}x^{i_1} & _{(1)}x^{i_2} \\ \phi^{i_1} & \phi^{i_2} \end{vmatrix} \begin{vmatrix} _{(1)}x_{i_1} & _{(1)}x_{i_2} \\ \phi_{i_1} & \phi_{i_2} \end{vmatrix}.$$

(68)

The right-hand side of (68) contains all contravariant and covariant components of $_{(1)}f$ at the same time. After reminding (61)–(63), we finally write

$$^2\Delta^2(_1X) = \frac{1}{2} \begin{vmatrix} 2\|_{(1)}\mathbf{x}\|^2_\alpha & 2 {}_{(1)}\bar{\mathbf{x}} \\ 2 {}_{(1)}\bar{\mathbf{x}} & 2 \end{vmatrix}.$$

(69)

We always associate $_{(1)}x^{i_1}$ with $_{(1)}x_{i_1}$, $_{(1)}x^{i_2}$ with ϕ_{i_2}, ϕ^{i_1} with $_{(1)}x_{i_1}$, and ϕ^{i_2} with ϕ_{i_2}. Nevertheless, there are two variable indices separately appearing twice in each single term (monomial). After computing the determinant appearing on the right-hand side of (69), it is then possible to obtain

$$^2\Delta^2(_1X) = \frac{4}{2} \left(\|_{(1)}\mathbf{x}\|^2_\alpha - {}_{(1)}\bar{\mathbf{x}}^2 \right) = 2\sigma^2_{_1X}.$$

(70)

We wrote the square of the relationship between the mean quadratic difference of $_1X$ denoted by $^2\Delta(_1X)$ and its standard deviation (Gerstenberger and Vogel 2015). □

The relationship between the mean quadratic difference of $_1X$ denoted by $^2\Delta(_1X)$ and its standard deviation has been established by Corrado Gini (Ji et al. 2017). We consider the square of it (Li et al. 2016). In this paper, a tensorial approach to the mean quadratic difference is dealt with. More generally, in our opinion, a tensorial approach to the theory of decision-making is well-grounded because of various reasons. First, the object of decision-maker choice naturally embraces various elements made clear in this research work and it is closely connected with the notion of ordinal utility from an operational point of view. Second, the space where a given decision-maker chooses has a precise mathematical structure. Its technical characteristics must be taken into account to try to find out new results. Third, the conditions of certainty are an extreme simplification. In our opinion, they may produce a sterilization of the connection of choice problems with their applications to reality. Fourth, axiomatic constructions generally lead to accepting for certain the alternative based on which a given decision-maker decides to act. Such constructions link choice problems to reality and to applications by replacing a well-founded probability issue with an impossible translation of it into the logic of certainty. In our opinion, this replacement must not take place.

6. A Variability Tensor Based on Deviations from a Mean Value

Let $_{(1)}\mathbf{d}$ be the deviation vector corresponding to the vector denoted by $_{(1)}\mathbf{x}$ identifying $_1X$. By taking (59) into account, we write

$$_{(1)}\psi^{(i_1 i_2)} = \begin{vmatrix} _{(1)}d^{i_1} & _{(1)}d^{i_2} \\ \phi^{i_1} & \phi^{i_2} \end{vmatrix}, \tag{71}$$

where only the first row of (71) is different from the one of (59). The second row of (71) is the same as the one of (59). Hence, we prove the following:

Theorem 5. *Given $_{(1)}\psi$, its α-norm denoted by $\|_{(1)}\psi\|_\alpha^2$ represents the mean quadratic difference of $_1X$.*

Proof. The strict covariant components of $\|_{(1)}\psi\|_\alpha^2$ are given by

$$_{(1)}\psi_{(i_1 i_2)} = \begin{vmatrix} _{(1)}d_{i_1} & _{(1)}d_{i_2} \\ \phi_{i_1} & \phi_{i_2} \end{vmatrix}, \tag{72}$$

so we can compute the α-norm of $_{(1)}\psi$ denoted by $\|_{(1)}\psi\|_\alpha^2$. We consequently write

$$\|_{(1)}\psi\|_\alpha^2 = {}_{(1)}\psi^{(i_1 i_2)} {}_{(1)}\psi_{(i_1 i_2)} = \frac{1}{2} {}_{(1)}\psi^{i_1 i_2} {}_{(1)}\psi_{i_1 i_2}, \tag{73}$$

where we have

$$\frac{1}{2} {}_{(1)}\psi^{i_1 i_2} {}_{(1)}\psi_{i_1 i_2} = \frac{1}{2} \begin{vmatrix} _{(1)}d^{i_1} & _{(1)}d^{i_2} \\ \phi^{i_1} & \phi^{i_2} \end{vmatrix} \begin{vmatrix} _{(1)}d_{i_1} & _{(1)}d_{i_2} \\ \phi_{i_1} & \phi_{i_2} \end{vmatrix}. \tag{74}$$

The right-hand side of (74) contains all contravariant and covariant components of $_{(1)}\psi$ at the same time. We note that $_{(1)}\mathbf{d}$ and $\boldsymbol{\phi}$ are α-orthogonal. We therefore write

$$d^{i_1} \phi_{i_1} = 0. \tag{75}$$

It follows that we obtain

$$\|_{(1)}\psi\|_\alpha^2 = \frac{1}{2}\begin{vmatrix} 2\,\|_{(1)}\mathbf{d}\|_\alpha^2 & 0 \\ 0 & 2 \end{vmatrix} = \frac{1}{2}\cdot 2\left(2\,\|_{(1)}\mathbf{d}\|_\alpha^2\right), \tag{76}$$

where the expression enclosed in parentheses represents twice the α-norm of $_{(1)}\mathbf{d}$. We always associate $_{(1)}d^{i_1}$ with $_{(1)}d_{i_1}$, $_{(1)}d^{i_2}$ with ϕ_{i_2}, ϕ^{i_1} with $_{(1)}d_{i_1}$, and ϕ^{i_2} with ϕ_{i_2}. There are two variable indices separately appearing twice in each single term. Thus, we write

$$\|_{(1)}\psi\|_\alpha^2 = 2\,\|_{(1)}\mathbf{d}\|_\alpha^2, \tag{77}$$

so we observe

$$^2\Delta^2(_1X) = \|_{(1)}\psi\|_\alpha^2 = 2\,\|_{(1)}\mathbf{d}\|_\alpha^2 = 2\,\sigma_{_1X}^2. \tag{78}$$

The mean quadratic difference of $_1X$ denoted by $^2\Delta(_1X)$ is evidently the same (Shalit and Yitzhaki 2005). We can use both $_{(1)}f$ and $_{(1)}\psi$ to obtain it. They are both of them variability tensors identifying the riskiness of $_1X$. $\square$

The mean quadratic difference of $_1X$ measures the spread of the nonparametric distribution of mass taken into account (Jasso 1979). It is a measure of how far the possible values for $_1X$ are from $\mathbf{P}(_1X)$ (La Haye and Zizler 2019).

7. The Sharpe Ratio Obtained Using Multilinear Measures

It is possible to write

$$\mathbf{P}(X_{12\ldots n}) = \begin{vmatrix} \mathbf{P}(_1X\,_1X) & \mathbf{P}(_1X\,_2X) & \ldots & \mathbf{P}(_1X\,_nX) \\ \mathbf{P}(_2X\,_1X) & \mathbf{P}(_2X\,_2X) & \ldots & \mathbf{P}(_2X\,_nX) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{P}(_nX\,_1X) & \mathbf{P}(_nX\,_2X) & \ldots & \mathbf{P}(_nX\,_nX) \end{vmatrix}, \tag{79}$$

and

$$\text{Var}(X_{12\ldots n}) = \begin{vmatrix} \frac{1}{2}\,^2\Delta^2(_1X) & \text{Cov}(_1X,\,_2X) & \ldots & \text{Cov}(_1X,\,_nX) \\ \text{Cov}(_2X,\,_1X) & \frac{1}{2}\,^2\Delta^2(_2X) & \ldots & \text{Cov}(_2X,\,_nX) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(_nX,\,_1X) & \text{Cov}(_nX,\,_2X) & \ldots & \frac{1}{2}\,^2\Delta^2(_nX) \end{vmatrix}, \tag{80}$$

where $\mathbf{P}(X_{12\ldots n})$ denotes the expected return on an n-risky asset portfolio, whereas $\text{Var}(X_{12\ldots n})$ denotes its riskiness. We do not observe $2^2 = 4$ pairs of risky assets anymore, but we deal with n^2 pairs of them. Regarding the budget sets of a given decision-maker, there exist n^2 budget lines. In particular, the slope of the budget line is always equal to -1 whenever the two risky assets taken into account are the same. In these cases, the budget sets of a given decision-maker always consist of points whose number is infinite. Nevertheless, the joint masses of $_1X\,_1X, _2X\,_2X, \ldots, _nX\,_nX$ must be estimated in such a way that all off-diagonal joint masses of each two-way table with the same number of rows and columns coincide with zero. An interesting study for bear markets is made by Scholz (2007). In this section, the Sharpe ratio is obtained using a multilinear approach (other return-risk ratios are dealt with by Cheridito and Kromer (2013)). Let r_f be the risk-free asset paying a fixed rate of return. The Sharpe ratio is accordingly given by

$$\text{SR} = \frac{\mathbf{P}(X_{12\ldots n}) - r_f}{\sqrt{\text{Var}(X_{12\ldots n})}}, \tag{81}$$

where $\mathbf{P}(X_{12\ldots n})$ and $\sqrt{\text{Var}(X_{12\ldots n})}$ are two determinants of two square matrices of order n connected with two tensors of the same order (Angelini and Maturo 2022b). It measures how risk and return can be traded off in making portfolio choices. Such choices are studied inside the budget set of the decision-maker (Dowd 2000). The marginal rate of substitution between risk and return is given by (81). The slope of the budget line measuring the cost of achieving a larger expected return on $X_{12\ldots n}$ in terms of the increased standard deviation of the return is given by (81), where we assume $\mathbf{P}(X_{12\ldots n}) > r_f$ (a specific model about uncertainty is studied by Pham et al. (2022)). Please note that $\text{Var}(X_{12\ldots n})$ is obtained through the notion of mean quadratic difference. In this section, an extension of the mean-variance model is computationally shown. Moreover, since the beta of a given stock i can statistically be defined by considering the covariance of the return on the stock with the market return divided by the variance of the market return, and specifically it is then possible to write

$$\beta_i = \frac{\text{Cov}(r_i, r_m)}{\text{Var}(r_m)}, \tag{82}$$

what is said in this section can operationally be associated with the Capital Asset Pricing Model, which has many uses in the study of financial markets. The expected market return r_m can accordingly be expressed using a measure with the same structure as (79).

8. Conclusions, Discussion, and Future Perspectives

This paper answers different questions. Two of them are essential. First, the number of points of the budget set of the decision-maker is infinite because all admissible (rational) choices at the first stage derive from masses that are subjectively established. In the second stage, the object of decision-maker choice depends on further hypotheses of an empirical nature, but the distribution of masses identifying this object of decision-maker choice is always characterized by subjective and objective elements. Each point of the budget set of the decision-maker is a metric measure. Every measure is obtained after summarizing a nonparametric joint distribution of mass. Different distributions of mass are different measures. Nevertheless, when talking in terms of measure one does not make of it something fixed, with a special status. A given decision-maker accordingly focuses on masses because there is always the physical perception of being able to move them in whatever way he or she likes. In our approach, a mechanical transposition of all the notions, procedures, and results of measure theory into the calculus of probability does not happen. Every measure is not directly visible inside the budget set of the decision-maker because it is a real number. It appears as a two-dimensional point. Second, the role played by objective alternatives is fundamental. Structures open to the adjunction of new entities as new circumstances arise are studied. They are linear spaces over $\mathbb{R}$. Their dimensions are different. We can know $\mathbf{P}(X_{12})$ and $\text{Var}(X_{12})$ using a multilinear and quadratic metric, where X_{12} is a two-risky asset portfolio. We can also know $\mathbf{P}(X_{12\ldots n})$ and $\text{Var}(X_{12\ldots n})$, where $X_{12\ldots n}$ is an n-risky asset portfolio. Since we use a quadratic metric, we always consider two random goods at a time. We never consider more than two goods at a time. The notion of ordinal utility is a metric measure as well. In this paper, a more general approach to the riskiness of random goods is proposed. We use the notion of mean quadratic difference put forward by Corrado Gini. We develop it using a tensorial approach. If the decision-maker uses mean quadratic differences, then he or she expresses, from time to time, the knowledge hypothesis underlying the variability of his or her choices. It is possible to understand that the notion of mean quadratic difference is also connected with the Bravais–Pearson correlation coefficient. Regarding random goods, this coefficient is intrinsically referred to a double random good denoted by X_{12}. If $_{(1)}\mathbf{d}$ and $_{(2)}\mathbf{d}$ are α-orthogonal vectors, then we obtain

$$\|_{12}\hat{d}\|_\alpha^2 = \begin{vmatrix} \|_{(1)}\mathbf{d}\|_\alpha^2 & 0 \\ 0 & \|_{(2)}\mathbf{d}\|_\alpha^2 \end{vmatrix} = \|_{(1)}\mathbf{d}\|_\alpha^2 \|_{(2)}\mathbf{d}\|_\alpha^2. \tag{83}$$

Since it is possible to write

$$-1 \le \left(1 - \frac{\|_{12}d\|_\alpha^2}{\|_{12}\hat{d}\|_\alpha^2}\right)^{1/2} \le +1, \tag{84}$$

the above expression within the parentheses coincides with the Bravais–Pearson correlation coefficient referred to X_{12}, where $\|_{12}d\|_\alpha^2$ and $\|_{12}\hat{d}\|_\alpha^2$ are two aggregate measures obtained using a multilinear and quadratic metric. In this paper, the origin of the variability of a nonparametric distribution of mass depends on the variable state of information and knowledge associated with a given decision-maker. It is susceptible to being continuously enriched by the flow of new pieces of information. It can also be enriched by the results that are gradually learned or observed in relation to more or less analogous situations and cases. For this reason, the riskiness of a two-risky asset portfolio is studied using the notion of α-norm of an antisymmetric tensor of order 2.

What is said in this paper can be extended. This is because $m + 1$ possible values for a risky asset have an objective nature in the same way as $m + 1$ sampling units that are observed regarding a specific population. Multilinear relationships between variables with parametric probability distributions such as normal distributions can be dealt with using measures of a multilinear nature. A multilinear regression model based on this multilinear approach has been made by us. The paper containing this model is currently under review by an international journal.

Given $m + 1$ possible values for a risky asset, they identify a vector belonging to E^{m+1}. Two linearly independent vectors of E^{m+1} generate a linear subspace of E^{m+1}. Its dimension is equal to 2. The Grassmann coordinates of this linear subspace over $\mathbb{R}$ are the components of a tensor of order 2. Two linearly independent vectors of E^{m+1} are transferred on two mutually orthogonal one-dimensional straight lines, on which an origin, the same unit of length, and an orientation are established. It is possible to show that at least mean quadratic differences, the correlation coefficient, Jensen's inequality, revealed preference theory viewed to be as a branch of the theory of decision-making, the least-squares model, and principal component analysis can be based on intrinsic conditions of uncertainty characterized by objective and subjective elements that are studied inside subsets of linear spaces over $\mathbb{R}$ provided with a specific dimension.

It is possible to overcome the limits of the current research by focusing one's attention on a stochastic view of bound choices. Such a view can be based on subjective opinions or attitudes of a given person. The subjective opinion, meant as something known by the decision-maker taken into account, is something objective in the sense that can be a reasonable object of a rigorous study. Even when one point of a specific convex set is chosen, there is no reason that would lead a given person to consider correct from a philosophical point of view this one, or that one, among the infinitely many possible opinions about the evaluations of probability. Thus, whenever a given decision-maker is indifferent to the exchange of $_1X$ for $\mathbf{P}(_1X)$, a finite number of deviations or errors which are normally distributed can be determined. Whenever he or she is indifferent to the exchange of $_2X$ for $\mathbf{P}(_2X)$, a finite number of deviations or errors which are normally distributed can be determined. Finally, since he or she is also indifferent to the exchange of X_{12} for $\mathbf{P}(X_{12})$, a finite number of deviations or errors can be dealt with in an aggregate fashion.

- **The authors can confirm that all relevant data are included in the article**

Notes

1.   In economics, normal and ordinary goods are nonrandom goods. What is demanded for them does not depend on a usual random process. Only a degenerate random process implicitly appears. Only a degenerate probability distribution is implicitly handled. We do not deal with a prevision, but we deal with a prediction. In other words, given a finite number of possible alternatives, a prediction always reduces to the choice of a point in the set of possible alternatives, and not the barycenter of masses distributed over this set. To choose the barycenter of masses distributed over this set is that which characterizes a prevision. In our opinion, it is necessary to make explicit the latter process with respect to choices being made under claimed conditions of certainty.

2.   Reductions of dimension are considered in this paper. Hence, we pass from m to 1. Accordingly, we pass from m^2 to 2. Regarding reductions of dimension, a theorem has elsewhere been proved by us. The paper containing this theorem is currently under review by an international journal.

3.   Given (x_1, x_2), we first handle a closed neighborhood of x_1 denoted by $[x_1 - \epsilon\,;\,x_1 + \epsilon']$ on the horizontal axis, as well as a closed neighborhood of x_2 denoted by $[x_2 - \epsilon\,;\,x_2 + \epsilon']$ on the vertical one, where both ϵ and ϵ' are two small positive quantities. Since the state of information and knowledge associated with a given decision-maker is assumed to be incomplete at the time of choice, m possible quantities which can be demanded for good 1 belong to $[x_1 - \epsilon\,;\,x_1 + \epsilon']$ and m possible quantities which can be demanded for good 2 belong to $[x_2 - \epsilon\,;\,x_2 + \epsilon']$. These quantities belong to two one-dimensional convex sets. One of m possible alternatives does not need to coincide with x_1. The same is true regarding x_2. It follows that m^2 possible quantities which can be demanded for good 1 and good 2 are handled. After determining $\{x_1^1, \ldots, x_1^m\}$, $\{x_2^1, \ldots, x_2^m\}$, and $\{x_1^1, \ldots, x_1^m\} \times \{x_2^1, \ldots, x_2^m\}$, two nonparametric marginal distributions of mass together with a nonparametric joint distribution of mass are estimated in such a way that (x_1, x_2) is their chosen summary. m possible quantities which can be demanded for good 1 are found between zero and the horizontal intercept of the budget line, whereas m possible quantities which can be demanded for good 2 are found between zero and the vertical intercept of it.

4.   This element is not directly visible because it is a real number. It appears as a two-dimensional point belonging to the two-dimensional convex set. The latter is the budget set of the decision-maker. The budget set of the decision-maker is, therefore, a right triangle belonging to the first quadrant of a two-dimensional Cartesian coordinate system, where the vertex of the right angle of the triangle taken into account coincides with the point given by $(0, 0)$.

5.   We do not use the term "random variable", but we use the term "random quantity" because to say random variable might suggest that we are thinking of the statistical interpretation of repeated events, where many trials in which the random quantity under consideration can vary are involved. The random quantity taken into account could assume different values from trial to trial according to the statistical interpretation of repeated events, but this interpretation is contrary to our way of understanding the problem. We do not use the word event in a generic sense. In this paper, an event is always a single event. The sense of it is not generic, but it is specific. A nonparametric distribution of probability characterizing a random quantity can vary from individual to individual. It can also vary with the state of information and knowledge associated with a given individual.

6.   Since a larger space containing points that are already known to be impossible is always considered by us within this context, if a set is empty, then it is empty of possible points.

7.   A unique symbol $\mathbf{P}$ denotes both probability and prevision, thus avoiding duplication. This is because we use the indicator of an event E expressed by $|E|$. The indicator of E is a random quantity I_E taking values 1 or 0 whenever uncertainty ceases. The mathematical expectation or prevision of the indicator of an event E is denoted by $\mathbf{M}(I_E)$. Since the mathematical expectation of the indicator of an event E is equal to the probability of the same event, we write $\mathbf{M}(I_E) = \mathbf{P}(E)$. If we write $\mathbf{M}(I_E) = \mathbf{P}(E)$, then we must observe $\mathbf{P}(E) = \mathbf{P}(E)$. It follows that a unique symbol $\mathbf{P}$ can be used.

8.   If $\mathbf{x}$ is a vector belonging to E^m, then all collinear vectors regarding $\mathbf{x}$ are expressed by $\lambda\,\mathbf{x}$, $\forall \lambda \in \mathbb{R}$.

9.   The prevision bundle $(\mathbf{P}(_1X), \mathbf{P}(_2X))$ is nothing but the object of decision-maker choice under conditions of uncertainty and riskiness.

10.   In our approach, to consider larger spaces containing, in addition, impossible points in the light of more recent information and knowledge is never wrong. With respect to $[(m+1) \cdot (m+1)]$ points dealt with by the function denoted by $_1X_2X$, only $m^2 + 2$ points of them are really uncertain. Thus, there are points in which the evaluation of the probability is predetermined, rather than permitting the subjective choice of any value in the interval from 0 to 1, endpoints included.

11.   Given the masses of all possible values which are finite in number, their barycenter is a function of them. With regard to a double risky asset, we are not interested in establishing its exact distribution, but we are interested in knowing its barycenter. Whenever

an aggregate choice is studied, the notion of the barycenter of masses is extended together with its properties which are stable equilibrium and minimum of the moment of inertia. The same is true regarding a multiple risky asset of order greater than 2.

References

Abdellaoui, Mohammed, Han Bleichrodt, Olivier l'Haridon, and Corina Paraschiv. 2013. Is there one unifying concept of utility? An experimental comparison of utility under risk and utility over time. *Management Science* 59: 2153–69. [CrossRef]

Angelini, Pierpaolo, and Fabrizio Maturo. 2020. Non-parametric probability distributions embedded inside of a linear space provided with a quadratic metric. *Mathematics* 8: 1901. [CrossRef]

Angelini, Pierpaolo, and Fabrizio Maturo. 2021a. The consumer's demand functions defined to study contingent consumption plans. *Quality & Quantity* 56: 1159–75. [CrossRef]

Angelini, Pierpaolo, and Fabrizio Maturo. 2021b. Summarized distributions of mass: A statistical approach to consumers' consumption spaces. *Journal of Intelligent & Fuzzy Systems* 41: 3093–105.

Angelini, Pierpaolo, and Fabrizio Maturo. 2022a. Jensen's inequality connected with a double random good. *Mathematical Methods of Statistics* 31: 74–90. [CrossRef]

Angelini, Pierpaolo, and Fabrizio Maturo. 2022b. The price of risk based on multilinear measures. *International Review of Economics and Finance* 81: 39–57. [CrossRef]

Anscombe, Francis J., and Robert J. Aumann. 1963. A definition of subjective probability. *The Annals of Mathematical Statistics* 34: 199–205. [CrossRef]

Berkhouch, Mohammed, Ghizlane Lakhnati, and Marcelo Brutti Righi. 2018. Extended Gini-type measures of risk and variability. *Applied Mathematical Finance* 25: 295–314. [CrossRef]

Berti, Patrizia, and Pietro Rigo. 2002. On coherent conditional probabilities and disintegrations. *Annals of Mathematics and Artificial Intelligence* 35: 71–82.

Berti, Patrizia, Eugenio Regazzini, and Pietro Rigo. 2001. Strong previsions of random elements. *Statistical Methods and Applications (Journal of the Italian Statistical Society)* 10: 11–28. [CrossRef]

Capotorti, Andrea, Giulianella Coletti, and Barbara Vantaggi. 2014. Standard and nonstandard representability of positive uncertainty orderings. *Kybernetika* 50: 189–215. [CrossRef]

Cassese, Gianluca, Pietro Rigo, and Barbara Vantaggi. 2020. A special issue on the mathematics of subjective probability. *Decisions in Economics and Finance* 43: 1–2. [CrossRef]

Chambers, Christopher P., Federico Echenique, and Eran Shmaya. 2017. General revealed preference theory. *Theoretical Economics* 12: 493–511. [CrossRef]

Cherchye, Laurens, Thomas Demuynck, and Bram De Rock. 2018. Normality of demand in a two-goods setting. *Journal of Economic Theory* 173: 361–382. [CrossRef]

Cheridito, Patrick, and Eduard Kromer. 2013. Reward-risk ratios. *Journal of Investment Strategies* 3: 3–18. [CrossRef]

Chudjakow, Tatjana, and Frank Riedel. 2013. The best choice problem under ambiguity. *Economic Theory* 54: 77–97. [CrossRef]

Coletti, Giulianella, Davide Petturiti, and Barbara Vantaggi. 2016. When upper conditional probabilities are conditional possibility measures. *Fuzzy Sets and Systems* 304: 45–64. [CrossRef]

Dowd, Kevin 2000. Adjusting for risk: An improved Sharpe ratio. *International Review of Economics & Finance* 9: 209–22.

Echenique, Federico 2020. New developments in revealed preference theory: Decisions under risk, uncertainty, and intertemporal choice. *Annual Review of Economics* 12: 299–316. [CrossRef]

Ekren, Ibrahim, and Johannes Muhle-Karbe. 2019. Portfolio choice with small temporary and transient price impact. *Mathematical Finance* 29: 1066–115. [CrossRef]

Furman, Edward, Ruodu Wang, and Ričardas Zitikis. 2017. Gini-type measures of risk and variability: Gini shortfall, capital allocation, and heavy-tailed risks. *Journal of Banking & Finance* 83: 70–84.

Gerstenberger, Carina, and Daniel Vogel. 2015. On the efficiency of Gini's mean difference. *Statistical Methods & Applications* 24: 569–96.

Gilio, Angelo, and Giuseppe Sanfilippo 2014. Conditional random quantities and compounds of conditionals. *Studia logica* 102: 709–29. [CrossRef]

Halevy, Yoram, Dotan Persitz, and Lanny Zrill. 2018. Parametric recoverability of preferences. *Journal of Political Economy* 126: 1558–93. [CrossRef]

Herdegen, Martin, and Nazem Khan. 2022. Mean-ρ portfolio selection and ρ-arbitrage for coherent risk measures. *Mathematical Finance* 32: 226–72. [CrossRef]

Jasso, Guillermina 1979. On Gini's mean difference and Gini's index of concentration. *American Sociological Review* 44: 867–70. [CrossRef]

Ji, Ran, Miguel A. Lejeune, and Srinivas Y. Prasad. 2017. Properties, formulations, and algorithms for portfolio optimization using Mean-Gini criteria. *Annals of Operations Research* 248: 305–43. [CrossRef]

Jurado, Kyle, Sydney C. Ludvigson, and Serena Ng. 2015. Measuring uncertainty. *American Economic Review* 105: 1177–216. [CrossRef]

La Haye, Roberta, and Petr Zizler. 2019. The Gini mean difference and variance. *Metron* 77: 43–52. [CrossRef]

Li, Ping, Yingwei Han, and Yong Xia. 2016. Portfolio optimization using asymmetry robust mean absolute deviation model. *Finance Research Letters* 18: 353–362. [CrossRef]

Machina, Mark J. 1987. Choice under uncertainty: Problems solved and unsolved. *Journal of Economic Perspectives* 1: 121–154. [CrossRef]

Markowitz, Harry 1952. The utility of wealth. *Journal of Political Economy* 60: 151–8. [CrossRef]

Maturo, Fabrizio, and Pierpaolo Angelini. 2023. Aggregate bound choices about random and nonrandom goods studied via a nonlinear analysis. *Mathematics* 11: 2498. [CrossRef]

Nishimura, Hiroki, Efe A. Ok, and John K.-H. Quah. 2017. A comprehensive approach to revealed preference theory. *American Economic Review* 107: 1239–63. [CrossRef]

Nunke, Ronald J., and Leonard J. Savage. 1952. On the set of values of a nonatomic, finitely additive, finite measure. *Proceedings of the American Mathematical Society* 3: 217–18.

Oderinu, Razaq A., Johnson A. Owolabi, and Musilimu Taiwo. 2023. Approximate solutions of linear time-fractional differential equations. *Journal of Mathematics and Computer Science* 29: 60–72. [CrossRef]

Pfanzagl, Johann 1967. Subjective probability derived from the Morgenstern-von Neumann utility theory. In *Essays in Mathematical Economics in Honor of Oskar Morgenstern*. Edited by Martin Shubik. Princeton: Princeton University Press, pp. 237–51.

Pham, Huyên, Xiaoli Wei, and Chao Zhou. 2022. Portfolio diversification and model uncertainty: A robust dynamic mean-variance approach. *Mathematical Finance* 32: 349–404. [CrossRef]

Pompilj, Giuseppe 1957. On intrinsic independence. *Bulletin of the International Statistical Institute 35*: 91–97.

Regazzini, Eugenio 1985. Finitely additive conditional probabilities. *Rendiconti del Seminario Matematico e Fisico di Milano 55*: 69–89. [CrossRef]

Rockafellar, R. Tyrrell, Stan Uryasev, and Michael Zabarankin. 2006. Generalized deviations in risk analysis. *Finance and Stochastics* 10: 51–74. [CrossRef]

Samuelson, Paul A. 1948. Consumption theory in terms of revealed preference. *Economica* 15: 243–53. [CrossRef]

Schmeidler, David 1989. Subjective probability and expected utility without additivity. *Econometrica* 57: 571–87. [CrossRef]

Scholz, Hendrik 2007. Refinements to the Sharpe ratio: Comparing alternatives for bear markets. *Journal of Asset Management* 7: 347–57. [CrossRef]

Shalit, Haim, and Shlomo Yitzhaki. 2005. The mean-Gini efficient portfolio frontier. *The Journal of Financial Research* 28: 59–75. [CrossRef]

Varian, Hal R. 1982. The nonparametric approach to demand analysis. *Econometrica* 50: 945–73. [CrossRef]

Varian, Hal R. 1983. Non-parametric tests of consumer behaviour. *The Review of Economic Studies* 50: 99–110. [CrossRef]

Viscusi, W. Kip, and William N. Evans. 2006. Behavioral probabilities. *Journal of Risk and Uncertainty* 32: 5–15. [CrossRef]

von Neumann, John 1936. Examples of continuous geometries. *Proceedings of the National Academy of Sciences of the United States of America* 22: 101–8. [CrossRef] [PubMed]

Wang, Mengyu, Hanumanthrao Kannan, and Christina Bloebaum. 2018. Beyond mean-variance: The Mean-Gini approach to optimization under uncertainty. *Journal of Mechanical Design* 140: 031401. [CrossRef]

Article

Interconnectedness of Cryptocurrency Uncertainty Indices with Returns and Volatility in Financial Assets during COVID-19

Awad Asiri [1], Mohammed Alnemer [2] and M. Ishaq Bhatti [3,4,*]

[1] College of Business Administration, Jazan University, Jizan 45142, Saudi Arabia; aasiri@jazanu.edu.sa
[2] College of Business Administration, Shaqra University, Shaqra 15526, Saudi Arabia; malnemer@su.edu.sa
[3] UBD School of Business and Economics, Universiti Brunie Darussalam, Gadong BE 1410, Brunei
[4] Brunei & La Trobe Business School, La Trobe University, Melbourne, VIC 3086, Australia
[*] Correspondence: i.bhatti@latrobe.edu.au

Abstract: This paper investigates the dynamic relationship between cryptocurrency uncertainty indices and the movements in returns and volatility across spectrum of financial assets, comprising cryptocurrencies, precious metals, green bonds, and soft commodities. It employs a Time-Varying Parameter Vector Autoregressive (TVP-VAR) connectedness approach; the analysis covers both the entire sample period spanning August 2015 to 31 December 2021 and the distinct phase of COVID-19 pandemic. The findings of the study reveal the interconnectedness of returns within these asset classes during the COVID-19 pandemic. In this context, cryptocurrency uncertainty indices emerge as influential transmitters of shocks to other financial asset categories and it significantly escalates throughout the crisis period. Additionally, the outcomes of the study imply that during times of heightened uncertainty, exemplified by events such as the COVID-19 pandemic, the feasibility of portfolio diversification for investors might be constrained. Consequently, the amplified linkages between financial assets through both forward and backward connections could potentially compromise financial stability. This research sheds light on the impact of cryptocurrency uncertainty on the broader financial market, particularly during periods of crisis. The findings have implications for investors and policymakers, emphasizing the need for a comprehensive understanding of the interconnectedness of financial assets and the potential risks associated with increased interdependence. By recognizing these dynamics, stakeholders can make informed decisions to enhance financial stability and manage portfolio risk effectively.

Keywords: COVID-19 pandemic; dynamic connectedness; TVP-VAR model; precious metals

JEL Classification: C22; D81; G15

1. Introduction

In recent years, research related to cryptocurrency investment has generated significant debate within the financial sphere. As these digital assets are continuing to grow and gain recognition, comprehending their dynamics and influence on the broader financial landscape is important. Compounding this complexity, the advent of the COVID-19 pandemic and Russia–Ukraine conflict has introduced an additional layer of uncertainty into global financial markets. This study aims to investigate into the interplay between cryptocurrency uncertainty indices and the dynamic movement of returns and volatility in various financial assets during the COVID-19 crisis. Through this paper, we attempt to examine the interdependence among cryptocurrency uncertainty indices, precious metals, green bonds, and soft commodities. Our aim is to investigate the transmission channels and potential spillover effects during times of heightened uncertainty. The findings derived from this study will contribute to the existing literature by fostering a deeper comprehension of the interconnectedness inherent in financial assets. Furthermore, these findings will

provide insights to investors, policymakers, and financial institutions with tools to manage risks and keep financial stability.

The volatility in financial markets is an important factor that plays a significant role in economics, serving as a pivotal risk indicator. A nuanced understanding of the drivers behind volatility across various financial assets holds great importance for stakeholders including academics, investors, regulators, and speculators. This understanding helps in assessing the potential risks that could undermine the stability of the financial system. Stakeholders closely monitor the propagation of both volatility and returns across diverse assets and markets. Although the correlation between returns and volatility has been extensively examined across different financial markets in the existing literature, the spillover effects of cryptocurrencies and their uncertainty on other financial assets have received limited scholarly attention within academia and the financiers. Therefore, there is a need to explore the linkages and interdependencies between cryptocurrency uncertainty indices and the volatility of other financial assets. By doing so, we can foster a deeper comprehension of the intricate dynamics at play within these markets.

The cryptocurrency uncertainty indices play a vital role in enabling investors to gauge uncertainty within the cryptocurrency market, an aspect not fully captured by conventional uncertainty measures such as economic policy uncertainty (Al-Yahyaee et al. 2019; Antonakakis et al. 2013; Demir et al. 2018), VIX volatility index (Alqahtani and Chevallier 2020; Fakhfekh et al. 2021), Investor Attention Index (Smales 2022), and Twitter Economic Uncertainty index (El Khoury and Alshater 2022; Gök et al. 2022). These indices often fall short in accurately reflecting the surge in uncertainty within the cryptocurrency market, a pivotal determinant of asset returns. Notably, heightened uncertainty within the cryptocurrency market directly influences investor returns (Bashir and Kumar 2023). Thus, cryptocurrency uncertainty indices act as instruments that highlight the shift in the cryptocurrency market in response to various events, such as COVID-19 pandemic (Khan et al. 2023). This research sheds light on the impact of cryptocurrency uncertainty on the broader financial market, particularly during periods of crisis. The findings have implications for investors and policymakers, emphasizing the need for a comprehensive understanding of the interconnectedness of financial assets and the potential risks associated with increased interdependence. By recognizing these dynamics, stakeholders can make informed decisions to enhance financial stability and manage portfolio risk effectively.

Nonetheless, the role of cryptocurrencies in the global financial system is increasing every day as this is an important investment asset for most retail and institutional investors. The total market capitalization of cryptocurrencies surpassed USD 1.29T in May 2022. Bitcoin leads the market with a market capitalization of USD 577B and 44% share of the cryptocurrency market. In recent times, Bitcoin's dominance has fallen with the rise in stable coins (Ghabri et al. 2022; Kristoufek 2021; Wang et al. 2020), asset-backed cryptocurrencies (Aloui et al. 2021; Jalan et al. 2021; Yousaf and Yarovaya 2022a), decentralized finance assets (DeFi) (Yousaf et al. 2022; Yousaf and Yarovaya 2022b), and non-fungible tokens (NFTs) (Aharon and Demir 2021; Yousaf and Yarovaya 2022b). Cryptocurrencies exhibited higher volatility in the global COVID-19 crisis, which also affected the cryptocurrency market. The World Health Organization (WHO) declared COVID-19 as global pandemic on 11 March 2020; after this announcement, the Bitcoin price was $3953 on 11 March, and it sharply rose during the pandemic as retail and institutional investors shifted their investments from equity markets to cryptocurrencies and other non-traditional financial assets due to the safe-haven role of cryptocurrencies (Bouri et al. 2020; Corbet et al. 2020b; Rubbaniy et al. 2021a).

The cryptocurrency market is highly volatile; many investors want to invest in the market in the hope of getting higher returns during financial turmoil. During turbulent periods, regulators, policymakers, and investors are interested in observing the return and volatility spillovers for: firstly, decisions about portfolio diversification; and secondly, implementing policies for financial stability. These issues are relevant to the COVID-19 pandemic, when the unemployment rate increased, halting economic activities as economic uncertainty results in financial chaos that disturbed the portfolio asset allocations and

reduced financial stability. The COVID-19 pandemic hugely disrupted financial markets and affected all economy sectors, which ultimately triggered the global recession. With the increase in systematic risk during the COVID-19 outbreak, market participants were interested in obtaining information about volatility transmission among various financial assets for portfolio diversification. Investors re-balanced their portfolios during the financial turmoil by switching from risky to safe-haven assets (Bouri et al. 2021b; Choudhry et al. 2015; Ha and Dai 2022; Khan et al. 2023; Ghouse et al. 2023).

The existing literature largely ignored the interaction of cryptocurrencies and their interaction with other relatively safe traditional financial assets. The COVID-19 pandemic also changed the co-movements between cryptocurrencies and traditional assets. Hence, this study focuses on uncovering the drivers of cryptocurrencies and traditional asset return and volatility spillovers as information transmission among financial markets is extensively studied (Bação et al. 2018; Forbes and Rigobon 2002; Kurka 2019). Existing studies also discussed the connectedness of financial assets during the financial crisis. The important works of Diebold and Yilmaz (2009, 2012, 2014) developed a quantitative measure of dynamic connectedness based on forecast error variance decomposition using VAR models. We try to contribute to the relevant literature by investigating dynamic connectedness of different assets and cryptocurrency uncertainty indices. Specifically, we are interested in examining the dynamic connectedness during the COVID-19 shock by considering its time-varying structure. Thus, we try to answer the following questions: What role has the COVID-19 pandemic played in exhibiting the return and cryptocurrency uncertainty connectedness of different financial assets? Are the asset returns time-varying in nature? Do the cryptocurrency uncertainty indices explain the return and volatility connectedness among financial assets?

We used indices for cryptocurrencies, precious metals, green bonds, and soft commodities and cryptocurrency uncertainty to apply the time-varying parameter vector auto-regressions (TVP-VAR) dynamic connectedness approach to answer the above research questions. To the best of our knowledge, this is the first study of its kind on the dynamic connectedness of returns and volatility of different assets during a financial crisis (Adekoya and Oliyide 2021; Bouri et al. 2021a; Corbet et al. 2020a; Kamal and Hassan 2022; So et al. 2020). This paper also investigates the response of financial assets to the COVID-19 pandemic by extending published studies conducted in different financial markets (Adekoya and Oliyide 2021; Baig et al. 2020; Le et al. 2021b; Rubbaniy et al. 2021b). The usage of the TVP-VAR approach by Antonakakis et al. (2020) overcomes the shortcomings (e.g., outlier sensitivity, short time, rolling window size) of the original connectedness approach by Diebold and Yilmaz (2009, 2012, 2014). The TVP-VAR approach also serves to measure cross-asset connectedness in the network.

Furthermore, existing studies have discussed the role of various uncertainty indices in shaping the dynamics of cryptocurrency returns and volatility. These different measures of uncertainty encompass the economic policy uncertainty index (Elsayed et al. 2022a; Foglia and Dai 2021; Yen and Cheng 2021), Twitter-based uncertainty index (Aharon et al. 2022; Wu et al. 2021), and the economic and political uncertainty (Colon et al. 2021; Kyriazis 2021) and cryptocurrency uncertainty indices (Elsayed et al. 2022b; Lucey et al. 2022).

In a recent study, Yousaf and Goodell (2023) investigated the central banks' digital currencies (CDBC), cryptocurrency policy uncertainty index as well as digital payments stocks by using the dynamic connectedness approach. Their findings highlight the transmission of shocks from UCRY policy and price to digital payment stocks. Moreover, they identified the limited interconnectedness between cryptocurrency uncertainty indices and digital payment stocks, indicating their potential hedging tools against cryptocurrency market volatility. Yan et al. (2022) investigated the impact of cryptocurrency uncertainties on sustainable and traditional mutual fund and found that traditional mutual funds' investments are influenced by uncertainty in the cryptocurrencies market.

Wei et al. (2023) delved into safe-haven properties of cryptocurrencies and forecasting ability of cryptocurrency uncertainty indices for volatility in precious metals. Employ-

ing the GARCH-MIDAS approach, their results show that the forecasting prowess of cryptocurrency uncertainty indices in the precious metals market.

In our contribution to the field, we leverage novel cryptocurrency uncertainty indices in conjunction with various asset classes—namely cryptocurrencies, precious metals, green bonds, and soft commodities. Furthermore, we extend the work initiated by Elsayed et al. (2022b) on dynamic connectedness between gold, cryptocurrency index and cryptocurrency uncertainty indices. Our findings show the higher returns and volatility connectedness in the overall sample and during the COVID-19 pandemic, and most financial assets are net receivers of shocks. The patterns of return and volatility spillover changed during the pandemic for most financial assets. Overall, our findings suggest that cryptocurrency uncertainty indices and transmitters of shocks extend to other financial assets. The COVID-19 pandemic resulted in a spike of risk in financial markets and the magnitude of dynamic connectedness increased during the first wave of COVID-19 which is like the finding of Bhatti and Ghouse (2023). Hence, risk-averse equity market investors can minimize such risks by investing in less-connected assets to diversify portfolios. The remainder of the study is as follows. Section 2 presents the relevant literature review. In Section 3, we describe the methodology and data. In Section 4, we show the findings of the study. Finally, Section 5 concludes the study.

2. Literature Review

Studies have discussed the return and volatility transmission across financial assets using different methods, for instance Granger causality (Adekoya and Oliyide 2021; Albulescu et al. 2019; Zhang and Broadstock 2020) and dynamic conditional correlation (Abuzayed and Al-Fayoumi 2021; Hassan et al. 2019). However, the existing literature highlights the usefulness of connectedness of financial assets using the dynamic connectedness approach (Shahzad et al. 2021a, 2021b). The higher inter-connectedness among financial assets indicates greater market risk, and investors minimize their risks by investing in weakly connected financial assets. The higher market risk in the network explains the instability of the financial markets. Dynamic return connectedness is used to identify the isolated assets so that these assets function as hedge or safe haven against the risk of other financial assets.

The literature on return and volatility connectedness among different financial assets, such as equity, bonds, and commodities, are scarce. Some authors studied the link between commodities, currency, and equity markets. For instance, Kang et al. (2017) studied the price transmission among crude oil, agricultural commodities and precious metals using the DECO-GARCH model. They detected an increase in spillover impacts during the financial crisis. Lundgren et al. (2018) also tested the connectedness and causality by using equities, currencies, oil, and US treasury bonds, as well as different proxies of uncertainty (EU and US EPU and VIX) using data from 2004–2016, and they found the uncertainty proxies were net transmitters of shocks during the financial crisis.

Mensi et al. (2017) investigated the spillovers between gold, Dow Jones, conventional, Islamic, technology, financial, and telecommunications sector and sustainable indices. These authors found that gold, energy, technology and telecom sectors and receivers of shocks and Dow Jones indices contribute to risk spillovers. Yoon et al. (2019) investigated dynamic and static returns connectedness among equity, bond, commodity, and currency markets. They identified the Shanghai stock exchange, Nikkei 225, and KOSPI are receivers of spillover shocks. Kumar et al. (2019) investigated volatility and correlation between stock prices, natural gas, and oil in India via the VARMA-DCC GARCH models. Their findings highlights highest short-term spillovers between oil and natural gas. Iglesias-Casal et al. (2020) discussed the volatility spillovers and diversification potential of oil, gold and clean energy indices in Brazil by using BEKK and A-DCC models. They emphasized gold's higher diversification potential and optimal portfolio weights.

A recent study by Mensi et al. (2020) explored the risk spillovers between energy futures and precious metals, noting increased volatility spillovers during the financial crisis. They observed that gold and oil transmit volatility to other financial assets.

Bouri et al. (2021a) explored return connectedness with crude oil, equities, currencies, and bonds during the COVID-19 pandemic using the TVP-VAR approach. They observed changes in connectedness network's structure and identified equity and USD indices as shock transmitters before COVID-19, while bond indices become the volatility shock transmitter during the outbreak.

Asl et al. (2021) analyzed volatility transmission between clean energy indices and energy commodities using an asymmetric BEKK-MGARCH(1,1) model. They found higher optimal weights and hedging effectiveness for clean energy indices, making them useful for hedging equity risks in the energy sector. They concluded that investors can invest in green assets to hedge the equity risk for stocks in energy sector. Further, Szczepanska-Przekota (2021) explored the impact of cryptocurrencies on economic conditions of different markets, and found that investors perceive the cryptocurrency market as more risky as compared to equity markets.

Shahid et al. (2023) explored the interconnectedness and risk transmission across global financial markets and assessed the portfolio diversification potential of socially responsible investments using DCC-GARCH and VAR-GARCH models. Their findings indicated negative correlation between traditional volatility indices and socially responsible investment indices. They also found that implied volatility indices of silver and golds hedge the risks against SRIs investments. Furthermore, Elsayed et al. (2022b) extended the above research by examining the return and volatility spillovers in gold, cryptocurrency index, and cryptocurrency price and policy uncertainty indices. They found that cryptocurrency policy uncertainty is the transmitter of shocks to other assets while gold is the receiver of shocks. We extend the above research by investigating the dynamic returns and volatility spillovers among various financial assets and cryptocurrency uncertainty indices developed by Lucey et al. (2022). The news-based uncertainty indices are relevant to cryptocurrencies and can better predict uncertainty in the cryptocurrency market.

3. Research Methodology

3.1. The Data

To test and study the dynamic connectedness of different financial asset returns and cryptocurrency price and policy uncertainty index constructed by Lucey et al. (2022), we collected the weekly data of cryptocurrency uncertainty indices from the authors' website[1]. The cryptocurrency uncertainty indices were constructed using news articles related to cryptocurrency on the Lexis Nexis database. We also gathered the daily closing price of Bitcoin and Ethereum from the Coin Market Cap website[2]. The closing price data for precious metals (gold, silver, platinum), S&P green bonds, and S&P GSCI soft commodities index were downloaded from the DataStream database provided by Thomson Reuters[3]. The final sample includes data from 7 August 2015 to 31 December 2021.

In the next step, we converted the daily closing prices of financial assets into log returns that were further converted into weekly returns to estimate dynamic returns and volatility connectedness at a weekly frequency. The dynamic connectedness requires that all series follow non-stationary unit root test processes. Hence, for implementing the dynamic connectedness approach, we transformed the data using the first log-difference of series: $yit = log(xit) - log(xit - 1)$. The selected financial assets are relatively stable during the period of extreme volatility (Le et al. 2021b; Mo et al. 2022; Su et al. 2022; Umar et al. 2021) and essential for the stability of financial markets due to their volatility to other markets. Hence, it would be interesting to study the connection between these financial assets with cryptocurrency uncertainty. In addition, well, our dataset includes data for the period of the COVID-19 pandemic, which is useful to observe the returns and volatility connectedness during it. We offer a snapshot of the data in the following Table 1, which includes the

descriptive statistics of cryptocurrency indices, cryptocurrencies, precious metals, green bonds, and soft commodities indices.

Table 1. Descriptive Statistics of Sample.

	Panel A: Descriptive Statistics Full Sample (7 August 2015 to 31 December 2021).								
Variables	UCRY_Policy	UCRY Price	Bitcoin	Ethereum	Gold	Silver	Platinum	SP Green Bonds	SP GSCI Softs
Mean	0.000	0.000	0.024	0.058	0.001	0.003	0.003	0.000	0.001
Variance	0.000	0.000	0.002	0.018	0.000	0.000	0.000	0.000	0.000
Skewness	8.081 ***	6.617 ***	4.795 ***	5.461 ***	4.930 ***	13.001 ***	10.276 ***	9.971 ***	4.146 ***
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ex.Kurtosis	76.285 ***	53.141 ***	29.936 ***	37.572 ***	36.607 ***	193.579 ***	120.734 ***	114.121 ***	28.471 ***
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
JB	84,622.361 ***	41,738.229 ***	13,751.343 ***	21,305.339 ***	20,002.199 ***	530,908.103 ***	208,736.085 ***	186,779.394 ***	12,237.996 ***
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ERS	−5.437 ***	−4.707 ***	−6.351 ***	−0.871	−6.624 ***	−5.640 ***	−6.281 ***	−4.318 ***	−4.953 ***
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q(10)	91.739 ***	126.632 ***	45.933 ***	63.807 ***	52.338 ***	52.491 ***	89.844 ***	124.134 ***	22.726 ***
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Q2(10)	15.345 ***	8.362	3.844	35.637 ***	13.855 ***	12.043 **	65.188 ***	70.642 ***	7.058
	−0.005	−0.146	−0.69	0	−0.01	−0.026	0	0	−0.248
	Panel B: Descriptive Statistics Full Sample COVID-19 (1 January 2020 to 31 December 2021)								
Mean	0.000	0.000	0.022	0.041	0.001	0.005	0.005	0.000	0.002
Variance	0.000	0.000	0.002	0.006	0.000	0.000	0.000	0.000	0.000
Skewness	4.606 ***	3.574 ***	7.074 ***	4.594 ***	4.417 ***	7.727 ***	5.964 ***	6.089 ***	3.909 ***
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Ex.Kurtosis	23.186 ***	14.875 ***	59.241 ***	26.666 ***	26.720 ***	64.121 ***	38.122 ***	38.670 ***	20.649 ***
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
JB	2697.329 ***	1180.259 ***	16,075.460 ***	3447.105 ***	3432.009 ***	18,851.366 ***	6914.146 ***	7122.527 ***	2112.400 ***
	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ERS	−3.596 ***	−2.898 ***	−3.750 ***	−3.743 ***	−3.391 ***	−4.079 ***	−3.402 ***	−4.030 ***	−3.774 ***
	−0.001	−0.005	0.000	0.000	−0.001	0.0000	−0.001	0.000	0.000
Q (10)	17.926 ***	19.297 ***	10.909 **	11.869 **	25.382 ***	16.323 ***	26.846 ***	42.238 ***	13.134 **
	−0.001	−0.001	−0.045	−0.028	0	−0.003	0.000	0.000	−0.015
Q2 (10)	4.088	1.439	0.644	2.008	5.521	3.597	19.948 ***	21.665 ***	2.264
	−0.651	−0.977	−0.998	−0.937	−0.43	−0.73	0.000	0.000	−0.912

Notes: The symbols ***, ** indicate significance at the 1%, 5% levels; the D'Agostino (1970) skewness test, Anscombe and Glynn (1983) kurtosis test, Jarque and Bera (1980) normality test, Elliott et al. (1996) ERS unit-root test, *Q* (10) & *Q2*(10), and Fisher and Gallagher (2012) weighted portmanteau tests are applied to the dataset.

Table 1 panel A includes the descriptive statistics of the whole sample (7 August 2015, to 31 December 2021) and panel B includes the data covering the COVID-19 pandemic (1 January 2020 to 31 December 2021). In panel A of Table 1, the full sample results show that average returns are positive for most of the series. From the selected financial assets, Bitcoin provides higher returns with a value of 0.058 and, unit-root test processes during the COVID-19 pandemic, gold, silver, platinum, and S&P GSCI soft commodities are increased. Gold returns remain stable in the overall sample during the COVID-19 pandemic. The returns of cryptocurrencies are reduced during the crisis with the average values of Bitcoin and Ethereum being 0.024 and 0.058, respectively, in the overall sample compared to average values of 0.022 and 0.041 during COVID-19. Overall, the weekly returns of these financial assets are not negative and provide better returns to investors in cryptocurrencies and commodities markets in the presence of cryptocurrencies uncertainty.

The difference between Bitcoin and Ethereum is higher in the full sample and during COVID-19, which shows that cryptocurrencies are riskier than precious metals, green bonds, and soft commodities. Further, the positive and significant rightward skewed returns series show that the mean is higher than median in different financial assets used in this study. The kurtosis and Jarque–Bera normality tests confirm that all returns series have fat tails and follow the leptokurtic distribution. Results support the non-normality of the data in line with Jarque and Bera (1980), in which they show that all financial assets

are not normally distributed. The stationarity is tested utilizing the Elliott et al. (1996) ERS unit root test, which shows that all returns series are significant and stationary at the 1% level of significance. Finally, we also checked the goodness-of-fit of financial time series using the Fisher and Gallagher (2012) weighted portmanteau test that is significant at 1% in most cases. It shows the autocorrelation between returns and squared returns is useful for examining the interconnectedness of these financial assets using the TVP-VAR dynamic connectedness approach.

The weekly log returns on financial assets and cryptocurrency uncertainty indices are displayed in Figure 1 below. We take the natural log by following (Demir et al. 2018; Hasan et al. 2021; Xu et al. 2023). Shown here is that prices of cryptocurrencies, precious metals, green bonds, and soft commodities indices show a sharp reduction during the first phase of the COVID-19 pandemic. Furthermore, the cryptocurrency policy and price uncertainty indices rapidly increased as the COVID-19 crisis progressed.
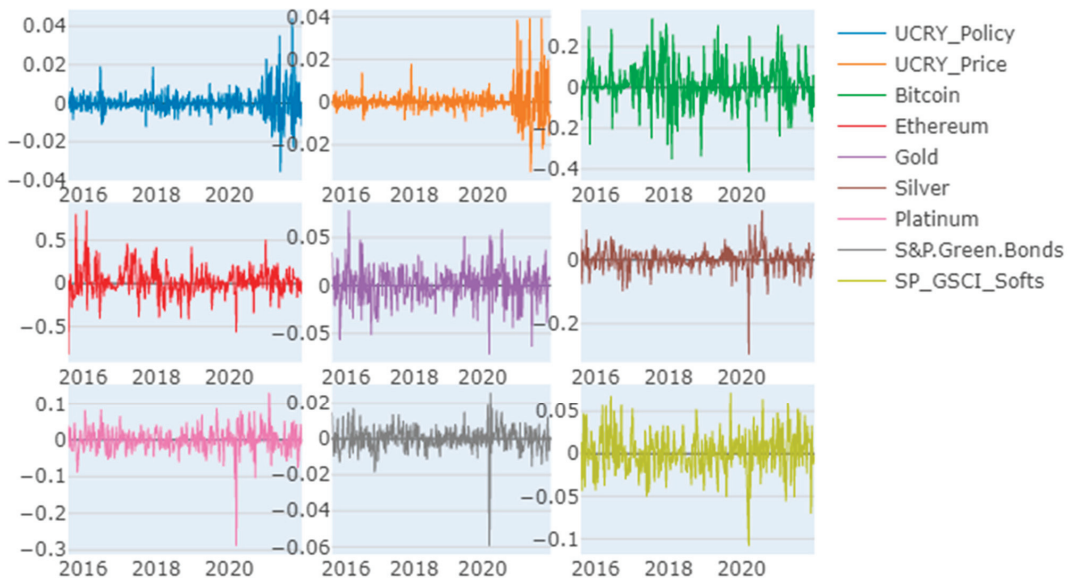
Figure 1. Log returns data of cryptocurrencies, precious metals, green bonds, and soft commodities.
Notes: Figure 1 displays the weekly log returns of cryptocurrencies, precious metals, green bonds, and soft commodities.

3.2. The Model

We investigated the impact of cryptocurrency uncertainty indices on the return and volatility connectedness with the Time-Varying Parameter Autoregressive (TVP-VAR) dynamic connectedness approach developed by Antonakakis et al. (2020). Their approach is essentially an extension of Diebold and Yilmaz (2009, 2012, 2014). Selecting this econometric method is based on existing studies for testing the dynamic connectedness among financial markets. This method is useful when correlation among financial markets rises during times of financial turmoil. The dynamic conditional correlation models provide biased results during the crisis as they are based on market state and macroeconomic factors, yet the real connectedness among markets is not affected by the financial turbulence. In contrast, the spillover approach is based on Forecast Error Variance Decomposition (FEVD), which originated from the Vector Autoregressive model (VAR) model and not affected by the conditional correlation (Elsayed et al. 2022b; Elsayed and Helmi 2021; Umar et al. 2022).

The spillover approach devised by Diebold and Yilmaz (2009, 2012, 2014) is used to obtain the information about shock in one variable to another variable based on FEVD.

However, the spillover method has several drawbacks: first, the arbitrary selection of rolling-window size is not useful in small samples; second, original connectedness approach by Diebold and Yilmaz (2009, 2012, 2014) is highly sensitivity to outliers and based on normalization techniques, which may provide biased results (Caloia et al. 2019); and third, the original dynamic connectedness approach may produce biased estimations due to sign and rank errors. These issues are solved using the TVP-VAR dynamic connectedness approach devised by Antonakakis et al. (2020). The window size is ideally estimated by mean squared prediction errors based on the multivariate Kalman Filters (Koop and Korobilis 2013, 2014). This technique is also useful for dealing with outliers, especially during the financial crisis. The short time series, and sign and rank errors are also reduced in the TVP-VAR approach with scaler-based normalization of Generalized Forecast Error Variance.

The return and volatility spillovers are discussed in existing literature using the original dynamic connectedness approach (Diebold and Yilmaz 2009, 2012, 2014), TVP-VAR approach (Adekoya and Oliyide 2021; Bouri et al. 2021a; Dai et al. 2022; Elsayed et al. 2022b) and extended joint connectedness based on TVP-VAR (Balcilar et al. 2021; Chen et al. 2022). Meanwhile the TVP-VAR approach is based on auto-regressive conditional heteroscedasticity (ARCH) process proposed by Engle (1982), and the generalized autoregressive conditional heteroscedasticity (GARCH) approach by Bollerslev (1986) to overcome problems in ARCH models. Consequently, the TVP-VAR approach of Antonakakis et al. (2020) examines the return and volatility transmission across cryptocurrencies, precious metals, green bonds, soft commodities, and cryptocurrency uncertainty indices.

3.3. Research Methods

3.3.1. Time-Varying Parameter Vector Autoregression (TVP-VAR)

The TVP-VAR dynamic connectedness approach of Antonakakis et al. (2020) is applied to assess the dynamic connectedness between various financial assets. The TVP-VAR model with the lag-length of order one is selected by the Bayesian information criterion (BIC), and we choose the rolling window of 52 weeks and weekly returns which are nth-transformed to control the stationarity issues:

$$
\begin{aligned}
y_t &= B_t y_{t-1} + \epsilon_t & \epsilon_t &\sim N(0, \Sigma_t) \\
\text{vec}(B_t) &= \text{vec}(B_{t-1}) + v_t & v_t &\sim N(0, R_t)
\end{aligned}
\tag{1}
$$

where Equation (1) shows that y_t, y_{t-1} and ϵ_t are $K \times 1$ dimensional vector and B_t and Σ_t are $K \times K$ dimensional matrices.

The symbols $\text{vec}(B_t)$ and v_t are $K^2 \times 1$ dimensional vectors, whereas R_t is a $K^2 \times K^2$ dimensional matrix. All parameters (B_t) are allowed to use the TVP-VAR approach, which is also helpful for examining the time-varying relationship and variance-covariance matrices; Σ_t and R_t.

Further, the Wold theorem is applied to transform the model to the TVP-VMA model in

$$
y_t = \sum_{h=0}^{\infty} A_{h,t} \epsilon_{t-i}
\tag{2}
$$

where Equation (2) shows that $A_0 = I_K$ and ϵ_t is a vector of white noise symmetric shocks with $K \times K$ time-varying covariance matrix of $E(\epsilon_t \epsilon_t') = \Sigma_t$.

Therefore, in the next step, the H-step forecast error is estimated in Equation (3);

$$
\begin{aligned}
\xi_t(H) &= y_{t+H} - E(y_{t+H} \mid y_t, y_{t-1}, \ldots) \\
&= \sum_{h=0}^{H-1} A_{h,t} \epsilon_{t+H-h},
\end{aligned}
\tag{3}
$$

with forecast error covariance matrix equal to in Equation (4):

$$
E\big(\xi_t(H)\xi_t'(H)\big) = A_{h,t} \Sigma_t A_{h,t}'.
\tag{4}
$$

3.3.2. The Generalized Dynamic Connectedness Approach

The generalized dynamic connectedness approach is based on the *H*-step ahead of generalized forecast error variance decomposition (GFEVD); $gSOT_{ij,t}$, is also applied and it can be interpreted as the effect the shock in variable *j* has on variable *i*. This process is explained in Equation (5) below:

$$
\begin{aligned}
\zeta_{ij,t}^{gen}(H) \quad &= \frac{E\left(\xi_{i,t}^2(H)\right)-E\left[\xi_{i,t}(H)-E\left(\xi_{i,t}(H)\right)|\epsilon_{j,t+1},\dots,\epsilon_{j,t+H}\right]^2}{E\left(\xi_{it}^2(H)\right)} \\
&= \frac{\sum_{h=0}^{H-1}\left(e_i'A_{ht}\Sigma_t e_j\right)^2}{\left(e_j'\Sigma_t e_j\right)\sum_{h=0}^{H-1}\left(e_i'A_{ht}\Sigma_t A_{ht}'e_i\right)} \\
gSOT_{ij,t} \quad &= \frac{\zeta_{ij,t}^{gen}(H)}{\sum_{j=1}^K \zeta_{ij,t}^{gen}(H)}
\end{aligned}
\tag{5}
$$

where e_i is a $K \times 1$ zero selection vector with unity on its *i*th position and $\zeta_{ij,t}^{gen}(H)$ denotes the proportional reduction of the *H*-step forecast error variance of variable *i* due to conditioning on future shocks of variable *j*.

The generalized dynamic spillover average table displays the total connectedness to demonstrate total connectedness among financial assets from shock in one variable to the whole network. This dynamic connectedness metric is explained in below in Equation (6):

$$
\begin{aligned}
S_{i\leftarrow\cdot,t}^{\text{gen},from} &= \sum_{j=1,i\neq j}^K gSOT_{ij,t} \\
S_{i\rightarrow 0,t}^{\text{gen},to} &= \sum_{j}^K gSOT_{ji,t}
\end{aligned}
\tag{6}
$$

Another measure is the net total directional connectedness of variable *i*, and it displays whether variable *i* influences the network more than being influenced by it and it is presented in Equation (7):

$$
S_{i,t}^{gen,net} = S_{i\rightarrow 0,t}^{gen,to} - S_{i\leftarrow\cdot,t}^{gen,from}
\tag{7}
$$

If $S_{i,t}^{gen,net} > 0\left(S_{i,t}^{gen,net} < 0\right)$, variable *i* is a net transmitter (receiver) of shocks which shows that variable *i* is driving (driven by) the network.

The next metric is TCI is average total directional connectedness from (to) others and we explain it in Equation (8), which is shown here:

$$
gSOI_t = \frac{1}{K}\sum_{i=1}^K S_{i\leftarrow,t}^{gen,from} = \frac{1}{K}\sum_{i=1}^K S_{i\rightarrow 0,t}^{gen,to},
\tag{8}
$$

A high value of average total directional connectedness (TCI) reveals an increased risk in the financial market and its low value highlights the low risk. This means shocks in one variable are influenced by its future values and shocks are not transmitted from one variable to another variables.

Finally, the dynamic connectedness approach provides information about net pairwise spillovers relationship between two variables and it is presented in Equation (9):

$$
S_{ij,t}^{gen,net} = gSOT_{ji,t}^{gen,to} - gSOT_{ij,t}^{gen,from}
\tag{9}
$$

If $S_{ij,t}^{gen,net} > 0\left(S_{ij,t}^{gen,net} < 0\right)$, variable *i* has a higher impact on variable *j* and vice versa, implying that variable *i* dominates variable *j*.

4. Empirical Results

In this section we present the results of the dynamic connectedness approach based on the TVP-VAR approach.

4.1. The Average Dynamic Connectedness

Table 2 displays the average dynamic connectedness based on the TVP-VAR approach. The diagonal elements in Table 2 are associated with their own contribution to volatility spillover while off-diagonal elements refer to the contribution 'from' or 'to' others. The rows are linked with the contribution of each asset and uncertainty index to forecast error variance of specific asset. Conversely, the columns are associated with the impact of shock in one financial asset to all other assets separately.

Table 2. Volatility and return connectedness of cryptocurrency uncertainty indices, cryptocurrencies, precious metals, green bonds, and soft commodities: Evidence using the TVP-VAR approach. **Notes:** TVP-VAR dynamic connectedness approach results with the lag-length of order one by criterion (BIC) with window size of 52 weeks. Panel A includes the dynamic connectedness in full sample, and we tested the dynamic connectedness during COVID-19 in panel B.

Panel A: Average Dynamic Connectedness Table (Full Sample)									
Variables	UCRY Policy	UCRY Price	Bitcoin	Ethereum	Gold	Silver	Platinum	S&P Green Bonds	SP GSCI Softs	FROM
UCRY Policy	43.02	47.18	1.41	0.25	0.78	2.65	2.07	1.86	0.79	56.98
UCRY Price	35.59	58.18	0.7	0.15	0.47	1.77	1.47	1.05	0.62	41.82
Bitcoin	8.35	14.55	44.24	4.25	2.82	6.47	5.69	10.06	3.55	55.76
Ethereum	5.13	6.93	5.29	58.93	7.18	6.2	2.75	5.37	2.23	41.07
Gold	7.02	10.05	3.89	6.49	40.23	10.14	9.25	11.49	1.45	59.77
Silver	7.76	10.84	4.66	4.7	7.73	37.94	17.02	7.12	2.23	62.06
Platinum	15.51	25.9	3.75	1.44	5.17	12.5	27.82	6.31	1.6	72.18
S&P Green Bonds	6.9	9.7	7.44	1.52	3.07	5.64	5.72	54.98	5.04	45.02
SP GSCI Softs	7.82	13.37	4.54	2.61	1.41	4.24	9.29	12.99	43.74	56.26
TO	94.07	138.52	31.66	21.4	28.63	49.6	53.26	56.25	17.51	490.91
Inc.Own	137.09	196.7	75.9	80.33	68.87	87.54	81.08	111.23	61.26	cTCI/TCI
NET	37.09	96.7	−24.1	−19.67	−31.13	−12.46	−18.92	11.23	−38.74	61.36/54.55
NPT	7	8	3	1	1	4	5	6	1	
Panel B: COVID-19 Pandemic (1 January 2020 to 31 December 2021)									
Variables	UCRY Policy	UCRY Price	Bitcoin	Ethereum	Gold	Silver	Platinum	S&P Green Bonds	SP GSCI Softs	FROM
UCRY Policy	47.47	35.4	0.88	3.03	1.62	0.91	3.61	2.62	4.44	52.53
UCRY Price	39.24	45.01	0.92	2.76	1.11	0.91	3.82	2.4	3.85	54.99
Bitcoin	5.4	6.18	18	11.6	4.5	12.28	14.38	24.81	2.85	82
Ethereum	4.6	5.41	10.45	29.06	4.19	11.65	9.57	22.61	2.46	70.94
Gold	4.05	2.5	5.12	7.32	24.81	15.83	16.42	21.19	2.76	75.19
Silver	2.28	3.19	7.78	9.06	8.91	19.39	21.61	25.52	2.26	80.61
Platinum	2.18	2.37	6.2	9.52	7.49	16.11	31.48	21.27	3.37	68.52
S&P Green Bonds	3.1	2.35	2.35	2.71	1.79	4.19	9.11	68.63	5.78	31.37
SP GSCI Softs	12.46	12.94	3.53	3.05	1.74	1.63	9.2	8.27	47.19	52.81
TO	73.3	70.34	37.24	49.04	31.36	63.52	87.71	128.69	27.77	568.97
Inc.Own	120.77	115.35	55.24	78.1	56.16	82.91	119.2	197.31	74.95	cTCI/TCI
NET	20.77	15.35	−44.76	−21.9	−43.84	−17.09	19.2	97.31	−25.05	71.12/63.22
NPT	7	5	2	3	0	3	7	7	2	

The findings of the dynamic connectedness network of weekly returns of cryptocurrencies, precious metals, green bonds, soft commodities, and cryptocurrency uncertainty indices display higher internal connectedness with an average total connectedness index (TCI) value of 54.55%. The value of TCI within the dynamic connectedness network explains the higher interconnectedness of these financial assets and cryptocurrency uncertainty indices. The cryptocurrency price uncertainty index transmits the shocks to other assets in the network with a forecast error variance value of 138.52%.

Similarly, during the COVID-19 pandemic, the TCI value is 63.22%, which is higher than full sample results. Moreover, during the pandemic, green bonds with a value of 128.69% transmit the shock to other assets and uncertainty indices. The connectedness is

increased during the COVID-19 pandemic. The increase in TCI during COVID-19 pandemic suggest that volatility spillover and transmission of risk increase during financial turmoil. These findings are consistent with previous research (Akhtaruzzaman et al. 2021; Boubaker et al. 2016; Bouri et al. 2021b; Costa et al. 2022), as they also find that total connectedness is increased during crisis times.

Overall, these findings strongly suggest that the selected financial assets are closely connected, and shocks are transmitted from cryptocurrency uncertainty indices to other financial assets. Hence, risk-averse investors should take these findings into account for investing in these financial assets, and they can diversify their portfolios by investing in assets with low interconnectedness. Investors should consider cryptocurrency uncertainty before investing in these financial assets, especially during the COVID-19 pandemic.

4.2. The Dynamic Total Connectedness

It is important to note that, in Table 2, results about dynamic connectedness across time are not included. Moreover, we cannot see the connectedness during the Global Financial Crisis (GFC), COVID-19 pandemic, and other influential and/or extreme events. Figure 2 illustrates the dynamic total connectedness (TCI) across time to explain the volatility transmission across financial assets. As shown in Figure 2, the total connectedness is within the 45% to 95% range. The TCI is higher during 2016 and it remained at 55% from 2018 to the first days of 2020. However, the TCI values sharply increase during the first wave of the COVID-19 pandemic, with a TCI value of around 77%.
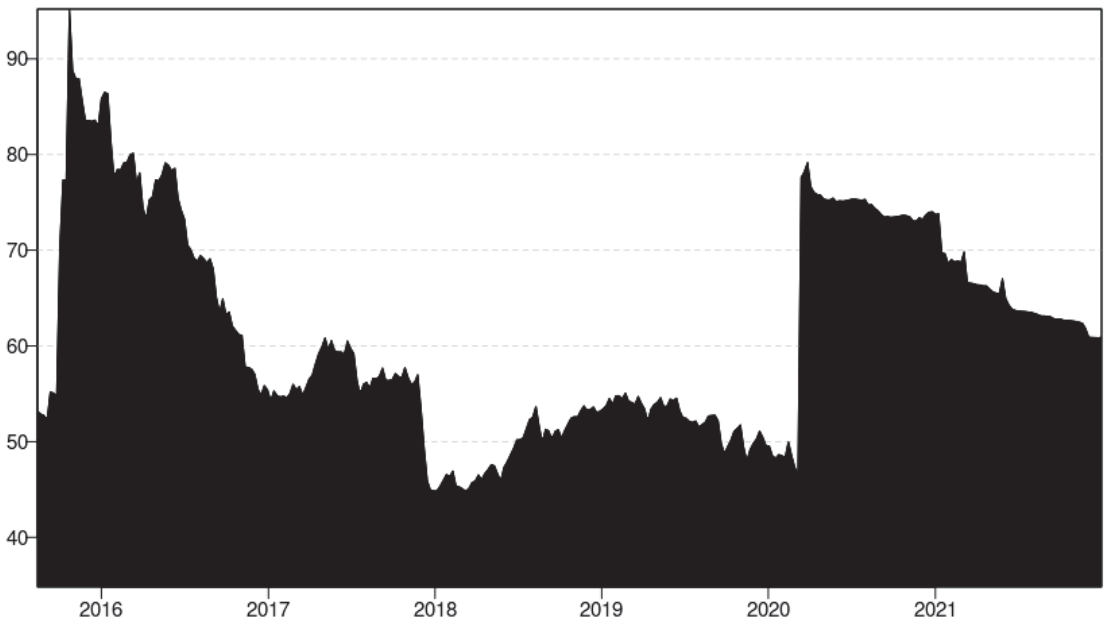
Figure 2. Dynamic total connectedness. **Notes:** Figure 2 shows the results of the TVP-VAR approach with lag-length of order one by BIC criterion and a 20 step-ahead generalized forecast error variance decomposition (FEVD). The black area in the figure represents the total connectedness (TCI).

The consistent presence of the TCI value above 50% across a majority of time frames provides substantial evidence for heightened return and volatility spillover within these financial assets. This trend signifies that elevated cryptocurrency uncertainty indices wield a notable influence over these financial components. The amplified dynamic connectedness values recorded during the COVID-19 pandemic can be attributed to the escalated apprehension among investors triggered by negative news related to the COVID-19 situation. Concurrently,

the surge in economic policy uncertainty further compounds this situation. It is noteworthy that cryptocurrencies were particularly susceptible to changes in the COVID-19 pandemic scenario (Allen 2022; Lahmiri and Bekiros 2020; Salisu and Vo 2020).

4.3. Net Total Directional Connectedness

Net total directional connectedness results are displayed in Figure 3 This figure presents the time-varying role of net receiving or net transmitting role of financial assets. Figure 4 shows that the UCRY policy and price indices remain stable before the COVID-19 pandemic and that the UCRY price index acts as a transmitter of shocks during the first days of COVID-19; these findings are consistent with Lundgren et al. (2018). UCRY policy index is a net receiver of the shocks. As shown in Figure 3, Ethereum, Bitcoin, gold, and soft commodities are net receivers of shocks. The silver, platinum, and green bonds are net receivers of shocks before COVID-19, but they become the net transmitter of shocks during it. Overall, most of the assets are net receivers of shocks and their spillover behavior changes during the pandemic.
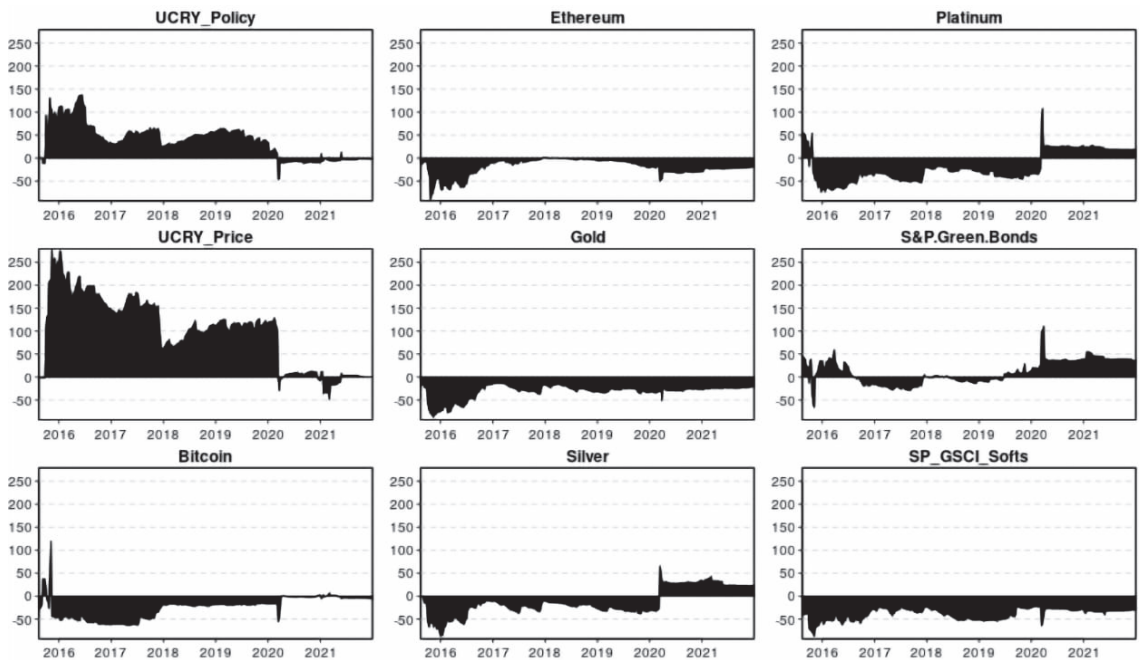
Figure 3. Dynamic net total directional connectedness. **Notes:** Figure 3 shows net total directional connectedness based on the TVP-VAR approach with lag-length of order one by BIC criterion and a 20 step-ahead generalized forecast error variance decomposition (FEVD). The black area represents the net total directional connectedness. Meanwhile the positive values show a net transmitter role, and the negative values indicate the net receiving role of financial assets.
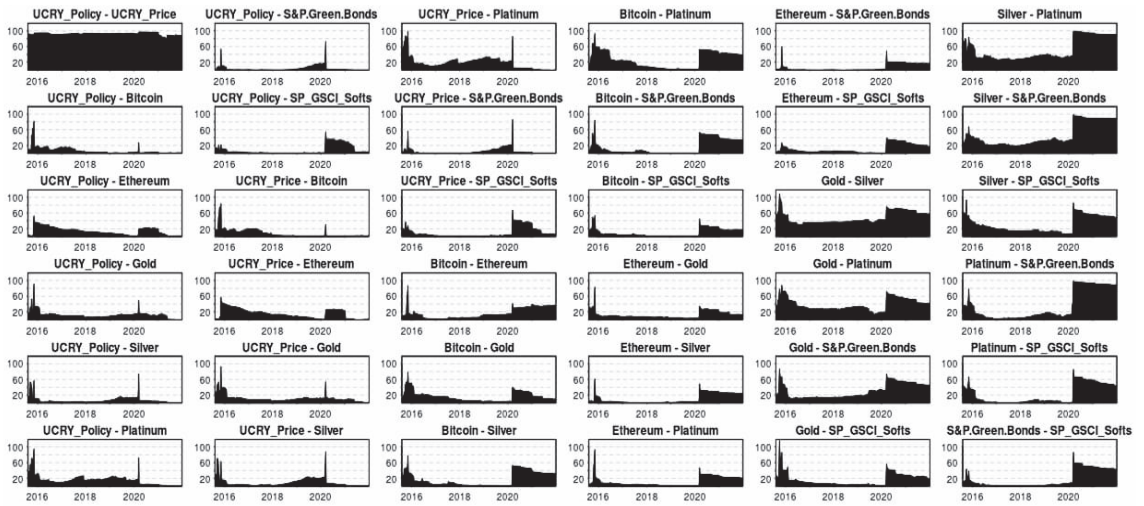
Figure 4. Dynamic net pairwise total directional connectedness. **Notes:** Figure 4 shows net pairwise total directional connectedness based on the TVP-VAR approach with lag-length of order one by BIC criterion and a 20- step-ahead generalized forecast error variance decomposition (FEVD). The black area represents the net total directional connectedness.

4.4. Net Pairwise Directional Connectedness

The results of time-varying net pairwise connectedness between cryptocurrency uncertainty indices and financial assets are presented below in Figure 4. The pairwise connectedness is higher between UCRY indices, cryptocurrencies, and precious metals, especially during the COVID-19 pandemic. The connectedness between precious metals, soft commodities, and green bonds is also higher, especially during the outbreak's spread. Overall, the magnitude of net pairwise connectedness is higher, and the spillover patterns were changed during the COVID-19 pandemic, which is consistent with findings of other studies (Bouri et al. 2021a; Elsayed et al. 2022a; Le et al. 2021a). These findings suggest that investors should consider persistence of asset before investing as patterns of spillovers were changed during COVID-19 pandemic.

4.5. Dynamic Connectedness Network Plot

Figure 5 illustrates the network plot of the return and volatility connectedness between cryptocurrency uncertainty indices and different financial assets. The UCRY price and policy indices are net transmitters of shocks to Bitcoin, Ethereum, gold, silver, platinum, and soft commodities. Moreover, the green bonds are net transmitter of shocks towards gold and soft commodities. The net total directional connectedness between UCRY price to Bitcoin, Platinum, and gold is higher because the node size is large. Our findings suggest that equity market investors should look for volatility spillovers from cryptocurrency uncertainty indices towards different financial assets before investing in these assets during a financial crisis. These findings suggest that investors in traditional markets should be cautious during financial turmoil and its influence on traditional assets as our findings show that cryptocurrency uncertainties transmit the shocks towards traditional assets market; hence, traditional investors experienced lower returns during the COVID-19 pandemic.
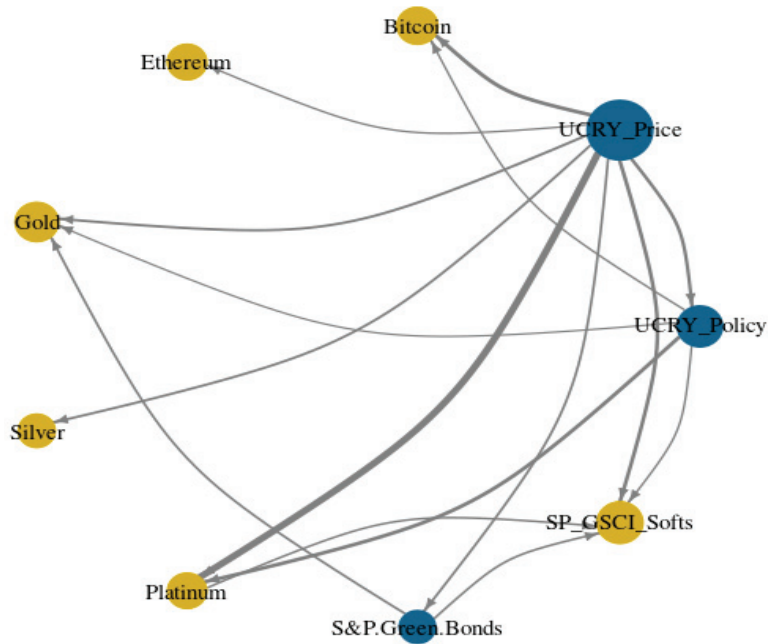
Figure 5. Dynamic connectedness network plot. Notes: Figure 5 displays the network plot using the TVP-VAR approach. The blue nodes represent the net transmitter role while the yellow nodes illustrate the net receiver of shocks. The node sizes show the weighted average net total directional connectedness.

5. Concluding Remarks

This study examines the dynamic connectedness of return and volatility spillover among cryptocurrency uncertainty, cryptocurrencies, green bonds, precious metals, and soft commodities. The investigation relies on weekly returns data from 7 August 2015 to 31 December 2021, using the TVP-VAR approach as detailed by Antonakakis et al. (2020). The total connectedness is higher, a trend particularly heightened during the COVID-19 pandemic. During this crisis, the cryptocurrency policy uncertainty index emerged as the primary transmitter of shocks to other financial assets, while the cryptocurrency price index assumed the role of shock receiver of shock during COVID-19. The pandemic has instigated shifts in returns and volatility connectedness across these financial assets. For instance, certain assets that were previously net shocks receivers transitioned into shock transmitters during the COVID-19 outbreak. Moreover, the pandemic has fostered heightened connection among precious metals, soft commodities, and green bonds. Precious metals and cryptocurrencies, as recipients of shocks, warrant particular attention from investors and practitioners who can opt for alternative assets as a strategy to hedge the cryptocurrency uncertainty and reduce the portfolio risk in times of financial turmoil.

These findings hold considerable implications, urging investors to carefully assess volatility spillovers from cryptocurrency uncertainty indices into traditional markets for comprehensive diversification insights across assets. Consequently, policymakers and investors are encouraged to scrutinize cryptocurrency uncertainty spillover patterns onto various traditional markets, enabling them to optimize returns through diversified global asset portfolios—especially crucial amidst financial disturbances.

To acknowledge this study's limitations, it is worth noting that the availability of UCRY uncertainty indices in a weekly frequency prompted the utilization of weekly data. However, this choice may omit some critical information. For future investigations, re-

searchers should consider constructing a daily cryptocurrency uncertainty index to explore its dynamic connectedness with other assets, thereby offering a more nuanced perspective.

Author Contributions: Conceptualization, M.A.; Methodology, A.A., and M.A.; Software, A.A.; Validation, A.A., and M.A.; Formal analysis, A.A.; Investigation, A.A.; Resources, M.A.; Data curation, M.A.; Writing—original draft preparation, A.A., and M.A.; Writing—review and editing, A.A., M.I.B., and M.A.; Visualization, A.A., M.I.B., and M.A.; Supervision, M.I.B.; Project administration, M.I.B. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are openly available in the cryptocurrency uncertainty index, the coinmarektcap.com, and thomsonreuters.com.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

[1] The cryptocurrency uncertainty index. Finance Research Letters, 102147. The latest UCRY Weekly Index data can be downloaded from: https://sites.google.com/view/cryptocurrency-indices/the-indices/crypto-uncertainty?authuser=0 (accessed on 8 January 2022).

[2] https://coinmarketcap.com/ (accessed 8 January 2022).

[3] https://www.thomsonreuters.com/en.html (accessed on 8 January 2022).

References

Abuzayed, Bana, and Nedal Al-Fayoumi. 2021. Risk spillover from crude oil prices to GCC stock market returns: New evidence during the COVID-19 outbreak. *The North American Journal of Economics and Finance* 58: 101476. [CrossRef]

Adekoya, Oluwasegun B., and Johnson A. Oliyide. 2021. How COVID-19 drives connectedness among commodity and financial markets: Evidence from TVP-VAR and causality-in-quantiles techniques. *Resources Policy* 70: 101898. [CrossRef]

Aharon, David Y., and Ender Demir. 2021. NFTs and asset class spillovers: Lessons from the period around the COVID-19 pandemic. *Finance Research Letters* 47: 102515. [CrossRef] [PubMed]

Aharon, David Y., Ender Demir, Chi Keung Marco Lau, and Adam Zaremba. 2022. Twitter-Based uncertainty and cryptocurrency returns. *Research in International Business and Finance* 59: 101546. [CrossRef]

Akhtaruzzaman, Md, Sabri Boubaker, and Ahmet Sensoy. 2021. Financial contagion during COVID–19 crisis. *Finance Research Letters* 38: 101604. [CrossRef]

Albulescu, Claudiu Tiberiu, Riza Demirer, Ibrahim D. Raheem, and Aviral Kumar Tiwari. 2019. Does the U.S. economic policy uncertainty connect financial markets? Evidence from oil and commodity currencies. *Energy Economics* 83: 375–88. [CrossRef]

Allen, David E. 2022. Cryptocurrencies, diversification and the COVID-19 pandemic. *Journal of Risk and Financial Management* 15: 103. [CrossRef]

Aloui, Chaker, Hela ben Hamida, and Larisa Yarovaya. 2021. Are Islamic gold-backed cryptocurrencies different? *Finance Research Letters* 39: 101615. [CrossRef]

Alqahtani, Abdullah, and Julien Chevallier. 2020. Dynamic Spillovers between Gulf Cooperation Council's Stocks, VIX, Oil and Gold Volatility Indices. *Journal of Risk and Financial Management* 13: 69. [CrossRef]

Al-Yahyaee, Khamis Hamed, Mobeen Ur Rehman, Walid Mensi, and Idries Mohammad Wanas Al-Jarrah. 2019. Can uncertainty indices predict Bitcoin prices? A revisited analysis using partial and multivariate wavelet approaches. *The North American Journal of Economics and Finance* 49: 47–56. [CrossRef]

Anscombe, Francis J., and William J. Glynn. 1983. Distribution of the kurtosis statistic b 2 for normal samples. *Biometrika* 70: 227–234. [CrossRef]

Antonakakis, Nikolaos, Ioannis Chatziantoniou, and David Gabauer. 2020. Refined measures of dynamic connectedness based on time-varying parameter vector autoregressions. *Journal of Risk and Financial Management* 13: 84. [CrossRef]

Antonakakis, Nikolaos, Ioannis Chatziantoniou, and George Filis. 2013. Dynamic co-movements of stock market returns, implied volatility and policy uncertainty. *Economics Letters* 120: 87–92. [CrossRef]

Asl, Mahdi Ghaemi, Giorgio Canarella, and Stephen M. Miller. 2021. Dynamic asymmetric optimal portfolio allocation between energy stocks and energy commodities: Evidence from clean energy and oil and gas companies. *Resources Policy* 71: 101982. [CrossRef]

Bação, Pedro, António Portugal Duarte, Helder Sebastião, and Srdjan Redzepagic. 2018. Information transmission between cryptocurrencies: Does bitcoin rule the cryptocurrency world? *Scientific Annals of Economics and Business* 65: 97–117. [CrossRef]

Baig, Ahmed S., Hassan Anjum Butt, Omair Haroon, and Syed Aun R. Rizvi. 2020. Deaths, panic, lockdowns and US equity markets: The case of COVID-19 pandemic. *Finance Research Letters* 38: 101701. [CrossRef] [PubMed]

Balcilar, Mehmet, David Gabauer, and Zaghum Umar. 2021. Crude Oil futures contracts and commodity markets: New evidence from a TVP-VAR extended joint connectedness approach. *Resources Policy* 73: 102219. [CrossRef]

Bashir, Hajam Abid, and Dilip Kumar. 2023. Investor attention, Twitter uncertainty and cryptocurrency market amid the COVID-19 pandemic. *Managerial Finance* 49: 620–42. [CrossRef]

Bhatti, M. Ishaq, and Ghulam Ghouse. 2023. The Trade Development among OIC Countries: Exploring Challenges, Opportunities, and the Impact of the Covid-19 Pandemic. *Muslim Business and Economic Review* 2: 1–25.

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]

Boubaker, Sabri, Jamel Jouini, and Amine Lahiani. 2016. Financial contagion between the US and selected developed and emerging countries: The case of the subprime crisis. *The Quarterly Review of Economics and Finance* 61: 14–28. [CrossRef]

Bouri, Elie, David Gabauer, Rangan Gupta, and Aviral Kumar Tiwari. 2021a. Volatility connectedness of major cryptocurrencies: The role of investor happiness. *Journal of Behavioral and Experimental Finance* 30: 100463. [CrossRef]

Bouri, Elie, Oguzhan Cepni, David Gabauer, and Rangan Gupta. 2021b. Return connectedness across asset classes around the COVID-19 outbreak. *International Review of Financial Analysis* 73: 101646. [CrossRef]

Bouri, Elie, Oguzhan Cepni, David Gabauer, and Rangan Gupta. 2020. Cryptocurrencies as hedges and safe-havens for US equity sectors. *The Quarterly Review of Economics and Finance* 75: 294–307. [CrossRef]

Caloia, Francesco Giuseppe, Andrea Cipollini, and Silvia Muzzioli. 2019. How do normalization schemes affect net spillovers? A replication of the Diebold and Yilmaz (2012) study. *Energy Economics* 84: 104536. [CrossRef]

Chen, Ruoyu, Najaf Iqbal, Muhammad Irfan, Farrukh Shahzad, and Zeeshan Fareed. 2022. Does financial stress wreak havoc on banking, insurance, oil, and gold markets? New empirics from the extended joint connectedness of TVP-VAR model. *Resources Policy* 77: 102718. [CrossRef]

Choudhry, Taufiq, Syed S. Hassan, and Sarosh Shabi. 2015. Relationship between gold and stock markets during the global financial crisis: Evidence from nonlinear causality tests. *International Review of Financial Analysis* 41: 247–56. [CrossRef]

Colon, Francisco, Chaehyun Kim, Hana Kim, and Wonjoon Kim. 2021. The effect of political and economic uncertainty on the cryptocurrency market. *Finance Research Letters* 39: 101621. [CrossRef]

Corbet, Shaen, Charles Larkin, and Brian Lucey. 2020a. The contagion effects of the COVID-19 pandemic: Evidence from gold and cryptocurrencies. *Finance Research Letters* 35: 101554. [CrossRef]

Corbet, Shaen, Yang Greg Hou, Yang Hu, Charles Larkin, and Les Oxley. 2020b. Any port in a storm: Cryptocurrency safe-havens during the COVID-19 pandemic. *Economics Letters* 194: 109377. [CrossRef]

Costa, Antonio, Paulo Matos, and Cristiano da Silva. 2022. Sectoral connectedness: New evidence from US stock market during COVID-19 pandemics. *Finance Research Letters* 45: 102124. [CrossRef] [PubMed]

D'Agostino, Ralph B. 1970. Transformation to normality of the null distribution of g1. *Biometrika* 57: 679–681. [CrossRef]

Dai, Zhifeng, Haoyang Zhu, and Xinhua Zhang. 2022. Dynamic spillover effects and portfolio strategies between crude oil, gold and Chinese stock markets related to new energy vehicle. *Energy Economics* 109: 105959. [CrossRef]

Demir, Ender, Giray Gozgor, Chi Keung Marco Lau, and Samuel A. Vigne. 2018. Does economic policy uncertainty predict the Bitcoin returns? An empirical investigation. *Finance Research Letters* 26: 145–49. [CrossRef]

Diebold, Francis X., and Kamil Yilmaz. 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal* 119: 158–71. [CrossRef]

Diebold, Francis X., and Kamil Yilmaz. 2012. Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting* 28: 57–66. [CrossRef]

Diebold, Francis X., and Kamil Yılmaz. 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics* 182: 119–34. [CrossRef]

El Khoury, Rim, and Muneer M. Alshater. 2022. Spillovers between Twitter Uncertainty Indexes and sector indexes: Evidence from the US. *Borsa Istanbul Review* 22: 961–74. [CrossRef]

Elliott, Graham, Thomas J. Rothenberg, and James H. Stock. 1996. Efficient Tests for an Autoregressive Unit Root. *Econometrica* 64: 813–36. [CrossRef]

Elsayed, Ahmed H., and Mohamad Husam Helmi. 2021. Volatility transmission and spillover dynamics across financial markets: The role of geopolitical risk. *Annals of Operations Research* 305: 1–22. [CrossRef]

Elsayed, Ahmed H., Giray Gozgor, and Chi Keung Marco Lau. 2022a. Risk transmissions between bitcoin and traditional financial assets during the COVID-19 era: The role of global uncertainties. *International Review of Financial Analysis* 81: 102069. [CrossRef]

Elsayed, Ahmed H., Giray Gozgor, and Larisa Yarovaya. 2022b. Volatility and return connectedness of cryptocurrency, gold, and uncertainty: Evidence from the cryptocurrency uncertainty indices. *Finance Research Letters* 47: 102732. [CrossRef]

Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society* 50: 987–1007. [CrossRef]

Fakhfekh, Mohamed, Ahmed Jeribi, Ahmed Ghorbel, and Nejib Hachicha. 2021. Hedging stock market prices with WTI, Gold, VIX and cryptocurrencies: A comparison between DCC, ADCC and GO-GARCH models. *International Journal of Emerging Markets*. *ahead-of-print*. [CrossRef]

Fisher, Thomas J., and Colin M. Gallagher. 2012. New weighted portmanteau statistics for time series goodness of fit testing. *Journal of the American Statistical Association* 107: 777–87. [CrossRef]

Foglia, Matteo, and Peng-Fei Dai. 2021. "Ubiquitous uncertainties": Spillovers across economic policy uncertainty and cryptocurrency uncertainty indices. *Journal of Asian Business and Economic Studies* 29: 35–49. [CrossRef]

Forbes, Kristin J., and Roberto Rigobon. 2002. No contagion, only interdependence: Measuring stock market co-movements. *The Journal of Finance* 57: 2223–61. [CrossRef]

Ghabri, Yosra, Oussama Ben Rhouma, Marjène Gana, Khaled Guesmi, and Ramzi Benkraiem. 2022. Information transmission among energy markets, cryptocurrencies, and stablecoins under pandemic conditions. *International Review of Financial Analysis* 82: 102197. [CrossRef]

Ghouse, Ghulam, Muhammad Ishaq Bhatti, Aribah Aslam, and Nawaz Ahmad. 2023. Asymmetric spillover effects of Covid-19 on the performance of the Islamic finance industry: A wave analysis and forecasting. *Journal of Economic Asymmetries* 27: e00280. [CrossRef]

Gök, Remzi, Elie Bouri, and Eray Gemici. 2022. Can Twitter-based economic uncertainty predict safe-haven assets under all market conditions and investment horizons? *Technological Forecasting and Social Change* 185: 122091. [CrossRef]

Ha, Le Thanh, and Nguyen Van Dai. 2022. Total and Net-Directional Connectedness of Cryptocurrencies During the Pre-and Post-COVID-19 Pandemic. *Journal of International Commerce, Economics and Policy* 13: 2250004. [CrossRef]

Hasan, Md Bokhtiar, M. Kabir Hassan, Md Mamunur Rashid, and Yasser Alhenawi. 2021. Are safe haven assets really safe during the 2008 global financial crisis and COVID-19 pandemic? *Global Finance Journal* 50: 100668. [CrossRef]

Hassan, Kamrul, Ariful Hoque, and Dominic Gasbarro. 2019. Separating BRIC using Islamic stocks and crude oil: Dynamic conditional correlation and volatility spillover analysis. *Energy Economics* 80: 950–69. [CrossRef]

Iglesias-Casal, Ana, María-Celia López-Penabad, Carmen López-Andión, and José Manuel Maside-Sanfiz. 2020. Diversification and optimal hedges for socially responsible investment in Brazil. *Economic Modelling* 85: 106–18. [CrossRef]

Jalan, Akanksha, Roman Matkovskyy, and Larisa Yarovaya. 2021. "Shiny" crypto assets: A systemic look at gold-backed cryptocurrencies during the COVID-19 pandemic. *International Review of Financial Analysis* 78: 101958. [CrossRef]

Jarque, Carlos M., and Anil K. Bera. 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics Letters* 6: 255–59. [CrossRef]

Kamal, Javed Bin, and M. Kabir Hassan. 2022. Asymmetric connectedness between cryptocurrency environment attention index and green assets. *Journal of Economic Asymmetries* 25: e00240. [CrossRef]

Kang, Sang Hoon, Ron McIver, and Seong-Min Yoon. 2017. Dynamic spillover effects among crude oil, precious metal, and agricultural commodity futures markets. *Energy Economics* 62: 19–32. [CrossRef]

Khan, Muhammad Asif, Juan E. Trinidad Segovia, M. Ishaq Bhatti, and Asif Kabir. 2023. Corporate vulnerability in the US and China during COVID-19: A machine learning approach. *The Journal of Economic Asymmetries* 27: e00302. [CrossRef]

Koop, Gary, and Dimitris Korobilis. 2013. Large time-varying parameter VARs. *Journal of Econometrics* 177: 185–98. [CrossRef]

Koop, Gary, and Dimitris Korobilis. 2014. A new index of financial conditions. *European Economic Review* 71: 101–16. [CrossRef]

Kristoufek, Ladislav. 2021. Tethered, or Untethered? On the interplay between stablecoins and major cryptoassets. *Finance Research Letters* 43: 101991. [CrossRef]

Kumar, Satish, Ashis Kumar Pradhan, Aviral Kumar Tiwari, and Sang Hoon Kang. 2019. Correlations and volatility spillovers between oil, natural gas, and stock prices in India. *Resources Policy* 62: 282–291. [CrossRef]

Kurka, Josef. 2019. Do cryptocurrencies and traditional asset classes influence each other? *Finance Research Letters* 31: 38–46. [CrossRef]

Kyriazis, Nikolaos A. 2021. The effects of geopolitical uncertainty on cryptocurrencies and other financial assets. *SN Business and Economics* 1: 1–14. [CrossRef]

Lahmiri, Salim, and Stelios Bekiros. 2020. The impact of COVID-19 pandemic upon stability and sequential irregularity of equity and cryptocurrency markets. *Chaos, Solitons and Fractals* 138: 109936. [CrossRef]

Le, Lan-TN, Larisa Yarovaya, and Muhammad Ali Nasir. 2021a. Did COVID-19 change spillover patterns between Fintech and other asset classes? *Research in International Business and Finance* 58: 101441. [CrossRef]

Le, TN-Lan, Emmanuel Joel Aikins Abakah, and Aviral Kumar Tiwari. 2021b. Time and frequency domain connectedness and spillover among fintech, green bonds and cryptocurrencies in the age of the fourth industrial revolution. *Technological Forecasting and Social Change* 162: 120382. [CrossRef]

Lucey, Brian M., Samuel A. Vigne, Larisa Yarovaya, and Yizhi Wang. 2022. The cryptocurrency uncertainty index. *Finance Research Letters* 45: 102147. [CrossRef]

Lundgren, Amanda Ivarsson, Adriana Milicevic, Gazi Salah Uddin, and Sang Hoon Kang. 2018. Connectedness network and dependence structure mechanism in green investments. *Energy Economics* 72: 145–53. [CrossRef]

Mensi, Walid, Mobeen Ur Rehman, and Xuan Vinh Vo. 2020. Spillovers and co-movements between precious metals and energy markets: Implications on portfolio management. *Resources Policy* 69: 101836. [CrossRef]

Mensi, Walid, Shawkat Hammoudeh, Idries Mohammad Wanas Al-Jarrah, Ahmet Sensoy, and Sang Hoon Kang. 2017. Dynamic risk spillovers between gold, oil prices and conventional, sustainability and Islamic equity aggregates and sectors with portfolio implications. *Energy Economics* 67: 454–75. [CrossRef]

Mo, Bin, Juan Meng, and Liping Zheng. 2022. Time and frequency dynamics of connectedness between cryptocurrencies and commodity markets. *Resources Policy* 77: 102731. [CrossRef]

Rubbaniy, Ghulame, Ali Awais Khalid, and Aristeidis Samitas. 2021a. Are cryptos safe-haven assets during Covid-19? Evidence from wavelet coherence analysis. *Emerging Markets Finance and Trade* 57: 1741–56. [CrossRef]

Rubbaniy, Ghulame, Ali Awais Khalid, Konstantinos Syriopoulos, and Aristeidis Samitas. 2021b. Safe-haven properties of soft commodities during times of Covid-19. *Journal of Commodity Markets* 27: 100223. [CrossRef]

Salisu, Afees A., and Xuan Vinh Vo. 2020. Predicting stock returns in the presence of COVID-19 pandemic: The role of health news. *International Review of Financial Analysis* 71: 101546. [CrossRef]

Shahid, Muhammad Naeem, Wajahat Azmi, Mohsin Ali, Muhammad Umar Islam, and Syed Aun R. Rizvi. 2023. Uncovering risk transmission between socially responsible investments, alternative energy investments and the implied volatility of major commodities. *Energy Economics* 120: 106634. [CrossRef]

Shahzad, Syed Jawad Hussain, Elie Bouri, Ladislav Kristoufek, and Tareq Saeed. 2021a. Impact of the COVID-19 outbreak on the US equity sectors: Evidence from quantile return spillovers. *Financial Innovation* 7: 14. [CrossRef]

Shahzad, Syed Jawad Hussain, Elie Bouri, Sang Hoon Kang, and Tareq Saeed. 2021b. Regime specific spillover across cryptocurrencies and the role of COVID-19. *Financial Innovation* 7: 1–24. [CrossRef]

Smales, Lee A. 2022. Investor attention in cryptocurrency markets. *International Review of Financial Analysis* 79: 101972. [CrossRef]

So, Mike K. P., Agnes Tiwari, Amanda M. Y. Chu, Jenny T. Y. Tsang, and Jacky N. L. Chan. 2020. Visualizing COVID-19 pandemic risk through network connectedness. *International Journal of Infectious Diseases* 96: 558–61. [CrossRef] [PubMed]

Su, Chi-Wei, Yuan Xi, Ran Tao, and Muhammad Umar. 2022. Can Bitcoin be a safe haven in fear sentiment? *Technological and Economic Development of Economy* 28: 268–89. [CrossRef]

Szczepanska-Przekota, Anna. 2021. Cryptocurrency versus other financial instruments: How a small market affects a large market. *Journal of Investment Strategies* 10: 2. [CrossRef]

Umar, Zaghum, Francisco Jareño, and Ana Escribano. 2021. Oil price shocks and the return and volatility spillover between industrial and precious metals. *Energy Economics* 99: 105291. [CrossRef]

Umar, Zaghum, Francisco Jareño, and Ana Escribano. 2022. Dynamic return and volatility connectedness for dominant agricultural commodity markets during the COVID-19 pandemic era. *Applied Economics* 54: 1030–54. [CrossRef]

Wang, Gang-Jin, Xin-yu Ma, and Hao-yu Wu. 2020. Are stablecoins truly diversifiers, hedges, or safe havens against traditional cryptocurrencies as their name suggests? *Research in International Business and Finance* 54: 101225. [CrossRef]

Wei, Yu, Yizhi Wang, Brian M. Lucey, and Samuel A. Vigne. 2023. Cryptocurrency uncertainty and volatility forecasting of precious metal futures markets. *Journal of Commodity Markets* 29: 100305. [CrossRef]

Wu, Wanshan, Aviral Kumar Tiwari, Giray Gozgor, and Huang Leping. 2021. Does economic policy uncertainty affect cryptocurrency markets? Evidence from Twitter-based uncertainty measures. *Research in International Business and Finance* 58: 101478. [CrossRef]

Xu, Danyang, Yang Hu, Shaen Corbet, and John W. Goodell. 2023. Volatility connectedness between global COVOL and major international volatility indices. *Finance Research Letters* 56: 104112. [CrossRef]

Yan, Lei, Nawazish Mirza, and Muhammad Umar. 2022. The cryptocurrency uncertainties and investment transitions: Evidence from high and low carbon energy funds in China. *Technological Forecasting and Social Change* 175: 121326. [CrossRef]

Yen, Kuang-Chieh, and Hui-Pei Cheng. 2021. Economic policy uncertainty and cryptocurrency volatility. *Finance Research Letters* 38: 101428. [CrossRef]

Yoon, Seong-Min, Md Al Mamun, Gazi Salah Uddin, and Sang Hoon Kang. 2019. Network connectedness and net spillover between financial and commodity markets. *The North American Journal of Economics and Finance* 48: 801–18. [CrossRef]

Yousaf, Imran, and John W. Goodell. 2023. Linkages between CBDC and cryptocurrency uncertainties, and digital payment stocks. *Finance Research Letters* 54: 103765. [CrossRef]

Yousaf, Imran, and Larisa Yarovaya. 2022a. Spillovers between the Islamic gold-backed cryptocurrencies and equity markets during the COVID-19: A sectorial analysis. *Pacific-Basin Finance Journal* 71: 101705. [CrossRef]

Yousaf, Imran, and Larisa Yarovaya. 2022b. Static and dynamic connectedness between NFTs, Defi and other assets: Portfolio implication. *Global Finance Journal* 53: 100719. [CrossRef]

Yousaf, Imran, Ramzi Nekhili, and Mariya Gubareva. 2022. Linkages between DeFi assets and conventional currencies: Evidence from the COVID-19 pandemic. *International Review of Financial Analysis* 81: 102082. [CrossRef]

Zhang, Dayong, and David C. Broadstock. 2020. Global financial crisis and rising connectedness in the international commodity markets. *International Review of Financial Analysis* 68: 101239. [CrossRef]

Article

Properties of VaR and CVaR Risk Measures in High-Frequency Domain: Long–Short Asymmetry and Significance of the Power-Law Tail

Tetsuya Takaishi

Hiroshima University of Economics, Department of Liberal Arts, Hiroshima 731-0192, Japan; tt-taka@hue.ac.jp

Abstract: This study investigates the properties of risk measure, value at risk (VaR) and conditional VaR (CVaR), using high-frequency Bitcoin data. These data allow us to conduct a high statistical analysis. Our findings reveal a disparity in VaR and CVaR values between the left and right tails of the return probability distributions. We refer to this disparity as "long–short asymmetry". In the high-frequency domain, the tail distribution can be accurately described by a power-law function. Moreover, the ratio of CVaR to VaR is expected to be determined solely by the power-law exponent. Through empirical analysis, we confirm that this ratio property holds true for high confidence levels. Furthermore, we investigate the relationship between risk measures (VaR and CVaR) and realized volatility. We observe that they trace a trajectory in a two-dimensional plane. This trajectory changes gradually, indicating periods of both high and low risk.

Keywords: risk measure; value at risk; conditional value at risk; expected shortfall; power-law function; realized volatility; Bitcoin; Rachev ratio

1. Introduction

Risk management plays a central role in various financial sectors. Its purpose is to prevent unexpected substantial losses in trading and operations by closely monitoring these risks. Although there are various sources of financial risk, such as credit and liquidity risks, we will focus specifically on market risk, which refers to the risk associated with market price changes. The widely accepted risk measure is known as the value at risk (VaR). VaR provides a single numerical value that summarizes the overall risk of a portfolio (see, for example, Abad et al. (2014); Duffie and Pan (1997); Gourieroux and Jasiak (2010); Linsmeier and Pearson (2000)). For our analysis, we will consider a simple portfolio consisting of a single asset, such as a stock. In this context, changes in asset prices are described as returns r, and historical return data form a return probability distribution $P(r)$. VaR is defined as the maximum loss at a given confidence level, denoted as $X\%$, over a given time horizon, denoted as T. Figure 1 shows a schematic drawing for the VaR approach. "$VaR(X)_L$" stands for the VaR for the long position at the confidence level of $X\%$ and is defined so that the probability of the left tail ($-\infty < r \leq VaR(X)_L$) becomes $(100\text{-}X)\%$. Similarly, "$VaR(X)_S$" for the short position at the confidence level of $X\%$ is defined as the right tail ($VaR(X)_S \leq r < \infty$) of the return probability distribution.

One drawback of VaR is its inability to provide information on potential losses beyond the VaR threshold. This limitation becomes particularly significant when dealing with the tail of the probability distribution. In scenarios where the tail is heavier than that of a normal distribution, the potential loss can exceed what would be expected under normal distribution assumptions. Empirical evidence shows that asset return distributions often exhibit fat tails, which are recognized as stylized facts (Cont 2001). To address this issue and incorporate tail information, an improved risk measure known as conditional VaR (CVaR), or expected shortfall with coherent properties, has been introduced (Acerbi and

Tasche 2002; Artzner et al. 1999). CVaR is defined as the average value of VaR that exceeds the VaR at a given confidence level, denoted as X%.
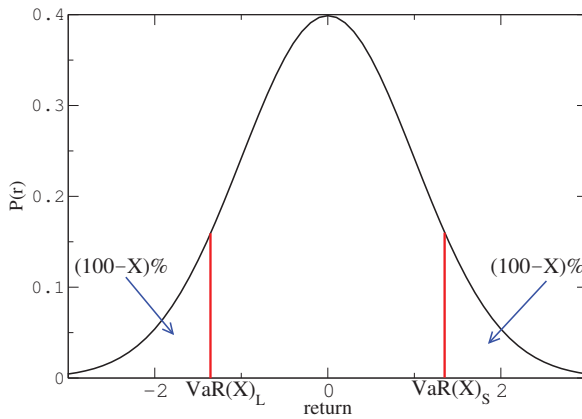
Figure 1. Schematic drawing of the VaR approach.

Usually, VaR and CVaR calculations in risk management focus on the left tail of the return probability distribution. This region corresponds to potential losses in a long position. Conversely, the right tail represents losses in a short position. When assuming a symmetrical return probability distribution, there would be no distinction between VaR and CVaR values in the left and right tails. However, it cannot be guaranteed that empirical distributions will exhibit symmetry. We refer to the difference between VaR and CVaR associated with asymmetric distributions as "long–short asymmetry". One of our objectives was to empirically investigate the presence of such long–short asymmetry. Measuring long–short asymmetry poses various challenges, primarily due to the limited availability of accurate probability distributions. This limitation becomes more pronounced when using one-day or longer time returns, as the shorter sampling period results in fewer statistics. For instance, if we collect daily returns for one year, we would have only 365 (or approximately 250) working days' worth of data. For risk measures in the cryptocurrency market, based on daily returns, see, e.g., Almeida et al. (2022).

To address this issue, we analyze the data in the high-frequency domain, which involves collecting a large amount of statistical data. In this study, we specifically utilize 1 min Bitcoin data. In cryptocurrency markets, Bitcoin is traded continuously for 24 h, allowing us to gather 52,560 1 min return data points over a 1-year period. This extensive dataset enables us to conduct a thorough statistical analysis.

Long–short asymmetry is closely associated with the reward–risk ratios that are defined as ratios between a reward measure and a risk measure (Cheridito and Kromer 2013). We calculate one of the reward–risk ratios, the Rachev ratio (Biglova et al. 2004), which is given by a ratio of the CVaR for the short position to the CVaR for the long position. Since the Rachev ratio deviates from one for asymmetrical distribution, it also quantifies the long–short asymmetry.

Our second objective is to explore the relationship between the tail exponent and risk measures. In the high-frequency domain, the tails of the return probability distributions are fat-tailed and exhibit power-law behavior (Gabaix 2009; Gopikrishnan et al. 1998, 1999; Pan and Sinha 2007; Plerou et al. 1999). Let α be the power-law exponent of the cumulative return distribution. As we will explain later, under the assumption of a power-law probability distribution, the ratio of CVaR to VaR is $\alpha / (\alpha - 1)$; we empirically verify the presence of this $\alpha / (\alpha - 1)$ relationship at high confidence levels.

We also investigate the relationship between realized volatility (RV) (Andersen and Bollerslev 1998; Andersen et al. 2003; McAleer and Medeiros 2008) and the risk measures.

Our findings reveal that these variables form a trajectory that exhibits periods of both high and low risk.

The remainder of this paper is organized as follows: Section 2 describes the methodology and data used in this study. Section 3 presents the empirical results. Finally, we discuss and conclude our findings in Section 4.

2. Methodology and Data

First, we calculate the VaR and CVaR for the return probability distribution $P(r)$. Here, for simplicity, we assume that the mean of the returns is zero. The VaR at a confidence level $X\%$ is denoted as $VaR(X)_L$ and is defined for the long position (left tail) as

$$p_X = \int_{-\infty}^{VaR(X)_L} P(r)dr, \tag{1}$$

where $p_X = 1 - X/100$. Similarly, the VaR for the short position, $VaR(X)_S$, is defined in the right tail as

$$p_X = \int_{VaR(X)_S}^{\infty} P(r)dr. \tag{2}$$

CVaR is defined as the average VaR that exceeds the VaR at confidence level $X\%$. The CVaR for the long position at confidence level $X\%$ is given by

$$CVaR(X)_L = \int_{-\infty}^{VaR(X)_L} rP(r)dr/p_X. \tag{3}$$

Similarly, the CVaR for the short position is given by

$$CVaR(X)_S = \int_{VaR(X)_S}^{\infty} rP(r)dr/p_X. \tag{4}$$

Next, we calculate the VaR and CVaR for specific forms of $P(r)$. Note that although we calculate VaR and CVaR for the long position, we obtain the same expression for both of the short positions, except with regard to the sign. First, we assume that $P(r)$ is a normal distribution with a standard deviation σ, that is, $P(r) = \exp(-\frac{r^2}{2\sigma^2})/\sqrt{2\pi\sigma^2}$. Then, Equation (1) is as follows (Hull 2018):

$$VaR(X)_L = -\sigma N^{-1}(X), \tag{5}$$

where $N^{-1}(X)$ denotes the inverse cumulative normal function. Similarly, we obtain Equation (3).

$$CVaR(X)_L = -\sigma \frac{\exp(-K^2/2)}{\sqrt{2\pi}p_X}, \tag{6}$$

where $K \equiv N^{-1}(X)$. Using Equations (5) and (6), the ratio of CVaR to VaR, denoted by R_{norm}, is

$$R_{norm} = \frac{\exp(-K^2/2)}{\sqrt{2\pi}Kp_X}. \tag{7}$$

Second, we assume that the total probability distribution $P(r)_t$, defined as $P(r)_t = P(r)_{tail} + P_o(r)$, consists of two parts: the tail $P(r)_{tail}$ and the other $P_o(r)$. $P(r)_{tail}$ is described by a power-law function as

$$P(r)_{tail} = c|r|^{-(\alpha+1)}, \tag{8}$$

where the constant c is the normalization factor determined such that $\int_{-\infty}^{\infty} P(r)_t = 1$. The actual value of c is not significant. At the tail, we assume that only $P(r)_{tail}$ contributes to the calculations of VaR and CVaR. Using Equation (8), Equation (1) is calculated to be[1]

$$p_X = \frac{c}{\alpha} |VaR(X)_L|^{-\alpha}. \tag{9}$$

Similarly, Equation (3) leads to the following:

$$CVaR(X)_L = \frac{c}{p_X(\alpha - 1)} |VaR(X)_L|^{-(\alpha - 1)}. \tag{10}$$

Using Equations (9) and (10), the ratio of $CVaR(X)_L$ to $VaR(X)_L$ becomes:

$$R_{power} = \frac{\alpha}{\alpha - 1}. \tag{11}$$

Interestingly, at any confidence level for the tail distributions described by a power-law function, R_{power} is determined by α only.

To quantify the differences in VaR and CVaR between the left and right tails, we compute

$$D_{VaR}(X) = VaR_S(X) - |VaR_L(X)|, \tag{12}$$

and

$$D_{CVaR}(X) = CVaR_S(X) - |CVaR_L(X)|, \tag{13}$$

at a given confidence level $X\%$. $D_{VaR}(X)$ and $D_{CVaR}(X)$ take zero values for symmetrical distributions. We also calculate the Rachev ratio (R-ratio) (Biglova et al. 2004) defined as

$$R\text{-}ratio = \frac{CVaR_S(X)}{|CVaR_L(X)|}. \tag{14}$$

Similarly, we can also define a ratio by VaR (V-ratio) as

$$V\text{-}ratio = \frac{VaR_S(X)}{|VaR_L(X)|}. \tag{15}$$

Both the R-ratio and V-ratio will take the value of 1 for symmetric distributions. The R-ratio and V-ratio are related to D_{VaR} and D_{CVaR} as follows:

$$R\text{-}ratio - 1 = D_{CVaR}/|CVaR_L|, \tag{16}$$

and

$$V\text{-}ratio - 1 = D_{VaR}/|VaR_L|. \tag{17}$$

For the tail distributions described by a power-law function, the ratio of R-ratio to V-ratio, denoted as R_{RV}, is given by

$$R_{RV} \equiv \frac{R\text{-}ratio}{V\text{-}ratio} = \frac{\alpha_S}{\alpha_S - 1} \frac{\alpha_L - 1}{\alpha_L}, \tag{18}$$

where α_S (α_L) is the power-law exponent at the right (left) tail of the return probability distribution.

In this study, we used Bitcoin data traded on Bitstamp exchanges[2] from 1 January 2015 to 21 May 2022. In the early stages of the Bitcoin market, characterized by low liquidity, we observed market properties that differed from those of liquid markets, such as developed-country stock markets (Di Matteo et al. 2005). For example, the Hurst exponent of the return time series in the early stages of the Bitcoin market was found to be less than 0.5, indicating the anti-persistence of the series (Urquhart 2016). It is argued that the anti-persistence seen in the cryptocurrency market can be attributed to the low liquidity of the market (Wei 2018).

The power-law exponents α of the tail return distributions were significantly lower than the expected values of 3 based on return distributions in developed countries (Begušić et al. 2018; Drożdż et al. 2018; Easwaran et al. 2015; Takaishi 2021a). Due to the low liquidity on the Bitstamp exchange before 2013 (Takaishi and Adachi 2020), we selected a period after 2015 when liquidity was sufficiently high.

From the 1 min price data p_t, we construct 1 min return data using $r_t = \ln p_t - \ln p_{t-1}$. Table 1 describes the descriptive statistics of the whole 1 min return data. The kurtosis was found to be about 98, which is considerably high, implying that the return distribution is fat-tailed. It is known that as the time scale of returns increases, the return distributions approach the Gaussian distribution. On the Bitcoin market, the kurtosis reaches the value of 3 (the kurtosis of the Gaussian distribution) at the time scale of two weeks (Takaishi 2018 2021b).

Table 1. Descriptive statistics for the whole sample of 1 min returns. The values in parentheses indicate one-sigma errors estimated by the Jackknife method.

Mean	Standard Deviation	Kurtosis	Skewness	Nobs
$1.2(8) \times 10^{-6}$	0.00135(8)	98(44)	$-0.3(2)$	3.87M

We analyze the data within a 1-year window containing 52,560 return data points. The window is then shifted by one day to capture time-varying properties and enable further investigation. To calculate VaR and CVaR, we first sort the 52,560 return data points in ascending (descending) order for long (short) positions. Then, $VaR(X)$ at the confidence level X is obtained from the $52,560 \times (X/100)$th value in the sorted data[3]. Similarly, $CVaR(X)$ is obtained from the average of the sorted data from the first to the $52,560 \times (X/100)$th data point.

3. Empirical Results

Figures 2 and 3 show the time evolution of VaR and CVaR, respectively. In the figures, positive (negative) values correspond to the VaR and CVaR regarding the short (long) position or the right (left) tail of the return probability distribution. The magnitude of the VaR and CVaR is found to be relatively small, i.e., an order of about 0.005∼0.01 since the variation in 1 min returns that we use here is also small. As described in Table 1, the standard deviation of 1 min returns is small, ∼0.00135. Here, note that the standard deviation of 1-day returns is calculated to be ∼4.63 (Takaishi 2021b), which is bigger than that of 1 min returns.
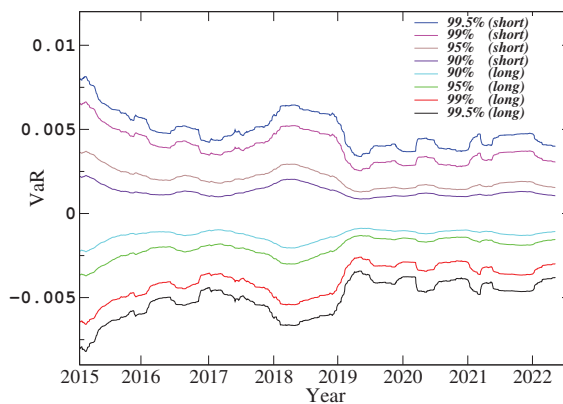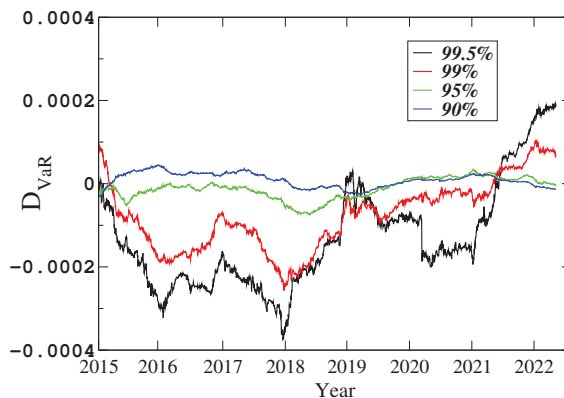

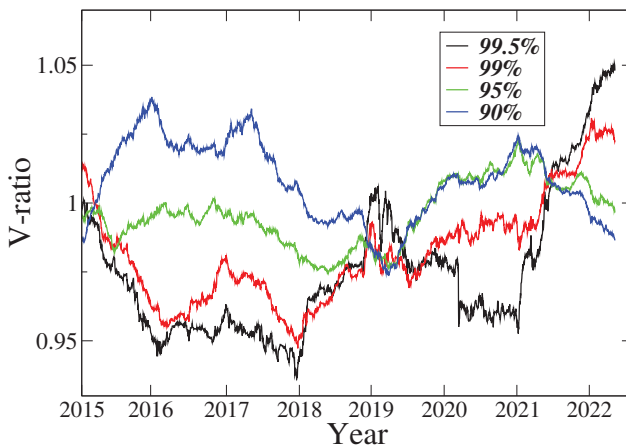**Figure 2.** Time evolution of VaR at confidence levels $X = 99.5\%, 99\%, 95\%$, and 90%.
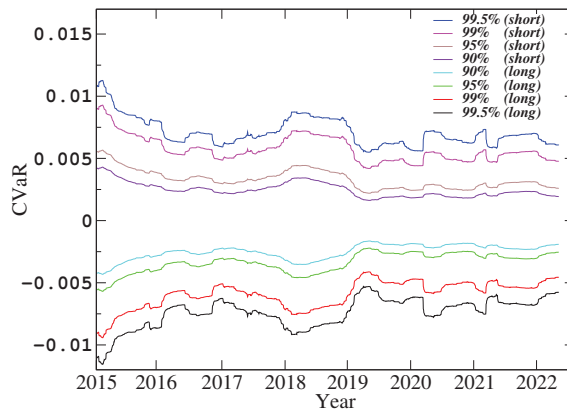
Figure 3. Time evolution of CVaR at confidence levels $X = 99.5\%, 99\%, 95\%$, and 90%.

It is evident from the figure that VaR and CVaR are not constant and vary over time on the Bitcoin market. The VaR and CVaR show similar time variation patterns. Namely, we observed that the magnitude of VaR and CVaR increases around 2015 and 2018, implying that the market risk was higher around 2015 and 2018 compared to other periods. This high and low risk pattern will be more clear when we analyze both risk measures (VaR and CVaR) and the RV simultaneously (we will return to this point later).

Figures 4 and 5 display the time evolutions of D_{VaR} and D_{CVaR}, respectively. D_{VaR} and D_{CVaR} quantify long–short asymmetry in VaR and CVaR. It is evident from the figures that D_{VaR} and D_{CVaR} predominantly take non-zero values, indicating the presence of long–short asymmetry in VaR and CVaR.

This asymmetry is more pronounced at high confidence levels, highlighting that risks can differ between the left and right tails at the same confidence level. At high confidence levels, D_{VaR} and D_{CVaR} take mostly negative values except for in some periods, which means that in the period we studied here, the long position is riskier than the short position.

The significance of long–short asymmetry should be compared to the magnitude of VaR or CVaR. For example, at the confidence level $X = 99.5\%$, the absolute value of CVaR (D_{CVaR}) around 2016 is about 0.01 (0.0005), which results in $|D_{CVaR}/CVaR| \simeq 0.05$. Thus, in this case, the significance of long–short asymmetry is about 5%. The significance of long–short asymmetry is also measured by directly comparing long and short positions. Such measurements are the V-ratio and R-ratio.

Figure 4. Time evolution of D_{VaR} at confidence levels of $X = 99.5\%, 99\%, 95\%$, and 90%.
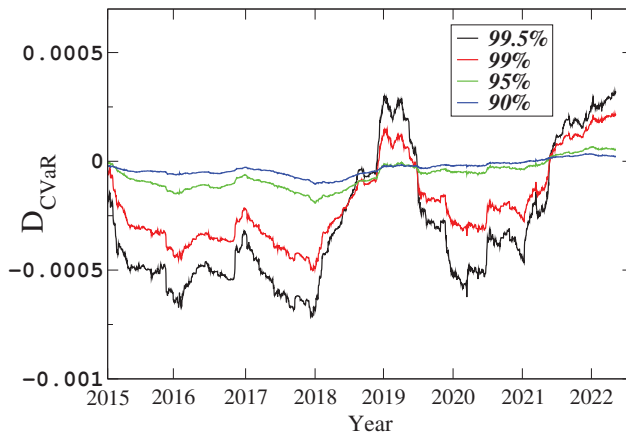
Figure 5. Time evolution of D_{CVaR} at confidence levels of $X = 99.5\%, 99\%, 95\%$, and 90%.

Figures 6 and 7 show the time evolutions of the V-ratio and R-ratio, respectively. These ratios exhibit similar time variations with D_{VaR} and D_{CVaR}, and predominantly deviate from 1, indicating the presence of long–short asymmetry. Moreover, since the V-ratio and R-ratio are defined by Equations (14) and (15), they could represent the significance of long–short asymmetry. For example, at the confidence level $X = 99.5\%$, the V-ratio from 2016 to 2018 is about 0.95, implying that there is about a 5% difference in the VaR between long and short positions. Although the magnitude of D_{VaR} is small at the confidence level $X = 90\%$, the V-ratio from 2016 to 2018 is around 1.03 (3% difference in VaR), which means that the long–short asymmetry could be significant at lower confidence levels.

"D_{VaR}, D_{CVaR}" and "V-ratio, R-ratio" can also identify the period in which the return probability distributions are approximately symmetrical. For symmetrical distributions, D_{VaR} and D_{CVaR} take values of zero at any confidence level. Similarly, the V-ratio and R-ratio take one at any confidence level. Therefore, the criterion that D_{VaR} and D_{CVaR} are zero and that the V-ratio and R-ratio are one at any confidence level can help to provide information on symmetrical distributions. For example, from Figures 4–7 we recognize that the middle of 2021 matches the criterion approximately, and thus, the return probability distribution is expected to be symmetrical approximately in the middle of 2021.

Figure 6. Time evolution of V-ratio at confidence levels of $X = 99.5\%, 99\%, 95\%$, and 90%.
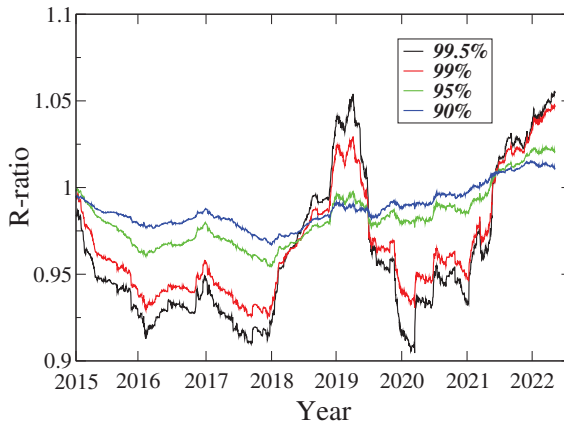
Figure 7. Time evolution of R-ratio at confidence levels of $X = 99.5\%, 99\%, 95\%$, and 90%.

To examine the significance of the power-law distribution with regard to the VaR and CVaR, we determine the power-law exponent α by fitting the tail data to $\sim |r|^{\alpha+1}$ using the Hill estimator (Hill 1975). The Hill estimator estimates the power-law exponent α with the equation

$$\frac{1}{\alpha+1} = \frac{1}{k}\sum_{i=1}^{k}(\ln r^{(i)} - \ln r^{(k)}), \tag{19}$$

where $r^{(1)} \geq r^{(2)} \geq \cdots \geq r^{(k)}$ is the order statistics for the tail data.

Figure 8 shows the time evolution of α obtained from the left and right tails of the return probability distribution. The power-law exponent α varies considerably over time around $\alpha = 3$. The values of α are higher than the value of $\alpha = 2$ observed in the early stage of the Bitcoin market (Easwaran et al. 2015). The low value of α could be related to the liquidity in the early stage of the Bitcoin market. The liquidity is also considered to be the origin of the low Hurst exponents observed in the cryptocurrency markets (Wei 2018). In the period we studied here, the liquidity of the Bitcoin market is expected to be high (Takaishi and Adachi 2020) and the power-law exponent α comes close to the value of $\alpha = 3$ observed in the stock market (Gopikrishnan et al. 1998, 1999).
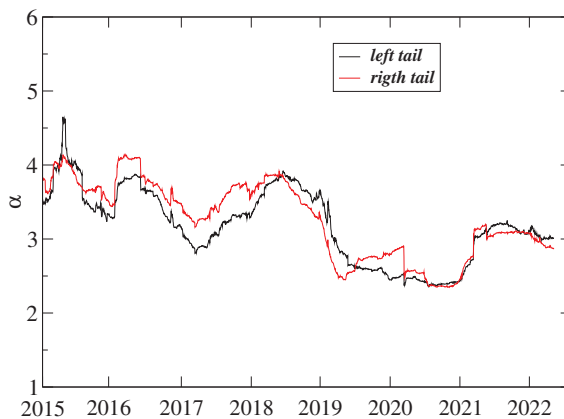
Figure 8. Time evolution of the power-law exponent α.

As suggested by Equation (11), the ratio CVaR/VaR can be expressed as $\alpha/(\alpha - 1)$ for the power-law tail. In Figures 9 and 10, we show the time evolution of the ratio alongside the corresponding values of $\alpha/(\alpha - 1)$. The dashed straight lines in the figures

represent the theoretical results under a normal distribution assumption obtained with Equation (7). The empirical ratios consistently exceed those derived from the normal assumption, suggesting that the empirical return distributions exhibit fatter tails compared to the normal distribution. The results at high confidence levels (99.5% and 99%) closely align with the $\alpha/(\alpha-1)$ law. This indicates that the probability distributions are well described by the power-law function in the region corresponding to high confidence levels. As the confidence level decreases, the results deviate from $\alpha/(\alpha-1)$, which implies that in the region with lower confidence levels, the probability distributions are not well approximated by the power-law function.
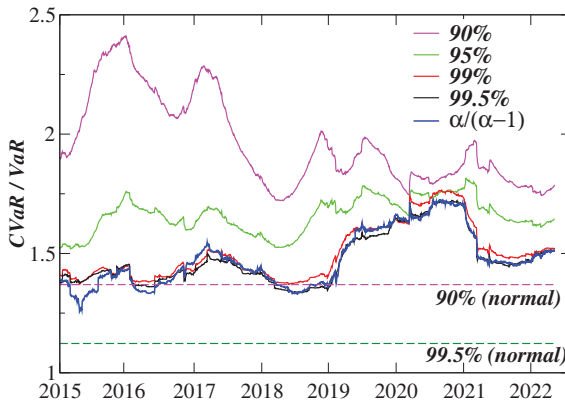
Figure 9. Time evolution of CVaR/VaR for the long position at confidence levels of $X = 99.5\%$, 99%, 95%, and 90%. The dashed lines show the theoretical values under the normal distributional assumption, obtained with Equation (7).
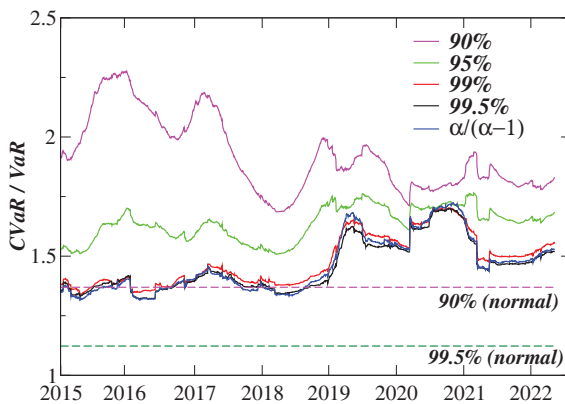
Figure 10. Time evolution of CVaR/VaR for the short position at confidence levels of $X = 99.5\%$, 99%, 95%, and 90%. The dashed lines show the theoretical values under the normal distributional assumption, obtained with Equation (7).

To visualize the $\alpha/(\alpha-1)$ law more clearly, we plot CVaR/VaR as a function of α in Figures 11 and 12. At high confidence levels ($X = 99.5\%$ and 99%), the ratio aligns well with the line $\alpha/(\alpha-1)$, indicating that the return probability distribution at these high confidence levels is consistent with the power-law distribution. For lower confidence levels, the ratio deviates from the $\alpha/(\alpha-1)$ law and moves above the curve of the $\alpha/(\alpha-1)$ law. The deviation from the curve of the $\alpha/(\alpha-1)$ law implies that the return probability distribution departs from the power-law distribution. Thus, it is concluded that the tail

distributions corresponding to lower confidence levels are not well described by the power-law function.

In the absence of a specific distributional assumption, such as the normal distribution, the variance alone does not provide accurate values for VaR and CVaR. However, the magnitudes of VaR and CVaR are correlated with the variance. To investigate the correlation between risk measures and variance, we use RV as a proxy for variance. The daily RV is constructed as a sum of squared intraday returns,

$$RV = \sum_{i=1}^{n} r_{i,\Delta}^2, \tag{20}$$

where $r_{i,\Delta}$ is the number of returns sampled at a Δ-minute sampling frequency and n is the number of returns sampled in a day. We calculate the daily RV from the 5 min returns (Liu et al. 2015). Since we consider the VaR and CVaR calculated over a 1-year window, we use an RV averaged over the same 1-year window.
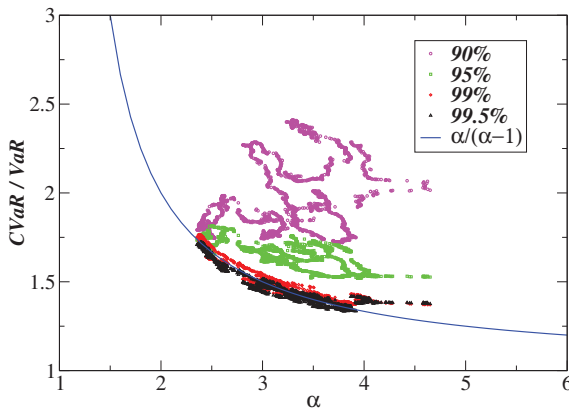
Figure 11. Time evolution of CVaR/VaR for long position at confidence levels of $X = 99.5\%, 99\%, 95\%$, and 90% as a function of α.
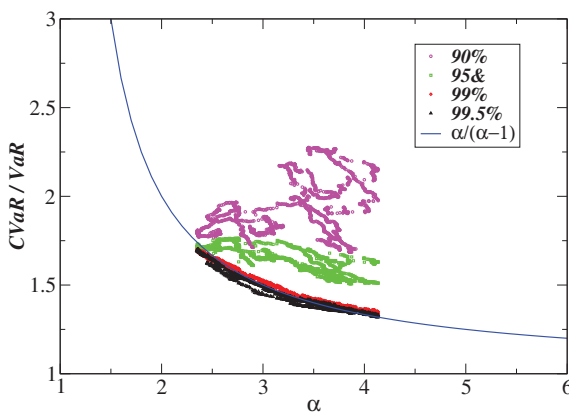
Figure 12. Time evolution of CVaR/VaR for short position at confidence levels of $X = 99.5\%$, $99\%, 95\%$, and 90% as a function of α.

Figure 13 shows the RV averaged over a 1-year window of data. The RV also varies over time and its time variation pattern is very similar to those of the VaR and CVaR, which indicates that the RV is correlated with the VaR and CVaR. Using the RV, we make two-dimensional plots of the risk measures and RV at confidence levels of $X = 99.5\%$ and

90%, as shown in Figures 14 and 15. These plots reveal a strong correlation between the VaR (CVaR) and RV. In our analysis of the 1-year window, we observed that the magnitude of the risk appears to change, resulting in trajectories representing changes in the VaR(CVaR)–RV plane. The trajectories found in the VaR–RV and CVaR–RV planes are very similar, and we also observed the similar trajectories for the short position (not shown here). These trajectories enable us to identify periods of high or low risk. The periods around 2015 and 2018 are classified as high risk, while the periods around 2017 and 2019 correspond to low-risk periods.
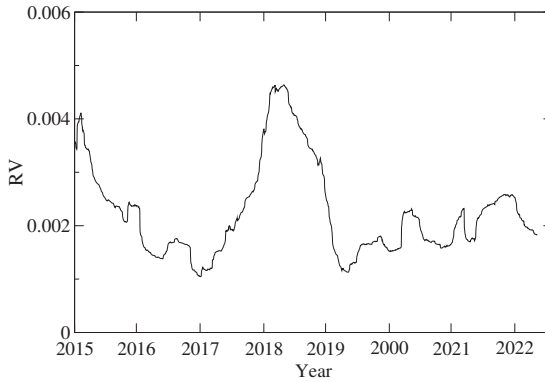
Figure 13. Time evolution of the daily RV averaged over a one-year window.
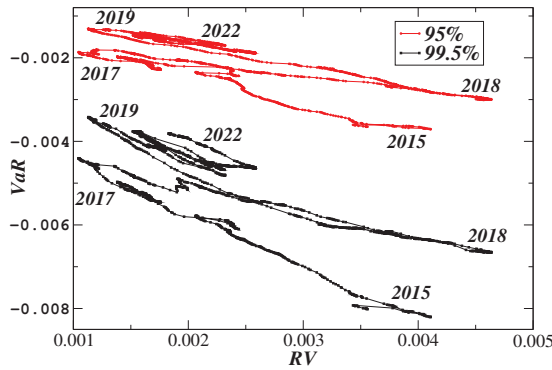
Figure 14. VaR for the long position versus RV at confidence levels of $X = 99.5\%$ and 95%.
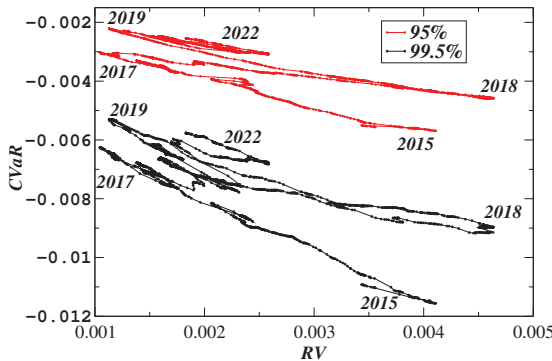
Figure 15. CVaR for the long position versus RV at confidence levels of $X = 99.5\%$ and 95%.

4. Discussion and Conclusions

Using high-frequency Bitcoin data, we conducted an investigation into the properties of the risk measures VaR and CVaR. By analyzing the risk measures in the left and right tails of the return probability distribution, we discovered "long–short asymmetry" for VaR and CVaR. This finding implies that the risk differs between long and short positions, particularly at high confidence levels. Such divergence in risks between long and short trades can offer valuable insights for trading strategies, particularly in the realm of high-frequency trading.

Furthermore, we observed that the ratios, CVaR/VaR, at high confidence levels align well with the $\alpha/(\alpha-1)$ law derived from the power-law distributional assumption. The presence of the $\alpha/(\alpha-1)$ law suggests that VaR and CVaR are no longer independent when the power-law distribution assumption holds.

Moreover, we observed a strong correlation between the risk measures and the RV, which resulted in the formation of trajectories in a two-dimensional plane. These trajectories unveiled periods of high and low risk. It would be intriguing to explore whether the high- and low-risk periods are associated with other measures, such as market efficiency (the Hurst exponent) (Bariviera 2017; Urquhart 2016), multifractality (Takaishi 2018), and inverted volatility asymmetry (Bouri et al. 2017; Katsiampa 2017; Stavroyiannis and Babalos 2017; Takaishi 2021b).

It is worth noting that our findings are based solely on Bitcoin data. Further investigations should be conducted using other assets to ascertain the universality of the "long–short asymmetry" and the $\alpha/(\alpha-1)$ law. Since previous studies (Gopikrishnan et al. 1998 1999) have already revealed that stock price returns exhibit the power-law probability distributions, we should expect the $\alpha/(\alpha-1)$ law for $CVaR/VaR$ on the stock markets.

In this study, we employed a 1-year window to investigate the time-varying properties. Consequently, the observed properties were averaged over the course of a year. To capture more dynamic changes within a year, smaller windows would be necessary. Although analyzing smaller windows poses some challenges due to reduced statistical data, it could yield interesting insights into more dynamic fluctuations.

Conflicts of Interest: The author declares no conflict of interest.

Notes

[1] In the context of the Pareto distribution, the similar expression can be found in Abad et al. (2014); Gourieroux and Jasiak (2010).

[2] http://www.bitstamp.net, accessed on 22 May 2022.

[3] When "$52,560 \times (X/100)$" is not a multiple of an integer, we interpolate two neighbor returns.

References

Abad, Pilar, Sonia Benito, and Carmen López. 2014. A comprehensive review of value at risk methodologies. *The Spanish Review of Financial Economics* 12: 15–32. [CrossRef]

Acerbi, Carlo, and Dirk Tasche. 2002. Expected shortfall: A natural coherent alternative to value at risk. *Economic Notes* 31: 379–88. [CrossRef]

Almeida, Dora, Andreia Dionísio, Isabel Vieira, and Paulo Ferreira. 2022. Uncertainty and risk in the cryptocurrency market. *Journal of Risk and Financial Management* 15: 532. [CrossRef]

Andersen, Torben G., and Tim Bollerslev. 1998. Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* 39: 885–905. [CrossRef]

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2003. Modeling and forecasting realized volatility. *Econometrica* 71: 579–625. [CrossRef]

Artzner, Philippe, Freddy Delbaen, Jean-Marc Eber, and David Heath. 1999. Coherent measures of risk. *Mathematical Finance* 9: 203–28. [CrossRef]

Bariviera, Aurelio F. 2017. The inefficiency of Bitcoin revisited: A dynamic approach. *Economics Letters* 161: 1–4. [CrossRef]

Begušić, Stjepan, Zvonko Kostanjčar, H. Eugene Stanley, and Boris Podobnik. 2018. Scaling properties of extreme price fluctuations in Bitcoin markets. *Physica A* 510: 400–6. [CrossRef]

Biglova, Almira, Sergio Ortobelli, Svetlozar T. Rachev, and Stoyan Stoyanov. 2004. Different approaches to risk estimation in portfolio theory. *The Journal of Portfolio Management* 31: 103–12. [CrossRef]

Bouri, Elie, Georges Azzi, and Anne Haubo Dyhrberg. 2017. On the return-volatility relationship in the Bitcoin market around the price crash of 2013. *Economics* 11: 1–16. [CrossRef]

Cheridito, Patrick, and Eduard Kromer. 2013. Reward-risk ratios. *Journal of Investment Strategies* 3: 3–18. [CrossRef]

Cont, Rama. 2001. Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance* 1: 223–36. [CrossRef]

Di Matteo, Tiziana, Tomaso Aste, and Michel M. Dacorogna. 2005. Long-term memories of developed and emerging markets: Using the scaling analysis to characterize their stage of development. *Journal of Banking & Finance* 29: 827–51.

Drożdż, Stanisław, Robert Gębarowski, Ludovico Minati, Paweł Oświęcimka, and Marcin Wątorek. 2018. Bitcoin market route to maturity? evidence from return fluctuations, temporal correlations and multiscaling effects. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28: 071101. [CrossRef]

Duffie, Darrell, and Jun Pan. 1997. An overview of value at risk. *Journal of Derivatives* 4: 7–49. [CrossRef]

Easwaran, Soumya, Manu Dixit, and Sitabhra Sinha. 2015. Bitcoin dynamics: The inverse square law of price fluctuations and other stylized facts. In *Econophysics and Data Driven Modelling of Market Dynamics*. Cham: Springer, pp. 121–28.

Gabaix, Xavier. 2009. Power laws in economics and finance. *Annual Review of Economics* 1: 255–94. [CrossRef]

Gopikrishnan, Parameswaran, Martin Meyer, L. A. Nunes Amaral, and H. Eugene Stanley. 1998. Inverse cubic law for the distribution of stock price variations. *The European Physical Journal B-Condensed Matter and Complex Systems* 3: 139–40. [CrossRef]

Gopikrishnan, Parameswaran, Vasiliki Plerou, Luis A. Nunes Amaral, Martin Meyer, and H. Eugene Stanley. 1999. Scaling of the distribution of fluctuations of financial market indices. *Physical Review E* 60: 5305. [CrossRef]

Gourieroux, Christian, and Joann Jasiak. 2010. Value at risk. In *Handbook of Financial Econometrics: Tools and Techniques*. Amsterdam: Elsevier pp. 553–615.

Hill, Bruce M. 1975. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics* 3: 1163–74. [CrossRef]

Hull, John. 2018. *Risk Management and Financial Institutions*. Hoboken: Wiley.

Katsiampa, Paraskevi. 2017. Volatility estimation for Bitcoin: A comparison of GARCH models. *Economics Letters* 158: 3–6. [CrossRef]

Linsmeier, Thomas J., and Neil D. Pearson. 2000. Value at risk. *Financial Analysts Journal* 56: 47–67. [CrossRef]

Liu, Lily Y., Andrew J. Patton, and Kevin Sheppard. 2015. Does anything beat 5-minute RV? a comparison of realized measures across multiple asset classes. *Journal of Econometrics* 187: 293–311. [CrossRef]

McAleer, Michael, and Marcelo C. Medeiros. 2008. Realized volatility: A review. *Econometric Reviews* 27: 10–45. [CrossRef]

Pan, Raj Kumar, and Sitabhra Sinha. 2007. Self-organization of price fluctuation distribution in evolving markets. *EPL (Europhysics Letters)* 77: 58004. [CrossRef]

Plerou, Vasiliki, Parameswaran Gopikrishnan, Luis A. Nunes Amaral, Martin Meyer, and H. Eugene Stanley. 1999. Scaling of the distribution of price fluctuations of individual companies. *Physical Review E* 60: 6519. [CrossRef]

Stavroyiannis, Stavros, and Vassilios Babalos. 2017. Dynamic properties of the Bitcoin and the US market. *SSRN Electronic Journal*. [CrossRef]

Takaishi, Tetsuya. 2018. Statistical properties and multifractality of Bitcoin. *Physica A* 506: 507–19. [CrossRef]

Takaishi, Tetsuya. 2021a. Recent scaling properties of Bitcoin price returns. *Journal of Physics: Conference Series* 1730: 012124. [CrossRef]

Takaishi, Tetsuya. 2021b. Time-varying properties of asymmetric volatility and multifractality in Bitcoin. *PLoS ONE* 16: e0246209. [CrossRef]

Takaishi, Tetsuya, and Takanori Adachi. 2020. Market efficiency, liquidity, and multifractality of Bitcoin: A dynamic study. *Asia-Pacific Financial Markets* 27: 145–54. [CrossRef]

Urquhart, Andrew. 2016. The inefficiency of Bitcoin. *Economics Letters* 148: 80–82. [CrossRef]

Wei, Wang Chun. 2018. Liquidity and market efficiency in cryptocurrencies. *Economics Letters* 168: 21–24. [CrossRef]

Article

The Six Decades of the Capital Asset Pricing Model:
A Research Agenda

Santosh Kumar [1], Ankit Kumar [2], Kamred Udham Singh [3,*] and Sujit Kumar Patra [4]

[1] Jaipuria Institute of Management, Jaipur 302033, India; talksant@gmail.com
[2] Department of Computer Engineering & Applications, GLA University, Mathura 281406, India;
    iiita.ankit@gmail.com
[3] School of Computing, Graphic Era Hill University, Dehradun 248002, India
[4] GITAM Institute of Management, GITAM (Deemed to Be) University, Vishakhapatnam 530045, India;
    patrasujitk@gmail.com
[*] Correspondence: 11004033@gs.ncku.edu.tw

Abstract: This paper re-examines the presence of the Sharpe–Treynor–Lintner–Mossin capital asset pricing model (CAPM) in the finance literature and is accompanied by a bibliometric summary analysis. The popular model is in its sixth decade; we summarized the relevance of the CAPM using publication and citation trends, as well as identifying its most prolific and impactful contributors. This paper is based on a systematic review of the literature and was completed with the help of various bibliometric techniques. During the study process, we presented a map of various themes and areas of the CAPM and its evolution. Our findings indicate that the extant literature on this topic (the cost of capital, asset pricing, portfolio, risk management, beta, systematic risk, and value premium) is based on the principles and assumptions of the CAPM. We are considering suggestions on the future use, trend, and direction of the CAPM, based on our summary of thematically developed clusters.

Keywords: capital asset pricing model (CAPM); bibliometric; systematic literature review research agenda

1. Introduction

The capital asset pricing model (CAPM) of Sharpe (1964), Treynor (1999), Lintner (1965a, 1965b), and Mossin (1966) rightfully occupies a central place in the asset pricing literature. Not surprisingly, an enormous research effort has been devoted to the testing of the CAPM over the past six decades due to its relatively simple and effective framework. The model is not only used to study the returns on shares but is also used for various conventional and non-conventional asset classes: USA painting (Agnello 2016); the oil market (Adekunle et al. 2020); credit market assets (Hwang et al. 2010); the real estate market (Coşkun et al. 2017); etc. No model is without its criticisms and the CAPM is subject to various limitations and challenges. To overcome such limitations, the CAPM has taken various forms in the last six decades of its journey: international CAPM (Black 1974; Stulz 1981), the inter-temporal capital asset pricing model (Merton 1973), expectile CAPM (Hu and Zheng 2020), and downside CAPM (Rutkowska-Ziarko et al. 2022), etc. The CAPM is changing in terms of its face and utilization across sets of literature in the field of finance; thus, providing a future direction for research has become of prime importance. The current study is an opportunity for the development of respective research fields. This study revolves around the following key ideas related to articles published with the keywords 'CAPM' in the last six decades:

- KI 1. The growth and pattern of publications and citations since its inception;
- KI 2. The detail provided by contributors and their affiliated institutions and countries to the related articles;
- KI 3. The frequency of article citations;
- KI 4. A detailed study of the most prominent themes and ideas published;

- KI 5. Details of a research agenda for future studies.

As suggested by Kraus et al. (2022), literature reviews can serve as a starting point for larger research projects. Literature reviews also present an understanding of a domain, providing a theoretical underpinning for empirical research. This review correspondingly provides an investigation into research conducted with the keyword 'CAPM', using this to present its current status and to suggest future directions. The explained key concerns of the research are answered systematically using bibliometric analysis tools. The rest of the paper has been structured in the following sections. Section 2 discusses the methodology used, while Section 3 presents the description of the data analysis. Sections 4 and 5 present the co-authorship, bibliographic coupling, and keyword analysis. Section 6 concludes by summarizing the findings, along with future directions, and Section 7 offers considerations for the implications, limitations, and future directions of this study.

2. Materials and Methods

Bibliometric analysis is a well-known research methodology in library and information sciences that uses published data and constructs meaningful summaries by quantifying the material. Otlet (1934) first explained the term bibliometric (bibliométrie) as "the measurement of all aspects related to the publication and reading of books and documents"; the credit of being the pioneer in the statistical analysis of bibliographic data is claimed by Pritchard (1969). A bibliometric study helps to identify intellectual contributions in the scientific field (Hota et al. 2019). A broad range of evidence for such a bibliometric study is available in a peer-reviewed journal's published data as a primary source of standardized information (Baker et al. 2020; Gil-Doménech et al. 2020). A bibliometric study is divided into two major nodes, initially explained in the literature of Kessler (1963). A bibliographic couple, as a common source of intelligence, is generated if two publications refer to one or more shared literature items. The literature in question cites two or more publications and each cited publication receives a co-citation. Co-citations represent the intellectual congruence and similarities in sources. Some frequently used bibliometric tools include the analysis of authorship and keywords as co-authorship and co-occurrence analyses (Peters and Van Raan 1991; Callon et al. 1983; Ravikumar et al. (2015). A bibliometric study also includes measures for productivity, with two dimensions: the number of publications and the influence through citations (Svensson 2010). Some other tools investigate each document according to citation count per publication and the h-index Alonso et al. (2009). The methodology is widely accepted for quantitative research in various domains (Ellegaard and Wallin 2015), including financial management Zupic and Cater (2015).

A structured bibliometric methodology was used in the current study to achieve the following objectives in line with the identified key ideas:

Objective 1: to examine the systematic growth pattern of publications and citations within the subject area since its inception;

Objective 2: to explore the contributions and characteristics of authors, affiliated institutions, and related countries;

Objective 3: to perform a detailed thematic analysis of the published articles and predict the future based on this.

The bibliometric study method has been improved with the empirical establishment of a hypothesis related to future trends related to the presence of the CAPM in research.

Ho. *There is no significant linear trend for the number of publications in the CAPM-related area.*

Ha. *There is a significant linear trend for the number of publications in the CAPM-related area.*

The research used bibliographic data to provide the significance of elements like the number of publications, the number of citations, authors' frequency, and contributing institutions and countries. The most prominent themes emerging from the publications were highlighted using keyword co-occurrence analysis and co-authorship analysis. The intellectual level of documents was revealed by analyzing keywords and network designing, primarily using "VOSviewer" software and the "Tableau" visual analytics tool. The

software was the primary logical instrument in the study of document citation, co-author networking, country networking, co-citation, and keyword search (Van Eck and Waltman 2009).

The data for the current bibliographic study were collected from the Scopus database using the keyword set ("Capital* Asset* Pricing* Model*" OR "CAPM*") (keywords correspond to the research title, abstract, and/or keywords of the article), and were published up to May-2023.

The Scopus database is considered to be the most suitable and most extensive available dataset for academic study (Bartol et al. 2014). The Scopus database is utilized for a variety of purposes: it is one of the most comprehensive databases of peer-reviewed academic literature; it is one of the most easily available databases containing the most recognized finance publications; it has extensive search capabilities, as well as bibliometric analysis tools such as exporting bibliographical data based on user requirements; and compared to other options such as Web of Science, it includes a substantial quantity of publications to fulfil the purposes of this study. Figure 1 explains the research design used in the study through a detailed flow chart. This visual representation explains the research's sequential and systematic approach, displaying each stage of the process from data collection to sorting, analysis, and interpretation. By following this well-structured research design, the study aimed to ensure rigor, reliability, and validity in the findings, ultimately contributing to the credibility and significance of the research outcomes.
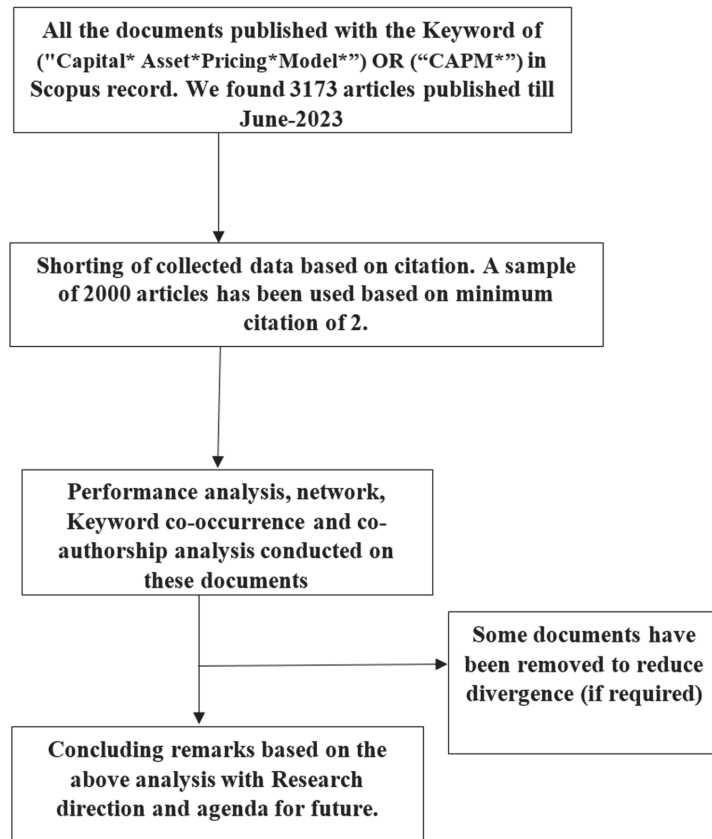

Figure 1. Design flow of the current study (author's own processing).

3. Description of the Data Collected

We found 3173 articles available in the Scopus database, with the selected keywords TITLE-ABS-KEY (capital AND asset AND pricing AND model OR capm) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "re")), up until June-2023. These articles were sorted based on their citations from the year 1972 until June-2023. We created a final sample of 2000 articles, with the elimination of articles with one or less than one citation.

The trend to use the CAPM from its inception showed exponential growth until 2021 for "number of publications" and until 2020 for "count of citations". The maximum number of articles was 171, published in 2020, with the maximum annual citations being 133 in the same year. The average annual publication number was nearly 61, with nearly 0.76 annual citations per article during our study period and sample, as explained in Figure 2. The wide acceptance of the CAPM as a fundamental model in finance can be easily explained via the participation of authors across the globe. Authors from 128 different countries have been contributing to the domain, and nearly 75% of the total contributions come from authors from 52 countries, with a minimum document number of five. Nearly 52% of contributions come from the top 10 countries, as shown in Figure 3 below. The maximum number of contributions comes from those of US affiliation, with a total of 738 articles and a total citation number of 42,694. The authors with Chinese affiliation come in second position, with a total contribution of 271 and 9739 citations. These numbers indicate a dominance of US-based authors in the field.
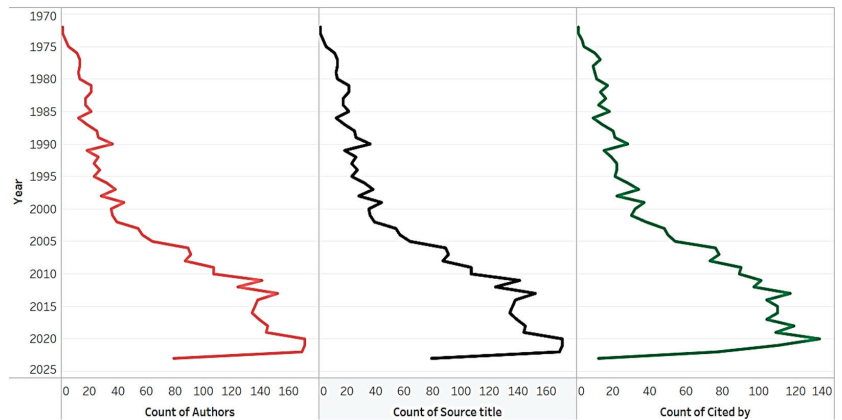

Figure 2. Annual trends in publications (distinct author count, number of titles, and count of citations using Tableau) (author's own processing).
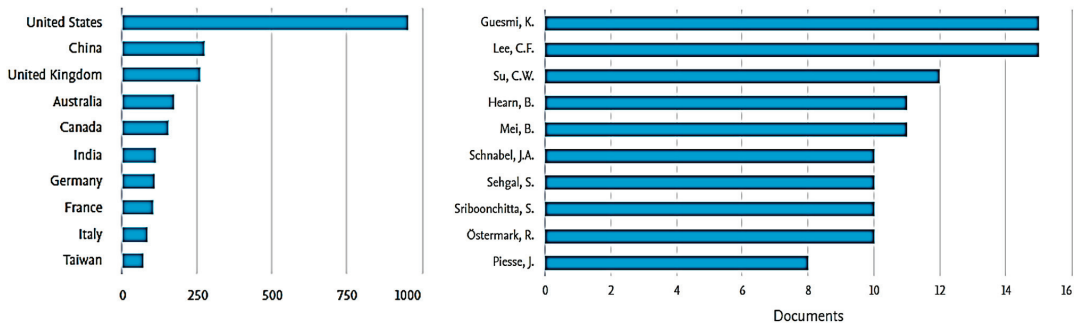

Figure 3. Top 10 countries/authors and their contributions (author's own processing).

Guesmi Khaled, from the Paris School of Business, France, and Lee Cheng Few from Rutgers University–New Brunswick, New Brunswick, US, are the highest contributing authors, with 15 articles each.

In terms of journal and subject-wise analysis of the keywords, as shown in Figure 4 below, we found that the "*Journal of Financial Economics*" has a maximum publication number of 83 documents, followed by the "*Journal of Finance*" with 78 documents. Below, Figure 4 explains a list of the top 10 journals in the field. Although the CAPM is considered to be the fundamental model in the field of economics, finance, and business management, we found that the model is used in 25 different subject areas. Even subject areas like agriculture and energy have good contributions, as shown in Figure 4 below. We have found the presence of the CAPM even in the subject areas of healthcare, immunology, and microbiology. This clearly indicates the wide acceptance of the CAPM in the academic universe.
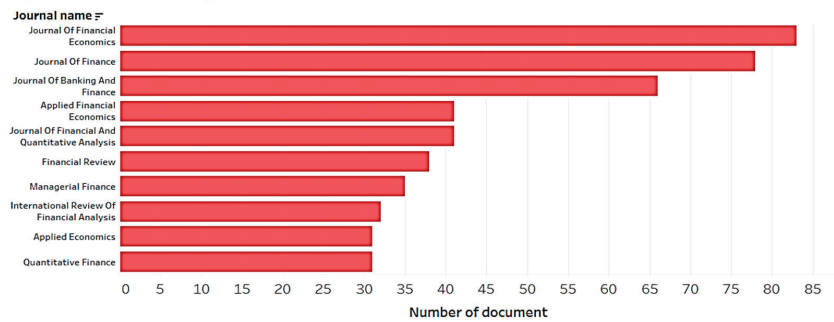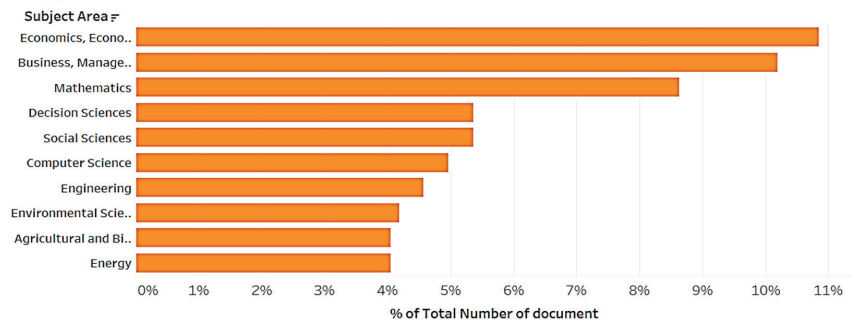
Figure 4. List of top 10 journals and subjects based on contributions with the help of Tableau (author's own processing).

The strong influence of the journal can also be supported by an increasing trend in the data of citations per cited publication.

4. Performance Analysis

The determination of the performance of a fundamental model in finance research can be easily justified by the acceptance trend, citation trend, and impactful publications. In this section, we present an analysis of the publications, citations, and network analysis, showing the relative performance.

Guesmi, Khaled, from the IPAG Business School, France, has the highest number of contributions of 15 documents, followed by Lee, Cheng-Few, from Rutgers University, the United States, with 15 documents. The total citations to date are 64,696, with an average citation score of 32 per article and 1269 citations per year in the selected sample and year.

Table 1 provides a list of the 15 most cited publications in the field, along with their annual citation indexes. Rolf W. Banz (1981) has had the maximum number of citations of 2520 since its publication, while Graham and Harvey (2001) have had the most effective citation score with 113 citations per year (C/Y score) since its publication date.

Table 1. Top 15 most cited articles (author's own processing).

Year	Citation Count	Authors' Full Names	Title	Source Title	C/Y
1981	2520	Banz, Rolf W.	*The relationship between return and market value of common stocks*	*Journal of Financial Economics*	60
2001	2482	Graham, John R.; Harvey, Campbell R.	*The theory and practice of corporate finance: Evidence from the field*	*Journal of Financial Economics*	113
1976	1402	Black, Fischer	*The pricing of commodity contracts*	*Journal of Financial Economics*	30
2005	1190	Acharya, Viral V.; Pedersen, Lasse Heje	*Asset pricing with liquidity risk*	*Journal of Financial Economics*	66
1979	1172	Breeden, Douglas T.	*An intertemporal asset pricing model with stochastic consumption and investment opportunities*	*Journal of Financial Economics*	27
2007	1121	Lambert, Richard; Leuz, Christian; Verrecchia, Robert E.	*Accounting information, disclosure, and the cost of capital*	*Journal of Accounting Research*	70
2001	1035	Hirshleifer, David	*Investor psychology and asset pricing*	*Journal of Finance*	47
1998	1002	Fama, Eugene F.; French, Kenneth R.	*Value versus growth: The international evidence*	*Journal of Finance*	40
2006	752	Barro, Robert J.	*Rare disasters and asset markets in the twentieth century*	*Quarterly Journal of Economics*	44
1997	735	Brav, Alon; Gompers, Paul A.	*Myth or reality? The long-run underperformance of initial public offerings: Evidence from venture and non-venture capital-backed companies*	*Journal of Finance*	28
2004	711	Fama, Eugene F.; French, Kenneth R.	*The Capital Asset Pricing Model: Theory and Evidence*	*Journal of Economic Perspectives*	37
1981	706	Reinganum, Marc R.	*Misspecification of capital asset pricing. Empirical anomalies based on earnings' yields and market values*	*Journal of Financial Economics*	17
2011	669	Cochrane, John H.	*Presidential Address: Discount Rates*	*Journal of Finance*	56
1976	653	Galai, Dan; Masulis, Ronald W.	*The option pricing model and the risk factor of stock*	*Journal of Financial Economics*	14
2000	617	Henry, Peter Blair	*Stock market liberalization, economic reform, and emerging market equity prices*	*Journal of Finance*	27

4.1. Author and Collaboration Study

The collaborative study between authors in the field explains the level of concentration and distribution of research ideas. The universe of the CAPM is highly diversified and well spread in terms of authorship, whereby only 38 authors from different affiliations show a co-authorship link strength of two and more, and the number of authors reduces to 23 with a limited co-authorship link strength of five, as shown in Table 2 below.

A low value for the co-authorship link strength indicates a very high level of acceptability of the CAPM, as the research idea is neither promoted by a limited number of authors nor a set of authors from the same organization.

Table 2. Top 15 high co-authorship link strengths (author's own processing).

Sr. No.	Author	TD	TC	TLS
1	Su, Chi-Wei	12	357	21
2	Sriboonchitta, S.	15	69	17
3	Umar, M.	11	263	17
4	Lee, A.C.	10	55	16
5	Qin, M.	8	313	16
6	Tao, R.	5	187	14
7	Lee, C.F.	15	72	13
8	Guesmi, K.	15	122	11
9	Tsai, C.-M.	5	16	10
10	Chanaim, S.	7	39	9
11	Autchariyapanitkul, K.	5	36	9
12	Chen, H.-Y.	6	34	8
13	Teulon, F.	8	54	7
14	Hearn, B.	11	155	6
15	Yamaka, W.	9	20	6

Note: TP = total publications; TC = total citations; TLS = total link strength.

The co-author network explains an author's association with others and the authorship pattern, as shown in Figure 5. There are 26 clusters dividing 566 contributing authors by network linkage using VoSviewer network software. The size of the node in the network reflects the number of collaborations with other nodes in the network. These clusters are created based on a linking network with the highest link of 38 items. The density of the links shows the strength of association between two nodes, i.e., a denser link between two nodes signifies a higher number of collaborations. The web reveals the growing collaboration and quality of collaboration among authors in the field.
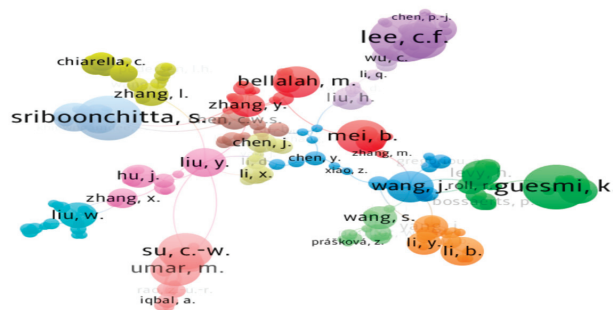
Figure 5. Co-authorship network visualization using VOSviewer (26 clusters) (author's own processing).

Figure 6 reveals the co-authorship coupling among the authors' affiliated countries. The CAPM model has wide acceptability, with authorship from more than 120 countries in the last six decades. To study the clustering effect of co-authorship with a limitation of a minimum of three documents from the country, a total of 61 qualified items with three clusters are shown in the network coupling diagram made using VoSviewer. The US shows a maximum link strength of 299, followed by the UK with a link strength of 177, and China with a link strength of 113. The three clusters have groups of 26, 22, and 13 countries, respectively.
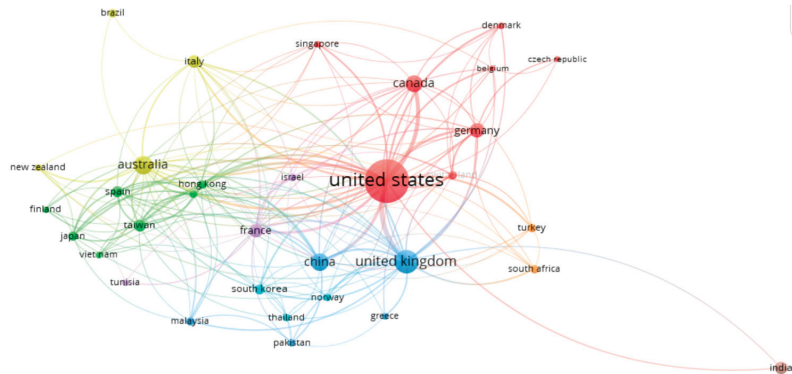
Figure 6. Co-authorship countries' network visualization using VOSviewer (3 clusters) (author's own processing).

The diagram showing the co-authorship coupling among the authors' affiliated institutions explains the relationships among contributing institutions, shown in Figure 7. Co-authorships in scientific documents are the formal way of having intellectual or scientific collaborations among scholars (Acedo et al. 2006). The coupling relation explains the existence of two clusters with a total of 131 institutions with different geographical locations. Cluster 1 has a group of 121 institutions, dominated by the NBER (National Bureau of Economic Research), US, with a link strength of 27. Cluster 2 has a group of 10 institutions, dominated by the CEPR (Centre for Economic Policy Research), UK, with a link strength of 22. The study reveals a stronghold of the CAPM as the keyword in the research of these institutions and shows high reach in every part of the globe.
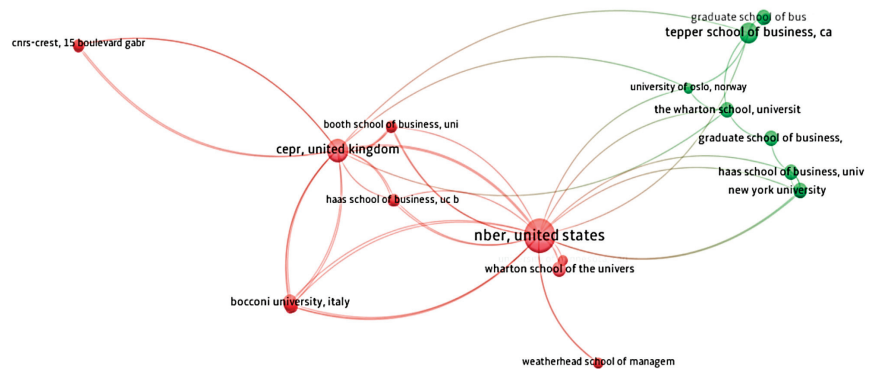
Figure 7. Co-authorship organization network visualization using VOSviewer (2 clusters) (author's own processing).

4.2. Keyword Coupling Study

Comerio and Strozzi (2019) explained the importance of the keyword co-occurrence study as an adequate explanation of the publication themes and trends in the research domain. The keyword coupling for "CAPM" and "Capital Asset Pricing Model" explains the strength and linkage of the words in other finance-related domains. The keyword coupling study based on the author's keyword with the minimum frequency of five divides all qualified documents into 10 clusters with two hundred eighty-seven keywords. In Figure 8, clusters are presented in different colors, along with the dimensions of the bubble representing the level of a keyword's connectedness. The "CAPM" and "Capital Asset Pricing Model" keywords have a combined link strength of 1033.

Figure 8. Co-occurrence keyword network with a frequency of 5 using VOSviewer (10 clusters) (author's own processing).

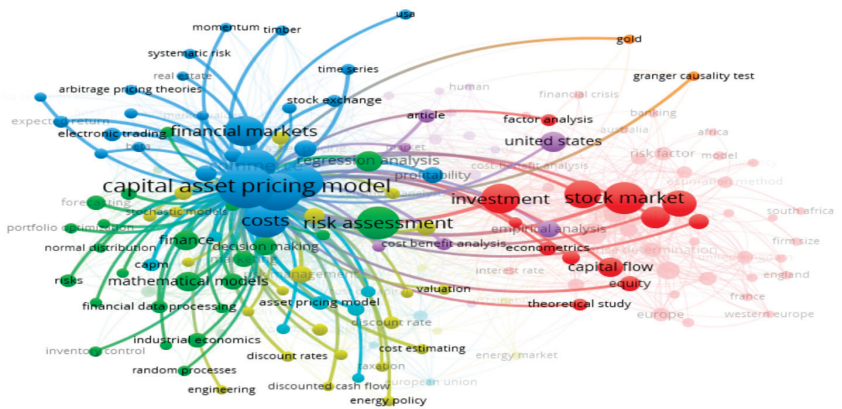Clusters 1 and 2, with forty-nine keywords each, dominated the publication data in the areas of asset pricing and investment. The importance of the CAPM as the center of the portfolio and asset pricing research has been proven by this keyword coupling study. Research areas like cost of capital, asset pricing, portfolio, risk management, beta, systematic risk, value premium, and many more have direct links with the CAPM.

5. Discussion and Future Trends

The current bibliometric study confirms the acceptance and growth of the CAPM in the finance research area. The CAPM keyword has entered its sixth decade and is used in various research fields and to promote different ideas. Figure 9, a keyword network with different frequency levels, shows the strength of keywords in the domain research. The CAPM has been used with nearly 5349 keywords at least once in the available sets of the literature. There were 110 research ideas that used the CAPM at least 10 times in the sample used for the study. Similarly, we can analyze the strength of the CAPM with other keywords and research items according to the frequency of usage from the data given below in Figure 9.

With frequency of 50 (8 items; 1 cluster)

With frequency of 40 (11 items; 1 cluster)

Figure 9. *Cont.*
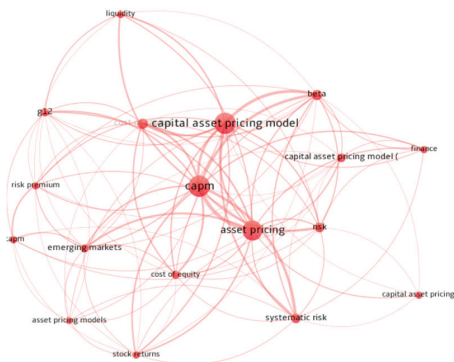
With frequency of 30 (18 items; 1 cluster)
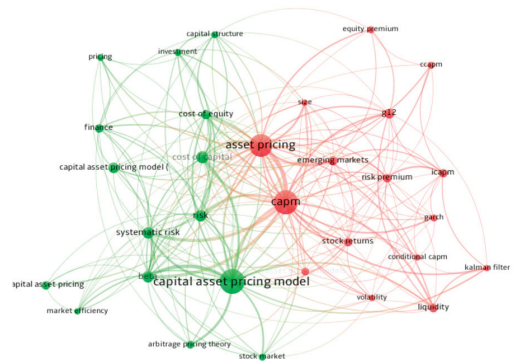
With frequency of 20 (31 items; 2 clusters)

With frequency of 15 (54 items; 2 clusters)
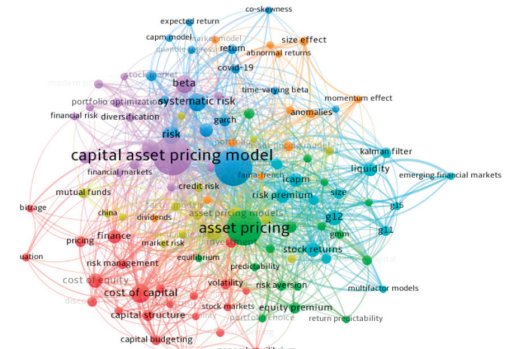
With frequency of 10 (110 items; 7 clusters)
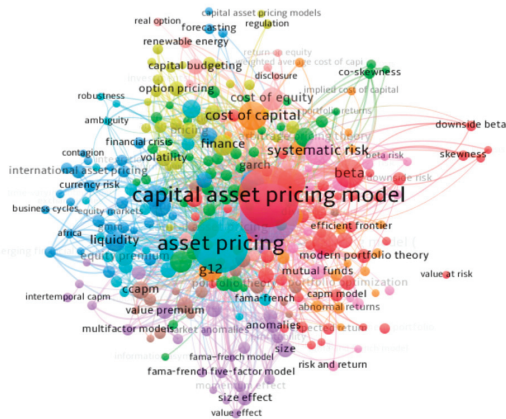
With frequency of 5 (287 items; 10 clusters)
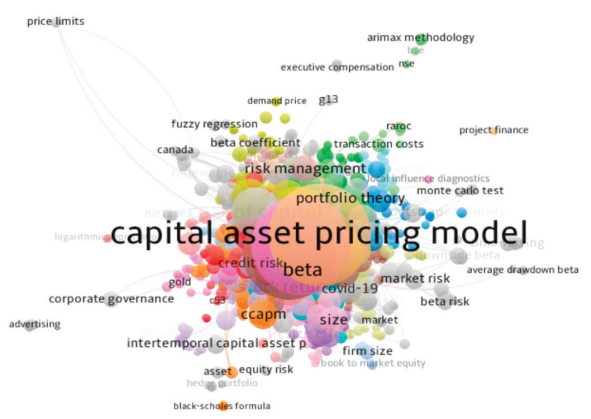
With frequency of 1 (5349 items; 35 clusters)

Figure 9. Keyword network with different frequency levels using VOSviewer (author's own processing).

The publication and research trend of the CAPM shows the dominance of five journals in the last six decades, as shown in Figure 10. One can obviously see the increase in the number of documents in the "Applied Finance and economic journal" between 2008 and 2014; similarly, the "Journal of Banking and Finance" shows spikes during the early years between 1974 and 1984 and later in the years from 2014 onwards. These journals neither show consistent growth nor downfall in the field. Silver (2012) supports trend analysis to

examine upward, downward, or cyclical movements in indicators in such cases based on historical data.
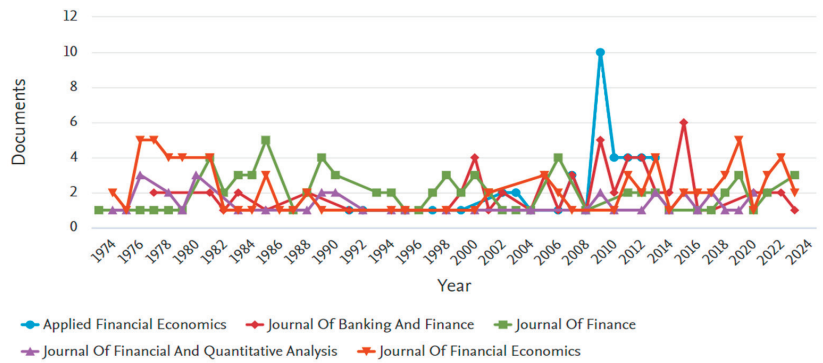
Figure 10. Year-wise publications in top 5 contributing journals in the field (author's own processing).

A linear trend model was computed for a distinct count of titles (actual and forecasted) in a given year. The model may be significant at $p \leq 0.05$. The factor of the forecast indicator may be significant at $p \leq 0.05$. The p-value statistics helped us to reject the null hypothesis (Ho) and accept the alternative hypothesis (Ha), "There is significant linear trend for number of publications with the CAPM related area", which helped us to achieve Objective 3 of the study. The outcome of the model, as outlined in the Appendix A, indicates a positive future trend in this research area. In outcome of forecasted trend line presented in Figure 11, we found a downfall in the number of annual publications in the last few years, good growth for the future is being seen in terms of numbers.
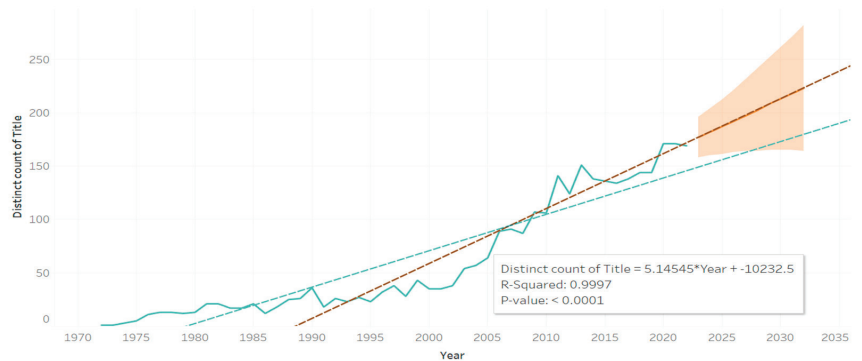
Figure 11. Future trends of research in the field (using Tableau) (author's own processing).

6. Conclusions

The result of the current bibliometric study confirms the acceptability and growth of the CAPM in the finance research area in the last six decades. The study helps to provide some meaningful insights for academicians, researchers, reviewers, and editorial board members about the research trend and contribution of the CAPM. The CAPM has been used in various research in different subject areas with increasing recognition by scholars of different institutions and countries. The study findings suggest that the primary research area for the CAPM is to be used in publishing original empirical research of a high quality from the financial sector. This is attracting high-quality research work from researchers from various domains, like cost of capital, asset pricing, portfolio, risk management, beta, systematic risk, value premium, and many more.

The data available from Scopus reveal the growth in citations and geographical penetration, but the research area is still dominated by authors from the US, UK, China, and Australia.

The study used various techniques to examine the impact of the CAPM since its inception between 1972 and 2023. The study highlighted development in both publications and citations to measure productivity and reputation. Over time, the CAPM has transitioned from focusing on promoting and developing systematic risk and portfolio management research to a broader array of mainly the economics and econometrics categories. The findings suggest that quantitative research with essential words related to the cost of capital, capital structure, emerging market, asset pricing, and value premium dominates publications. The bibliographic coupling of articles established by the NBER (National Bureau of Economic Research) from the US and the CEPR (Centre for Economic Policy Research) from the UK are the top two contributing institutions.

An analysis of the evolution of new research in the area of finance and its relation with the CAPM explains its bright future. The result of the trend analysis is in favor of the CAPM and its acceptability in the research domain, even in the upcoming future. The *p*-value-statistics-based rejection of the null hypothesis (Ho) supports the relevance of the CAPM as an important research topic in the domain in the future.

7. Implications, Limitations, and Future Directions

This study has put forward and achieved three objectives to streamline the implications: Objective 1 helped to identify the trends in article publications over time and their key influence, along with the citation dynamics; Objective 2 provided a detailed analysis of the collaborative networks and research productivity to shed light on the global acceptance of the CAPM within the subject area; and Objective 3 aimed to provide the significant breakthroughs within the subject area, offering valuable insights into the intellectual landscape and the major future research directions.

By achieving these objectives, the study contributes valuable knowledge about the subject area's evolution, the contributions of various stakeholders, and the prevailing themes and ideas that have influenced its development. The current study based on the utilization and importance of the CAPM in the finance literature demonstrates the relevance of the concept in research, mainly in the fields of investment management, portfolio management, return, and risk aspects. Using the bibliometric method of highlighting the changes in the CAPM, the outcome can be used by researchers and academicians in the field to set their research agendas and plan their course delivery. The research outcome can also be suitable for professionals in the field of asset management to decide upon their dependency on the CAPM or any fundamental asset pricing model. Although the frequency of the CAPM usage in the literature has shown a downward trend in the last few years, the concept is still valid and helpful for the development of new portfolio management strategies. Additionally, the findings can serve as a foundation for future research agendas, inform decisions in academia, policy, and industry related to the fundamental investment model, contribute to the body of knowledge, and foster evidence-based decision making for stakeholders.

While our research has yielded valuable insights, it is essential to acknowledge some limitations that may have influenced the findings. The primary limitation lies in the selection of research articles for our study. We only considered 2000 articles out of a total of 317,300 available in the subject area. This sampling constraint could potentially introduce a bias in the results and limit the generalizability of our findings to the entire population of published articles.

To address the limitations mentioned above and enhance the robustness of the study, future research in this area can explore the following directions: the inclusion of all published articles, and extending the use of the model to other related studies or subject areas can be beneficial to the field and can advance one's understanding of the subject area. Ad-

ditionally, such efforts will strengthen the credibility and applicability of research findings in diverse contexts.

Appendix A

Appendix A.1. Trend Lines Model

A linear trend model was computed for a distinct count of titles (actual and forecasted) in a given year. The model may be significant at $p \leq 0.05$. The factor of the forecast indicator may be significant at $p \leq 0.05$.

Model Formula:	Forecast Indicator × (Year + Intercept)
Number of modeled observations:	61
Number of filtered observations:	0
Model degrees of freedom:	4
Residual degrees of freedom (DFs):	57
SSE (sum squared error):	17,334
MSE (mean squared error):	304.105
R-Squared:	0.940849
Standard error:	17.4386
p-value (significance):	<0.0001

Appendix A.2. Analysis of Variance

Field	DF	SSE	MSE	F	*p*-Value
Forecast indicator	2	10,184.821	5092.41	16.7456	<0.0001

Appendix A.3. Individual Trend Lines

Panes		Color	Line		Coefficients				
Row	Column	Forecast Indicator	*p*-Value	DF	Term	Value	StdErr	t-Value	*p*-Value
Distinct count of titles	Year	Estimate	<0.0001	8	Year	5.87273	0.0340151	172.651	<0.0001
					Intercept	−11,710.3	68.9656	−169.798	<0.0001
Distinct count of titles	Year	Actual	<0.0001	49	Year	3.13231	0.178921	17.5067	<0.0001
					Intercept	−6198.77	357.314	−17.3482	<0.0001

References

Acedo, Francisco José, Carmen Barroso, Cristobal Casanueva, and José Luis Galán. 2006. Co-authorship in management and organizational studies: An empirical and network analysis. *Journal of Management Studies* 43: 957–83. [CrossRef]

Adekunle, Wasiu, Abubakar M. Bagudo, Monsuru Odumosu, and Suraj B. Inuolaji. 2020. Predicting stock returns using crude oil prices: A firm-level analysis of Nigeria's oil and gas sector. *Resources Policy* 68: 101708. [CrossRef]

Agnello, J. Richard. 2016. Do U.S. paintings follow the CAPM? Findings disaggregated by subject, artist, and value of the work. *Ricerche Economiche [Research in Economics]* 70: 403–11. [CrossRef]

Alonso, Sergio, Francisco Javier Cabrerizo, Enrique Herrera-Viedma, and Francisco Herrera. 2009. H-index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics* 3: 273–89. [CrossRef]

Baker, H. Kent, Satish Kumar, and Debidutta Pattnaik. 2020. Twenty-five years of the journal of corporate finance: A scientometric analysis. *Journal of Corporate Finance* 66: 101572. [CrossRef]

Banz, Rolf W. 1981. The relationship between return and market value of common stocks. *Journal of Financial Economics* 9: 3–18. [CrossRef]

Bartol, Tomaz, Gordana Budimir, Doris Dekleva-Smrekar, Miro Pusnik, and Primoz Juznic. 2014. Assessment of research fields in Scopus and web of science in the view of national research evaluation in Slovenia. *Scientometrics* 98: 1491–504. [CrossRef]

Black, Fischer. 1974. International capital market equilibrium with investment barriers. *Journal of Financial Economics* 1: 337–52. [CrossRef]

Callon, Michel, Jean-Pierre Courtial, William A. Turner, and Serge Bauin. 1983. From translations to problematic networks: An introduction to co-word analysis. *Social Science Information* 22: 191–235. [CrossRef]

Comerio, Niccolo, and Fernanda Strozzi. 2019. Tourism and its economic impact: A literature review using bibliometric tools. *Tourism Economics* 25: 109–31. [CrossRef]

Coşkun, Yener, A. Sevtap Selcuk-Kestel, and Bilgi Yilmaz. 2017. Diversification benefit and return performance of REITs using CAPM and Fama-French: Evidence from Turkey. *Borsa Istanbul Review* 17: 199–215. [CrossRef]

Ellegaard, Ole, and Johan A. Wallin. 2015. The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics* 105: 1809–31. [CrossRef]

Gil-Doménech, Dolors, Jasmina Berbegal-Mirabent, and José M. Merigó. 2020. STEM education: A bibliometric overview. In *International Conference on Modelling and Simulation in Management Science*. Cham: Springer. [CrossRef]

Graham, John R., and Campbell R. Harvey. 2001. The theory and practice of corporate finance: Evidence from the field. *Journal of Financial Economics* 60: 187–243. [CrossRef]

Hota, Pradeep Kumar, Balaji Subramanian, and Gopalakrishnan Narayanamurthy. 2019. Mapping the intellectual structure of social entrepreneurship research: A citation/co-citation analysis. *Journal of Business Ethics* 166: 89–114. [CrossRef]

Hu, Wei, and Zhenlong Zheng. 2020. Expectile CAPM. *Economic Modelling* 88: 386–97. [CrossRef]

Hwang, Young-Soon, Hong-Ghi Min, Judith A. McDonald, Hwagyun Kim, and Bong-Han Kim. 2010. Using the credit spread as an option-risk factor: Size and value effects in CAPM. *Journal of Banking & Finance* 34: 2995–3009. [CrossRef]

Kessler, M. Myer. 1963. Bibliographic coupling between scientific articles. *American Documentation* 14: 123–31. [CrossRef]

Kraus, Sascha, Matthias Breier, Weng Marc Lim, Marina Dabić, Satish Kumar, Dominik Kanbach, Debmalya Mukherjee, Vincenzo Corvello, Juan Piñeiro-Chousa, Eric Liguori, and et al. 2022. Literature reviews as independent studies: Guidelines for academic practice. *Review of Managerial Science* 16: 2577–95. [CrossRef]

Lintner, John. 1965a. Security Prices, Risk and Maximal Gains from Diversification. *Journal of Finance* 20: 587–615.

Lintner, John. 1965b. The Valuation of Risk Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets. *Review of Economics and Statistics* 47: 13–37. [CrossRef]

Merton, C. Robert. 1973. An inter-temporal capital asset pricing model. *Econometrica: Journal of the Econometric Society* 41: 867–87. [CrossRef]

Mossin, Jan. 1966. Equilibrium in a Capital Asset Market. *Econometrica* 35: 768–83. [CrossRef]

Otlet, Paul. 1934. *Traité de Documentation: Le Livre Sur le Livre, Théorie et Pratique*. Brussels: Editiones Mundaneum.

Peters, Hugo, and A. Van Raan. 1991. Structuring scientific activities by co-author analysis: An exercise on a university faculty level. *Scientometrics* 20: 235–55. [CrossRef]

Pritchard, Alan. 1969. Statistical bibliography or bibliometrics? *Journal of Documentation* 25: 348–49.

Ravikumar, Sharina, Ashutosh Agrahari, and Sri Nivas Singh. 2015. Mapping the intellectual structure of scientometrics: A co-word analysis of the journal scientometrics (2005–2010). *Scientometrics* 102: 929–55. [CrossRef]

Rutkowska-Ziarko, Anna, Lesław Markowski, Christopher Pyke, and Saqib Amin. 2022. Conventional and downside CAPM: The case of London stock exchange. *Global Finance Journal* 54: 100759. [CrossRef]

Sharpe, William F. 1964. Capital Asset Prices: A Theory of Market Equilibrium Under Conditions of Risk. *Journal of Finance* 19: 425–42.

Silver, Nate. 2012. *The Signal and the Noise: Why So Many Predictions Fail—But Some Don't*. London: Penguin Books.

Stulz, M. Rene. 1981. A model of international asset pricing. *Journal of Financial Economics* 9: 383–406. [CrossRef]

Svensson, Goran. 2010. SSCI and its impact factors: A 'prisoner's dilemma? *European Journal of Marketing* 44: 23–33. [CrossRef]

Treynor, L. Jack. 1999. Toward a Theory of Market Value of Risky Assets. In *Asset Pricing and Portfolio Performance*. Edited by Robert A. Korajczyk. London: Risk Books, pp. 15–22.

Van Eck, Nees Jan, and Ludo Waltman. 2009. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics* 84: 523–38. [CrossRef]

Zupic, Ivan, and Tomaz Cater. 2015. Bibliometric methods in management and organization. *Organizational Research Methods* 18: 429–72. [CrossRef]

Communication

Exploring Implied Certainty Equivalent Rates in Financial Markets: Empirical Analysis and Application to the Electric Vehicle Industry

Yifan He * and Svetlozar Rachev

Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX 79409-1042, USA;
zari.rachev@ttu.edu
* Correspondence: yifan.he@ttu.edu

Abstract: In this paper, we mainly study the impact of the implied certainty equivalent rate on investment in financial markets. First, we derived the mathematical expression of the implied certainty equivalent rate by using put-call parity, and then we selected some company stocks and options; we considered the best-performing and worst-performing company stocks and options from the beginning of 2023 to the present for empirical research. By visualizing the relationship between the time to maturity, moneyness, and implied certainty equivalent rate of these options, we have obtained a universal conclusion—a positive implied certainty equivalent rate is more suitable for investment than a negative implied certainty equivalent rate, but for a positive implied certainty equivalent rate, a larger value also means a higher investment risk. Next, we applied these results to the electric vehicle industry, and by comparing several well-known US electric vehicle production companies, we further strengthened our conclusions. Finally, we give a warning concerning risk, that is, investment in the financial market should not focus solely on the implied certainty equivalent rate, because investment is not an easy task, and many factors need to be considered, including some factors that are difficult to predict with models.

Keywords: put-call parity; implied put-call parity certainty equivalent rate; electric vehicle industry

1. Introduction

The certainty equivalent rate is a measure derived from the certainty equivalent[1], which plays a pivotal role in financial investment. Investors usually need to refer to the changing trend of this value to decide whether a certain company is worth investing in or if certain companies are worth investing in, that is, it is used to determine the priority of investment. The main purpose of this paper is to solve these two problems.

First, we give the mathematical expression of the certainty equivalent rate using the put-call parity formula. From the mathematical expression, we can find the factors that cause changes in the certainty equivalent rate.

Second, we select the stocks and options of three companies with the best performance from the beginning of 2023 to the current time and the stocks and options of three companies with the worst performance for empirical research. Through data visualization, we obtain a general conclusion.

Third, we apply the general conclusions drawn to the US electric vehicle industry. Specifically, we select three well-known US electric vehicle companies and explore whether they are worth investing in from the perspective of the certainty equivalent rate and the priority of investment.

Finally, we summarize this paper and emphasize that investment is an extremely complicated matter that requires the consideration of many factors, not just the certainty equivalent rate. When many factors are considered, investors are more likely to make the optimal decision.

2. Theoretical Support

The key theorem that we will use is put-call parity. A detailed explanation of put-call parity can be found in Hull (2022). Overall, Hull (2022) considers the interest and dividends to be paid in accordance with continuous compounding, but in a real financial market, interest and dividends are more likely to be paid at specific points in time rather than every second. Thus, we prefer to use the discrete-compounding version of put-call parity when we consider problems in a real financial market. Hence, the following is the detailed mathematical expression of put-call parity that we will use in this paper:

$$C + \frac{K}{(1+r)^T} = P + \frac{S}{(1+q)^T},$$ (1)

where

$$\begin{cases} T & \stackrel{\text{def}}{=} \text{The time to maturity,} \\ C & \stackrel{\text{def}}{=} \text{A given company's call option price with respect to the maturity date,} \\ P & \stackrel{\text{def}}{=} \text{A given company's put option price with respect to the maturity date,} \\ K & \stackrel{\text{def}}{=} \text{A given company's option strike price with respect to the maturity date,} \\ S & \stackrel{\text{def}}{=} \text{A given company's stock price with respect to the start date,} \\ q & \stackrel{\text{def}}{=} \text{A given company's dividend yield,} \\ r & \stackrel{\text{def}}{=} \text{A given company's put-call parity certainty equivalent rate.} \end{cases}$$

Based on (1), we can obtain the mathematical expression for r:

$$r = \left[\frac{K(1+q)^T}{S + (P-C)(1+q)^T} \right]^{\frac{1}{T}}.$$ (2)

In the following sections, we will mainly use (2) to explore the relationship between the time to maturity T, the moneyness S/K, and the implied company-specific put-call parity certainty equivalent rate r.

3. Empirical Research

3.1. Preparation

Before conducting our empirical research, we had to figure out how to obtain the values of the arguments in (2):

- Argument S: Since our purpose is to explore the stocks' behavior in 2023, we choose the start date as 3 January 2023, which is the first business day in 2023. Therefore, S in (2) will be the company's stock price on 3 January 2023. We can find these values in every stock's "historical data" section on Yahoo Finance.
- Argument q: We will consider the value of the "forward annual dividend yield"; the relevant data can be found in the "statistics" section on Yahoo Finance.
- Arguments T and K: On CBOE, we can find a given company's stock option's strike T and its maturity date. Then, we subtract the start date (3 January 2023) from the maturity date and convert the result into years[2]. Finally, we obtain the value of the time to maturity T.
- Arguments C and P: Although we cannot obtain the values of these two arguments directly from CBOE, we can obtain the "bid" and "ask" of every option. Here, we calculate the mid-price of the bid and ask, and we consider it to be the corresponding call option price and put option price.

3.2. Data Visualization and Explanation

Based on financial news from Saul (2023), Son (2023), Shinn and Velasquez (2023), and Khederian (2023), we can select three of the best-performing stocks, which come from Apple, Nvidia, and Meta, respectively. On the other hand, the three worst-performing stocks that we select are from First Republic Bank, Signature Bank, and Charles Schwab, respectively. Next, we use MATLAB to create figures that describe the relationship between the time to maturity T^3, the moneyness S/K^4, and the implied put-call parity certainty equivalent rate r for the companies we selected, and we explain some key values from these figures. (Figures 1–3 correspond to the best-performing companies, while Figures 4–6 correspond to the worst-performing companies).

Figure 1. Relationship between the time to maturity, moneyness, and the implied put-call parity certainty equivalent rate for Apple stock.

Figure 2. Relationship between the time to maturity, moneyness, and the implied put-call parity certainty equivalent rate for Nvidia stock.

From Table 1[5], it is clear that the best-performing companies have strictly positive implied put-call parity certainty equivalent rates, regardless of the maximum value, minimum value, or mean value. On the other hand, the worst-performing companies have strictly negative implied put-call parity certainty equivalent rates.

Figure 3. Relationship between the time to maturity, moneyness, and the implied put-call parity certainty equivalent rate for Meta stock.

Figure 4. Relationship between the time to maturity, moneyness, and the implied put-call parity certainty equivalent rate for First Republic Bank stock.
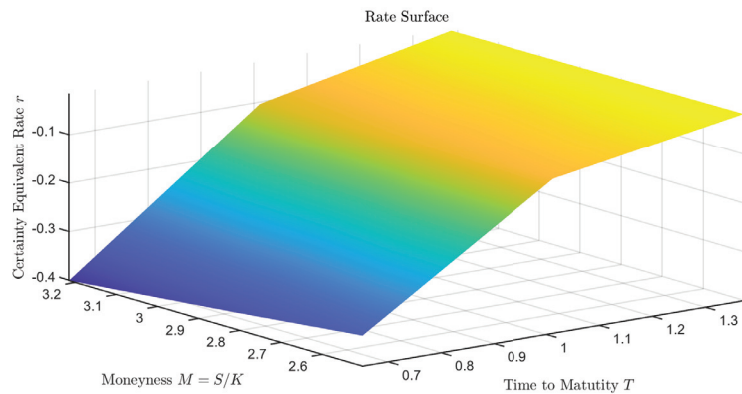
Figure 5. Relationship between the time to maturity, moneyness, and the implied put-call parity certainty equivalent rate for Signature Bank stock.
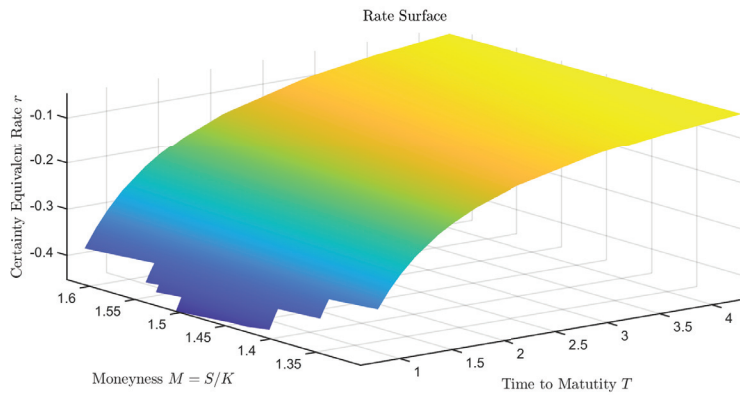
Figure 6. Relationship between the time to maturity, moneyness, and the implied put-call parity certainty equivalent rate for Charles Schwab stock.

Table 1. Implied put-call parity certainty equivalent rate (best- and worst-performing companies).

Company	Maximum Value	Minimum Value	Mean Value
Apple	0.8644	0.1181	0.4437
Nvidia	4.0908	0.3308	1.8897
Meta	2.6418	0.2238	1.2135
First Republic Bank	−0.8600	−0.9998	−0.9731
Signature Bank	−0.0166	−0.4008	−0.1601
Charles Schwab	−0.0481	−0.4523	−0.2625

A positive rate means that investors have a high probability of obtaining a return if they invest in the company's stock, while a negative rate means that investors may lose their money if they try to invest in the company's stock. By combining the data and the financial news, we find that the implied put-call parity certainty equivalent rate is quite useful; it can help investors determine which company's stock option is worth investing in.

Now, let us consider the shape of the graph. For the best-performing companies, we can observe that when the time to maturity T is small, the corresponding rate is high, which means that in the near future the return of the stock option is quite high, so investors may make money by investing in the option. Of course, they may have to take on some amount of risk. In this case, we would suggest that the investor consider Apple first, then Meta, and finally Nvidia. A high rate represents a high risk, and normal investors definitely do not want to take on a high risk when they decide to invest in something.

As time goes by, i.e., as the value of the time to maturity T becomes larger, the implied put-call parity certainty equivalent rate will become smaller because it is reasonable to consider the long-term rate as the riskless rate of the financial market. Additionally, it is clear that the riskless rate of the financial market should be lower than the near-future implied put-call parity certainty equivalent rates of the best-performing companies.

Let us consider the worst-performing companies. We can see that the values of the certainty equivalent rates of these companies are negative, which means that investors have a high risk of losing money if they decide to invest in these companies' stock options.

Finally, from Figures 1–6, we can see that if we fix the value of the time to maturity T and change the value of the moneyness S/K, the value of the implied certainty equivalent rate r hardly changes, which tells us that the implied certainty equivalent rate is almost independent of moneyness.

4. Application: Electric Vehicle Industry

Today, more and more people are paying attention to environmental issues. To reduce the pollution released by vehicles, people are considering driving electric vehicles[6] instead of traditional oil-powered vehicles. In the US, there are several well-known companies that produce electric vehicles, such as Tesla, General Motors, and Ford Motor Company[7]. We can apply the results we obtained in the previous section to these electric vehicle companies. Figures 7–9 visualize the data of these companies, and the key values derived from these figures are given in Table 2.
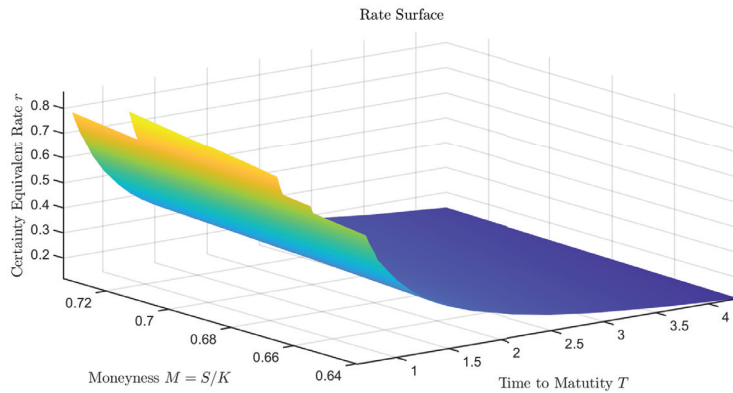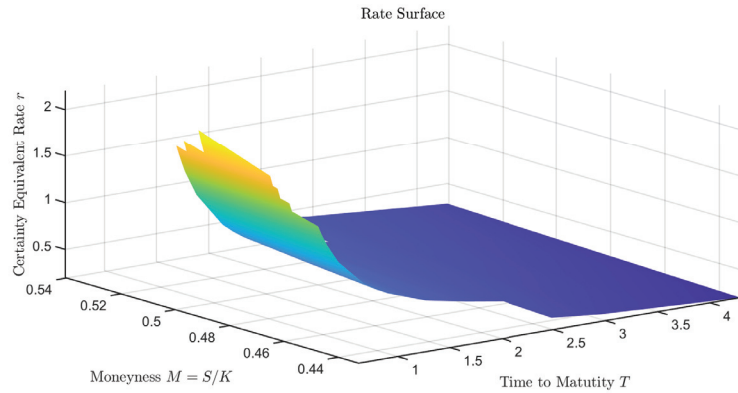

Figure 7. Relationship between the time to maturity, moneyness, and the implied put-call parity certainty equivalent rate for Tesla stock.
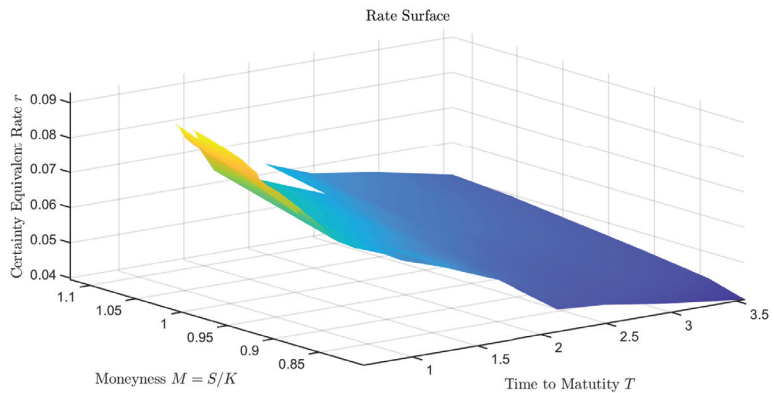
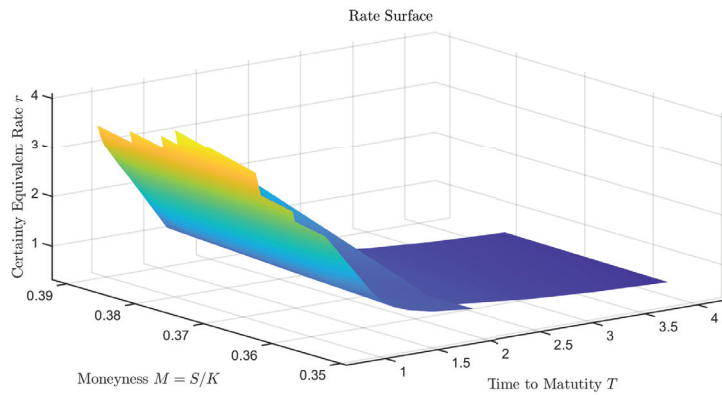**Figure 8.** Relationship between the time to maturity, moneyness, and the implied put-call parity certainty equivalent rate for General Motors stock.

Table 2. Implied put-call parity certainty equivalent rate (electric vehicle companies).

Company	Maximum Value	Minimum Value	Mean Value
Tesla	2.1970	0.1994	1.0320
General Motors	0.0927	0.0395	0.0708
Ford Motor Company	0.4549	0.0615	0.2626

Based on the shapes of the graphs shown in Figures 7–9, we can see that these three companies are all doing well. In the near future, we recommend that investors invest in these companies' stock options. According to the data from Table 2, we can see that if we

plot the three surfaces on the same coordinate axis, the order of these three surfaces from top to bottom is Tesla, then Ford Motor Company, and finally General Motors. Based on the positional relationships, we can see that investors should prefer to invest in General Motors, then Ford Motor Company, and finally Tesla.
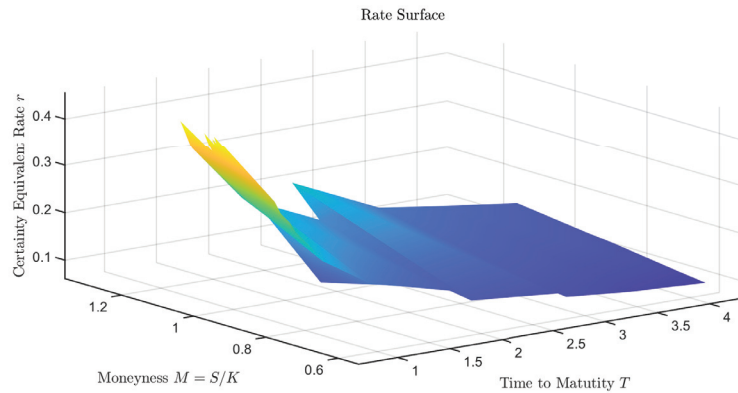
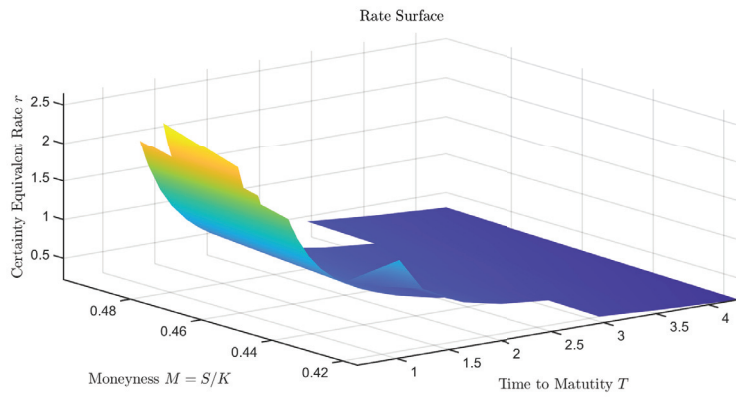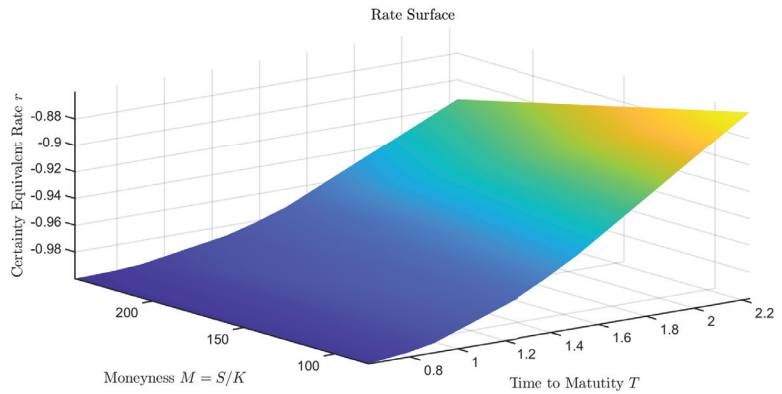**Figure 9.** Relationship between the time to maturity, moneyness, and the implied put-call parity certainty equivalent rate for Ford Motor Company stock.

5. Summary

In this paper, we have used put-call parity to derive the implied put-call parity certainty equivalent rate. We have also considered the meaning of positive rates and negative rates. Then, we utilize the idea[8] of the implied volatility surface to construct the implied put-call parity certainty equivalent rate surface. From the relative positions of these surfaces, we can determine which stock options we should consider investing in first.

However, the certainty equivalent rate is only one factor that should be considered in investing. It is obvious that investors cannot consider this factor alone. When deciding to invest in a product, investors should also consider other factors, such as political factors. To be more specific, at the end of May 2023, Tesla CEO Elon Musk's visit to China (See the details in Kharpal (2023)) caused Tesla's stock to soar and made him the world's richest man. Thus, investors who invested in Tesla before this event made a large amount of money. It is clear that we cannot predict such an outcome via a mathematical model. Hence, when deciding to invest in a product, investors should also consider other factors, so that they have a better chance of making optimal decisions.

Author Contributions: Conceptualization, S.R.; methodology, S.R.; software, Y.H.; formal analysis, Y.H. and S.R.; investigation, Y.H. and S.R.; resources, Y.H. and S.R.; data curation, Y.H.; writing—original draft preparation, Y.H.; writing—review and editing, Y.H. and S.R.; visualization, Y.H.; supervision, S.R.; project administration, S.R. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1   A detailed explanation of the certainty equivalent can be found in Kenton (2021).

2   This is because the time to maturity T in the put-call parity expression has units of years, and we assume that one calendar year has 252 business days in this paper.

3   The options' maturity dates for Apple, Nvidia, Meta, and Charles Schwab range from 9 June 2023 to 19 December 2025; the options' maturity dates for First Republic Bank range from 9 June 2023 to 19 July 2024; and the options' maturity dates for Signature Bank range from 16 June 2023 to 15 December 2023.

4   The strike prices for these six companies' options correspond to the maturity dates.

5   Via put-call parity, we can obtain different rates with respect to different time to maturity values, and thus we can determine the maximum value, minimum value, and mean value of these rates. The values in Table 2 are similar.

6   The current electric vehicle market situation is described in Sanguesa et al. (2021), and the reasons that electric vehicles can reduce pollution can be found in Thomas (2012).

7   The options' maturity dates for Tesla and Ford Motor Company range from 9 June 2023 to 19 December 2025; the options' maturity dates for General Motors range from 9 June 2023 to 20 June 2025. The options' strike prices correspond to the maturity dates.

8   The implied volatility surface is a three-dimensional surface that explores the relationship between the time to maturity T, moneyness S/K, and volatility σ.

References

Hull, John C. 2022. Put-call parity. In *Options, Futures and Other Derivatives*, 11th ed. New York: Pearson, p. 368.

Kenton, Will. 2021. What a Certainty Equivalent Is, What It Tells You, How to Use It. *Investopedia*, February 22. Available online: https://www.investopedia.com/terms/c/certaintyequivalent.asp (accessed on 22 February 2021).

Kharpal, Arjun. 2023. Elon Musk Wrapped Up His First Visit to China in Years. Here's What the Tesla CEO Was Up To. *CNBC*, June 1. Available online: https://www.cnbc.com/2023/06/01/elon-musk-china-visit-heres-what-the-tesla-ceo-was-up-to.html (accessed on 1 June 2023).

Khederian, Henry. 2023. Why Charles Schwab (SCHW) Stock Is Falling Sharply? *BENZINGA*, May 2. Available online: https://www.benzinga.com/news/23/05/32168058/why-charles-schwab-schw-stock-is-falling-sharply (accessed on 2 May 2023).

Sanguesa, Julio A., Vicente Torres-Sanz, Piedad Garrido, Francisco J. Martinez, and Johann M. Marquez-Barja. 2021. A review on electric vehicles: Technologies and challenges. *Smart Cities* 4: 372–404. [CrossRef]

Saul, Derek. 2023. These Are the Best (and Worst) Performing Stocks of 2023. *Forbes*. Available online: https://www.forbes.com/sites/dereksaul/2023/02/20/these-are-the-best-and-worst-performing-stocks-of-2023/?sh=39a59b3d85a6 (accessed on 20 February 2023).

Shinn, Lora, and Vikki Velasquez. 2023. What Happened to Signature Bank? *Investopedia*, May 1. Available online: https://www.investopedia.com/what-happened-to-signature-bank-7370710 (accessed on 1 May 2023).

Son, Hugh. 2023. JPMorgan Chase Takes over First Republic after Biggest U.S. Bank Failure Since 2008. *CNBC*, May 1. Available online: https://www.cnbc.com/2023/05/01/first-republic-bank-failure.html (accessed on 1 May 2023).

Thomas, C.E. Sandy. 2012. How green are electric vehicles? *International Journal of Hydrogen Energy* 37: 6053–62. [CrossRef]

Article

A New Entropic Measure for the Causality of the Financial Time Series

Peter B. Lerner [1,2]

1   SUNY-Brockport, Brockport, NY 14420, USA; pblerner18@gmail.com or pblerner@syr.edu
2   School of Business Administration, Anglo-American University, ul. Letnikow, 120 Prague, Czech Republic

Abstract: A new econometric methodology based on deep learning is proposed for determining the causality of the financial time series. This method is applied to the imbalances in daily transactions in individual stocks and also in exchange-traded funds (ETFs) with a nanosecond time stamp. Based on our method, we conclude that transaction imbalances of ETFs alone are more informative than transaction imbalances in the entire market despite the domination of single-issue stocks in imbalance messages.

Keywords: causality; market microstructure; market imbalance; TAQ-ARCA; C-GAN neural network

1. Introduction

A conventional method for determining the causality of the financial time series was developed by Clive Granger in the 1980s, who was awarded the Nobel Memorial Prize in 2003. The essence of the method is that a subset of an information set is excluded from analysis, and probability distributions are evaluated on a smaller information set (Diebold 2006; Diks and Panchenko 2006). The generation of probability distributions usually requires fitting the vector autoregressive model (VAR) to the time series and excluding some of the explanatory variables.

Nonparametric versions of the Granger causality tests were developed later, especially in the papers by Baek and Brock (1992), and extended by many authors (Diks and Panchenko 2006, and op. cit.). The Baek and Brock test and its variants are computationally very intensive for the large datasets prevailing in modern securities studies. They require state space coverage with cells of the size $\varepsilon > 0$ and computing correlations between the cells for the decreasing epsilon.

Since the 2010s, transactions in the stock market began to carry nanosecond time stamps. This change requires new methods of analysis adapted to the new realities.[1] The emergence of the big data framework and attempts to use deep learning methods created the following challenge: Regressions became nonlinear, and may contain hundreds of thousands of parameters in the case of this paper—and trillions in the case of Google datasets. Furthermore, deep learning algorithms usually present a "black box", and it is hard to attribute the input changes to the output differences.

The capacity of the human mind to analyze multidimensional time series consisting of billions of market events has remained largely unchanged. Because of our evolution in three-dimensional space, humans have the best grasp of two-dimensional information. Consequently, the methods of image analysis are among the best developed in the whole discipline of signal processing.

My paper adapts deep learning methods developed for image processing to the causality of the financial time series. Comparing two datasets, the one which requires more information to produce a deepfake using a neural network is considered more informative. A precise formulation of these criteria is provided in Section 4.

C-GANs (convolutional generational adversarial neural networks) appeared in 2015. The original purpose of the method was the image analysis and/or generation of deepfakes

(Goodfellow 2017). The essence of the C-GAN is that the network is divided into two parts: generator and discriminator, or critic (Rivas 2020). The generator net produces fake images from random noise and learns to improve them with respect to the training file. The discriminator tries to distinguish fake from real images based on training statistics.

To demonstrate this method's utility, I use it to analyze trading imbalances in New York Stock Exchange (NYSE) trading, which the Securities and Exchange Commission (SEC) requires to be stored with a nanosecond time stamp. These images, for different days, are standardized to ensure their comparability. The imbalance events constitute a situation when the counterparty does not instantly deliver the stock to close its position. The number of these events per day is several million. The time series are preprocessed into two-dimensional images of realistic size to be analyzed using a PC.

Why is this problem important? Given the instances of "flash crashes" in the market, the first and largest of those reported being the Flash Crash of 2010 on the NYSE, the question of whether exchange-traded funds stabilize or destabilize the market became increasingly important. In particular, the Flash Crash was attributed to the toxicity of liquidity of S&P minis orders (Easley et al. 2013). Because of the explosive growth in ETF markets (more information in Section 2), the traded volume in the compound portfolios representing ETF shares can easily exceed trading volume in the underlying securities. Intuitively, this can cause a problem in the delivery of the underlying, which can propagate through the system and, in rare cases, cause a crash. Alternatively, the Mini-Flash crash of 24 August 2015 demonstrated a significant deviation in market index price—subject to the circuit breaking several times—and weighted ETF prices (for a detailed description, see Moise 2023, especially Figure 1).
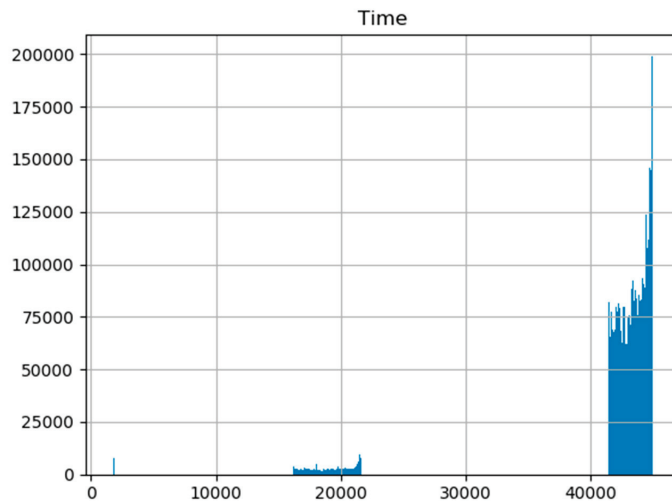
Figure 1. The rate of daily imbalance messages in the TAQ ARCA database. During 45,000 s of the day-to-day operation of the system, there were around 4 million messages, the maximum coming at or around 4:00 p.m. The maximum rate during a 100 s interval typically reached 200,000. Most events are concentrated at the beginning, noontime, and the end of the day.

In the current paper, I explore whether the market imbalance events drive ETF imbalances or vice versa. This problem is a good testbed for the proposed deep learning methodology.

The paper is structured as follows. In Section 2, I briefly outline commonly referred information on the ETF market. In Section 3, I describe the structure of the database. Section 4 describes preprocessing of SEC data into two-dimensional images. In Section 5, we establish the causality (information link) between ETF and market transaction imbalances.

Section 6 has an additional robustness check of the proposed results. Section 7 is the summary of the results. The Appendix A describes different kinds of imbalance messages according to NYSE. I review a possible theoretical justification for ETF informational dominance in Appendix B.

2. Formulation of the Problem and the Literature Review

The ETF market has exhibited explosive growth in recent years (NYSE 2021). Invented to securitize stock inventories of large asset managers, it developed into a major asset class in its own right (Gastineau 2010). Different anomalies in the markets, particularly the Flash Crash in May 2010, were partly attributed to the instability of exchange-traded products (ETPs), especially the S&P minis (Easley et al. 2013).

The question of whether the introduction of ETFs increase or decrease market quality has been discussed in many authoritative publications (Bhattacharya and O'Hara 2018, and op. cit.), Israeli et al. (2017) mentioned that there are two main opposing reasons for expectations of market quality change. On the positive side, exposure to ETFs provides more information on the underlying stocks, particularly stocks with lesser liquidity exposed to little coverage by analysts. The question of whether the use of ETFs increase or decrease the information efficiency of markets was also posed by Glosten et al. (2020). They suggested that the inclusion of ETFs increases information efficiency for low-liquidity stocks.

On the negative side, uninformed traders tend to exit the market in the underlying securities in favor of ETFs, thus depressing liquidity. Furthermore, much ETF activity happens at market close (Shum et al. 2016). Shum et al. noticed that ETFs typically have larger mispricing and wider spreads during end-of-trading, especially on the days of most volatile trading.

If the influence of ETFs is so prominent, can they be a catalyst for extreme events in the market? Some authors, e.g., A. Madhavan, have answered positively (Madhavan and Sobczyk 2019). At least, there is a recognition that new kinds of risks are inherent in the proliferation of ETFs (Pagano et al. 2019). If that is true, can big data analyses and deep learning instruments provide some warning about whether extreme events may be coming? And, what is the direction of the information flow—from ETF to the broader market or vice versa? (Glosten et al. 2020).

To answer this question, we develop a structured methodology, which allows us to determine with some certainty whether ETF illiquidity results from market fluctuations or if it is the other way around. The hypothetical mechanism is as follows: ETF trading initiates the delivery of ETF constituents ("in-kind" transaction) or a cash equivalent ("in-cash" transaction) if the underlying assets are easily tradable (Bhattacharya and O'Hara 2018). If the aggregate volume of ETF executions were small and/or evenly spread in time, this would introduce friction in orderly market execution.

And indeed, there are a few inherent problems. First, there needs to be more clarity as to whether trade imbalance results from the actual economic events in one or more underlying stocks or from the effects of stock aggregation by the ETFs, for instance, changes in the credit risk of the ETF swap counterparty.

Second, because ETF transactions are prominent in hedges, they are highly nonuniform throughout the day (Shum et al. 2016). A company that misses a delivery can wait until the end of the day to close the deal when the market is more stable. This paper does not judge whether ETFs are a "good" or "bad" influence on market liquidity. It strives to clarify the enormous influence ETFs have on the stock market, particularly the direction of information transmission.

According to the foundational models of market microstructure, the principal driver of price changes is the imbalances in supply and demand for a given security (Kyle 1985; Glosten and Milgrom 1985). Currently, imbalance messages can be followed with nanosecond precision. There is probably little value added to further increasing accuracy because signals only propagate a few meters—i.e., the size of the trading room—with already achievable latency (Bartlett and MacCrory 2019). One of the first studies going up to nanosecond granularity was "Price discovery in high resolution" (Hasbrouck 2021). This

wealth of available data creates many problems of its own. The human mind is poorly adapted to rationalize such amounts of data. Furthermore, our senses evolved in 3D space and have difficulty comprehending multidimensional datasets.

One of the principal channels of influence of the ETF market on overall market stability is using ETF shares for shorting and hedging. ETF shares "fail-to-deliver" on settlements for different reasons. Fail-to-deliver could be a signal of actual economic troubles in a company, shorting in expectation of a real price movement by the AP (authorized provider), the analog of the market makers (MM) for stock, or "operational shorting" (Evans et al. 2018). The description of operational shorting in the above-cited paper by Evans, Moussawi, Pagano, and Sedunov is so exhaustive that I provide a somewhat long quote.[2] Before a security is classified as "fail-to-deliver", an imbalance record is created. Usually, the imbalance is cleared by end-of-day trading or the next day before trading hours. The reputation penalty for being cited in imbalances is typically small (Evans et al. 2018).

The reason markets and the SEC do not regulate intraday deliveries with harsher penalties is obscure. We hypothesize that the inherent optionality involved in paying for order flow (PFOF) is partially responsible for this market feature (for PFOF analysis, see, e.g., (Lynch 2022)). If there were substantial fines or a negative reputation associated with the "failure-to-deliver", the PFOF mechanism would suffer disruptions. Consider that an expected negative return for a penalty would overcome the price improvement offered by the wholesaler. Because the non-delivery probability is nonzero, only relatively large price improvements would justify the counterparty risk, and a large volume of trade would miss the wholesaler. A detailed discussion of the issue is outside the scope of this paper.

This work is dedicated to researching methods to rationalize imbalance datasets with nanosecond time stamps. We compress them into two-dimensional "fingerprints", for which a rich array of algorithms developed for analyzing the visual images is already available. The dataset we use is the list of imbalance messages provided by NYSE Arca. "NYSE Arca is the world-leading ETF exchange in terms of volumes and listings. In November 2021, the exchange had a commanding 17.3% of the ETF market share in the US" (Hayes 2022). The special significance of the data for our problem setting is illustrated by the fact that a glitch in the NYSE Arca trading system influenced hundreds of ETF funds in March 2017 (Loder 2017).

Messages in our database have the following types: type "3", type "34", and type "105". Message type 3 is a symbol index mapping (reset) message. Message 34 is a security status message, which can indicate "opening delay", "trading halt", and "no open/no resume status". Finally, message type 105 is an imbalance message. More information about the format and content of the messages and the datasets can be found in Appendix A and (NYSE Technologies 2014).

To make use of the large statistics of the nanosecond time stamps of the 105 messages, we selected them for our analysis. Our choice is justified because the daily stream of 105 messages is in the millions, while 3 and 34 messages are in the tens of thousands.

The number of imbalance messages (type 105) for each trading day is around four million, each comprising 15–20 standardized fields. TAQ NYSE Arca equities—TAQ NYSE imbalance files provide "buy and sell imbalances sent at specified intervals during auctions throughout the trading day for all listed securities" (NYSE Technologies 2014).

3. Preprocessing—Formation of the State Variables Database

We selected the following variables: (1) the number of messages per unit time, and price, (2) the dollar imbalance at the exception message, and (3) the remaining imbalance at settlement. The latter is rarely different from zero because a failure to rectify stock imbalances at the close of a trading session indicates a significant failure in market discipline and may entail legal consequences.

Our data can be divided into two unequal datasets: market messages in their totality and ETF-related messages, and the first group encompasses the second. Because of the large volume of the data, we used an algorithmic selection of data for the ETF group. The

messages in the datasets contain the identifier "E" for exchange-traded funds, but in other places, "E" can indicate corporate bonds, but it is too common a letter to filter for it in a text file. Instead, we chose separate data on market participants provided by the NYSE, of which we filtered the names explicitly containing the words "exchange-traded" or "fund". This identification is not 100% accurate because some closed-end mutual funds, which are not ETFs, could have entered our list, but they are expected to be dominated by ETFs.

We were left with 1061 names automatically selected from the messages file. The number of daily events related to our list can be half a million or more, so sorting by hand would be difficult, if possible at all.

We further grouped our data as follows: First, the number of type 105 messages per 100 s. Second, the cumulative imbalance every 100 s in a 12 1/2 h trading day[3,4]. The number of price bins chosen was approximately equal to the number of time intervals. Dollar imbalances are calculated by the following formula:

$$\$Imb_t = p \cdot (Imb_t - Settle_{4:00}) \tag{1}$$

where p is the last market price, Imb_t is the undelivered number of shares, and $Settle_{4:00}$ is the number of shares unsettled by the end of the trading session, usually at 4:00 p.m.

The 100 s intervals were chosen arbitrarily but intended to have a two-dimensional data tensor processed on a laptop and have sufficiently acceptable statistics. The imbalances are distributed quite irregularly at around 45,000 s and can be visually grouped into the "beginning of the day settlement", "midday settlement", and "end-of-day settlement" (see Figure 1).

As expected, most of the dollar imbalances are small. To avoid data being swamped into a trivial distribution—a gigantic zeroth bin—and a uniformly small right tail, we used a logarithmic transformation for all variables, including time: $\tilde{x}_t = \ln(1 + x_t)$. Unity was added to deal with the zeroes in the database. Nonlinear transformation distorts probability distributions, but given their sharply concentrated and peaked shape, we did not expect it to influence the results too much.

The plot of the summary statistics for a typical day is shown in Figure 2. We observed that the maximum number of imbalance messages created by ETFs for each 100 s during a trading day is about one-eighth of the total number of exceptions (~25,000:200,000) in the market, but the cumulative value of imbalances created by the ETFs is about 60% of the total. This disparity suggests that average imbalances are much higher when ETF shares are involved.
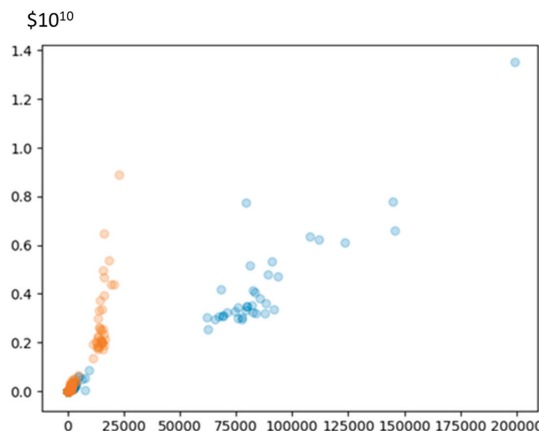
Figure 2. Cumulative dollar imbalances as a function of the number of imbalance messages per 100 s period. Orange dots are the transaction imbalances involving only ETF shares, and blue dots are the total market imbalance. We observe that the message rate of ETFs is approximately one-eighth of the total market, but their dollar value exceeds 60% of the cumulative market imbalances.

The final step was making our data amenable to deep learning algorithms, mostly designed to deal with visual data ("fake images"). We further compressed our nonuniform rectangular matrices—in some cases, messaging did not begin at exactly 3:30 a.m., etc. One data file included only NYSE into 96 × 96 squares, which we call "fingerprints" of daily trading.[5] (Figure 3). The fingerprints do not have an obvious interpretation; rather, they take the form of a machine-readable image, like a histogram. The transaction rate was plotted on the vertical scale. We plotted an accompanying dollar value for an imbalance on the horizontal scale. This compression method allows for treating daily imbalances as uniform images, which can be subjected to processing using deep learning algorithms.

Five randomly selected trading days (7–8 October 2019, 9 September 2020, and 4–5 October 2020) produced ten daily samples: one with total market imbalance messages, the other with ETF data only. We constructed five testing and five training samples from them according to the protocol exhibited in Figure 4.
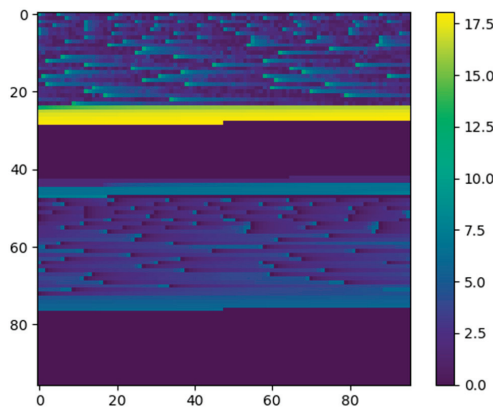
Figure 3. Example of the "fingerprint" of a trading day on a natural logarithmic scale. The 105-type message rate and the cumulative dollar amount of imbalances are placed into 96 bins.
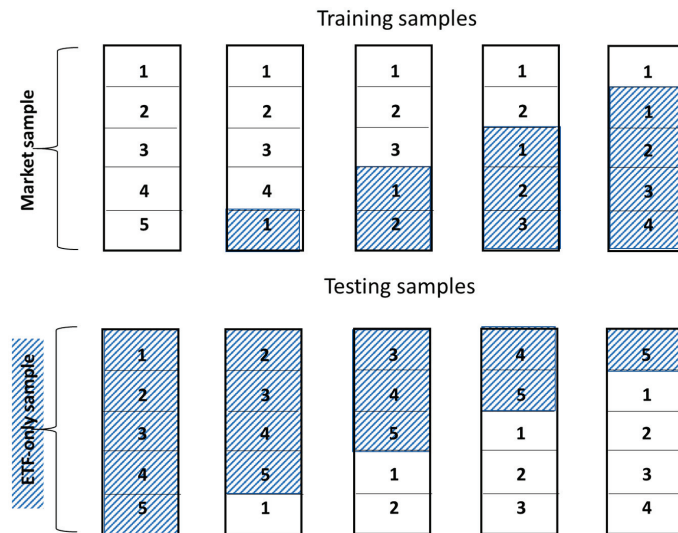
Figure 4. Composition of the training and testing samples. Each number indicates a fingerprint of a given day in chronological order (7 October 2019, 8 October 2019, 9 September 2020, 4 October 2021, and 5 October 2021). Note that only the first samples in corresponding columns are mirror images of each other. Blue diagonal pattern indicates ETF-only samples.

A relative proportion of all-market and ETF-only samples, according to Figure 3, is provided in Table 1.

Table 1. The proportion of the all-market and ETF-only samples in each simulation.

Training File	Market:ETF	Test File	Market:ETF
tr1	100%:0%	tes1	0%:100%
tr2	80%:20%	tes2	20%:80%
tr3	60%:40%	tes3	40%:60%
tr4	40%:60%	tes4	60%:40%
tr5	20%:80%	tes5	80%:20%

4. Distances on a State Space

Because of their "black box" nature, the output of neural networks is hard to rationalize. First, the human mind has evolved to analyze two- or three-dimensional images in three-dimensional space. Most humans cannot directly comprehend tensor inputs, intermediate results, and outputs typical for neural networks. Second, the results of neural network analyses are necessarily stochastic and depend on the large number of estimated intrinsic parameters, which are frequently inaccessible, but in any case, too numerous to rationalize. Third, deep learning results can depend on how the training and testing samples are organized, even if they represent identical datasets. All of this can indicate the failure of a deep learning procedure (Brownlee 2021), but it can also show additional information we fail to recognize.[6] Because neural networks are "black boxes", instead of the interpretation of a hundred thousand—in my case trillions of—parameters in the case of Google and Microsoft deep learning networks, one has to design numerical experiments and analyze the output from a deep learning algorithm in its entirety.

To systematize the results, we propose two measures of divergence of images as follows: After the C-GAN generated fake images ("fingerprints") of the session, we considered these images as (1) matrices and (2) nonnormalized probability distributions.

The first approach is to treat arrays as matrices (tensors). We computed the pseudo-metric cosine between the image arrays X and Y according to the following formula:

$$C_{XY} = \frac{\|X + Y\|^2 - \|X - Y\|^2}{4\|X\|\cdot\|Y\|} \tag{2}$$

In the above formula, the norm $\|\cdot\|$ is a Frobenius matrix norm representing each image array. In the first stage, we computed the distance as the average of each twentieth of the last 400 images in the sequence. Because, sometimes, the fake image is an empty list having a zero norm, we modified this formula according to the following prescription:

$$C_{train,fake} = \frac{\|train + fake\|^2 - \|train - fake\|^2}{4\|train\|\cdot\|test\|} \tag{3}$$

$$C_{test,fake} = \frac{\|test + fake\|^2 - \|test - fake\|^2}{4\|train\|\cdot\|test\|}$$

Equation (3) provides answers close to the correct geometric Formula (2), but it does not fail in the case of an empty fake image. The pseudo-metric measure, calculated according to Equation (3), provides a fair picture of the affinity of the fake visual images to the originals (see Figure 5), but it is still unstable with respect to different stochastic realizations of the simulated images.

(**A**)

(**B**)
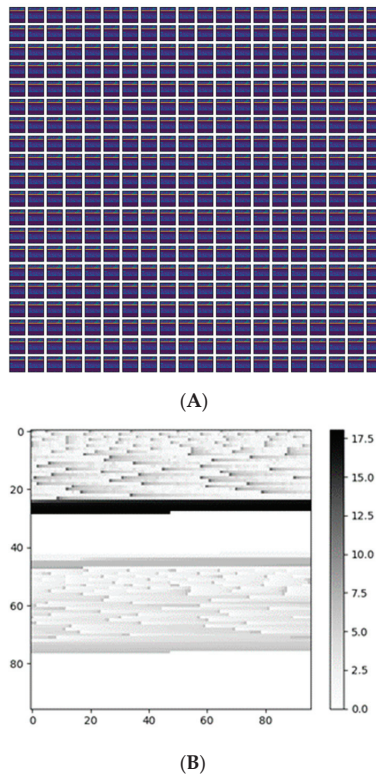
Figure 5. The output of C-GAN. (**A**) Set of 400 fake images generated by the generator part of the deep learning network during 1600 epochs. (**B**) An individual fake image (compare Figure 3).

So, we applied a second stage averaging according to the formula for the mutual information:

$$MInfo = \frac{1}{N} \sum_{i=1}^{N} log_2 \left(\frac{C_{train,fake,i}}{C_{test,fake,i}} \right) \tag{4}$$

In Equation (4), N is the number of independent network runs. Note that this formula does not depend on whether we use a "geometrically correct" Equation (2) or a computationally convenient Equation (3). Unlike separate $C_{train,fake}$, and $C_{test,fake}$ norms, which may vary widely between consecutive runs of the C-GAN, their ratio is reasonably stable for a given training and testing sample. Furthermore, if one exchanges training and test samples, the argument of the summation of Equation (4) only changes its sign.

The second option is to treat output arrays as quasiprobability distributions. We used Kullback–Leibler (KL) divergences $D_{KL}(P||Q)$ between the two distributions (Burnham and Anderson 2002). As is well known, Kullback–Leibler divergence is not a real metric and is asymmetrical with respect to its arguments. The intuitive meaning of $D_{KL}(P||Q)$ is an information gain achieved when one replaces distribution Q (usually meaning sample) with distribution P (usually meaning model). In our context, distribution P is a training or test dataset, and distribution(s) Q are fakes generated using the neural net. A final information criterion is:

$$r = \frac{D_{KL}\left(X_{test}||X_{fake}\right)}{D_{KL}\left(X_{train}||X_{fake}\right)} \tag{5}$$

The values $r > 1$ suggest that the test file is difficult to reproduce by training the net. The values $r < 1$ indicate that reproducing the dataset—the neural network does not know anything about trading, which generates observable data—is relatively easy.

There are also several variations for the design of the samples we analyzed. In the first preliminary design, we compared data of the entire market, which includes ETFs, against the data files, which contain only our sample(s) of ETFs. In that case, we expected a positive sign of mutual information when we trained the network using the entire market file—because the information from the ETFs is already contained in a training file. We display these results in the next section.

In the second design, we separated data files into single-issue and ETF stocks. The positive sign of the mutual information must appear if the training occurs on the more informative information subset of the two. We identified a more informative data file as causally determining a less informative data file.

Despite the complexity of the described procedure, the intuition behind it is quite simple. A convolutional generative adversarial network generates many fake images inspired by a training sample. These images are compared with a test sample. If the fake images perfectly fit both the training and test samples, the mutual information between them is exactly zero. Vice versa, very divergent training and test distributions suggest much additional information, which must be known to reproduce a dataset.

Positive mutual information or a higher correlation between a fake distribution and training one than between a fake and test distribution means that a training file is easier to fake than the test. On the contrary, small or negative mutual information suggests that the C-GAN's operation produces fakes, which are relatively easily distinguishable from the training file.

5. Preliminary Results of C-GAN Analysis

The results of the fingerprint analysis using the C-GAN are displayed in Tables 2 and 3. In that analysis, we used our vocabulary of 1060 funds automatically selected by words in their name. For the robustness of this choice, we also tested the list of 1753 US ETF funds in the database: https://www.etf.com/channels/us-etfs (accessed on 20 May 2021).[7] We display five runs of the C-GAN network with the construction of samples according to the first line from Table 1. The comparison with the first cell of Table 3 suggests that with this second sample of ETFs, mutual information is only strengthened in the direction of "training by the overall market" and "testing by the ETF-only file".

Each cell in Table 2 is the binary logarithm of the ratio of the distances of the generator-devised fingerprint between training and testing samples, respectively. Table 3 shows that the mutual information generally decreases with a diminishing fraction of market samples and increasing ETF samples. The fraction of the market vs. ETF samples in the testing samples demonstrates no visible tendency, with one exception. When the testing file is almost entirely composed of the market samples, mutual information becomes zero irrespective of the training samples (Tables 2 and 3).[8]

The column and row averages are provided in Table 3. Testing sample 2 is an outlier. We tentatively attribute it to one of the day's data being exceptional. And indeed, the file for 9 September 2020 contained, probably, only stocks listed on NYSE, not all traded stocks. All averages are positive. This suggests that all-market files are easier approximated by the generator-produced fakes than the ETF-only files. We consider this evidence that ETF-only files have more distinguishing features than the all-market files and, consequently, are more distant from fakes than the training files, i.e., if one used ETF data as a training set, it was relatively easy to train the network to have a relatively high correlation of fakes with all-market samples. On the contrary, the all-market training set was insufficient to train the network to distinguish fakes from the original ETF data.

In a further robustness check, we tested another sample of 1753 ETFs selected using an online database in the same setup. The results are similar to our ad-hoc tests of the 1060 computer-selected funds (Table 4).

Table 2. The results of measuring MInfo (Equations (2) and (3)) between the 20 fake images created using the generator and the training and test images. C-GAN was run for 1600 epochs, and the fake images were taken uniformly from the last 400 images.

	tr1		tr2		tr3		tr4		tr5	
tes1	0.7855	0.6538	0.1511	0.7556	−0.4958	−0.0203	0.2263	−0.1734	−0.6959	0.8837
tes2	0.5147	-	1.0126	-	−0.2086	-	0.5735	-	−1.4379	-
tes3	2.5042	2.6258	2.2150	2.5771	2.2438	2.0920	2.5458	2.3841	2.0923	2.2976
tes4	0.3850	0.3785	0.4250	0.7784	0.7484	0.0000	0.1193	0.9425	−0.0687	−0.4286
tes5	0.0000	-	0.0000	-	0.0000	-	0.0000	-	0.0000	-

Table 3. Mutual information (Equation (3)) for the training and test samples.

Training File	Test File Average	Test File	Training File Average
tr1	0.9809	tes1	0.2071
tr2	0.9894	tes2	0.0909
tr3	0.5449	tes3	2.3578
tr4	0.8273	tes4	0.3280
tr5	0.3303	tes5	0

Table 4. Mutual information for the selection of 1753 ETFs from etf.com. The arrow shows the direction from the training to the test file in the C-GAN network. SU denotes "NYSE Stock Universe" data, and "ETF" denotes ETF-only data. The arrow indicates the direction from training to test files.

Runs	MI SU→ETF	MI ETF→SU
1	0.8605	0.2059
2	1.0595	−1.0925
3	0.8900	0.2510
4	0.6670	−1.0926
5	1.0403	0.1261
Average	0.9035	−0.3204
Std. dev.	0.1589	0.7063

6. The Test of Single-Issue Stocks against ETF Data

We conducted a test wherein we excluded the data for 1753 ETFs from the market data. The results of the test are displayed in Table 5.

We observed that one can teach the network by feeding it with ETF-only data. Our network successfully interpolated single-issue stock data by the "learned fakes" but not vice versa. We tentatively make the case that ETF trading provides more information for traders in single-issue stocks, and henceforth, the direction of causality is from ETFs to single issues. A possible economic explanation for this phenomenon is provided in Appendix B.

The treatment of the neural network outputs as probability distributions allows for using another measure on the state space, namely Kullback–Leibler distance (see Section 4). The results of computing the parameter r from Equation (5) are shown in Table 6.[9] The layout of this table reflects the difficulty of expressing relations between tensors in a human-readable format.

In each of the cells, except one—the upper second from the left—a lower asymmetry between the proportions of ETF and single-issue samples in test and training files corresponds to a higher value of r. Average values of r in the cells are the largest for a low (25% and 50%) fraction of the ETF samples in the test file. Henceforth, a GAN network can falsify the single-issue data more successfully.

Table 5. Mutual information for the selection of 1753 ETFs from etf.com. The arrow shows the direction from the training to the test file in the C-GAN network. SI denotes "Single-issue stock" data, and "ETF" denotes ETF-only data. The arrow indicates the direction from training to test files. We observed that the ETF-trained network successfully teaches SI-enabled data, while the obverse is impossible.

Runs	MI ETF→SI	MI SI→ETF
1	0.6857	−0.0495
2	0.9169	0.2002
3	0.9599	−0.0520
4	0.9822	−0.0670
5	1.0789	0.0675
Average	0.9247	0.0199
Std. dev.	0.1462	0.1143

Table 6. The ratio of Kullback–Leibler distances r (Equation (5)) as a function of the share of ETF samples in test data (1) and the difference between the proportion of ETF and single-issue samples in the test and training files, respectively (2). Averages in each group are indicated by (3). In each cell except one, lower asymmetry entails a higher r index. Color coding is added for visibility.

Averages (Rows)	ETF Fraction	ETF Fraction Difference between Training and Test Sets								ETF Fraction	Averages (Rows)
		100%	50%	0%	−50%	100%	50%	0%	−50%		
2.252	75%		4.251		0.802		9.865		3.15	25%	4.455
	75%	2.665		2.023		1.983		2.823		25%	
1.892	100%		2.330		1.440		7.787		41.549	50%	13.142
	100%	2.031		1.331		1.927		1.305		50%	
Averages (columns)		2.819		1.399		5.391		12.207			Averages (columns)

7. Conclusions

In this paper, I presented a new econometric methodology to investigate the causality of the financial time series. In variance to original Granger causality, this methodology does not rely on any explicit model of the stochastic process by which the input data were generated. It is preferable to nonparametric Granger causality techniques in the case of extra-large or multidimensional datasets because it does not rely on the computation of correlations between multiple subsets of the original data.

The proposed method was applied to solve an important question: whether individual stocks or ETFs drive the liquidity of markets. I chose the information content of the number of imbalances to measure liquidity. The latter indicates the inability to instantly fill the quote at a given price and the dollar value of incomplete transactions. The information content was measured as a pseudodistance between the time series in a two-dimensional state space (the number of a price bucket and its dollar imbalance).

The preliminary answer is that both the rate of imbalance arrivals and the dollar value of resulting imbalances of the ETFs are more informative—in the sense of finer features nonreproducible by fakes—than the individual stocks with ETFs counted as separate stocks. Higher information content of ETF imbalances is not surprising. Indeed, the imbalance messages produced by 1000+ ETFs constitute about one-eighth of the totality of exceptions in the entire database on average, but the dollar value of their imbalances is about two-thirds of the entire dollar value of the market imbalance. Theoretically, this is not surprising because, as pointed out by (Shum et al. 2016) and (Evans et al. 2018), ETF securities are used for hedging much more frequently than individual stocks. An explanation of this phenomenon based on extant economic theory is provided in Appendix B.

Appendix A

Description of the TAQ ARCA messages

2.11 SYMBOL INDEX MAPPING MESSAGE (MSG TYPE '3')

FIELD NAME	FIELD ORDER	FORMAT	DESCRIPTION
MsgType	1	Numeric	This field identifies the type of message. '3' – Sequence Number Reset message
SequenceNumber	2	Numeric	Message sequence number by channel. Must derive sequence number based on leading sequence number for each packet.
Symbol	3	NYSE Symbology	This field displays the symbol in NYSE symbology
SymbolIndex	4	Numeric	This field identifies the numerical representation of the NYSE symbol. This field is unique for products within each respective market and cannot be used to cross reference a security between markets.
Market ID	5	Numeric	ID of the Originating Market: ■ '1' - NYSE Cash

TAQ NYSE Arca Integrated Feed /V1.8   20

2.15 SECURITY STATUS MESSAGE (MSG TYPE '34')

FIELD NAME	FIELD ORDER	FORMAT	DESCRIPTION
MsgType	1	Numeric	This field identifies the type of message. '34' – Security Status Message
SequenceNumber	2	Numeric	Message sequence number by channel. Must derive sequence number based on leading sequence number for each packet.
SourceTime	3	HH:MM:SS.nnnnnn	This field specifies the time when the msg was generated in the order book. The number represents the number of seconds at microsecond accuracy in UTC time (since EPOCH)
Symbol	4	NYSE Symbology	This field displays the symbol in NYSE symbology
SymbolSeqNum	5	Numeric	This field contains the symbol sequence number
Security Status	6	ASCII	The following are Halt Status Codes: ■ '3' - Opening Delay

TAQ NYSE Arca Integrated Feed /V1.8   25

2.9 IMBALANCE MESSAGE (MSG TYPE '105')

FIELD NAME	FIELD ORDER	FORMAT	DESCRIPTION
Msg Type	1	Numeric	This field identifies the type of message.

TAQ NYSE Arca Integrated Feed /V1.8

FIELD NAME	FIELD ORDER	FORMAT	DESCRIPTION
			'105' – Imbalance Message
SequenceNumber	2	Numeric	Message sequence number by channel. Must derive sequence number based on leading sequence number for each packet.
SourceTime	3	HH:MM:SS.nnnnnn	This field specifies the time when the msg was generated in the order book. The number represents the number of seconds at microsecond accuracy in UTC time (since EPOCH)
Symbol	4	NYSE Symbology	This field displays the symbol in NYSE symbology
SymbolSeqNum	5	Numeric	This field contains the symbol sequence number
ReferencePrice	6	Numeric	The Reference Price is the Last Sale if the last sale is at or between the current best quote. Otherwise the Reference Price is the Bid Price if last sale is lower than Bid price, or the Offer price if last sale is higher than Offer price.
PairedQty	7	Numeric	This field contains the paired off quantity at the reference price point
TotalImbalanceQty	8	Numeric	This field contains the total imbalance quantity at the reference price point
MarketImbalanceQty	9	Numeric	This field indicates the total market order imbalance at the reference price
AuctionTime	10	hh:mm	Projected Auction Time
AuctionType	11	Alpha	■ 'O' – Open (4am) Arca Only ■ 'M' – Market (9:30am) ■ 'H' - Halt ■ 'C' – Closing ■ 'R' – Regulatory Imbalance **Note:** For the NYSE/MKT, the opening imbalance will have an "M" Auction Type

Message type '105'

FIELD NAME	FIELD ORDER	FORMAT	DESCRIPTION
ImbalanceSide	12	Alpha	This field indicates the side of the imbalance Buy/sell. Valid Values: ■ 'B' – Buy ■ 'S' – Sell ■ ' ' – No imbalance Note: This field is a future enhancement for NYSE Arca and will have a '0' value until such time.
ContinuousBook ClearingPrice	13	Numeric	The Continuous Book Clearing Price is defined as the price closest to last sale where imbalance is zero. If a Book Clearing Price is not reached, the Clearing Price, a zero will be published in the Book Clearing Price Field Note: This field is a future enhancement for NYSE Arca and will have a '0' value until such time.
ClosingOnly ClearingPrice	14	Numeric	This field contains the indicative price against closing only order only Note: This field is a future enhancement for NYSE Arca and will have a '0' value until such time.
SSRFilingPrice	15	Numeric	This field contains the SSR Filing Price. This price is the price at which Sell Short interest will be filed in the matching in the event a Sell Short Restriction is in effect for the security. Note: The SSR Filing price is based on the National Best Bid at 9:30am. This price remains static after the SSR Filing price has been determined. Note: This field is a future enhancement for NYSE Arca and will have a '0' value until such time.

The format of TAQ ARCA messages with a nanosecond
time stamp

34,10,00:24:58.796044288,CBO,1,P,~,,,,,,~,P

3,11,BANC,1,51,N,C,100,11.47,0,0,N,.0001,1

105,273294,09:29:34.061214976,UHT,33,66.21,170,1141,0,
0930,M,B,67.67,0,0,0,0,0,0,0,0,0,0, ,

Appendix B

The problem of ETFs as potential market drivers was explored by Semyon Melamed (Melamed 2015), who provided a dynamic equilibrium theory of the interaction of ETFs and the general market. In particular, he proved that if approved participants (AP, analog of market makers) act as arbitrageurs, correlations in a broad market increase. Moreover, stocks included in physical ETFs influence correlations more than stocks in synthetic-only ETFs. This effect can be amplified by less liquid securities in some of the physical ETF portfolios (Marta and Riva 2022). These securities are inevitably present in the ETF tied to MSCI World, Russell 2000, and other popular indexes. Then, ETF participants can emulate not the entire portfolio but a representative subset (see also (Koont et al. 2022)). However, Melamed's theory is difficult to use for practical estimation because of many unobservable parameters. A more parsimonious model was proposed by Pan and Zeng (2017). Technically, their treatment calls a risky illiquid asset a "bond", but for intraday trading, it makes no difference with stock. In particular, we can use Theorem 6.1 to elucidate the information content of the action of APs (Koont et al. 2022; Melamed 2015; Pan and Zeng 2017) in ETFs. One of the principal results of Pan and Zeng was Equation (A2) for the optimal number of shares issued z^*. If we assume a high correlation between ETF and illiquid securities lying in the foundation of its portfolio, it is convenient to make a substitution for the correlation factor $\rho \to 1 - \epsilon \leq 1$. This assumption is certainly true when applied to the entire ETF universe. Then, we can rewrite Equation (A2) uisng a slightly different notation:

$$z^* = \frac{(\lambda + \theta\sigma^2)(\pi + c_L) + \lambda\epsilon(\pi - c_E)}{\lambda(\lambda + 2\mu) + 2\epsilon\sigma^2(\lambda + \mu)} \tag{A1}$$

Here, in Equation (1), λ is the Kyle–Amihud illiquidity factor, μ is the flotation cost of the ETF share, θ is the risk avoidance factor, σ is the volatility, $c_L = c_B + \lambda x_{B-} - c_E \geq 0$ is the difference between price movement of the illiquid (c_B) and liquid asset (c_E) for the customer order x_{B-}.

The current study quantitatively measures the enhanced informational content of the ETF imbalances by Equation (A1). For instance, for the pair tes1–tr1 (Table 2), this suggests that per dollar imbalance basis, the ETF transaction is $2^{0.7855} = 1.724$ times more informative than an action of the entire market, ETFs included.

The measure of the informativeness of AP trades is hardto design. Yet, if we assume that it is some version of entropy, i.e., a convex functional on the optimal number of newly issued shares, then the first-order condition will be

$$\frac{\delta S(z)}{\delta z}\Big|_{z=z^*}\delta z^* = 0$$

An approximate formula for the FOC becomes

$$\frac{\pi - c_E}{\pi + c_L} = \frac{2\theta\sigma^2(\lambda + \mu + \theta\sigma^2)}{\lambda(\lambda + 2\mu)} \tag{A2}$$

The price mismatch near the equilibrium has an order of $\frac{2\theta\sigma^2}{\lambda}$, i.e., the volatility is greatly amplified if the risk avoidance factor is nonzero and the Kyle–Amihud constant is

small.[10] Given the same level of risk aversion, higher liquidity indicates higher information content. All of the above supports the conclusion of higher information brought upon by the AP orders.

Notes

1. Hasbrouck, https://www.youtube.com/watch?v=EZCgW1mFRP8 2010 (accessed on 20 May 2021).
2. "One important objective of APs in the primary ETF market is to harvest the difference between ETF market price and its NAV . . . As demand for the ETF grows from investors in the secondary market, the ETF's market price should increase [Increasing the possibility of market arbitrage—P. L.] However . . . selling ETF shares and buying the underlying basket/creating the ETF shares are not necessarily instantaneous. The AP sells the new ETF shares to satisfy bullish order imbalances but can opt to delay the physical share creation until a future date. By selling ETF shares that have not yet been created, the AP incurs a short position for operational reasons . . . that we hereafter call an "operational short" position." The paper (Evans et al. 2018) also lists "directional shorting", i.e., speculation on the changing market price as a reason for "fail-to-deliver.".
3. Our messages begin at 3:30 a.m. and end at 4 p.m., 45,000 s in total, usually, but not always, 449,100 s intervals starting with zero. Each time interval contains ~9000 messages on average. Yet, the highest message rate in each price bin can be more than twenty times as high.
4. The price bin methodology is reminiscent of the VPIN measure of Easley et al. (2013). We experimented with linear as well as the logarithmic scale of our data. In this paper, we use a logarithmic scale.
5. Each fingerprint contains 9216 pixels. We compress our ~400 MB daily database into ~200 K text file, a compression of 2000 times.
6. The output from C-GAN indicates a deep learning failure called "mode collapse" (Brownlee 2021). Yet, the look-alike of the fake images remains excellent.
7. We used the data from only four days in our sampling because, for unknown reasons, one of the TAQ ARCA files has no overlap with the second ETF list.
8. We must issue a caution that most applications of the GAN networks suffer from overfitting and mode drop (Yazici et al. 2020). Visual inspection of the losses by the critic and the generator suggests that it can take place but, currently, we can do nothing about it. Variation in the number of epochs, discrimination tolerance and other standard remedies do not change the qualitative picture.
9. Data in Table 6 results from a single run of the C-GAN network.
10. The small Kyle-Amihud constant is a reasonable assumption given the relatively high liquidity of the NYSE-traded shares.

References

Baek, Ehung, and William Brock. 1992. *A General Test for Nonlinear Granger Causality*. Working Paper. Ames: Iowa State University.

Bartlett, Robert P., III, and Justin M. MacCrory. 2019. How rigged are stock markets? Evidence from microsecond timestamps. *Journal of Financial Markets* 45: 37–60. [CrossRef]

Bhattacharya, Ayan, and Maureen O'Hara. 2018. Can ETFs Increase Market Fragility? Available online: https://www1.villanova.edu/content/dam/villanova/VSB/department/finance/fall17seminar/Bhattacharya-Can-ETFs-Increase-Market-Fragility.pdf (accessed on 5 May 2021).

Brownlee, Jason. 2021. *Machine Learning Mastery with Python: Understand Your Data, Create Accurate Models and Work Projects End-to-End*. San Francisco: Machine Learning Mastery.

Burnham, Kenneth P., and David R. Anderson. 2002. *Model Selection and Multi-Model Inference*. Berlin/Heidelberg: Springer.

Diebold, Francis X. 2006. *Elements of Forecasting*. Cincinnati: Cengage.

Diks, Cees, and Valentyn Panchenko. 2006. A new statistic and practical guidelines for nonparametric Granger causality testing. *Journal of Economic Dynamics and Control* 30: 1647–49. [CrossRef]

Easley, David, Marcos M. López de Prado, and Maureen O'Hara. 2013. The exchange of flow toxicity. *The Journal of Trading* 6: 8–13. [CrossRef]

Evans, Richard B., Rabih Moussawi, Michael S. Pagano, and John Sedunov. 2018. ETF Short Interest and Failures-to-Deliver. Available online: https://jacobslevycenter.wharton.upenn.edu/wp-content/uploads/2018/08/ETF-Short-Interest-and-Failures-to-Deliver.pdf (accessed on 5 May 2021).

Gastineau, Gary L. 2010. *The Exchange-Traded Funds Manual*. New York: Wiley.

Glosten, Lawrence R., and Paul R. Milgrom. 1985. Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics* 14: 71–100. [CrossRef]

Glosten, Lawrence, Suresh Nallareddy, and Yuan Zou. 2020. ETF Activity and Informational Efficiency of Underlying Securities. *Management Science* 67: 22–47. [CrossRef]

Goodfellow, Ian. 2017. Available online: https://arxiv.org/pdf/1701.00160.pdf (accessed on 5 May 2021).

Hasbrouck, Joel. 2021. Price discovery in high resolution. *Journal of Financial Econometrics* 19: 395–430. [CrossRef]

Hayes, Adam. 2022. Available online: https://www.investopedia.com/terms/n/nyse-arca.asp (accessed on 5 May 2021).

Israeli, Doron, Charles M. C. Lee, and Suhas A. Sridharan. 2017. Is There a Dark Side to Exchange Traded Funds? An Information Perspective. *Revue of Accounting Studies* 22: 1048–83. [CrossRef]

Koont, Naz, Yiming Ma, Lubos Pastor, and Yao Zeng. 2022. *Steering a Ship in Illiquid Waters: Active Management of Passive Funds*. ESRB Working Paper Series; Florham Park: Jacobs Levy Management Center for Quantitative Financial Research.

Kyle, Albert S. 1985. Continuous auctions and insider trading. *Econometrica* 53: 1315–35. [CrossRef]

Loder, Asjylyn. 2017. NYSE Arca suffers glitch during Closing auction. *Wall Street Journal*.

Lynch, Bradford. 2022. Price Improvement and Payment for Order Flow: Evidence from A Randomized Controlled Trial. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4189658 (accessed on 2 July 2023).

Madhavan, Ananth, and Aleksander Sobczyk. 2019. Does trading by ETF and mutual fund investors hurt performance? *Journal of Investment Management* 7: 4–20.

Marta, Thomas, and Fabrice Riva. 2022. *Do ETFs Increase the Comovements of Their Underlying Assets? Evidence from a Switch in ETF Replication Technique*. Dauphine Research Paper. Paris: Universite Paris.

Melamed, Semyon. 2015. *A Dynamic Equilibrium Model of ETFs*. Genève: Swiss Finance Institute Research.

Moise, Claudia E. 2023. High-Frequency Arbitrage and Market Illiquidity. Available online: https://ssrn.com/abstract=3768926 (accessed on 2 July 2023).

NYSE. 2021. Available online: https://www.nyse.com/etf/exchange-traded-funds-quarterly-report (accessed on 1 May 2021).

NYSE Technologies. 2014. TAQ NYSE ARCA Integrated Feed Client Specification. Available online: https://www.nyse.com/publicdocs/nyse/data/TAQ_NYSE_Arca_Integrated_Feed_Client_Spec.pdf (accessed on 1 May 2021).

Pagano, Marco, Antonio Sánchez Serrano, and Josef Zechner. 2019. *Can ETFs Contribute to Systemic Risk?* Reports of the Advisory Scientific Committee. Frankfurt am Main: ERSB.

Pan, Kevin, and Yao Zeng. 2017. *ETF Arbitrage under Liquiidty Mismatch*. ESRB Working Paper Series; Frankfurt am Main: ERSB.

Rivas, Pablo. 2020. *Deep Learning for Beginner*. Birmingham: Packt Publishing.

Shum, Pauline, Walid Hejazi, Edgar Haryanto, and Arthur Rodier. 2016. Intraday share price volatility and leverage ETF rebalancing. *Review of Finance* 2: 2379–409. [CrossRef]

Yazici, Yasin, Chuan-Sheng Foo, Stefan Winkler, Kim-Hui Yap, and Vijay Chandrasekhar. 2020. Empirical Analysis of Overfitting and Mode Drop in Gan Training. Paper presented at IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, October 25–28; pp. 1651–55.

Article

Multicriteria Portfolio Choice and Downside Risk

Anna Rutkowska-Ziarko [1,*] and Pawel Kliber [2]

1   Faculty of Economics, University of Warmia and Mazury in Olsztyn, 10-719 Olsztyn, Poland
2   Institute of Informatics and Quantitative Economics, Poznan University of Economics and Business, 61-875 Poznań, Poland; p.kliber@ue.poznan.pl
*   Correspondence: aniarek@uwm.edu.pl

Abstract: In this study, we investigated some extensions of the classical portfolio theory and try to evaluate them in a situation of crisis. We studied some additional criteria for portfolio selection, based on market multiples representing the overall situation of companies. Additionally, we investigated semi-variance as an alternative measure of risk. We developed a range of portfolios that were built using different criteria for risk and the fundamental values of companies from the Polish stock market. Then, we compared their returns during the crisis that occurred after the outbreak of the COVID-19 pandemic. The results of empirical research on the major companies traded on the Warsaw Stock Exchange reveal that investors can achieve better investment results by augmenting the standard Markowitz model with an additional criterion connected with the fundamental standing of companies, such as book-to-market or earnings-to-market ratios. The second result is that using nonclassical risk measures such as semi-variance instead of variance provides better results, and this method of measuring risk is especially essential in periods characterized by the collapse of the capital market.

Keywords: portfolio analysis; fundamental value; multicriterial choice; market multiple; downside risk

1. Introduction

The classical methods for selecting an investment portfolio, developed by Markowitz (1952, 1959) and Sharpe (1963), only take into account the market performance of companies, measured by assessing changes in their prices. In the classical model, the potential portfolios of investment are evaluated according to two criteria: expected return (which describes the potential level of profitability from an investment) and risk. The first criterion is measured using the expected rate of return, and the second one using the variance or standard deviation of returns. No other criteria are considered that might give some additional information about the financial standing and prospects of a company that could influence the prices of its shares.

In recent years, however, there has been growing interest in portfolio analysis methods with alternative ways of constructing portfolios. An article by Kolm et al. (2014) contains a review of major developments in portfolio theory since its origin, and a book by Doumpos and Zopounidis (2014) draws attention to the multicriteria methods used in this field. Most innovations depend on using criteria of risk other than variance or the standard deviation of returns, for example, semi-variance or conditional value at risk. An article by Fabozzi et al. (2007) presents a variety of risk measures that are currently used in the practice of portfolio investments. In other approaches, some characteristics of the distribution of returns on assets are used as additional criteria for evaluating portfolio performance. Examples of such characteristics include skewness or kurtosis. Expanded portfolio analysis is presented by Briec et al. (2007) and Rodríguez et al. (2011).

There are several studies that include criteria that are not based on returns on assets. A branch of the literature takes into account ethical, social, or environmental criteria in

portfolio construction, for example, the so-called socially responsible investments approach described in Steuer et al. (2007). Articles by Ballestero et al. (2012), Bilbao-Terol et al. (2013), and Burchi and Włodarczyk (2020) are a few more examples illustrating this approach.

Lo et al. (2003) considered the liquidity of stocks as an additional criterion in the portfolio construction process. There are only a few papers that also take into consideration the fundamental values of companies. Xidonas et al. (2010) considered the sum of dividends paid by companies. Jacobs and Levy (2013) took into account the risk associated with leverage. The utility function of an investor includes the costs of margin calls, which can force borrowers to liquidate securities at adverse prices due to their illiquidity; losses exceeding the capital invested; and the possibility of bankruptcy.

In accordance with the theoretical concept and empirical research (Fama and French 1992, 2015, 2017; Lam 2002; Zaremba and Czapkiewicz 2017), fundamental factors are important in shaping returns on capital markets. Therefore, it seems rational to include them in a stock portfolio model. The current extensive research on the main financial markets of Eastern Europe has corroborated the significant impact of fundamental information concerning companies on their rates of return. This was also found to hold true for the book-to-market ratio as an indicator of the financial standing of companies. The research sample included five countries: the Czech Republic, Hungary, Poland, Russia, and Turkey (Zaremba and Czapkiewicz 2017).

There have been several attempts to combine portfolio analysis with the fundamental analysis of companies from the Polish stock markets. Tarczyński (2002) developed a synthetic measure to evaluate the economic and financial standing of a company, which he called the taxonomic measure of attractiveness of investment (TMAI), and applied this measure as an additional criterion in the evaluation of possible portfolios. The portfolio constructed using TMAI was called a fundamental portfolio. This model has been modified in recent years, for example, by substituting variance with semi-variance as a risk measure (Rutkowska-Ziarko and Garsztka 2014). In Rutkowska-Ziarko (2013), the Mahalanobis distance was used to determine the TMAI due to the possible correlations between diagnostic financial variables. Another method was proposed by Pośpiech (2019) in their research on financial ratios, and market indicators were applied to guide the initial selection of companies.

In this work, in addition to the classic measure of risk (variance), we also used semi-variance. The use of the downside risk measure in choosing an investment portfolio seems to be particularly useful in times of strong declines in the financial markets, such as those in February and March 2020. These were caused by a decline in investor optimism caused by the development of the COVID-19 pandemic. In the calculation of semi-variance, one takes into account only negative deviations below a certain level. Upward deviations, which are connected with higher returns than expected, are not taken into account in determining this measure. Another advantage of semi-variance is that there is no need to make any assumptions about the distribution of rates of return and investors' utility functions (Harlow and Rao 1989). The quadratic utility function has some undesirable properties and therefore misrepresents the actual behavior of investors. First, it reaches a maximum for a certain rate of return, and then, its value decreases with an increase in returns, which is in direct contradiction to the preferences of investors, who always prefer to have more than less. Building an effective portfolio for semi-variance is more complicated than the approach in which variance is used as a risk measure. It is impossible to use standard solver software to find a minimum semi-variance portfolio. In the calculation of semi-covariances, one has to know in which periods the rate of return of the entire portfolio was lower than the target value, and this depends on the composition of the portfolio.

This article is organized as follows: After this introduction, in Section 2, we present a brief description of commonly used market multiples and give some reasons why they could be used as an additional criterion for portfolio choice. In Section 3, downside risk measures are described. Section 4 contains a mathematical formulation of portfolio optimization problems and presents the algorithms that were used to solve them. Section 5

presents the results of empirical research concerning the Polish stock market during the crisis of the COVID-19 pandemic.

2. Market Multiples

Market multiples provide an indication of how the market values a publicly traded company. To calculate the values of these multiples, market data and financial results of a company are used. Breen (1968) and Basu (1977) analyzed the effect of market multiples on the future profitability of companies. They found that portfolios of companies with lower P/E multiples had higher annual returns in the following year than portfolios formed from companies with higher P/E multiples. The article of Basu (1977) is frequently cited as the first publication in which the impact of market multiples on the future profitability of the companies is analyzed. However, similar research was carried out even earlier, for example, by William Breen (1968). He examined companies from index S&P500 indexed over the period from 1953 to 1966, using the COMPUSTAT database. This is a source of fundamental and market information on active and inactive companies and covers around 99% of the world's total market capitalization. For certain years, equally weighted portfolios (of 10 and 50 companies) were constructed using the stocks of companies with the lowest and highest P/E multiples. The results indicate that portfolios built with stocks of companies with lower P/E multiples had higher annual returns in the following year than portfolios built with stocks of companies with higher P/E multiples.

Barbee et al. (2008) investigated the impact of market multiples values on future prices of stocks. They analyzed the profitability of equally weighted portfolios built using the shares of companies with various values of different market multiples.

The most popular indicator of the market's valuation of a company is the P/E multiple, which relates the earnings per one ordinary share to its market price:

$$P/E = \frac{market\ price\ per\ share}{profit\ per\ share},$$

In this study, four different measures of the ability of a company to generate profit were considered: net profit (EAT); gross profit (GP); earnings before interest, taxes, depreciation, and amortization (EBITDA); and operating profit (EBIT).

Net profit is the last position in the Profit and Loss Account, it is calculated as follows:

$$EAT = net\ sales\ - cost\ of\ goods\ sold\ - operating\ expense\ -\ taxes\ -\ interest.$$

Gross profit is earnings before taxation. EBITDA is earnings before interest, taxes, depreciation, and amortization. EBITDA can be used to describe a company's financial performance without taking into account its capital structure. The operating profit is an accounting measure that measures the profits that a company generates from its operating activities. Interest and taxes are not considered here.

One can relate a share price not only to different profit categories but also to other characteristics that describe the economic situation of a company. It may be important for an investor to relate the market valuation of the company's share capital to the net value of its assets. The P/BV multiple relates the price of an ordinary share to the book value of the company, estimated per ordinary share. This multiple shows the market value of the company in relation to its book value.

$$P/BV = \frac{market\ value\ per\ share}{book\ value\ per\ share}$$

where

$$book\ value\ per\ share = \frac{assets\ - liabilities}{number\ of\ shares}.$$

The book value refers to the total amount the company would be worth if it liquidated its assets and paid back all its liabilities, and it is also the net asset value of the company.

A positive relationship between the book-to-market ratio and average returns was described by Rosenberg et al. (1985). This phenomenon was also observed in Japanese stocks (Chan et al. 1991). Based on this research, Fama and French (1992) suggested that the book-to-market ratio would be an important risk factor explaining the variability of stock rate of returns.

In this paper, instead of using classical market multiples, we used their reciprocals, i.e., the values of financial indicators (expressed per one share) divided by the current market price of a share. The reason is the additivity of such indicators. The value of these indicators for the portfolio as a whole is a weighted average of the values of the indicators of individual companies. Thus, P/BV is replaced by the book-to-market ratio (BV/M):

$$BV/M = \frac{book\ value\ of\ the\ company}{market\ value\ of\ the\ company} = \frac{book\ value\ per\ share}{price\ of\ the\ share}.$$

The same was used for the P/E multiple, in fact, for the whole group of market multiples based on various methods for calculating the company's profits. Instead of the price-to-equity ratio, the earnings-to-price ratio (E/M) was used, calculated using the equation below:

$$E/M = \frac{profit\ per\ share}{market\ price\ of\ a\ share}.$$

3. Downside Risk in Portfolio Choice

In portfolio theory, variance has been a commonly used measure of risk in capital market analysis from its inception to the present day (Markowitz 1952). At the same time, there have been doubts about the validity of using this risk measure for almost as long (Markowitz 1959). The main disadvantage of variance as a measure of risk is that it treats negative and positive deviations from a mean return in the same way. In fact, negative deviations are undesirable, and positive deviations create an opportunity for greater profit. To measure only negative deviations, Markowitz (1959) proposed semi-variance, which is an average of deviations below a certain level. Semi-variance and lower moments consider only the variability on the left side of a distribution. The reference point can be the mean, as in the case of variance, but another value can also be used as the reference point. Using semi-variance as a measure of risk is consistent with investors' intuitive perception of risk (Boasson et al. 2011).

Variance is assumed to be an appropriate risk measure when the distribution of returns is normal, or at least symmetric, or when an investor has a quadratic utility function. The classical Markowitz model (Markowitz 1952) is ineffective in selecting portfolios that comprise assets with skewed returns. The traditional mean–variance model, which treats deviations above and below the target return equally, tends to overestimate risk and imposes unnecessary conditions that exclude portfolios that are downside efficient.

Pla-Santamaria and Bravo (2013) constructed portfolios of blue-chip stocks from the Dow Jones Industrial Average. Their results show significant differences between the portfolios obtained by mean–semi-variance efficient frontier model and those with the same expected returns obtained using the classical Markowitz mean–variance efficient frontier model.

An investor who does not wish the return of their portfolio to fail below the target rate of return would tend to compose portfolios that minimize downside risk measures (Klebaner et al. 2017).

It is believed that, in symmetrical distributions, variance as a risk measure is no worse than semi-variance (Estrada and Serra 2005; Galagedera and Brooks 2007). However, research on capital markets shows that the distributions of rates of return of many companies are not normal or at least symmetrical (Adcock and Shutes 2005; Estrada and Serra 2005; Markowski 2001; Post and van Vliet 2006; Sun and Yan 2003). Then, the use of lower-risk measures becomes important. In the case of right-skewed distributions of returns, the main part of the variance includes upper deviations, which mean the achievement of high

returns. The impact of lower deviations is relatively small. For this reason, investors are looking for companies with right-skewed distributions of returns (Galagedera and Brooks 2007; Peiro 1999), which suggests that the issue of skewness cannot be ignored in the risk analysis, even if the distribution of the returns of some companies are symmetrical. Also, according to the perspective theory (Kahneman and Tversky 1979), it is more appropriate to use semi-variance instead of variance as a risk measure.

The above arguments speak in favor of lower-risk measures when compared with their classic counterparts. Lower-risk measures, such as a semi-variance, allow for a universal approach to risk analysis and equity portfolio construction, regardless of the empirical distribution of returns. One also does not have to assume a specific analytical form of the utility function. It is sufficient to make the obvious assumption that an investor prefers to earn more than less, and therefore higher rates of return are better than lower rates of return.

A semi-variance, defined by Markowitz (1959), is a lower counterpart of a variance. This lower-risk measure is the sum of the squared of lower deviations from the target rate of return γ. It is calculated using the following formula:

$$dS^2(\gamma) = \frac{\sum_{t=1}^{m} d_t^2(\gamma)}{m-1}, \ t = (1, 2, \ldots, m),$$

where

$$d_t(\gamma) = \begin{cases} 0 & for \ z_t \geq \gamma \\ z_t - \gamma & for \ z_t < \gamma \end{cases}$$

z_t—Rate of return of company i in period t;
$dS^2(\gamma)$—Semi-variance for company i;
m—The number of time periods;
γ—The mean rate of return or any target rate of return chosen by an investor.

Extensions of semi-variance as a risk measure are lower partial moments, introduced by Bawa (1975) and Fishburn (1977). According to these authors, the lower partial moment of order n is given by

$$LPM_i^n = \frac{1}{m-1} \sum_{t=1}^{m} lpm_{it}^n,$$

where

$$lpm_{it} = \begin{cases} 0 & for \ z_t \geq \gamma \\ z_t - \gamma & for \ z_t < \gamma \end{cases}.$$

Notice that for $n = 2$, the lower partial moment is equal to semi-variance.

The semi-variance of an investment portfolio $dS_P^2(\gamma)$ is given by

$$dS_P^2(\gamma) = \sum_{i=1}^{k} \sum_{j=1}^{k} x_i x_j d_{ij}(\gamma)$$

where x_i is the share of stock i in the portfolio, and $d_{ij}(\gamma)$ is the semi-covariance of the rate of return for the i-th and the j-th shares, which is defined by

$$d_{ij}(\gamma) = \frac{1}{m-1} \sum_{t=1}^{m} d_{ijt}(\gamma), \tag{1}$$

where

$$d_{ijt}(\gamma) = \begin{cases} 0 & for \ z_{pt} \geq \gamma \\ (z_{it} - \gamma)(z_{jt} - \gamma) & for \ z_{pt} < \gamma \end{cases}$$

$$z_{pt} = \sum_{i=1}^{k} x_i z_{it}, \ t = 1, 2, \ldots, m.$$

4. Problems Related to Portfolio Choice

We consider a portfolio of k different assets. Let μ_i be a mean return of an asset i, estimated from the last m observations of

$$\mu_i = \frac{\sum_{t=1}^{m} z_{it}}{m}.$$

where σ_{ij} denotes a covariance between the returns of asset i and asset j:

$$\sigma_{ij} = \frac{1}{m-1} \sum_{t=1}^{m} (z_{it} - \mu_i)(z_{jt} - \mu_j).$$

where x_i denotes the proportion of the wealth invested in asset i. The mean return of the portfolio is then given using the following formula:

$$\mu_P = \sum_{i=1}^{k} x_i \mu_i$$

and the variance of the portfolio's rate of return is given by

$$S_P^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} x_i x_j \sigma_{ij}.$$

The semi-variance of the portfolio form given by

$$dS_P^2(\gamma) = \sum_{i=1}^{k} \sum_{j=1}^{k} x_i x_j d_{ij}(\gamma),$$

where semi-covariances of assets' returns are given in (1).

We assume that some market multiples are also considered. It can be connected with the book-to-market value per share or with the earnings-to-price ratio of a share. Let β_i denote the value of this criterion for the asset i. The value of this multiple for the whole portfolio is given by

$$\beta_P = \sum_{i=1}^{k} x_i \beta_i.$$

In the empirical part of the work, we consider portfolios that are the solutions to the following optimization problems:

A portfolio minimizing the variance, i.e., a portfolio that is the solution of

$$\min_{x_1, \ldots, x_k} S_P^2 = \sum_{i=1}^{k} \sum_{j=1}^{k} x_i x_j \sigma_{ij} \tag{2}$$

with the constraint that

$$\sum_{i=1}^{k} x_i = 1, \tag{3}$$

$$x_1, x_2, \ldots x_k \geq 0. \tag{4}$$

A portfolio that minimizes variance with a constraint on the mean return: In this case, we assume that the mean return of the portfolio should be no smaller than the required rate of return μ_0. The optimization problem is given in (2)–(4) with an additional constraint.

$$\mu_P = \sum_{i=1}^{k} x_i \mu_i \geq \mu_0 \tag{5}$$

A portfolio that minimizes variance with constraints on the mean return and its fundamental value: We assume that the fundamental value of the portfolio (measured with one of the market multiples) should not be lower than its required value β_0. In the empirical part of this study, we assume that the minimal value of the portfolio multiple

should be equal to the average of the multiples for all the companies considered. This yields problems (2)–(5) with an additional constraint.

$$\beta_P = \sum_{i=1}^{k} x_i \beta_i \geq \beta_0. \tag{6}$$

Mathematically, variance minimization problems are problems of quadratic optimization problems with linear constraints, which are determined using equations and inequalities. They can be solved using standard algorithms. We solved them using the method of Goldfarb and Idnani (1983) implemented in the R package quadprog.

The second set of portfolios are those that were optimized with the use of semi-variance as a measure of risk. The optimization problems were defined as follows:

A portfolio minimizing the unconditional semi-variance, i.e., a portfolio that is the solution of

$$\min_{x_1,\dots,\,x_k} dS_P^2(\gamma) = \sum_{i=1}^{k} \sum_{j=1}^{k} x_i x_j d_{ij}(\gamma), \tag{7}$$

with the constraints determined using (3) and (4).

A portfolio minimizing semi-variance with a constraint on mean return, which should be no lower then μ_0: the optimization problem is given with the set of conditions in (7) and (3)–(5).

A portfolio that minimizes semi-variance with a constraint on mean return and its fundamental value: the optimization problem is given with the set of conditions (7) and (3)–(6).

To solve these problems, the following numerical algorithm was used: We started with an initial portfolio (in this case, it was a portfolio minimizing variance). Then, we solved each of the problems as a problem of quadratic programming, using the Goldfarb and Idnani (1983) method. After each iteration, we re-estimated semi-covariances $d_{ij}(\gamma)$ and solved the problems with the new input data. We repeated this process until convergence, i.e., until changes in the portfolio structure between each iteration were sufficiently small. In the calculations, we used procedures written in R and the R package quadprog.

5. Data and Empirical Results

The studies covered 20 of the largest and most liquid companies listed on the Warsaw Stock Exchange, excluding financial companies. Close share prices from the period 1 April 2016–4 September 2020 were taken for analysis. In the estimation, we assessed the portfolios using their monthly rate of returns. The parameters (mean returns, variances, and semi-variances) used in constructing the portfolios investigated in this research were estimated using a period of three years before the start of an investment. Figures from financial statements were used, namely the net profit (EAT); the gross profit (GP); earnings before interest, taxes, depreciation, and amortization (EBITDA); the operating profit (EBIT); and the book value (BV). They changed with the publication of the quarterly financial statements of the companies. We calculated the appropriate indicators for each company according to its financial statement. For the calculation of financial indicators, we always used the latest available data, according to the date of publication. We considered the financial statements for the period from Q3 2019 to Q2 2020. Information on financial results is usually published with a delay of 60 to 120 calendar days. It was assumed that portfolios purchased on a given day were sold after a month (four weeks). The data were taken from the Thomson Reuters database—Refinitiv Eikon.

In economics, it is not possible to carry out repetitive experiments, as in, for example, physics or chemistry. Thus, in this article, the COVID-19 pandemic was used as a natural experiment. During the pandemic, there was a sharp collapse in stock exchanges, which allows us to test the usefulness of various risk diversification methods in times of sharp drops in prices in financial markets.

In this paper, we considered 15 types of portfolios. Table 1 lists the descriptions of these types and the symbols used to refer to them.

Table 1. The list of the types of constructed portfolios.

Portfolio	Description
Equally weighted	A portfolio with an equal share of each asset
MinV	A portfolio minimizing variance (unconditionally)
MinV-E	A portfolio minimizing variance with a constraint on mean return
MinV-E-EBIT	A portfolio minimizing variance with a constraint on mean return and EBIT/M
MinV-E-GP	A portfolio minimizing variance with a constraint on mean return and GP/M
MinV-E-EBITA	A portfolio minimizing variance with a constraint on mean return and EBITDA/M
MinV-E-EAT	A portfolio minimizing variance with a constraint on mean return and EAT/M
MinV-E-B	A portfolio minimizing variance with a constraint on mean return and B/M
MinSV	A portfolio minimizing semi-variance (unconditionally)
MinSV-E	A portfolio minimizing semi-variance with a constraint on mean return
MinSV-E-EBIT	A portfolio minimizing semi-variance with a constraint on mean return and EBIT/M
MinSV-E-GP	A portfolio minimizing semi-variance with a constraint on mean return and GP/M
MinSV-E-EBITA	A portfolio minimizing semi-variance with a constraint on mean return and EBITDA/M
MinSV-E-EAT	A portfolio minimizing semi-variance with a constraint on mean return and EAT/M
MinSV-E-BV	A portfolio minimizing semi-variance with a constraint on mean return and BV/M

Altogether, we developed 2655 portfolios during 177 trading days. We assumed that the investment period is one month. However, to assess the performance of the strategies, we calculated portfolios for each trading day. Thus, the first analyzed set of portfolios was created on 21 November 2019, and its performance was calculated based on one-month returns (i.e., price changes until 19 December 2019). The next set of portfolios was created on the next trading day (22 November 2019), and the performance was evaluated based on price changes until 20 December 2019, etc.

During the research period, four research subperiods of different lengths were specified. The division criterion was the changes in the situation of the capital market, which was reflected in the changes in the values of the WIG Index, the main index on the Warsaw Stock Exchange. The key aspect for identifying subperiods was the situation of the market during buying and selling a portfolio. Table 2 outlines the descriptions of the subperiods.

Table 2. Research periods.

The Time of Buying a Portfolio	The Situation of the Capital Market
21 November 2019–29 January 2020	Buying and selling before the collapse of the market
30 January 2020–10 March 2020	Selling during the collapse of the market
11 March 2020–13 May 2020	Buying and selling during the growth of the market
14 May 2020–07 August 2020	Buying and selling during the stabilization of the market

Due to the very large number of the considered portfolios (2655), in this study, we omitted the factors related to the structure of these portfolios and other elements of ex ante analysis, such as the expected portfolio risk, the average rate of return, or the average market ratio. All the characteristics of the distribution of return presented in Tables 3–7 refer to realized returns. That is, they describe the actual returns of investors, as well as the risk they bear. For the individual subperiods and the entire research period, the following characteristics were calculated: the mean rate of return, the minimal rate of return, the value-at-risk (VaR) measure, semi-deviation, standard deviation, and skewness.

Table 3. Summary statistics of the realized rates of return for the portfolios bought during the 21 November 2019–29 January 2020—I period (44 trade days)—before the collapse of the market.

Portfolios	Mean	Min	VaR 0.1	VaR 0.05	Semi-Dev.	Std. Dev.	Skewness
Equally weighted	0.034	−0.018	0.006	−0.011	0.003	0.021	−0.335
MinV	0.044	−0.018	0.011	−0.005	0.072	0.027	−0.117
MinV-E	0.034	−0.013	−0.003	−0.011	0.064	0.029	0.307
MinV-E-EBIT	0.035	−0.015	−0.003	−0.011	0.062	0.030	0.341
MinV-E-GP	0.040	−0.009	0.002	−0.007	0.062	0.029	0.292
MinV-E-EBITA	0.041	−0.013	0.004	−0.002	0.061	0.029	0.237
MinV-E-EAT	0.046	−0.009	0.010	0.007	0.060	0.029	0.221
MinV-E-BV	0.044	−0.015	0.004	0.001	0.061	0.029	−0.019
MinSV	0.033	−0.037	−0.004	−0.025	0.067	0.034	0.037
MinSV-E	0.040	−0.006	0.004	−0.004	0.060	0.028	0.383
MinSV-E-EBIT	0.041	−0.006	0.003	−0.003	0.058	0.029	0.393
MinSV-E-GP	0.043	−0.003	0.006	−0.001	0.058	0.028	0.372
MinSV-E-EBITA	0.045	−0.006	0.010	0.006	0.056	0.029	0.348
MinSV-E-EAT	0.051	−0.005	0.016	0.013	0.056	0.030	0.295
MinSV-E-BV	0.050	−0.009	0.011	0.009	0.056	0.029	0.144

In the first research subperiod, the negative effects of the COVID-19 pandemic had not yet affected the Polish stock market. We observed that all the portfolios minimizing the semi-variance had right-skewed distributions of rates. Returns of equally weighted portfolios and portfolios minimizing the variance were left-skewed. The highest average rate of return occurred for portfolios minimizing semi-variance and with a fundamental criterion. However, it is difficult to unequivocally determine which type of diversification was the most effective in reducing the risk during this period.

Table 4. Summary statistics of the realized rates of return for the portfolios bought during the 30 January 2020–10 March 2020—II period (29 trade days)—during the collapse of the market.

Portfolios	Mean	Min	VaR 0.1	VaR 0.05	Semi-Dev.	Std. Dev.	Skewness
Equally weighted	−0.169	−0.321	−0.287	−0.313	0.194	0.090	0.145
MinV	−0.158	−0.296	−0.265	−0.292	0.180	0.082	−0.008
MinV-E	−0.132	−0.298	−0.247	−0.277	0.160	0.089	0.142
MinV-E-EBIT	−0.127	−0.293	−0.241	−0.270	0.155	0.088	0.180
MinV-E-GP	−0.127	−0.292	−0.240	−0.269	0.155	0.087	0.141
MinV-E-EBITA	−0.121	−0.294	−0.239	−0.269	0.152	0.092	0.192
MinV-E-EAT	−0.121	−0.292	−0.236	−0.267	0.150	0.087	0.090
MinV-E-BV	−0.119	−0.296	−0.240	−0.269	0.151	0.094	0.197
MinSV	−0.146	−0.284	−0.250	−0.275	0.167	0.078	0.039
MinSV-E	−0.120	−0.289	−0.235	−0.267	0.150	0.089	0.133
MinSV-E-EBIT	−0.115	−0.282	−0.228	−0.259	0.145	0.087	0.156
MinSV-E-GP	−0.115	−0.281	−0.227	−0.259	0.145	0.086	0.137
MinSV-E-EBITA	−0.109	−0.283	−0.227	−0.257	0.141	0.089	0.134
MinSV-E-EAT	−0.108	−0.282	−0.225	−0.256	0.139	0.087	0.053
MinSV-E-BV	−0.106	−0.280	−0.223	−0.253	0.138	0.089	0.113

Table 5. Summary statistics of the realized rates of return for the portfolios bought during the 11 March 2020–13 May 2020—III period (43 trade days)—during the growth after the collapse of the market.

Portfolios	Mean	Min	VaR 0.1	VaR 0.05	Semi-Dev.	Std. Dev.	Skewness
Equally weighted	0.123	0.017	0.021	0.018	0.005	0.090	0.370
MinV	0.077	−0.020	−0.007	−0.013	0.007	0.073	0.202
MinV-E	0.104	0.000	0.013	0.003	0.006	0.082	0.301
MinV-E-EBIT	0.104	0.000	0.013	0.003	0.006	0.082	0.301
MinV-E-GP	0.104	0.000	0.013	0.003	0.006	0.082	0.301
MinV-E-EBITA	0.110	0.000	0.012	0.003	0.006	0.086	0.358
MinV-E-EAT	0.104	0.000	0.013	0.003	0.006	0.082	0.301
MinV-E-BV	0.122	0.006	0.024	0.013	0.008	0.085	0.340
MinSV	0.117	0.014	0.032	0.022	0.007	0.068	0.227
MinSV-E	0.124	0.014	0.032	0.022	0.007	0.071	0.170
MinSV-E-EBIT	0.124	0.014	0.032	0.022	0.007	0.071	0.170
MinSV-E-GP	0.124	0.014	0.032	0.022	0.007	0.071	0.170
MinSV-E-EBITA	0.116	0.014	0.028	0.021	0.008	0.072	0.395
MinSV-E-EAT	0.124	0.014	0.032	0.022	0.007	0.071	0.170
MinSV-E-BV	0.120	0.018	0.036	0.030	0.012	0.069	0.472

At the end of January 2020, the financial markets collapsed, and this situation lasted until mid-March. It affected not only Poland but practically all the most important world exchanges. During this period, all average rates of return were negative. The least effective strategy at this time was to select an equally weighted portfolio. It was the least profitable and the riskiest one, taking into account, among others, VaR 0.1, VaR 0.05, and semi-deviation. The most secure and at the same time most profitable portfolios were fundamental portfolios that minimized semi-variance. The distributions of the realized returns were right-skewed (with only one exception for the portfolio minimizing variance), but the strength of this asymmetry was small.

In the third subperiod, the quotations of the WIG Index slowly began to rise. During this period, purchasing an equally weighted portfolio proved to be a fairly effective method of risk diversification. The highest average rates of return, as in the second subperiod, were achieved using fundamental portfolios that minimized semi-variance. It is worth noting that the portfolios minimizing the semi-variance had higher values of this risk measure than the portfolios minimizing the variance. At the same time, portfolios minimizing the variance had higher realized variances. However, taking into account extreme values such as minimal return, VaR 0.05, and VaR 0.01, there is a clear advantage of portfolios with minimized semi-variance. In the third subperiod, all portfolios were characterized by right-hand asymmetry, which was stronger than in the second subperiod.

In the fourth subperiod, the Warsaw Stock Exchange stabilized, and price increases were small, as it is shown in Figure 1. The realized rates of return for equally weighted portfolios and those with minimum semi-variance were generally characterized by left-hand asymmetry, and those with minimum variance were characterized by right-hand asymmetry. During this period, for many fundamental portfolios, the condition imposed on a given market ratio for the portfolio was not active, especially for market ratios based on various measures of a company's profitability.

Over this period, the most profitable portfolios were those selected according to the minimum variance criterion. The average rate of return for the equally weighted portfolios was quite high, and the risk was lower than in Markowitz portfolios. In the fourth subperiod, advanced models of building stock portfolios had similar usefulness for stock exchange investors as the simple method of selecting an equally weighted portfolio.

In the entire research period (Table 7), portfolios with three criteria (average return, risk, and a fundamental criterion) allow for achieving higher realized returns than equally weighted portfolios, portfolios minimizing risks (measured either with the variance or semi-variance), or average return–risk portfolios. Additionally, it can be seen that the

introduction of a fundamental criterion reduced the risk borne by an investor, measured with both standard deviation and semi-deviation. An analysis of extreme values (minimum return, VaR 0.1, and VaR 0.05) also shows the advantage of fundamental portfolios. As can be seen, the whole group of portfolios minimizing semi-variance was characterized by a lower ex post risk than those minimizing variance. Fundamental portfolios had less left asymmetry, especially for portfolios that were designed to minimalize semi-variance. It should be emphasized that having less left asymmetry is beneficial for investors, as it means they are less exposed to very low rates of returns.

Table 6. Summary statistics of the realized rates of return for the portfolios bought during the 14 May 2020—7 August 2020—IV period (61 trade days)—during the stabilization of the market.

Portfolios	Mean	Min	VaR 0.1	VaR 0.05	Semi-Dev.	Std. Dev.	Skewness
Equally weighted	0.044	−0.028	0.011	−0.002	0.004	0.025	−0.457
MinV	0.059	−0.039	0.008	−0.013	0.006	0.042	0.044
MinV-E	0.040	−0.036	0.013	−0.007	0.005	0.024	−0.419
MinV-E-EBIT	0.040	−0.036	0.013	−0.007	0.005	0.024	−0.419
MinV-E-GP	0.040	−0.036	0.013	−0.007	0.005	0.024	−0.419
MinV-E-EBITA	0.040	−0.036	0.013	−0.007	0.005	0.024	−0.434
MinV-E-EAT	0.040	−0.036	0.013	−0.007	0.005	0.024	−0.419
MinV-E-BV	0.028	−0.041	0.000	−0.022	0.007	0.026	0.326
MinSV	0.018	−0.023	−0.011	−0.020	0.006	0.022	0.178
MinSV-E	0.018	−0.023	−0.011	−0.020	0.006	0.022	0.178
MinSV-E-EBIT	0.018	−0.023	−0.011	−0.020	0.006	0.022	0.178
MinSV-E-GP	0.018	−0.023	−0.011	−0.020	0.006	0.022	0.178
MinSV-E-EBITA	0.017	−0.025	−0.012	−0.020	0.006	0.022	0.161
MinSV-E-EAT	0.018	−0.023	−0.011	−0.020	0.006	0.022	0.178
MinSV-E-BV	0.016	−0.035	−0.020	−0.026	0.010	0.026	−0.025

Table 7. Summary statistics of the realized rates of return for the portfolios bought during 21 November 2019–7 August 2020—the whole research period (177 trade days).

Portfolios	Mean	Min	VaR 0.1	VaR 0.05	Semi-Dev.	Std. Dev.	Skewness
Equally weighted	0.026	−0.321	−0.142	−0.227	0.077	0.111	−0.814
MinV	0.024	−0.296	−0.139	−0.224	0.072	0.099	−1.289
MinV-E	0.026	−0.298	−0.100	−0.210	0.064	0.095	−0.814
MinV-E-EBIT	0.027	−0.293	−0.094	−0.200	0.062	0.093	−0.777
MinV-E-GP	0.028	−0.292	−0.098	−0.200	0.062	0.093	−0.802
MinV-E-EBITA	0.031	−0.294	−0.085	−0.195	0.061	0.094	−0.681
MinV-E-EAT	0.031	−0.292	−0.089	−0.195	0.060	0.092	−0.808
MinV-E-BV	0.031	−0.296	−0.083	−0.196	0.061	0.097	−0.501
MinSV	0.019	−0.284	−0.110	−0.202	0.067	0.097	−0.702
MinSV-E	0.026	−0.289	−0.082	−0.193	0.060	0.093	−0.562
MinSV-E-EBIT	0.028	−0.282	−0.080	−0.185	0.058	0.092	−0.509
MinSV-E-GP	0.028	−0.281	−0.080	−0.184	0.058	0.092	−0.516
MinSV-E-EBITA	0.028	−0.283	−0.071	−0.179	0.056	0.089	−0.544
MinSV-E-EAT	0.031	−0.282	−0.073	−0.176	0.056	0.091	−0.523
MinSV-E-BV	0.029	−0.280	−0.069	−0.177	0.056	0.090	−0.539
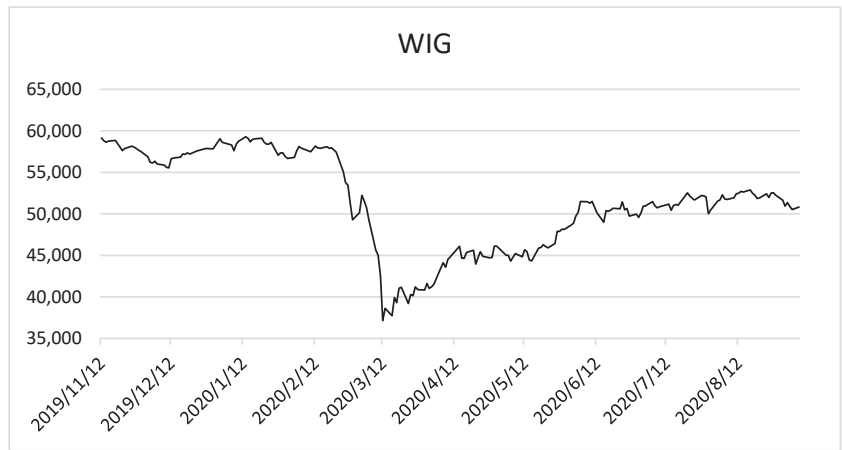
Figure 1. WIG closing prices.

In order to test whether there were differences in returns between the different portfolio selection methods, we performed appropriate statistical tests. Since the realized returns were not normally distributed (which we evaluated using the Shapiro–Wilk test and the Jarque–Bera test) we used the nonparametric Kruskal–Wallis test with post hoc Dunn's test to determine the differences between each pair of the portfolios. To assess the differences, we used the realized returns for the different types of portfolios from the entire research period. The test statistics for the Kruskal–Wallis test was 21.64, which indicates that the hypothesis of equal mean returns should be rejected, with a *p*-value lower than 0.1 (*p*-value = 0.086). This result shows that there were differences in the distributions between different groups of the realized returns. To analyze these differences, we carried out a post hoc analysis based on Dunn's tests, in which we assessed each pair of groups. Table 8 shows the results of these tests. A statistical difference was observed between the returns of equally weighted portfolios (i.e., portfolios constructed without using any theoretical methods) and portfolios that minimized the semi-variance. The semi-variance-minimizing portfolios were also statistically superior to the variance-minimizing portfolios.

Table 8. Dunn's test for the pairs of portfolios' returns.

	Equally Weighted	MinV	MinV-E	MinV-E-EBIT	MinV-E-GP	MinV-E-EBITA	MinV-E-EAT	MinV-E-BV	MinSV	MinSV-E	MinSV-E-EBIT	MinSV-E-GP	MinSV-E-EBITA	MinSV-E-EAT
MinV	0.21	-												
MinV-E	0.81	0.60	-											
MinV-E-EBIT	0.79	0.58	−0.02	-										
MinV-E-GP	0.49	0.27	−0.32	−0.31	-									
MinV-E-EBITA	0.74	0.53	−0.07	−0.05	0.25	-								
MinV-E-EAT	0.16	−0.05	−0.64	−0.63	−0.32	−0.57	-							
MinV-E-BV	0.87	0.66	0.06	0.08	0.38	0.13	0.70	-						
MinSV	**1.67** **	**1.46** *	0.86	0.88	1.18	0.93	**1.51** *	0.80	-					
MinSV-E	**1.58** *	**1.37** *	0.77	0.79	1.10	0.85	**1.42** *	0.72	−0.09	-				
MinSV-E-EBIT	**1.61** **	**1.40** *	0.80	0.82	1.13	0.87	**1.45** *	0.74	−0.06	0.03	-			
MinSV-E-GP	**1.43** *	1.21	0.62	0.63	0.94	0.69	1.26	0.56	−0.24	−0.16	−0.19	-		
MinSV-E-EBITA	**1.54** *	**1.33** *	0.74	0.75	1.06	0.81	**1.38** *	0.68	−0.13	−0.04	−0.07	0.12	-	
MinSV-E-EAT	1.03	0.82	0.22	0.24	0.54	0.29	0.87	0.16	−0.64	−0.55	−0.58	−0.40	−0.51	-
MinSV-E-BV	1.04	0.82	0.23	0.24	0.55	0.30	0.87	0.17	−0.63	−0.55	−0.58	−0.39	−0.51	0.01

p-values: * < 0.1, ** < 0.05, and *** < 0.01. The results with *p*-values lower than 0.1 are bolded.

6. Conclusions

This paper involves the development of fundamental portfolios using both variance and semi-variance approaches. An iterative algorithm written in R software (R.4.3.1) was used to construct the portfolios in the semi-variance framework.

For fundamental portfolios, three criteria were considered: profitability (measured with the expected return), risk (measured using the variance or semi-variance of returns), and the market ratio of the companies in the portfolio. Five different market ratios were used in this study. The usefulness of portfolio selection models during the COVID-19 pandemic was analyzed. This period was divided into four subperiods due to the changing situation of the Warsaw Stock Exchange.

During this period characterized by the collapse of the financial market, the worse strategy was to select an equally weighted portfolio. It was the least profitable and the most risky one, taking into account, the value-at-risk measure and semi-deviation. It can be seen that the safest and at the same time most profitable portfolios were fundamental portfolios that minimized semi-variance.

Throughout the entire period under review, the portfolios with three criteria (average return, risk, and a fundamental criterion) allowed higher realized returns to be achieved than equally weighted portfolios, portfolios minimizing the risk (measured using either the variance or semi-variance), or the average return–risk portfolios. In addition, it was found that the introduction of the fundamental criterion reduced the risk borne by an investor, measured with both standard deviation and semi-deviation. The analysis of the value-at-risk measure also shows the advantage of fundamental portfolios. It was revealed that the whole group of portfolios that minimized semi-variance were characterized by a lower ex post risk than those that minimized the variance.

- The empirical research for the largest companies traded on the Warsaw Stock Exchange reveals the following findings:
- Investors can obtain better investment results by adding a criterion associated with market ratios, such as book-to-market or earnings-to-market ratios, to the Markowitz model;
- The use of semi-variance instead of variance yields better results for investors, as can be clearly seen in the period of the collapse of the capital market;
- Fundamental portfolios with minimum semi-variance seem to be a useful tool to choose an investment strategy during the COVID-19 pandemic.

Author Contributions: Conceptualization, A.R.-Z. and P.K.; methodology, A.R.-Z. and P.K.; software, P.K.; validation, A.R.-Z. and P.K.; formal analysis, A.R.-Z.; investigation, A.R.-Z. and P.K.; resources, A.R.-Z.; data curation, A.R.-Z.; writing—original draft preparation, A.R.-Z. and P.K.; writing—review and editing, P.K.; visualization, P.K.; supervision, A.R.-Z.; project administration, A.R.-Z.; funding acquisition, A.R.-Z. All authors have read and agreed to the published version of the manuscript.

References

Adcock, Christopher J., and Karl Shutes. 2005. An analysis of skewness and skewness persistence in three emerging markets. *Emerging Markets Review* 6: 396–418. [CrossRef]

Ballestero, Enrique, Mila Bravo, Blanca Pérez-Gladish, Mar Arenas-Parra, and David Pla-Santamaria. 2012. Socially Responsible Investment: A Multicriteria Approach to Portfolio Selection Combining Ethical and Financial Objectives. *European Journal of Operational Research* 216: 487–94. [CrossRef]

Barbee, William C., Jr., Jin-Gil Jeong, and Sandip Mukherji. 2008. Relations between Portfolio Returns and Market Multiples. *Global Finance Journal* 19: 1–10. [CrossRef]

Basu, Sanjoy. 1977. Investment Performance of Common Stocks in Relation to their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis. *Journal of Finance* 32: 663–82. [CrossRef]

Bawa, Vijay S. 1975. Optimal Rules for Ordering Uncertain Prospects. *Journal of Financial Economics* 2: 95–121. [CrossRef]

Bilbao-Terol, Amelia, Mar Arenas-Parra, Verónica Cañal-Fernández, and Celia Bilbao-Terol. 2013. Selection of Socially Responsible Portfolios Using Hedonic Prices. *Journal of Business Ethics* 115: 515–29. [CrossRef]

Boasson, Vigdis, Emil Boasson, and Zhao Zhou. 2011. Portfolio optimization in a mean-semivariance framework. *Investment Management and Financial Innovations* 8: 58–68.

Breen, William. 1968. Low Price-Earnings Ratios and Industry Relatives. *Financial Analyst Journal* 24: 125–27. [CrossRef]

Briec, Walter, Kristiaan Kerstens, and Octave Jokung. 2007. Mean-Variance-Skewness Portfolio Performance Gauging: A General Shortage Function and Dual Approach. *Management Science* 53: 135–49. [CrossRef]

Burchi, Alberto, and Bogdan Włodarczyk. 2020. 'Best in class' socially responsible investment. The actual performance evaluation between the US and Europe. *Journal of Sustainable Finance & Investment* 12: 275–98. [CrossRef]

Chan, Louis K. C., Yasushi Hamao, and Josef Lakonishok. 1991. Fundamentals and Stock Returns in Japan. *The Journal of Finance* 46: 1739–64. [CrossRef]

Doumpos, Michael, and Constantin Zopounidis. 2014. *Multicriteria Analysis in Finance*. New York: Springer.

Estrada, Javier, and Ana Paula Serra. 2005. Risk and return in emerging markets: Family matters. *Journal of Multinational Financial Management* 15: 257–72. [CrossRef]

Fabozzi, Frank J., Sergio Focardi, and Caroline Jonas. 2007. Trends in Quantitative Equity Management: Survey Results. *Quantitative Finance* 7: 115–22. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 2015. A five-factor asset pricing model. *Journal of Financial Economics* 116: 1–22. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 2017. International tests of a five-factor asset pricing model. *Journal of Financial Economics* 123: 441–63. [CrossRef]

Fama, Eugene F., and Kenneth R. French. 1992. The Cross-Section of Expected Stock Returns. *The Journal of Finance* 47: 427–65. [CrossRef]

Fishburn, Peter C. 1977. Mean-Risk Analysis with Risk Associated With Below-Target Returns. *American Economic Review* 67: 116–26.

Galagedera, Don U. A., and Robert D. Brooks. 2007. Is co-skewness a better measure of risk in the downside than downside beta? Evidence in emerging market data. *Journal of Multinational Financial Management* 17: 214–30. [CrossRef]

Goldfarb, Donald, and Ashok Idnani. 1983. A numerically stable dual method for solving strictly convex quadratic programs. *Mathematical Programming* 27: 1–33. [CrossRef]

Harlow, W. Van, and Ramesh K. S. Rao. 1989. Asset pricing in a generalized mean-lower partial moment framework: Theory and evidence. *Journal of Financial and Quantitative Analysis* 24: 285–311. [CrossRef]

Jacobs, Bruce I., and Kenneth N. Levy. 2013. Leverage aversion, efficient frontiers, and the efficient region. *Journal of Portfolio Management* 39: 54–64. [CrossRef]

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47: 263–91. [CrossRef]

Klebaner, Fima, Zinoviy Landsman, Udi Makov, and Jing Yao. 2017. Optimal portfolios with downside risk. *Quantitative Finance* 17: 315–25. [CrossRef]

Kolm, Petter N., Reha Tütüncü, and Frank J. Fabozzi. 2014. 60 Years of Portfolio Optimization: Practical Challenges and Current Trends. *European Journal of Operational Research* 234: 356–71. [CrossRef]

Lam, Keith S. K. 2002. The relationship between size, book-to-market equity ratio, earnings-price ratio, and return for the Hong Kong stock market. *Global Finance Journal* 13: 163–79. [CrossRef]

Lo, Andrew W., Constantin Petrov, and Martin Wierzbicki. 2003. Its 11pm—Do You Know Where Your Liquidity Is? The Mean-Variance-Liquidity Frontier. *Journal of Investment Management* 1: 55–93.

Markowitz, Harry M. 1952. Portfolio Selection. *The Journal of Finance* 7: 77–91.

Markowitz, Harry M. 1959. *Portfolio Selection: Efficient Diversification of Investments*. New York: Wiley.

Markowski, Lesław. 2001. Risk analysis of capital investment on the Warsaw Stock Exchange in the context of portfolio theory. *Economic Science* 3: 159–77.

Peiro, Amado. 1999. Skewness in financial returns. *Journal of Banking and Finance* 23: 847–62. [CrossRef]

Pla-Santamaria, David, and Mila Bravo. 2013. Portfolio optimization based on downside risk: A mean-semivariance efficient frontier from Dow Jones blue chips. *Annals of Operations Research* 205: 189–201. [CrossRef]

Pośpiech, Ewa. 2019. Effective Portfolios—An Application of Multi-Criteria and Fuzzy Approach. *Folia Oeconomica Stetinesia* 18: 126–39. [CrossRef]

Post, Thierry, and Pim van Vliet. 2006. Downside risk and asset pricing. *Journal of Banking and Finance* 30: 823–49. [CrossRef]

Rodríguez, Rafael, Mariano Luque, and Mercedes González. 2011. Portfolio Selection in the Spanish Stock Market by Interactive Multiobjective Programming. *Top* 19: 213–31. [CrossRef]

Rosenberg, Barr, Kenneth Reid, and Ronald Lanstein. 1985. Persuasive evidence of market inefficiency. *The Journal of Portfolio Management* 11: 9–16. [CrossRef]

Rutkowska-Ziarko, Anna. 2013. Fundamental Portfolio Construction Based On Semi-Variance. *Olsztyn Economic Journal* 8: 151–62. [CrossRef]

Rutkowska-Ziarko, Anna, and Przemysław Garsztka. 2014. Diversification of risk of a fundamental portfolio based on semi-variance. *Poznań University of Economics Review* 14: 80–96.

Sharpe, Wiliam F. 1963. A Simplified Model for Portfolio Analysis. *Management Science* 9: 277–93. [CrossRef]

Steuer, Ralph E., Yue Qi, and Markus Hirschberger. 2007. Suitable-portfolio investors, nondominated frontier sensitivity, and the effect of multiple objectives on standard portfolio selection. *Annals of Operational Research* 152: 297–317. [CrossRef]

Sun, Qian, and Yuxing Yan. 2003. Skewness persistence with optimal portfolio selection. *Journal of Banking and Finance* 27: 1111–21. [CrossRef]

Tarczyński, Waldemar. 2002. *Portfel Fundamentalny Papierów Wartościowych [Fundamental Portfolio of Assets]*. Warsaw: Polish Economic Editorial Office (PWE). (In Polish)

Xidonas, Panagiotis, George Mavrotas, and John Psarras. 2010. Equity Portfolio Construction and Selection Using Multiobjective Mathematical Programming. *Journal of Global Optimization* 47: 185–209. [CrossRef]

Zaremba, Adam, and Anna Czapkiewicz. 2017. Digesting anomalies in emerging European markets: A comparison of factor pricing models. *Emerging Markets Review* 31: 115. [CrossRef]

Article

The Effects of Option Trading Behavior on Option Prices

Han-Sheng Chen [1] and Sanjiv Sabherwal [2,*]

[1] Department of Accounting, Finance, and Economics, Lipscomb University, 1 University Park Dr, Nashville, TN 37204, USA; hchen1@lipscomb.edu

[2] Department of Finance & Real Estate, University of Texas at Arlington, 701 S. West Street, Arlington, TX 76019, USA

[*] Correspondence: sabherwal@uta.edu

Abstract: This paper investigates the relationship between option trading behavior and option pricing patterns. We argue that greater active trading in the options market due to investor overconfidence leads to higher volatility and larger discrepancies in option pricing, which may be captured by implied volatility spread and implied volatility skewness. Using two different measures of excess option trading, we find that trading activities are correlated in different ways with volatility, volatility spread, and volatility skewness. We also find that these relationships exist both over time and cross-sectionally. We suggest that options investors tend to chase "hot" stocks, as we find evidence of a positive relationship between option trading activities and past underlying equity returns. Heavier trading in the options market also tends to make out-of-the-money call options more (less) expensive than the at-the-money counterparts over time (cross-sectionally). Because trading activities do not predict future equity returns, investor overconfidence, and not informed trading, seems to be a more plausible explanation for our findings.

Keywords: overconfidence; options market; option turnover; volatility spread; volatility smirk; behavioral finance

1. Introduction

Trading behavior in the options market has drawn increasing attention from financial economists as the importance of the options market has increased, especially in recent years. While, theoretically, options can be replicated in a complete market, and some deem them redundant securities (Black and Scholes 1973), markets for these financial instruments have only grown larger. For example, CBOE reported a total trading volume of 3.4 billion as the number of contracts across all options products in North America in 2022, compared to 1.05 billion in 2015 and 254 million in 1999. Many financial scholars have focused on the study of the options market. Researchers have been striving to interpret the information content embedded in trading behavior. However, interpreting the information correctly is difficult if we fail to identify the actual motives for trading.

Scholars have proposed two main reasons why investors trade. One is differences of opinion, while the other is superior information. Both arguments suggest that investors trade because they hold beliefs that differ from those of general market participants, the latter of which are reflected in the current market prices. The two hypotheses are distinct in that the information-driven hypothesis assumes that investors who trade possess private information, whereas the differences-in-opinion hypothesis suggests that investors interpret the same information differently. While we observe the same increase in trading activities, the two hypotheses generate distinct inferences regarding how asset prices react to this information. Therefore, an understanding of the reasons for trading is crucial.

Studies identifying the reasons for trading options have shown mixed results. Starting from Black (1975), who argues that the leverage effects in options can attract informed traders, Amin and Lee (1997), Easley et al. (1998), Cao et al. (2005), and Pan and Poteshman

(2006) have found evidence supporting the information-driven hypothesis. On the other hand, Stephen and Whaley (1990), Vijh (1990), Chan et al. (1993, 2002), Muravyev et al. (2013), and Choy and Wei (2012) present evidence against informed trading.

In attempts to find reasons for excess trading activities in the stock market, researchers have found behavioral factors to be a logical fit. Grinblatt and Keloharju (2009) report that overconfidence and sensation-seeking lead to more frequent stock trading activities. More recently, Ülkü et al. (2023) have presented evidence supporting the idea that retail investors generally exhibit contrarian traits. Using trading data obtained from several countries during the COVID-19 pandemic, they also show that the net-trading direction between retail traders and institutional trades may diverge for an extended time period. While it has been shown in the literature that the options market plays an important informational role (e.g., Chakravarty et al. 2004), one may wonder what effects behavioral factors may have in this market. According to Scheinkman and Xiong (2003), investor overconfidence may intensify differences of opinion in the form of over-optimism for overconfident agents, and consequently create a price bubble. Also, bubbles are associated with large trading volumes and high price volatility. Empirically, Choy (2015) shows that retail investors speculate and are willing to pay a premium for future expected volatility, which provides evidence supporting a behavioral theory in the options market. A similar phenomenon can also be established during a negative bubble. Baig et al. (2022) studied the increased role played by retail investors during stressful times, including the 2008–2009 financial crisis and the COVID-19 pandemic. They document a negative impact of retail trading on the stability of stock prices that was particularly strong during the 2008–2009 financial crisis and the pandemic. The findings of Baig et al. (2022) and Ülkü et al. (2023) suggest that empirically examining whether there is a linkage between trading activities and measures of price bubbles and between trading and volatility is meaningful for understanding option trading and its information content.

This paper addresses two main research questions. First, we investigate the relationship between option trading activities and option prices and volatility over time. That is, we examine whether higher or lower volatility or discrepancy levels in option prices are associated with a higher option turnover rate. According to Scheinkman and Xiong (2003), volatility and price bubbles would intensify when there was an increase in trading activities. It is therefore natural to examine how volatility and option pricing patterns develop over time due to excessive trading.

Second, we investigate whether investor sentiment affects option pricing cross-sectionally. Options with higher turnover rates may behave differently than would those with lower turnover rates, should behavioral factors play a significant role in option prices. As Grinblatt and Keloharju (2009) point out, behavioral factors such as overconfidence and sensation-seeking tend to drive up trading activities. Regardless of the market they choose to trade in, overconfident agents may try to take advantage of their information (or beliefs) and consequently trade more frequently. Therefore, we use option turnover rate as a proxy for investor overconfidence and test the hypotheses of there being relationships between overconfidence and option volatility and option pricing. Cremers and Weinbaum (2010) use the difference in implied volatility between pairs of call and put options (volatility spread) to measure the relative expensiveness of call options over put options. Volatility spread may serve as a good indicator in examining whether call options become more expensive relative to the corresponding put options when the market presents evidence of overconfidence. As stated above, we expect to observe a positive relationship between the overconfidence measure and volatility spread.

The rest of the paper is organized as follows. The next section includes a review of the related literature. Section 3 discusses the research questions and empirical methodology. Section 4 provides the empirical results. The last section includes a discussion of the paper's findings and their implications.

2. Literature Review

2.1. Overconfidence and Momentum

Both momentum (Jegadeesh and Titman 1993) and reversals (DeBondt and Thaler 1985) are well documented in the stock market literature. While a momentum strategy that buys winning stocks and short-sells losing stocks generates superior average returns in the short run, it results in negative average returns in the long run (reversals). A simple but popular explanation that fits both phenomena is behavioral. The behavioral theories that try to address the issue include Barberis et al. (1998), Daniel et al. (1998), Hong and Stein (1999), and George and Hwang (2004). Among those theories, Daniel et al. (1998) attribute the phenomena to the behavior of overconfident agents. In their framework, investors bear self-attribution bias; that is, they tend to attribute their success in investment to their trading skills and knowledge and blame their failure on bad luck or unpredictable noises. This theory is empirically supported by Lee and Swaminathan (2000), Statman et al. (2006), and Cremers and Pareek (2014), to name a few. While evidence from stock markets generally supports the self-attribution bias, even for institutional investors (Cremers and Pareek 2014), the existence and the potential influence of such a bias are largely not discussed in the literature.

2.2. Price Patterns in the Options Market

Options market pricing has intrigued financial economists in various ways for decades. One of the heavily discussed topics is the existence of arbitrage opportunities. Initially, options were deemed to be redundant securities (Black and Scholes 1973; Cox et al. 1979), and investors should have no reason to trade such financial instruments. However, scholars have empirically identified deviations from basic option pricing rules such as put–call parity (e.g., Ofek et al. 2004) which provide incentives for investors to trade in the options market.

The argument about the existence of arbitrage opportunities goes on, as other researchers have shown evidence against those findings (e.g., Battalio and Schultz 2006). Even if deviations exist from no-arbitrage relations, most will agree that the arbitrage opportunities dissipate fairly quickly. They can hardly account for the extensive trading activities in the options market. Cremers and Weinbaum (2010), on the other hand, argue that the relative expensiveness of put and call options, as paired with strike price and that of the underlying security, may predict future stock performance. Their findings provide further reasons for trading, suggesting that predictability comes from the mispricing of options. Coinciding with Cremers and Weinbaum (2010), Xing et al. (2010) found that the shape of a volatility smirk predicts future stock returns. Some papers suggest that informed traders may lead the trading in the options market due to leverage (Black 1975) and reveal their information within option prices. While others argue that differences of opinion are the main reason for trading (Choy and Wei 2012), the causes of differences of opinion remain largely uninvestigated.

3. Research Questions and Empirical Methodology

This study aims to test the relationship between trading activities and option prices empirically. Grinblatt and Keloharju (2009) show the connections between more frequent trading activities, overconfidence, and sensation seeking. As suggested in Scheinkman and Xiong (2003), investor overconfidence intensifies differences of opinion and therefore causes heavier trading. Higher volatilities, as well as price bubbles, accompany heavier trading. When investors trade, not based on information, but on behavioral factors, we expect trading activities to be higher than usual. Also, it is more likely to observe a price bubble in the corresponding market. Building upon the abovementioned expectations, this paper empirically tests the relationship between excessive trading activities and option price patterns.

3.1. Empirical Methodology

Our first objective is to capture excessive trading activities. We propose two measures for this purpose. Following Statman et al. (2006), two control variables are used to account for normal trading motives. The first control variable is market volatility, misg, based on the research by Karpoff (1987) on the contemporaneous volume–volatility relationship. The second control variable is dispersion, disp, which is associated with the idiosyncratic risk of the underlying stock and therefore accounts for trading activities related to portfolio rebalancing. In addition, we include proportional effective spread, sprd, to control for liquidity. Specifically, the proportional effective spread (a measure of illiquidity) for underlying equity *j* on day *D* is calculated as follows:

$$sprd_{D,j} = \frac{1}{Vol_{D,j}}\sum\nolimits_{k=1}^{n} Vol_{D,j,k} \times 100 \times \frac{2 \times \left(Offer_{D,j,k} - Bid_{D,j,k}\right)}{\left(Offer_{D,j,k} + Bid_{D,j,k}\right)} \tag{1}$$

where $Vol_{D,j} = \sum_{k=1}^{n} Vol_{D,j,k}$, and *k* stands for different strike prices. The primary measure of trading volume used in this study is option trading turnover, TO_O, which is defined as option trading volume multiplied by 100, scaled by open interest. The model is as follows:

$$TO\_O_t = a + b_1 \times misg_t + b_2 \times disp_t + b_3 \times sprd_t + \varepsilon_t \tag{2}$$

We extract the residuals from the above regression and use them as our first measure of overconfidence over time.

In addition to the above measure, we apply the stochastic frontier analysis (SFA) technique to isolate the potential trading behavior due to overconfidence from the behavior based on random information flows. The rationale behind using SFA in this study is that we treat overconfidence as a systematic bias for investors, which constantly drives up trading volume. Since the standard ordinary least square method does not distinguish between systematic bias in trading and purely stochastic component in trading activities, SFA's capability to capture the systematic bias via skewness in residuals would help extract trading activities due to overconfidence. We use the following regression model:

$$TO\_O_t = a + b_1 \times misg_t + b_2 \times disp_t + b_3 \times sprd_t + v_t + u_t \tag{3}$$

where u_t is a one-sided error half normally distributed $N(0^+, \sigma_u^2)$.

We adopt two inefficiency measures in this study. Both are based on technical efficiency measures. That is, $OC_i = 1 - TE_i$, where $i = 1, 2$. TE_1 is the technical efficiency measure used by Battese and Coelli (1988), and TE_2 is the technical efficiency measure used by Jondrow et al. (1982).

Once the overconfidence measure is obtained, we test the relationships between overconfidence and price volatility and price bubble measures. We use the following two measures for volatility: VIX and realized volatility over the past 30 days. The change in volatility is also included as a dependent variable. We use volatility spread and volatility smirk (skewness) for price bubble measures. Volatility spread, proposed by Cremers and Weinbaum (2010), measures the relative expensiveness of calls and puts with the same strike price. Cremers and Weinbaum find that the stocks with relatively expensive calls outperform stocks with relatively expensive puts. They also document that this finding is likely due to the information risk that the underlying stocks face. If investors in the options market are overly optimistic about the performance of the underlying stock, we should observe more expensive call options relative to put options with the same strike. In such a case, the subsequent superior performance they have documented may be explained as the confidence building up in the options market and then spilling over to the underlying stock market.

A similar argument can be applied to volatility smirk, which measures the relative expensiveness of in-the-money and out-of-the-money calls (puts). As in the case of calls,

the general explanation is that in-the-money call options offer leverage and, therefore, a more promising strategy for investors who wish to take long positions. As investors become overly optimistic, we should observe more expensive in-the-money call options and less expensive out-of-the-money put options (fewer hedging activities using puts). Consequently, the volatility smirk for calls (puts) will become steeper (flatter).

3.2. Hypotheses

Based on our discussion above, we test the following hypotheses in this paper:

H1. *Higher investor overconfidence leads to both higher expected volatility and higher subsequently realized volatility.*

H2. *Higher investor overconfidence makes call options more expensive than the corresponding put options with the same strike price.*

H3. *Higher investor overconfidence results in more expensive in-the-money/at-the-money call options relative to the out-of-the-money options and less expensive out-of-the-money put options relative to the in-the-money/at-the-money options.*

Time series regressions were conducted to test the above hypotheses. In addition to examining the relationships between investor overconfidence and asset prices over time, we also examined the impacts across firms. Both Cremers and Weinbaum (2010) and Xing et al. (2010) find that the differences in implied volatilities predict future equity returns. While Cremers and Weinbaum indicate that mispricing is the main reason for this finding, Xing et al. argue that informed traders may be the driving force.

In both studies, the authors first sorted the sample firms into portfolios according to volatility spread/skew/smirk, and then showed differences in future performance across portfolios. If investor overconfidence played a specific role in their findings, one should expect that overconfidence measures would be associated with volatility spread/skew/smirk cross-sectionally. For instance, Cremers and Weinbaum found that stocks with more expensive calls or with calls becoming more expensive than in the previous periods earned abnormal positive returns, while the ones with more expensive puts or with puts becoming more expensive than in the previous period earned abnormal negative returns. If firms with more frequent trading activities generally have more expensive calls, the subsequent abnormal returns documented by Cremers and Weinbaum may be the price bubble suggested by Scheinkman and Xiong (2003). A similar argument applies to the predictability of future stock returns according to volatility skew/smirk, as argued by Xing et al. Therefore, we conducted a second series of tests to examine the relationships between trading activities and volatility spread/skew/smirk across firms.

3.3. Data

The option data was retrieved from OptionMetrics (New York, NY, USA) via WRDS. End-of-day bid and ask quotes, open interests, trading volume, and implied volatility were obtained from the database for the period ranging from January 1996 to December 2011. The sample included 2779 unique firms listed on NYSE/AMEX, and with options traded. In addition to the option turnover rate, the O/S ratio was also used as a measure of trading activities. Since different practices in reporting trading volume in dealers' markets may cause inconsistency in the O/S ratio, the sample in this study consisted only of firms listed on NYSE/AMEX with options. VIX, a forward volatility index proposed by CBOE, was used as a measure of volatility for the entire market. End-of-day stock prices and trading volume were extracted from the Center for Research in Security Prices (CRSP).

Table 1 contains descriptive statistics for all the variables used in the empirical analyses. Panel A shows the characteristics of the primary dependent variables used in the empirical studies. Note that the percentage change in 30-day volatility has a mean and median close

to zero. Volatility spread has a negative mean and median, as reported in Cremers and Weinbaum (2010), while volatility skew has a positive mean and median, consistent with Xing et al. (2010).

Table 1. Sample characteristics.

Panel A: Volatility Measures and Price Discrepancy Measures			
Measure	**Mean**	**Median**	**Std. Dev.**
% Change in VIX	0.0164	−0.0145	0.1800
% Change in 30-day Volatility	0.0002	0.0001	0.0130
30-day Realized Volatility	0.4617	0.4193	0.1914
Volatility Spread	−0.0091	−0.0079	0.0076
Volatility Skew	0.0458	0.0387	0.0237
Panel B: Overconfidence Measures			
Measure	**Mean**	**Median**	**Std. Dev.**
Option Turnover	3.8552	3.7904	0.9073
OLS Residual	0.0000	−0.1522	0.8498
OC1	7.3858	5.7913	5.3460
OC2	7.6567	5.9459	5.6454

This table shows summary statistics for variables used in this study. Panel A summarizes the mean, median, and standard deviation of the primary dependent variables used in the empirical analysis. The numbers shown for all variables, except for the percentage changes of volatility measures, are the daily averages over the sample period. Volatility change is the percentage changes in the daily average volatility of the corresponding month from that of the previous month. Volatility spread is the weighted average difference in implied volatility of paired call and put options with the same strike price and the same underlying equity, as specified in Cremers and Weinbaum (2010). Volatility skew is the difference between the implied volatility of out-of-the-money (OTM) put options and the implied volatility of at-the-money (ATM) call options. Panel B summarizes the mean, median, and standard deviation for the explanatory variables, which are used as a proxy of investor overconfidence. OC1 and OC2 are (1) residuals from ordinary least square regressions; and (2) overconfidence measures from stochastic frontier analysis (SFA), respectively.

Panel B summarizes the mean, median, and standard deviation for the explanatory variables, which are used as a proxy of investor overconfidence. OC1 and OC2 are the inefficiency measures derived from stochastic frontier analysis (SFA) and are very similar qualitatively and quantitatively. We expect they would yield similar results in the main empirical analyses.

4. Results

As discussed in the previous section, the main question addressed in this study is whether investor overconfidence plays a role in option pricing. To investigate this issue, we conducted two series of tests. The first set of tests ran regressions of trading activities, which is used as a proxy for investor overconfidence, on volatility measures and relative expensiveness across options. Before running this set of tests, we checked that the variables were stationary, in order to avoid spurious regressions. Specifically, we used augmented Dickey–Fuller and Phillip–Perron tests to check for the stationarity of all dependent and independent variables used in our regression analysis. The results of the tests on all independent variables (i.e., OLS residuals, OC1, and OC2) reject the null hypothesis of a unit root at the 1% level. The results of the tests on all dependent variables reject the null hypothesis of unit root at the 1% level, except for volatility skew (IV_SKEW) and volatility smirk (IV_SMIRK), for which the results reject the null hypothesis at the 5% level.

The second set of tests involved sorting sample firms into portfolios based on trading activities and examining the differences in volatility spread/skew/smirk across portfolios. This section provides the results of these two sets of tests.

4.1. Time-Series Regressions

To construct the measure of option trading activities, we aggregated daily trading volumes and open interests across all options for the entire sample of firms, and then divided the aggregated trading volume by the end-of-the-day aggregated open interest. We defined this ratio as the option turnover rate.

For changes in VIX, we obtained the daily VIX from WRDS and then took the average of the daily VIX over a calendar month. The changes in VIX are the percentage changes in daily average VIX in the current month from that of the previous month. Daily realized volatilities for sample firms were obtained from OptionMetrics. For each day of a given month, volatilities realized during the past 30 calendar days were extracted and averaged over the month. The changes in realized volatility are the percentage changes of average realized 30-day volatility in a given month as compared to those of the previous month.

The volatility spread was calculated daily for each sample firm and averaged over a month. Following Cremers and Weinbaum (2010), we paired call and put options with the same underlying equity, strike, and maturity, and then calculated volatility spread as the difference between the implied volatilities of the call and put options. Daily volatility spread was defined for each trading day as the weighted average spread for each pair of call and put options with the same strike price and maturity. Following Xing et al. (2010), implied volatility skew was calculated as the difference between the implied volatilities of OTM puts and ATM calls.

There are several ways to determine the moneyness of options. In this study, following Xing et al. (2010), an option is defined as OTM when the absolute delta of the option is at least 0.125 but less than 0.375. It is defined as ATM when the absolute delta is at least 0.375 but less than 0.625, and finally, it is defined as ITM when the absolute delta is at least 0.625 but less than 0.875. A simpler way to define moneyness is to use the ratio of the strike price to the stock price (K/S). Ni (2007) uses the total volatility-adjusted strike-to-stock-price ratio as another moneyness measure. However, these alternative methods yield quantitatively similar results.

Daily volatility skew was averaged across sample firms in a day, weighted by the end-of-the-day open interests. We computed monthly volatility skew by averaging the daily volatility skew over a month.

Table 2 presents the results of the first empirical test for all options (calls and puts). As mentioned in the previous section, the explanatory variables are derived from the first stage regression. The residuals are extracted from the first stage regression using the ordinary least square method, controlling for market volatility, idiosyncratic risk, and proportional effective spread. OC1 and OC2 are inefficiency measures derived from stochastic frontier analysis, assuming half normal distribution in inefficiency. Specifically, they are one minus the technical efficiency measures, as suggested by Battese and Coelli (1988) and Jondrow et al. (1982), respectively.

It is apparent that OLS residual and OC1/OC2 paint different pictures in this table. Focusing first on the results of the second-stage regression using OLS residuals as the explanatory variable, we find that OLS residuals are positively related to the percentage changes in expected and realized volatility measures from the previous month, with the F-statistics of the regressions being 2.95 and 5.18, respectively. These results serve as a piece of evidence supporting the theory in Scheinkman and Xiong (2003) that investor overconfidence intensifies differences of opinions and consequently causes higher volatility. On the other hand, OLS residuals and volatility spread are negatively correlated, with a regression F-statistic of 9.89. This suggests that an increase in the frequency of trading activities tends to make put options more expensive than call options.

OLS residuals and volatility skew are negatively correlated (the F-statistic is 3.23). This result is intriguing, as it indicates the presence of fewer hedging activities using OTM put options. Therefore, we further investigated the difference in implied volatilities across the moneyness of options. In the options market, implied volatility skew is negatively sloped across strike prices (higher implied volatility for ITM call options and OTM put options,

relative to OTM call options and ITM put options). As shown in Figure 1, the pattern is clear throughout the sample period, while it tends to be more severe during a financial crisis. In both crises during the sample period, i.e., the post-dot-com bubble era and the 2007–2009 financial crisis, there were large spikes. Also, there is a tendency towards steeper slopes over time.

Table 2. Regression analysis—volatility measures and price discrepancy measures against unexpected turnovers on *all* options.

Dependent Variables	Explanatory Variables			
	OLS Residual	OC1	OC2	Adj. R^2
Percentage Change in VIX	0.0280 **	−0.0050 **	−0.0047 **	0.0144
	(0.0115)	(0.0020)	(0.0019)	
Past 30 Days' Realized Volatility	0.0000	−0.0018	−0.0017	−0.0015
	(0.0205)	(0.0030)	(0.0028)	
Volatility % Change—Past 30 Days	0.0044 ***	−0.0072 ***	−0.0007 ***	0.0778
	(0.0010)	(0.0014)	(0.0001)	
Volatility Spread	−0.0018 *	0.0003 ***	0.0003 ***	0.0384
	(0.0011)	(0.0001)	(0.0001)	
Changes in Volatility Spread	−0.0007	0.0001	0.0001	0.0045
	(0.0006)	(0.0001)	(0.0001)	
Volatility Skew	−0.0049 **	0.0009 **	0.0008 **	0.0264
	(0.0024)	(0.0004)	(0.0004)	
Changes in Volatility Skew	0.0006	−0.0002	−0.0002	0.0034
	(0.0010)	(0.0001)	(0.0001)	
Call Volatility Smirk (ATM–OTM)	−0.0042 **	0.0007 ***	0.0007 ***	0.0396
	(0.0018)	(0.0002)	(0.0002)	
Call Volatility Smirk (ITM–OTM)	−0.0077 *	0.0013 **	0.0012 **	0.0223
	(0.0040)	(0.0006)	(0.0005)	
Put Volatility Smirk (ATM–OTM)	0.0048 *	−0.0009 **	−0.0008 **	0.0256
	(0.0025)	(0.0004)	(0.0004)	
Put Volatility Smirk (ITM–OTM)	0.0095 **	−0.0017 ***	−0.0016 ***	0.0475
	(0.0038)	(0.0006)	(0.0005)	

The regressions use monthly aggregated market observations. Explanatory variables are overconfidence measures, using option turnovers from *all* options, and controlling for market volatility, idiosyncratic risk of the underlying stock, and proportional effective spread. Specifically, the overconfidence measures, OC1 and OC2, are (1) residuals from ordinary least square regressions; and (2) overconfidence measures from stochastic frontier analysis (SFA), respectively. Dependent variables are volatility measures and price discrepancy measures. The volatility spread is from Cremers and Weinbaum (2010), while the volatility skew is from Xing et al. (2010). Numbers in parentheses are standard errors. The adjusted R^2 values are the averages of the corresponding values of the three regressions. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

In tests of the volatility skew/smirk slope, we found that OLS residuals were associated with flatter slopes, which means less expensive ITM calls and OTM puts. The findings are indicated by negative (positive) coefficients on volatility smirk for call (put) options, and the coefficients are statistically significant at the 10% level. The F-statistics for these regressions range from 2.90 to 6.67, which implies the validity of the models at the 10% and 5% levels.

A natural explanation for this finding may be that overconfident agents try to take their chances in the options market, generating a higher demand for OTM call options. In comparison, they are less worried about market crashes, creating less demand for put options. While the finding from the slopes of the volatility smirk is consistent with the one from the volatility skew, it still does not explain the lower volatility spread. One possibility is that the volatility spread is weighted by open interests, reflecting the relative expensiveness of ATM call and put options. That is, ATM call options become less expensive than ATM put options. This might be due to the standard trading strategy of a covered call, which sells short ATM call options instead of dumping underlying equity into the market to increase portfolio returns.
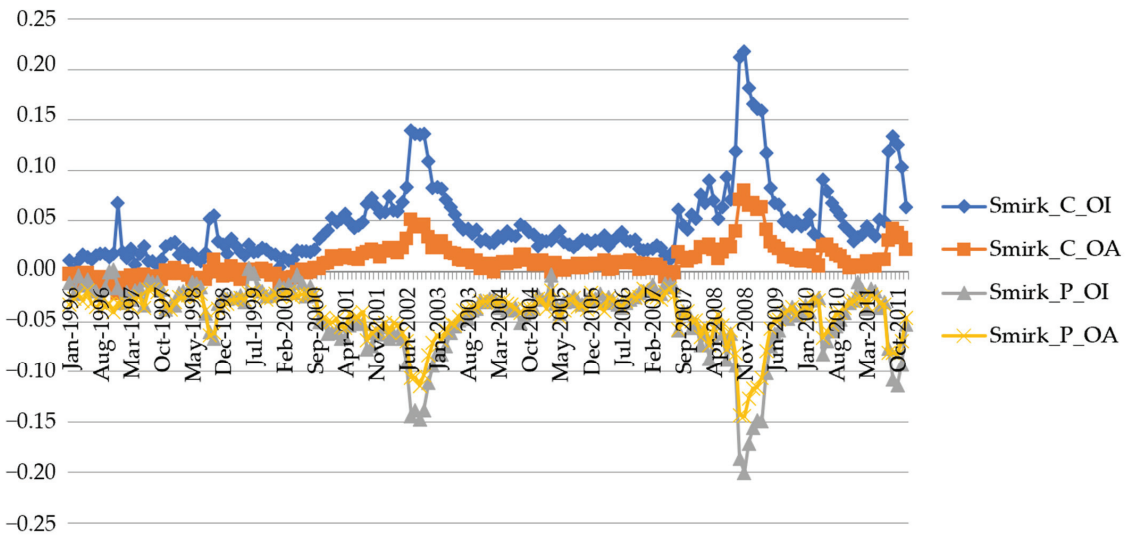
Figure 1. Volatility smirk over time. This figure exhibits an implied volatility smirk over time. Smirk_C_OA is the difference in implied volatility between ATM call options and OTM call options. Smirk_C_OI is the difference in implied volatility between ITM and OTM call options. Smirk_P_OA is the difference in implied volatility between ATM and OTM put options. Smirk_C_OA is the difference in implied volatility between ITM and OTM put options.

When we use inefficiency measures from SFA as a measure of investor overconfidence, we find a different picture. Both OC1 and OC2 are negatively correlated with changes in volatility measures from the previous month, while they are positively correlated with volatility spread (with F-statistics ranging from 7.96 to 18.12). In addition, there is a positive correlation between investor overconfidence measures and the steepness of volatility smirk across strike prices (with F-statistics ranging from 6.45 to 12.51). As the methodology section explains, OC1 and OC2 are technical inefficiency measures derived from SFA. Therefore, we see them as overly aggressive trading activities and as proxies for investor overconfidence. The results, in sum, do not agree with the argument.

First, we find negative and statistically significant coefficients for volatility measures. This suggests that OC1 and OC2 capture trading activities when option prices are relatively stable and expected to stay stable. The findings from volatilities are consistent with the ones from volatility skew/smirk. A general argument for the existence of volatility skew/smirk is that investors are worried about a market crash and, therefore, would like to protect their holdings by buying more OTM put options. Another popular explanation is that investors use ATM/ITM call options instead of their stock investments to enhance rates of return. Both explanations are supported in this line of tests, given that OC1 and OC2 are positively correlated with volatility skew (more expansive OTM puts than ATM calls) and with the slope of the volatility smirk. Again, volatility spread positively correlates with OC1 and OC2, which may seem to contradict the previous argument. As explained above, ATM call and put options may be driving this finding.

We conducted similar tests using call and put option turnover ratios, as described in Tables 3 and 4, respectively. The results are qualitatively similar across the three tables, as most coefficients appear in the same signs with their corresponding peers in all three tables, and no surprisingly larger or smaller coefficient is identified. The only noticeable difference is that put option turnover seems to have better explanatory power for volatility skew/smirk (and also with significantly higher F-statistics of 8.80 to 14.11 and higher adjusted R^2 of 0.0251 to 0.0453). This is consistent with the argument that investors in the options market favor using put options to avoid massive losses in a significant market

crash. The findings are more pronounced when OC1 and OC2 are used as measures of excess trading, which may suggest that the inefficiency trading measures derived from SFA capture investors' fears of market crashes.

Table 3. Regression analysis—volatility measures and price discrepancy measures against unexpected turnovers on *call* options.

	Explanatory Variables			
Dependent Variables	**OLS Residual**	**OC1**	**OC2**	**Adj. R^2**
Percentage Change in VIX	0.0266 **	−0.0063 ***	−0.0060 ***	0.0168
	(0.0135)	(0.0023)	(0.0022)	
Past 30 Days' Realized Volatility	0.0161	−0.0014	−0.0013	−0.0027
	(0.0200)	(0.0037)	(0.0036)	
Volatility % Change—Past 30 Days	0.0036 ***	−0.0008 ***	−0.0008 ***	0.0690
	(0.0010)	(0.0002)	(0.0002)	
Volatility Spread	−0.0025 **	0.0004 ***	0.0003 ***	0.0520
	(0.0012)	(0.0001)	(0.0001)	
Volatility Skew	−0.0022	0.0006	0.0006	0.0066
	(0.0023)	(0.0005)	(0.0004)	
Call Volatility Smirk (ATM–OTM)	−0.0019	0.0005 *	0.0005 **	0.0124
	(0.0017)	(0.0003)	(0.0003)	
Call Volatility Smirk (ITM–OTM)	−0.0027	0.0009	0.0008	0.0033
	(0.0039)	(0.0007)	(0.0007)	
Put Volatility Smirk (ATM–OTM)	0.0021	−0.0006	−0.0006	0.0066
	(0.0024)	(0.0005)	(0.0005)	
Put Volatility Smirk (ITM–OTM)	0.0053	−0.0013 **	−0.0013 **	0.0195
	(0.0036)	(0.0007)	(0.0006)	

The regressions use monthly aggregated market observations. Explanatory variables are overconfidence measures, using option turnovers from call options, and controlling for market volatility, idiosyncratic risk of the underlying stock, and proportional effective spread. Specifically, the overconfidence measures, OC1 and OC2, are (1) residuals from ordinary least square regressions; and (2) overconfidence measures from stochastic frontier analysis (SFA), respectively. Dependent variables are volatility measures and price discrepancy measures. The volatility spread is from Cremers and Weinbaum (2010), while the volatility skew is from Xing et al. (2010). Numbers in parentheses are standard errors. The adjusted R^2 values are the averages of the corresponding values for the three regressions. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

To further explore the above findings, we divided all sample firms into two groups according to the percentage of institutional holdings of the firm. Since institutional investors are less likely to be subject to behavioral biases, if a pattern is more pronounced in the group with lower institutional ownership, the pattern is more likely due to behavioral biases, such as investor overconfidence.

To form the two portfolios, we set the cutoff point at the median percentage of institutional holdings of the entire sample. This sorting resulted in each group having an equal number of firms. By comparing Panel A and Panel B in Table 5, we find very similar results in most of the tests, except the one for volatility spread. All trading measures exhibit a lack of explanatory power as to volatility spread for the group with higher institutional ownership. In comparison, they appear to be highly correlated with volatility spread for the group with lower institutional ownership. Again, OLS residuals are negatively correlated with volatility spread in this table, while OC1 and OC2 are positively correlated with volatility spread. Given that volatility spread is dominated by the demand for ATM call options relative to put options, one may conclude that OLS residuals capture demands on put options while OC1 and OC2 capture demands on call options.

Table 4. Regression analysis—volatility measures and price discrepancy measures against unexpected turnovers on *put* options.

	Explanatory Variables			
Dependent Variables	OLS Residual	OC1	OC2	Adj. R^2
Percentage Change in VIX	0.0406 ***	−0.0056 ***	−0.0053 ***	0.0224
	(0.0157)	(0.0021)	(0.0020)	
Past 30 Days' Realized Volatility	0.0289	−0.0018	−0.0017	0.0013
	(0.0212)	(0.0033)	(0.0031)	
Volatility % Change—Past 30 Days	0.0053 ***	−0.0008 ***	−0.0008 ***	0.0962
	(0.0011)	(0.0001)	(0.0001)	
Volatility Spread	−0.0012	0.0002 *	0.0002 *	0.0099
	(0.0009)	(0.0001)	(0.0001)	
Volatility Skew	−0.0017	0.0009 **	0.0009 **	0.0236
	(0.0027)	(0.0004)	(0.0004)	
Call Volatility Smirk (ATM–OTM)	−0.0026	0.0008 ***	0.0008 ***	0.0439
	(0.0021)	(0.0002)	(0.0002)	
Call Volatility Smirk (ITM–OTM)	−0.0040	0.0016 ***	0.0015 ***	0.0283
	(0.0048)	(0.0006)	(0.0006)	
Put Volatility Smirk (ATM–OTM)	0.0013	−0.0010 **	−0.0009 **	0.0251
	(0.0030)	(0.0004)	(0.0004)	
Put Volatility Smirk (ITM–OTM)	0.0050	−0.0018 ***	−0.0017 ***	0.0453
	(0.0045)	(0.0006)	(0.0006)	

The regressions use monthly aggregated market observations. Explanatory variables are overconfidence measures, using option turnovers from put options, and controlling for market volatility, idiosyncratic risk of the underlying stock, and proportional effective spread. Specifically, the overconfidence measures, OC1 and OC2, are (1) residuals from ordinary least square regressions; and (2) overconfidence measures from stochastic frontier analysis (SFA), respectively. Dependent variables are volatility measures and price discrepancy measures. The volatility spread is from Cremers and Weinbaum (2010), while the volatility skew is from Xing et al. (2010). Numbers in parentheses are standard errors. The adjusted R^2 values are the averages of the corresponding values for the three regressions. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

Table 5. Unexpected turnovers on all options against volatility measures and price discrepancy measures—sorted by institutional ownership.

Panel A: High Institutional Ownership				
	Explanatory Variables			
Dependent Variables	OLS Residual	OC1	OC2	Adj. R^2
Percentage Change in VIX	1.4638	−0.3828 **	−0.3653 **	0.0033
	(1.4622)	(0.1949)	(0.1848)	
Past 30 Days' Realized Volatility	−0.0008	0.0006	0.0006	−0.0051
	(0.0200)	(0.0030)	(0.0028)	
Volatility % Change—Past 30 Days	0.4590 ***	−0.0740 ***	−0.0704 ***	0.0983
	(0.1090)	(0.0197)	(0.0186)	
Volatility Spread	0.0674	−0.0022	−0.0021	−0.0020
	(0.0456)	(0.0060)	(0.0057)	
Volatility Skew	−0.3179	0.0719 *	0.0686 *	0.0108
	(0.2778)	(0.0398)	(0.0378)	
Call Volatility Smirk (ATM–OTM)	−0.3389 *	0.0636 ***	0.0606 ***	0.0307
	(0.1805)	(0.0229)	(0.0217)	
Call Volatility Smirk (ITM–OTM)	−0.0062	0.0012 **	0.0012 **	0.0196
	(0.0041)	(0.0006)	(0.0005)	
Put Volatility Smirk (ATM–OTM)	0.3506	−0.0782 **	−0.0746 **	0.0179
	(0.2586)	(0.0362)	(0.0344)	
Put Volatility Smirk (ITM–OTM)	0.0076 *	−0.0016 ***	−0.0016 ***	0.0386
	(0.0040)	(0.0006)	(0.0005)	

Table 5. *Cont.*

Panel B: Low Institutional Ownership				
Explanatory Variables				
Dependent Variables	**OLS Residual**	**OC1**	**OC2**	**Adj. R^2**
Percentage Change in VIX	1.5253	−0.4060 **	−0.3833 **	0.0086
	(1.4289)	(0.1781)	(0.1670)	
Past 30 Days' Realized Volatility	0.0317 *	−0.0036	−0.0034	0.0101
	(0.0190)	(0.0029)	(0.0027)	
Volatility % Change—Past 30 Days	0.4071 ***	−0.0647 ***	−0.0607 ***	0.0561
	(0.1063)	(0.0144)	(0.0136)	
Volatility Spread	−0.3161 **	0.0356 ***	0.0332 ***	0.0539
	(0.1479)	(0.0133)	(0.0124)	
Volatility Skew	−0.3141	0.0746 *	0.0703 *	0.0197
	(0.2449)	(0.0414)	(0.0388)	
Call Volatility Smirk (ATM–OTM)	−0.2928 *	0.0603 ***	0.0568 ***	0.0329
	(0.1736)	(0.0212)	(0.0199)	
Call Volatility Smirk (ITM–OTM)	−0.0045	0.0011 *	0.0011 *	0.0171
	(0.0040)	(0.0006)	(0.0005)	
Put Volatility Smirk (ATM–OTM)	0.2967	−0.0809 **	−0.0762 **	0.0244
	(0.2488)	(0.0403)	(0.0378)	
Put Volatility Smirk (ITM–OTM)	0.0063 *	−0.0014 **	−0.0013 **	0.0388
	(0.0036)	(0.0006)	(0.0005)	

We sorted the sample into two subsamples according to institutional ownership. Firms with a percentage of institutional ownership above (below) the sample median are considered high (low) institutional ownership. This table reports the regression results for these two subsamples using monthly aggregated market observations. Explanatory variables are overconfidence measures, using option turnovers from all options, and controlling for market volatility, idiosyncratic risk of the underlying stock, and proportional effective spread. Specifically, the overconfidence measures, OC1 and OC2, are (1) residuals from ordinary least square regressions; and (2) overconfidence measures from stochastic frontier analysis (SFA), respectively. Dependent variables are volatility measures and price discrepancy measures. The volatility spread is from Cremers and Weinbaum (2010), while the volatility skew is from Xing et al. (2010). Numbers in parentheses are standard errors. The adjusted R^2 values are the averages of the corresponding values for the three regressions. *, **, and *** indicate statistical significance at 10%, 5%, and 1% levels, respectively.

4.2. Cross-Sectional Analysis

As suggested in Cremers and Weinbaum (2010) and Xing et al. (2010), differences in implied volatility may predict future equity returns. While informed traders, as shown in both studies, may well be the driving force in the findings, we wanted to explore whether there might be an alternative explanation. Unlike some demand-based trading activity measures used in studies such as Pan and Poteshman (2006), option turnover ratios are publicly available information. It would be challenging to argue that informed traders are fully accountable for the predictability of volatility spread/skew/smirk if the volatility patterns are directly tied to observable trading activities. Therefore, we conducted a set of simple tests to examine if there was a cross-sectional connection between volatility patterns and trading activities.

First, we sorted the sample firms into deciles based on monthly average trading turnover and calculated the volatility patterns for each decile. All of the volatility patterns for each decile were weighted based on open interest. Table 6 depicts various trading measures, including all (calls and puts) option turnover, call option turnover, put option turnover, O/S ratio, and O/S ratio in USD value (DOS). Regardless of which trading measure is used, we observe a monotonic pattern on volatility spread across trading deciles, where more heavily traded portfolios have a more negative volatility spread. Also, the differences in volatility spread between the most and the least active portfolios are statistically significant across all measures.

Table 6. Cross-Sectional analyses—trading activities against price discrepancy measures.

Variables	0	1	2	3	4	5	6	7	8	9	*Diff*	*t*-Stat
Panel A: All Option Turnover												
VS	−0.0060	−0.0069	−0.0068	−0.0074	−0.0072	−0.0078	−0.0078	−0.0080	−0.0086	−0.0091	−0.0031	−6.08
IV_Skew	0.0460	0.0466	0.0448	0.0441	0.0433	0.0433	0.0430	0.0449	0.0454	0.0489	0.0029	0.93
Smirk_C_OA	0.0020	0.0044	0.0061	0.0071	0.0073	0.0083	0.0089	0.0085	0.0091	0.0092	0.0071	3.44
Smirk_C_OI	0.0311	0.0328	0.0324	0.0337	0.0332	0.0344	0.0352	0.0360	0.0371	0.0392	0.0082	2.47
Smirk_P_OA	−0.0327	−0.0355	−0.0345	−0.0343	−0.0340	−0.0341	−0.0341	−0.0349	−0.0358	−0.0365	−0.0038	−1.28
Smirk_P_OI	−0.0254	−0.0292	−0.0289	−0.0301	−0.0312	−0.0327	−0.0333	−0.0344	−0.0350	−0.0362	−0.0108	−2.96
Return	−0.0079	−0.0014	0.0018	0.0053	0.0081	0.0102	0.0112	0.0137	0.0175	0.0210	0.0319	0.0398	5.60
Future_Return	0.0133	0.0114	0.0105	0.0093	0.0093	0.0102	0.0097	0.0088	0.0093	0.0089	0.0078	−0.0056	−0.83
Panel B: Call Option Turnover												
VS	−0.0060	−0.0068	−0.0070	−0.0071	−0.0075	−0.0077	−0.0079	−0.0081	−0.0083	−0.0092	−0.0032	−6.18
IV_Skew	0.0500	0.0458	0.0452	0.0443	0.0444	0.0443	0.0440	0.0435	0.0454	0.0472	−0.0028	−0.83
Smirk_C_OA	0.0038	0.0056	0.0060	0.0073	0.0076	0.0082	0.0086	0.0088	0.0085	0.0086	0.0048	2.22
Smirk_C_OI	0.0313	0.0327	0.0337	0.0335	0.0340	0.0344	0.0354	0.0358	0.0371	0.0379	0.0065	1.88
Smirk_P_OA	−0.0323	−0.0352	−0.0355	−0.0343	−0.0343	−0.0346	−0.0351	−0.0344	−0.0356	−0.0352	−0.0029	−0.96
Smirk_P_OI	−0.0249	−0.0286	−0.0279	−0.0312	−0.0321	−0.0331	−0.0336	−0.0345	−0.0355	−0.0344	−0.0095	−2.88
Return	−0.0140	−0.0065	−0.0024	0.0014	0.0059	0.0097	0.0147	0.0203	0.0273	0.0450	0.0590	8.49
Future_Return	0.0128	0.0105	0.0099	0.0095	0.0110	0.0102	0.0094	0.0087	0.0089	0.0081	−0.0047	−0.69
Panel C: Put Option Turnover												
VS	−0.0055	−0.0070	−0.0073	−0.0077	−0.0076	−0.0077	−0.0079	−0.0081	−0.0084	−0.0082	−0.0027	−5.27
IV_Skew	0.0428	0.0456	0.0437	0.0437	0.0425	0.0425	0.0435	0.0448	0.0472	0.0502	0.0074	2.44
Smirk_C_OA	0.0005	0.0041	0.0054	0.0069	0.0074	0.0081	0.0082	0.0095	0.0103	0.0102	0.0097	5.34
Smirk_C_OI	0.0293	0.0321	0.0325	0.0326	0.0337	0.0345	0.0354	0.0363	0.0382	0.0404	0.0111	3.77
Smirk_P_OA	−0.0360	−0.0343	−0.0330	−0.0332	−0.0330	−0.0341	−0.0342	−0.0356	−0.0363	−0.0380	−0.0020	−0.71
Smirk_P_OI	−0.0268	−0.0285	−0.0288	−0.0298	−0.0307	−0.0322	−0.0329	−0.0343	−0.0351	−0.0386	−0.0118	−3.01
Return	0.0081	0.0110	0.0130	0.0119	0.0110	0.0128	0.0128	0.0087	0.0067	0.0051	−0.0029	−0.43
Future_Return	0.0128	0.0117	0.0105	0.0112	0.0095	0.0091	0.0086	0.0090	0.0085	0.0081	−0.0048	−0.72
Panel D: O/S Ratio (in the Number of Shares)												
VS	−0.0036	−0.0042	−0.0051	−0.0056	−0.0063	−0.0070	−0.0082	−0.0089	−0.0102	−0.0153	−0.0117	−18.27
IV_Skew	0.0396	0.0429	0.0433	0.0425	0.0428	0.0432	0.0428	0.0432	0.0446	0.0519	0.0123	2.60
Smirk_C_OA	−0.0023	0.0044	0.0043	0.0051	0.0067	0.0069	0.0080	0.0087	0.0093	0.0105	0.0128	4.01
Smirk_C_OI	0.0226	0.0333	0.0304	0.0316	0.0316	0.0324	0.0338	0.0359	0.0380	0.0421	0.0194	4.33
Smirk_P_OA	−0.0251	−0.0334	−0.0361	−0.0312	−0.0335	−0.0337	−0.0338	−0.0343	−0.0352	−0.0383	−0.0132	−2.23
Smirk_P_OI	−0.0266	−0.0215	−0.0237	−0.0282	−0.0285	−0.0302	−0.0315	−0.0338	−0.0356	−0.0384	−0.0118	−2.36
Return	−0.0004	0.0056	0.0076	0.0098	0.0100	0.0106	0.0136	0.0139	0.0147	0.0157	0.0161	2.51
Future_Return	0.0138	0.0119	0.0128	0.0105	0.0104	0.0091	0.0097	0.0085	0.0080	0.0045	−0.0093	−1.43
Panel E: O/S Ratio (in Dollar Value)												
VS	−0.0043	−0.0045	−0.0049	−0.0056	−0.0060	−0.0064	−0.0073	−0.0083	−0.0091	−0.0186	−0.0144	−17.31
IV_Skew	0.0397	0.0404	0.0400	0.0408	0.0402	0.0412	0.0417	0.0428	0.0455	0.0604	0.0206	4.96
Smirk_C_OA	0.0025	0.0045	0.0060	0.0069	0.0074	0.0083	0.0082	0.0094	0.0090	0.0083	0.0059	3.02
Smirk_C_OI	0.0262	0.0295	0.0291	0.0306	0.0306	0.0328	0.0349	0.0373	0.0391	0.0434	0.0172	4.59
Smirk_P_OA	−0.0282	−0.0306	−0.0321	−0.0299	−0.0315	−0.0329	−0.0333	−0.0344	−0.0369	−0.0423	−0.0141	−3.29
Smirk_P_OI	−0.0196	−0.0249	−0.0239	−0.0274	−0.0289	−0.0303	−0.0331	−0.0355	−0.0367	−0.0376	−0.0179	−4.03
Return	0.0024	0.0052	0.0072	0.0096	0.0110	0.0109	0.0129	0.0135	0.0154	0.0128	0.0105	1.42
Future_Return	0.0123	0.0099	0.0112	0.0118	0.0109	0.0094	0.0095	0.0108	0.0085	0.0049	−0.0074	−1.07

This table summarizes the price discrepancy measures across ten portfolios sorted by trading activities, where the portfolios with larger numbers represent more frequently traded firms. For example, portfolio 9 includes the sample firms whose average option turnovers fall within the top decile. All of the numbers are the averages of the corresponding variable over time. Each panel includes the analysis using a specific trading activity measure. Option turnovers are trading volumes over open interests. O/S ratios are option trading volumes relative to trading volumes of the underlying equity. Diff is the difference between portfolio 0 and portfolio 9. VS is the weighted average difference in implied volatility between paired call and put options with the same strike price, as in Cremers and Weinbaum (2010). IV_Skew is the weighted average difference in implied volatility between OTM and ATM put options, as in Xing et al. (2010). Smirk_C_OA is the difference in implied volatility between ATM call options and OTM call options. Smirk_C_OI is the difference in implied volatility between ITM and OTM call options. Smirk_P_OA is the difference in implied volatility between ATM and OTM put options. Smirk_C_OA is the difference in implied volatility between ITM and OTM put options. Paired differences are used to derive *t*-statistics.

In addition to volatility spread, we also find a pattern suggesting that option traders tend to be more active in trading stocks with better performance during the same time span. We find this by examining the Return variable, which is the monthly return during the month in which firms are sorted based on trading turnover. The above phenomenon is especially prominent for the trading of call options. In both Panels A and B, the difference in concurrent returns between the highest trading turnover decile and the lowest one is statistically significant, with the portfolio with the highest trading turnover earning better return than the one with the lowest trading turnover. This phenomenon suggests that option traders tend to chase "hot" firms in the options market.

According to the above findings, one can conclude that option traders are more active when the volatility spread is low and the underlying stock performs well. However, the subsequent returns on the portfolios with more active trading activities are not any better. Future_Return is the raw monthly return for the same portfolio over the subsequent calendar month; it shows a decreasing trend from the lowest trading decile to the highest one. However, the difference between the top and bottom deciles is not statistically significant.

As discussed above, the pattern of more active trading associated with better concurrent equity returns is mainly driven by call-option traders. The pattern appears in Panels A (all options) and B (call options), but not in Panel C (put options). Two implications may be derived from this finding. First, it is consistent with the general expectation that put options are used for hedging, and therefore, the trading activities of put options are not correlated with recent equity performance. Second, it supports the investor overconfidence hypothesis, in that call option traders are more active when the underlying equities are performing well on average. Note that our analysis here differs from Chen and Sabherwal (2019), as we are examining the characteristics of heavily traded options.

A positive relationship between trading turnover and underlying stock returns is less likely because of informed trading. Given that the short sale constraint is more of an issue in the equity market, investors who hold private information and expect future performance of certain stocks to be bad should tend to take advantage of their private information in the put-options market.

The above does not appear to be the case, however. It is rather difficult to argue that this finding captures investors' accurate forecasts if this pattern only applies to call option trading. If call options are being used for momentum or contrarian strategies, the pattern is inconsistent with the negative (but insignificant) relationship between trading turnover and future returns. Consequently, this finding makes investor overconfidence a more plausible explanation.

Another candidate explanation is the disposition effect. If investors tend to hold on to their losing stakes while liquidating winning ones, the supply of in-the-money options may increase, while that of out-of-the-money options decreases. The phenomenon should lead to less-expensive ITM call options and more-expensive OTM call options. Again, this does not appear to be the case, as Smirk_C_OA and Smirk_C_OI are positively correlated with trading turnover. These two variables measure the relative expensiveness between ATM/ITM options and OTM options, and larger figures mean more expensive ATM/ITM options relative to OTM ones. Therefore, the figures show that heavily traded call options generally have more expensive ATM/ITM options than OTM options. This is not consistent with the disposition-effect hypothesis.

It is worth noting that put option turnover and O/S ratio are positively correlated with implied volatility skew (IV_Skew), which is consistent with the argument that investors tend to utilize out-of-the-money put options to protect their investments in the underlying equity market and therefore make OTM put options more expensive.

Although this study does not rebalance portfolios in a way similar to Cremers and Weinbaum (2010) and Xing et al. (2010), we do consider trading activity and future stock performance. Panels A, B, and C do not show any significant patterns in future stock returns, despite the significant pattern found in volatility spread (VS). Nevertheless, Panels D and E, which use the O/S ratio to capture option trading activities, show some predictability

of future equity performance. In Panel D, the O/S ratio is based on the number of shares. It is negatively and significantly correlated with VS, and also negatively correlated with future stock returns. These findings suggest that when the options market is more active than its underlying equity market, the underlying equity tends to have worse performance in the future. This result is consistent with Cremers and Weinbaum (2010) and Xing et al. (2010). However, the direct connection between the O/S ratio and future equity returns may suggest that the options market reveals better information than does the underlying equity market. In addition, although we still find that option investors tend to pursue stocks with higher concurrent returns, this tendency is not as strong as in Panel B. In Panel E, the O/S ratio is based on the USD value of shares. Panel E shows the same pattern as Panel D.

The O/S ratio can be considered a measure of the focus of investors on the options market relative to the equity market, where a higher O/S ratio means more focus on the options market. Since a more active options market predicts worse future equity returns, we may conclude from our findings above that the options market reacts faster to negative signals. This is not inconsistent with the observations from Panels A through C that option traders might have difficulty processing positive signals as indicated by stronger recent performance.

Two potential factors could be driving the findings above, namely, underlying risks and liquidity. To examine whether these factors explain the findings above, Table 7 has analysis similar to that of Table 6, but controls for the above factors. We first ran time-series regressions of option turnovers and O/S ratios against the return volatility of the underlying equity, the proportional effective spread of options, and the illiquidity measure proposed by Amihud (2002) for each sample firm, and then extracted residuals from the regressions. According to Gopalan et al. (2012), this measure is highly skewed, and they use its square root version (p. 342). We also used the same adjusted measure. Then we sorted the sample into deciles according to the excess trading activities captured by OLS residuals.

Table 7. Cross-sectional analyses—excess trading activities against price discrepancy measures.

Variables	0	1	2	3	4	5	6	7	8	9	Diff	*t*-Stat
Panel A: All Option Turnover												
VS	−0.0065	−0.0065	−0.0065	−0.0066	−0.0065	−0.0066	−0.0067	−0.0068	−0.0070	−0.0073	−0.0008	−1.75
IV_SKEW	0.0432	0.0418	0.0427	0.0416	0.0414	0.0423	0.0418	0.0416	0.0427	0.0450	0.0017	0.67
SMIRK_C_OA	0.0104	0.0092	0.0089	0.0097	0.0092	0.0089	0.0089	0.0092	0.0090	0.0086	−0.0018	−1.11
SMIRK_C_OI	0.0366	0.0357	0.0367	0.0357	0.0353	0.0359	0.0359	0.0351	0.0366	0.0371	0.0005	0.17
SMIRK_P_OA	−0.0340	−0.0330	−0.0352	−0.0338	−0.0341	−0.0346	−0.0343	−0.0337	−0.0343	−0.0354	−0.0014	−0.63
SMIRK_P_OI	−0.0327	−0.0348	−0.0346	−0.0345	−0.0354	−0.0339	−0.0342	−0.0338	−0.0340	−0.0330	−0.0003	−0.09
Return	0.0083	0.0067	0.0057	0.0082	0.0085	0.0085	0.0125	0.0149	0.0169	0.0237	0.0154	2.30
Future_Return	0.0130	0.0116	0.0101	0.0117	0.0109	0.0122	0.0105	0.0101	0.0118	0.0102	−0.0028	−0.45
Panel B: Call Option Turnover												
VS	−0.0066	−0.0067	−0.0067	−0.0064	−0.0067	−0.0065	−0.0066	−0.0066	−0.0070	−0.0072	−0.0006	−1.30
IV_SKEW	0.0419	0.0432	0.0413	0.0428	0.0431	0.0424	0.0417	0.0423	0.0426	0.0434	0.0015	0.61
SMIRK_C_OA	0.0107	0.0098	0.0099	0.0092	0.0096	0.0092	0.0087	0.0088	0.0084	0.0082	−0.0026	−1.63
SMIRK_C_OI	0.0369	0.0355	0.0361	0.0356	0.0366	0.0364	0.0362	0.0358	0.0359	0.0359	−0.0010	−0.31
SMIRK_P_OA	−0.0327	−0.0338	−0.0341	−0.0344	−0.0354	−0.0354	−0.0346	−0.0342	−0.0342	−0.0339	−0.0012	−0.54
SMIRK_P_OI	−0.0328	−0.0324	−0.0345	−0.0353	−0.0353	−0.0353	−0.0347	−0.0339	−0.0343	−0.0316	0.0012	0.34
Return	0.0057	0.0041	0.0050	0.0051	0.0057	0.0085	0.0126	0.0135	0.0224	0.0316	0.0259	3.99
Future_Return	0.0129	0.0108	0.0110	0.0099	0.0111	0.0117	0.0117	0.0102	0.0132	0.0096	−0.0033	−0.55

Table 7. *Cont.*

Panel C: Put Option Turnover												
Variables	0	1	2	3	4	5	6	7	8	9	Diff	*t*-Stat
VS	−0.0054	−0.0067	−0.0066	−0.0068	−0.0068	−0.0068	−0.0067	−0.0073	−0.0071	−0.0072	−0.0018	−5.13
IV_SKEW	0.0460	0.0433	0.0417	0.0416	0.0410	0.0406	0.0411	0.0413	0.0430	0.0468	0.0008	1.18
SMIRK_C_OA	0.0078	0.0088	0.0085	0.0084	0.0087	0.0091	0.0095	0.0102	0.0098	0.0106	0.0028	2.20
SMIRK_C_OI	0.0345	0.0358	0.0355	0.0346	0.0359	0.0355	0.0360	0.0362	0.0368	0.0390	0.0045	1.93
SMIRK_P_OA	−0.0354	−0.0349	−0.0336	−0.0336	−0.0337	−0.0331	−0.0339	−0.0337	−0.0350	−0.0369	−0.0015	−0.67
SMIRK_P_OI	−0.0330	−0.0334	−0.0348	−0.0338	−0.0347	−0.0341	−0.0343	−0.0339	−0.0332	−0.0358	−0.0028	−0.62
Return	0.0159	0.0174	0.0149	0.0132	0.0122	0.0132	0.0077	0.0098	0.0047	0.0035	−0.0124	−1.84
Future_Return	0.0143	0.0131	0.0107	0.0120	0.0112	0.0094	0.0098	0.0102	0.0102	0.0107	−0.0037	−0.59
Panel D: O/S Ratio (in the Number of Shares)												
Variables	0	1	2	3	4	5	6	7	8	9	Diff	*t*-Stat
VS	−0.0075	−0.0063	−0.0056	−0.0057	−0.0050	−0.0055	−0.0061	−0.0068	−0.0073	−0.0111	−0.0035	−6.84
IV_SKEW	0.0387	0.0391	0.0381	0.0406	0.0410	0.0434	0.0429	0.0448	0.0450	0.0495	0.0108	4.35
SMIRK_C_OA	0.0108	0.0098	0.0090	0.0091	0.0076	0.0083	0.0077	0.0085	0.0095	0.0100	−0.0008	−0.52
SMIRK_C_OI	0.0364	0.0348	0.0350	0.0348	0.0344	0.0336	0.0334	0.0355	0.0379	0.0411	0.0047	1.41
SMIRK_P_OA	−0.0317	−0.0316	−0.0312	−0.0332	−0.0337	−0.0347	−0.0344	−0.0362	−0.0367	−0.0392	−0.0075	−3.38
SMIRK_P_OI	−0.0341	−0.0336	−0.0319	−0.0317	−0.0340	−0.0322	−0.0333	−0.0331	−0.0363	−0.0387	−0.0046	−1.47
Return	0.0125	0.0135	0.0098	0.0087	0.0078	0.0097	0.0109	0.0121	0.0130	0.0157	0.0031	0.49
Future_Return	0.0136	0.0134	0.0133	0.0120	0.0116	0.0096	0.0088	0.0095	0.0087	0.0117	−0.0019	−0.29
Panel E: O/S Ratio (in USD Value)												
Variables	0	1	2	3	4	5	6	7	8	9	Diff	*t*-Stat
VS	−0.0082	−0.0064	−0.0058	−0.0055	−0.0049	−0.0051	−0.0058	−0.0063	−0.0072	−0.0117	−0.0035	−5.44
IV_SKEW	0.0414	0.0391	0.0388	0.0395	0.0397	0.0405	0.0421	0.0429	0.0442	0.0530	0.0116	4.23
SMIRK_C_OA	0.0105	0.0092	0.0096	0.0093	0.0091	0.0081	0.0075	0.0081	0.0093	0.0098	−0.0007	−0.44
SMIRK_C_OI	0.0376	0.0355	0.0345	0.0337	0.0339	0.0317	0.0333	0.0353	0.0382	0.0427	0.0050	1.40
SMIRK_P_OA	−0.0341	−0.0314	−0.0323	−0.0326	−0.0324	−0.0324	−0.0324	−0.0346	−0.0369	−0.0414	−0.0072	−2.98
SMIRK_P_OI	−0.0359	−0.0347	−0.0341	−0.0328	−0.0318	−0.0298	−0.0315	−0.0329	−0.0355	−0.0386	−0.0027	−0.78
Return	0.0161	0.0123	0.0102	0.0098	0.0094	0.0086	0.0107	0.0122	0.0123	0.0121	−0.0039	−0.52
Future_Return	0.0140	0.0126	0.0126	0.0113	0.0108	0.0096	0.0092	0.0102	0.0102	0.0116	−0.0024	−0.35

This table summarizes the price discrepancy measures across ten portfolios sorted by excess trading activities while controlling for stock return volatility, proportional effective spread of options, and effective spread of underlying equity. The portfolios with large numbers represent firms with more excess-trading activities. For example, portfolio 9 includes the sample firms whose excess option trading measures fall in the top decile in the sample. All of the numbers are the averages of the corresponding variable over time. Each panel includes the analysis using a specific trading activity measure. Option turnovers are trading volumes over open interests. O/S ratios are option trading volumes relative to trading volumes of the underlying equity. Diff is the difference between portfolio 0 and portfolio 9. Paired differences are used to derive *t*-statistics.

At first sight, all five measures have less explanatory power cross-sectionally, except for volatility spread. Again, O/S ratios are positively correlated with volatility skew. However, the statistical significance is consumed by the control variables. It is intuitive to argue that the shift from the equity market to the options market is due to liquidity in corresponding markets, especially when it comes to the processing of negative information. Again, pricing negative information more efficiently in the equity market than in the options market might be relatively tricky. The illiquidity measures in both markets may well account for the difference and therefore consume the predictability. However, the finding that call option traders pursue "hot" stocks but do not predict future performance in the underlying equity market remains intact despite less-significant results.

In sum, Tables 6 and 7 generally support the investor overconfidence hypothesis. Although we also find some evidence supporting informed trading, it is more likely to be due to greater liquidity in the options market relative to the underlying equity market.

To further investigate the role of liquidity in options trading, we performed a double sorting by trading activities and liquidity in Table 8. The model in Easley et al. (1998) suggests that informed traders are more likely to trade in the options market when the liquidity of the options market is high.

Table 8. Cross-sectional analyses—double sorting by trading activities and liquidity measure.

		Panel A: Option Turnover as Trading Measure						
1: Volatility Spread		**Option Turnover**						
		1	**2**	**3**	**4**	**5**	**Diff**	*t*-**Stat**
	1	−0.0046	−0.0050	−0.0058	−0.0055	−0.0074	−0.0028	−3.15
	2	−0.0072	−0.0072	−0.0074	−0.0076	−0.0078	−0.0005	−0.86
Illiquidity	**3**	−0.0074	−0.0082	−0.0085	−0.0086	−0.0095	−0.0022	−3.42
	4	−0.0073	−0.0080	−0.0080	−0.0089	−0.0095	−0.0022	−4.16
	5	−0.0088	−0.0077	−0.0074	−0.0080	−0.0090	−0.0002	−0.18
	Diff	−0.0042	−0.0027	−0.0017	−0.0025	−0.0016		
	t-**Stat**	−4.40	−4.46	−3.14	−4.39	−1.98		
2: Volatility Skew		**Option Turnover**						
		1	**2**	**3**	**4**	**5**	**Diff**	*t*-**Stat**
	1	0.0585	0.0622	0.0567	0.0581	0.0676	0.0090	1.36
	2	0.0512	0.0489	0.0485	0.0498	0.0544	0.0031	0.92
Illiquidity	**3**	0.0428	0.0454	0.0457	0.0465	0.0499	0.0072	2.57
	4	0.0436	0.0413	0.0414	0.0432	0.0466	0.0030	1.10
	5	0.0401	0.0382	0.0380	0.0395	0.0427	0.0026	1.04
	Diff	−0.0185	−0.0240	−0.0187	−0.0186	−0.0249		
	t-**Stat**	−3.73	−6.83	−4.35	−4.05	−4.88		
3: Call Volatility Smirk (ATM–OTM)		**Option Turnover**						
		1	**2**	**3**	**4**	**5**	**Diff**	*t*-**Stat**
	1	−0.0096	−0.0053	−0.0046	−0.0055	−0.0040	0.0056	1.40
	2	0.0035	0.0021	0.0020	0.0018	−0.0001	−0.0036	−1.87
Illiquidity	**3**	0.0055	0.0077	0.0077	0.0071	0.0057	0.0002	0.11
	4	0.0079	0.0102	0.0106	0.0107	0.0102	0.0023	1.38
	5	0.0126	0.0129	0.0128	0.0132	0.0144	0.0018	1.07
	Diff	0.0222	0.0182	0.0174	0.0187	0.0184		
	t-**Stat**	6.43	7.79	7.03	8.11	6.92		
4: Call Volatility Smirk (ITM–OTM)		**Option Turnover**						
		1	**2**	**3**	**4**	**5**	**Diff**	t-**Stat**
	1	0.0311	0.0274	0.0245	0.0284	0.0338	0.0027	0.60
	2	0.0310	0.0312	0.0280	0.0298	0.0302	−0.0009	−0.28
Illiquidity	**3**	0.0318	0.0324	0.0332	0.0333	0.0328	0.0010	0.34
	4	0.0369	0.0342	0.0358	0.0360	0.0380	0.0011	0.33
	5	0.0363	0.0370	0.0381	0.0404	0.0437	0.0075	2.08
	Diff	0.0052	0.0096	0.0136	0.0120	0.0100		
	t-**Stat**	1.37	2.99	3.91	3.16	2.31		
5: Put Volatility Smirk (ATM–OTM)		**Option Turnover**						
		1	**2**	**3**	**4**	**5**	**Diff**	*t*-**Stat**
	1	−0.0402	−0.0431	−0.0398	−0.0514	−0.0468	−0.0066	−1.02
	2	−0.0361	−0.0353	−0.0364	−0.0379	−0.0397	−0.0035	−1.18
Illiquidity	**3**	−0.0343	−0.0351	−0.0346	−0.0345	−0.0363	−0.0020	−0.85
	4	−0.0357	−0.0330	−0.0332	−0.0327	−0.0357	0.0000	0.01
	5	−0.0306	−0.0311	−0.0319	−0.0325	−0.0349	−0.0044	−1.70
	Diff	0.0096	0.0121	0.0080	0.0189	0.0118		
	t-**Stat**	1.81	3.67	2.03	4.48	2.66		

Table 8. *Cont.*

6: Put Volatility Smirk (ITM–OTM)		Option Turnover						
		1	2	3	4	5	Diff	*t*-Stat
	1	−0.0207	−0.0184	−0.0232	−0.0278	−0.0275	−0.0069	−0.82
	2	−0.0226	−0.0247	−0.0229	−0.0242	−0.0195	0.0031	0.61
Illiquidity	3	−0.0281	−0.0279	−0.0302	−0.0301	−0.0290	−0.0009	−0.32
	4	−0.0364	−0.0328	−0.0337	−0.0332	−0.0353	0.0011	3.42
	5	−0.0370	−0.0344	−0.0363	−0.0396	−0.0427	−0.0057	−1.47
	Diff	−0.0163	−0.0160	−0.0132	−0.0118	−0.0151		
	t-Stat	−2.45	−3.98	−2.63	−1.97	−2.39		
Panel B: O/S Ratio as Trading Measure								
1: Volatility Spread		O/S Ratio						
		1	2	3	4	5	Diff	*t*-Stat
	1	−0.0032	−0.0048	−0.0067	−0.0121	−0.0211	−0.0179	−7.99
	2	−0.0046	−0.0053	−0.0071	−0.0105	−0.0200	−0.0155	−12.96
Illiquidity	3	−0.0044	−0.0055	−0.0068	−0.0097	−0.0186	−0.0142	−13.62
	4	−0.0044	−0.0056	−0.0067	−0.0081	−0.0141	−0.0097	−12.85
	5	−0.0042	−0.0059	−0.0066	−0.0075	−0.0100	−0.0058	−4.74
	Diff	−0.0010	−0.0011	0.0001	0.0045	0.0111		
	t-Stat	−0.86	−1.94	0.18	4.48	4.96		
2: Volatility Skew		O/S Ratio						
		1	2	3	4	5	Diff	*t*-Stat
	1	0.0541	0.0510	0.0606	0.0660	0.0878	0.0337	4.44
	2	0.0446	0.0446	0.0476	0.0530	0.0652	0.0206	5.11
Illiquidity	3	0.0410	0.0438	0.0428	0.0454	0.0586	0.0176	4.17
	4	0.0370	0.0378	0.0398	0.0416	0.0498	0.0128	2.88
	5	0.0300	0.0365	0.0360	0.0378	0.0425	0.0125	2.31
	Diff	−0.0241	−0.0145	−0.0246	−0.0282	−0.0453		
	t-Stat	−3.58	−3.69	−7.55	−5.53	−7.01		
3: Call Volatility Smirk (ATM–OTM)		O/S Ratio						
		1	2	3	4	5	Diff	*t*-Stat
	1	−0.0048	−0.0025	−0.0024	−0.0060	−0.0195	−0.0147	−2.89
	2	0.0033	0.0029	0.0022	0.0026	−0.0014	−0.0046	−2.02
Illiquidity	3	0.0071	0.0068	0.0074	0.0073	0.0051	−0.0020	−0.94
	4	0.0084	0.0088	0.0099	0.0107	0.0104	0.0019	0.98
	5	0.0066	0.0095	0.0120	0.0124	0.0145	0.0079	1.14
	Diff	0.0114	0.0120	0.0145	0.0184	0.0340		
	t-Stat	1.50	4.09	6.42	6.45	8.66		
4: Call Volatility Smirk (ITM–OTM)		O/S Ratio						
		1	2	3	4	5	Diff	*t*-Stat
	1	0.0340	0.0291	0.0308	0.0283	0.0285	−0.0055	−0.89
	2	0.0271	0.0305	0.0298	0.0305	0.0283	0.0013	0.38
Illiquidity	3	0.0291	0.0302	0.0321	0.0341	0.0337	0.0045	1.26
	4	0.0337	0.0321	0.0340	0.0362	0.0392	0.0055	1.48
	5	0.0332	0.0315	0.0321	0.0370	0.0444	0.0113	1.25
	Diff	−0.0008	0.0023	0.0014	0.0087	0.0160		
	t-Stat	−0.09	0.77	0.42	2.24	2.81		

Table 8. *Cont.*

5: Put Volatility Smirk (ATM–OTM)		O/S Ratio						
		1	2	3	4	5	Diff	*t*-Stat
	1	−0.0370	−0.0490	−0.0459	−0.0433	−0.0419	−0.0049	−0.63
	2	−0.0369	−0.0356	−0.0372	−0.0365	−0.0406	−0.0037	−0.94
Illiquidity	3	−0.0319	−0.0327	−0.0339	−0.0360	−0.0371	−0.0052	−1.48
	4	−0.0282	−0.0285	−0.0310	−0.0337	−0.0366	−0.0083	−2.19
	5	−0.0255	−0.0290	−0.0276	−0.0306	−0.0355	−0.0100	−2.39
	Diff	0.0115	0.0200	0.0184	0.0127	0.0064		
	t-Stat	1.87	4.45	5.11	2.75	0.99		

6: Put Volatility Smirk (ITM–OTM)		O/S Ratio						
		1	2	3	4	5	Diff	*t*-Stat
	1	−0.0272	−0.0269	−0.0259	−0.0179	−0.0047	0.0225	2.36
	2	−0.0257	−0.0265	−0.0251	−0.0228	−0.0176	0.0082	1.58
Illiquidity	3	−0.0245	−0.0276	−0.0283	−0.0313	−0.0268	−0.0023	−0.52
	4	−0.0378	−0.0278	−0.0315	−0.0345	−0.0357	0.0021	0.43
	5	−0.0131	−0.0298	−0.0311	−0.0360	−0.0429	−0.0298	−3.10
	Diff	0.0141	−0.0029	−0.0052	−0.0181	−0.0382		
	t-Stat	1.29	−0.69	−1.10	−3.35	−4.77		

This table summarizes the price discrepancy measures across 25 portfolios sorted independently by trading activities and liquidity, where the portfolios with larger numbers represent more-frequently-traded and more illiquid firms. For example, Portfolio 5, 5 includes the sample firms whose average option turnovers fall in the top quintile and whose options are the least liquid in the sample. All of the numbers are the averages of the corresponding variable over time. Each panel concludes the analysis using a specific trading activity measure. Option turnovers are trading volumes over open interests. O/S ratios are relative option trading volumes over trading volumes of the underlying equity. Proportional effective spread is used as the liquidity measure. Diff is the difference between Portfolio 1 and Portfolio 5. Paired differences are used to derive *t*-statistics.

Interestingly, after controlling for liquidity, option turnover only explains differences in volatility spread, and only to a much lower degree in volatility smirk. On the other hand, we find that options with higher liquidity tend to have a higher volatility spread and higher volatility skew. It is widely accepted that informed traders may actively trade on put options due to short-sale constraints in the equity market. The finding that higher volatility skew is associated with higher liquidity in both Panels A and B supports the argument. It is somewhat confusing to see a positive correlation between liquidity and volatility spread, controlling for option turnover, as volatility spread and volatility skew predict the opposite direction of future stock returns. In Panel B, when the O/S ratio is used as a trading measure, the results from volatility spread and volatility skew reconcile, especially for firms with more heavily traded options. This finding is consistent with Roll et al. (2010), who argue that O/S indicates informed trading. It is even more interesting to see the relative expensiveness of ATM and OTM call options in Panel B. OTM call options are more expensive for firms with higher O/S ratios and more liquid options. Consistent with the investor overconfidence theory, OTM call options become more expensive when overconfident agents create higher demand for them.

4.3. Momentum and Contrarain Strategies

Some may argue that an explanation for our above findings is that investors are conducting momentum or contrarian strategies in the options market. To further investigate this possibility, we sorted the sample into deciles based on the past one month's return on the underlying equity. If momentum or contrarian strategies are the main driving forces, we should observe a tendency in which the top and bottom deciles exhibit more activity, while the middle deciles are less active. In other words, the trading activities should present a U-shaped pattern across deciles.

We found that the above pattern exists, as described in Table 9. All five turnover measures (TO_O, TO_C, TO_P, OS, and DOS) exhibit similar patterns, especially OS. In addition, the current month's return and the one-month forward return reverse from the

previous month. That is, the top performers have lower average rates of returns in the following two months, while the bottom performers have higher average rates of returns.

Table 9. Momentum Portfolios.

Variable	0	1	2	3	4	5	6	7	8	9	Diff	*t*-Stat
VS	−0.0073	−0.0065	−0.0067	−0.0068	−0.0070	−0.0070	−0.0071	−0.0076	−0.0083	−0.0116	−0.0043	−6.07
IV_SKEW	0.0551	0.0463	0.0438	0.0418	0.0422	0.0417	0.0424	0.0415	0.0430	0.0505	−0.0045	−1.54
IV_SMIRK_C_OA	0.0045	0.0071	0.0078	0.0084	0.0083	0.0085	0.0088	0.0086	0.0090	0.0075	0.0030	1.69
IV_SMIRK_C_OI	0.0400	0.0362	0.0349	0.0343	0.0343	0.0343	0.0340	0.0342	0.0344	0.0355	−0.0044	−1.26
IV_SMIRK_P_OA	−0.0435	−0.0368	−0.0348	−0.0326	−0.0331	−0.0324	−0.0322	−0.0324	−0.0335	−0.0372	0.0062	2.55
IV_SMIRK_P_OI	−0.0357	−0.0328	−0.0307	−0.0308	−0.0321	−0.0309	−0.0323	−0.0330	−0.0336	−0.0353	0.0004	0.11
Return	0.0121	0.0128	0.0107	0.0104	0.0106	0.0103	0.0105	0.0072	0.0077	0.0085	−0.0036	−0.42
Future_Return	0.0138	0.0127	0.0120	0.0118	0.0100	0.0088	0.0091	0.0068	0.0067	0.0076	−0.0062	−0.73
TO_O	3.6789	3.7161	3.6557	3.6003	3.6092	3.6525	3.7487	3.9153	4.0726	4.5460	0.8671	9.63
TO_C	3.8779	3.8660	3.7970	3.8022	3.8743	3.9382	4.0091	4.2475	4.4557	4.9205	1.0426	6.76
TO_P	3.7581	3.7082	3.7853	3.6492	3.5696	3.5657	3.5870	3.7105	3.8437	4.2072	0.4492	4.59
OS	0.0754	0.0709	0.0671	0.0674	0.0657	0.0693	0.0691	0.0732	0.0776	0.0901	0.0147	5.38
DOS	0.0090	0.0058	0.0051	0.0052	0.0046	0.0052	0.0052	0.0054	0.0061	0.0085	−0.0006	−1.43
CP	4.3839	5.2058	5.0695	5.1324	5.6455	5.6580	6.3074	5.9968	5.8670	6.9154	2.5315	3.74
L1RET	−0.1760	−0.0843	−0.0498	−0.0252	−0.0037	0.0167	0.0388	0.0654	0.1037	0.2215	0.3976	37.75

The table shows summary statistics for various variables in each of the momentum portfolios. The sample firms are sorted into ten portfolios based on the past month's rate of return in the underlying stocks. The average past one month return for each portfolio is shown in row L1RET. TO_O, TO_C, and TO_P are option turnovers for all options, call options, and put options, respectively, and are defined as trading volume over open interest. O/S ratios are option trading volumes relative to trading volumes of the underlying equity. OS is based on the number of shares and DOS is based on the USD value of shares. Diff is the difference between portfolio 0 and portfolio 9. VS is the weighted average difference in implied volatility between paired call and put options with the same strike price, as in Cremers and Weinbaum (2010). IV_Skew is the weighted average difference in implied volatility between OTM and ATM put options, as in Xing et al. (2010). Smirk_C_OA is the difference in implied volatility between ATM call options and OTM call options. Smirk_C_OI is the difference in implied volatility between ITM and OTM call options. Smirk_P_OA is the difference in implied volatility between ATM and OTM put options. Smirk_C_OA is the difference in implied volatility between ITM and OTM put options.

While above findings in Table 9 suggest that contrarian strategies may be one of the reasons for the previous findings, the differences between the middle and bottom deciles are trivial. For example, the difference in option turnover (TO_O) between the worst performers (portfolio 0) and the decile with the lowest turnover rate (portfolio 3) is only approximately 0.07. On the other hand, the difference in the same measure between the top performers (portfolio 9) and the worst performers (portfolio 0) is 0.8671, and the difference is statistically significant. This suggests that while momentum or contrarian strategies might be an explanation for the phenomena, their contributions are not substantial. Also, significantly higher trading activities among the past top performers further strengthen the investor overconfidence argument, in that traders are pursuing "hot" stocks but do not seem to succeed much. The subsequent returns show a negative relationship with trading activities, but the relationship does not have statistical support (the differences are not statistically significant).

4.4. Interactions between Options and Stock Markets

Although this paper focuses on the trading activities in the options market and their potential impact on option prices, it is worthwhile to look at the underlying stock market. In the last analysis, OS is the most influential indicator among all trading activity measures. Since many stock market investors also trade in the options market, it is not surprising to see stock-market trading activities correlated with option pricing in certain ways. To examine the extent to which stock trading behaviors affect both option trading and option pricing, we conducted a cross-sectional analysis similar to that in Section 4.2, using both one-way and two-way sorting.

Table 10 summarizes the empirical results. In Panel A, sample firms are sorted into deciles based solely on stock trading turnover, defined as the stock trading volume divided by the number of shares outstanding. By comparing Panel A in Tables 6 and 10, we find similar patterns across all rows. However, there are a few distinctions.

Table 10. Cross-sectional analyses—interactions between stock and options markets.

	Panel A: Stock Trading Turnover											
Variables	0	1	2	3	4	5	6	7	8	9	Diff	*t*-Stat
VS	−0.0068	−0.0061	−0.0064	−0.0068	−0.0064	−0.0066	−0.0071	−0.0075	−0.0085	−0.0139	−0.0071	−3.31
IV_SKEW	0.0401	0.0384	0.0384	0.0401	0.0411	0.0425	0.0433	0.0451	0.0481	0.0567	0.0166	6.41
SMIRK_C_OA	0.0072	0.0074	0.0080	0.0072	0.0078	0.0075	0.0085	0.0077	0.0079	0.0090	0.0018	1.09
SMIRK_C_OI	0.0288	0.0306	0.0304	0.0288	0.0320	0.0340	0.0344	0.0359	0.0383	0.0448	0.0160	4.77
SMIRK_P_OA	−0.0269	−0.0280	−0.0287	−0.0269	−0.0320	−0.0333	−0.0352	−0.0358	−0.0379	−0.0434	−0.0165	−7.52
SMIRK_P_OI	−0.0233	−0.0266	−0.0281	−0.0233	−0.0314	−0.0316	−0.0324	−0.0331	−0.0357	−0.0411	−0.0178	−5.81
Return	0.0041	0.0068	0.0087	0.0041	0.0103	0.0121	0.0119	0.0119	0.0143	0.0112	0.0072	0.89
Future_Return	0.0085	0.0085	0.0100	0.0085	0.0099	0.0118	0.0099	0.0092	0.0109	0.0107	0.0022	0.30

	Panel B: Double Sorting By Stock Turnover and Option Turnover						
1: Volatility Spread		Option Turnover					
		1	2	3	4	5	Diff	*t*-Stat
	1	−0.0057	−0.0066	−0.0068	−0.0065	−0.0069	−0.0012	−1.95
	2	−0.0058	−0.0063	−0.0067	−0.0066	−0.0070	−0.0012	−2.18
Stock	3	−0.0065	−0.0064	−0.0067	−0.0066	−0.0064	0.0000	0.03
Turnover	4	−0.0073	−0.0071	−0.0071	−0.0076	−0.0074	0.0000	−0.07
	5	−0.0110	−0.0110	−0.0106	−0.0109	−0.0118	−0.0008	−0.64
	Diff	−0.0053	−0.0044	−0.0038	−0.0044	−0.0049	−0.0061	
	t-Stat	−4.59	−5.55	−6.26	−8.37	−7.44		−10.47

This table summarizes the price discrepancy measures across portfolios sorted by stock market trading activities, where the portfolios with larger numbers represent more frequently traded firms. All of the numbers are the averages of the corresponding variable over time. In Panel A, all the monthly observations are sorted into ten portfolios based on stock trading turnover. In Panel B, they are sorted independently, based on option turnover and stock turnover. Diff is the difference between the most- and least-frequently traded portfolios (portfolio 9–portfolio 0 in single sorting, and portfolio 5–portfolio 1 in double sorting). VS is the weighted average difference in implied volatility between paired call and put options with the same strike price, as in Cremers and Weinbaum (2010). IV_Skew is the weighted average difference in implied volatility between OTM and ATM put options, as in Xing et al. (2010). Smirk_C_OA is the difference in implied volatility between ATM call options and OTM call options. Smirk_C_OI is the difference in implied volatility between ITM and OTM call options. Smirk_P_OA is the difference in implied volatility between ATM and OTM put options. Smirk_C_OA is the difference in implied volatility between ITM and OTM put options. Paired differences are used to derive *t*-statistics.

First, the differences between the most- and least-frequently traded portfolios in all option pricing measures are statistically significant in Table 10, except for the volatility smirk between OTM and ATM call options. In Table 6, the differences in volatility skewness and in volatility smirk between OTM and ATM put options are not statistically significant. Second, the concurrent returns across portfolios increase monotonically with trading frequency in Table 6, but this phenomenon does not appear in Table 10. In addition, the t-test suggests no significant difference in contemporaneous return between the most and the least frequently traded portfolios in Table 10. These findings may be due to the use of momentum or contrarian strategy in the options market. From Table 6, we may attribute this finding more to the momentum traders, as ATM calls tend to be more expensive in the portfolio with more frequent option trading. However, it is less so in Table 10. Instead, a much steeper volatility skew for the most frequently traded portfolio in Table 10 suggests that OTM put options are much more expensive. Looking at Panel B in Table 10, we also find that stocks with less frequent option trading drive the steeper volatility skew. This conflicts with the notion in Xing et al. (2010) that informed traders use OTM put options to take advantage of negative information, but it is more in line with the investor overconfidence hypothesis of the stock market.

5. Discussion

This paper examines the relationship between trading activities and option pricing patterns. If investor overconfidence causes heavier trading activities, the option pricing patterns should strongly correlate with trading activities. Furthermore, market volatility should also be positively correlated with the trading activity. We present evidence showing that both relationships do exist. The relationships hold both over time and cross-sectionally.

The negative relationship between volatility spread and trading activity suggests that options traders are contrarians overall. The supporting evidence is also provided, by sorting the sample into deciles based on past equity returns. However, the findings also

suggest that the differences in trading activities and volatility spread and volatility skew do not predict future equity returns.

Our findings in this study differ from those of Cremers and Weinbaum (2010) and Xing et al. (2010) regarding the predictability of volatility spread and volatility skew, and therefore serve as evidence against theories of informed trading or superior information in the options market. Instead, our findings support the investor overconfidence theory in that options traders also tend to pursue top performers, strengthening the argument in Chen and Sabherwal (2019) that the positive relationship between past market return and option trading activities may be due to investor overconfidence.

This study adds to the discussion in the literature regarding the role played by behavioral biases. While the debate between efficient market advocates and behavioral finance supporters is still active in the equity market, this paper extends the debate to the options market. This focus on the options market not only provides insights to market speculators trying to exploit opportunities in the options market, but also serves as a caution to investors who heavily hedge their portfolios in the equity options market. If behavioral biases play an important role in the options market, the effectiveness of using options to hedge equity portfolios might be degraded. This study shows that options traders should pay attention to the behavioral patterns in the options market regardless of their purposes in trading. However, in this study, we focus on the generalized patterns using market-wide data, which may limit our interpretation of the empirical results. While retail investors may exhibit a higher degree of behavioral biases (Choy 2015; Baig et al. 2022; Ülkü et al. 2023), we do not attempt to differentiate between the sources of trading (retail versus institutional) in this study. We leave to future research the enhancement of our understanding of retail investors' role in option trading. Also, this study does not focus on the COVID-19 period, during which retail investor participation played a particularly important role. It would be interesting to examine the role of behavioral biases in the options market exclusively during this period.

References

Amihud, Yakov. 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5: 31–56. [CrossRef]

Amin, Kaushik I., and Charles M. C. Lee. 1997. Option Trading and Earnings New Dissemination. *Contemporary Accounting Research* 43: 153–92. [CrossRef]

Baig, Ahmed S., Benjamin M. Blau, Hassan A. Butt, and Awaid Yasin. 2022. Do Retail Traders Destabilize Financial Markets? An Investigation Surrounding the COVID-19 Pandemic. *Journal of Banking and Finance* 144: 106627. [CrossRef] [PubMed]

Barberis, Nicholas, Andrei Shleifer, and Robert Vishny. 1998. A Model of Investor Sentiment. *Journal of Financial Economics* 49: 307–43. [CrossRef]

Battalio, Robert, and Paul Schultz. 2006. Options and the Bubble. *Journal of Finance* 61: 2071–102. [CrossRef]

Battese, George E., and Timothy J. Coelli. 1988. Prediction of Firm-Level Technical Efficiencies with a Generalized Frontier Production Function and Panel Data. *Journal of Econometrics* 38: 387–99. [CrossRef]

Black, Fischer. 1975. Fact and Fantasy in the Use of Options. *Financial Analysts Journal* 31: 36–41, 61–72. [CrossRef]

Black, Fischer, and Myron Scholes. 1973. The Pricing of Options and Corporate Liabilities. *Journal of Political Economy* 81: 637–54. [CrossRef]

Cao, Charles, Zhiwu Chen, and John M. Griffin. 2005. Informational Content of Option Volume Prior to Takeovers. *Journal of Business* 78: 1073–109. [CrossRef]

Chakravarty, Sugato, Huseyin Gulen, and Stewart Mayhew. 2004. Informed Trading in Stock and Option Markets. *Journal of Finance* 59: 1235–57. [CrossRef]

Chan, Kalok, Y. Peter Chung, and Herb Johnson. 1993. Why Option Prices Lag Stock Prices: A Trading-Based Explanation. *Journal of Finance* 48: 1957–67. [CrossRef]

Chan, Kalok, Y. Peter Chung, and Wai-Ming Fong. 2002. The Information Role of Stock and Option Volume. *Review of Financial Studies* 15: 1049–75. [CrossRef]

Chen, Han-Sheng, and Sanjiv Sabherwal. 2019. Overconfidence among Options Traders. *Review of Financial Economics* 37: 61–91. [CrossRef]

Choy, Siu-Kai. 2015. Retail Clientele and Option Returns. *Journal of Banking and Finance* 51: 26–42. [CrossRef]

Choy, Siu-Kai, and Jason Wei. 2012. Option Trading: Information or Differences of Opinion. *Journal of Banking and Finance* 36: 2299–322. [CrossRef]

Cox, John C., Stephen A. Ross, and Mark Rubinstein. 1979. Option Pricing: A Simplified Approach. *Journal of Financial Economics* 7: 229–63. [CrossRef]

Cremers, Martijn, and Ankur Pareek. 2014. Short-term Trading and Stock Return Anomalies: Momentum, Reversal, and Share Issuance. *Review of Finance* 19: 1649–701. [CrossRef]

Cremers, Martijn, and David Weinbaum. 2010. Deviations from Put-Call Parity and Stock Return Predictability. *Journal of Financial and Quantitative Analysis* 45: 335–67. [CrossRef]

Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam. 1998. Investors Psychology and Security Market Under- and Over-Reaction. *Journal of Finance* 53: 1839–85. [CrossRef]

DeBondt, Werner F. M., and Richard Thaler. 1985. Does the Stock Market Overreact? *Journal of Finance* 40: 793–805. [CrossRef]

Easley, David, Maureen O'Hara, and P. S. Srinivas. 1998. Option Volume and Stock Prices: Evidence on Where Informed Traders Trade. *Journal of Finance* 53: 431–65. [CrossRef]

George, Thomas J., and Chuan-Yang Hwang. 2004. The 52-Week High and Momentum Investing. *Journal of Finance* 59: 2145–76. [CrossRef]

Gopalan, Radhakrishnan, Ohad Kadan, and Mikhail Pevzner. 2012. Asset Liquidity and Stock Liquidity. *Journal of Financial and Quantitative Analysis* 47: 333–64. [CrossRef]

Grinblatt, Mark, and Matti Keloharju. 2009. Sensational Seeking, Overconfidence, and Trading Activity. *Journal of Finance* 64: 549–78. [CrossRef]

Hong, Harrison, and Jeremy C. Stein. 1999. A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets. *Journal of Finance* 54: 2143–84. [CrossRef]

Jegadeesh, Narasimhan, and Sheridan Titman. 1993. Returns to Buying Winners and Selling Losers: Implications for Market Efficiency. *Journal of Finance* 48: 65–91. [CrossRef]

Jondrow, James, C. A. Knox Lovell, Ivan S. Materov, and Peter Schmidt. 1982. On the Estimation of Technical Inefficiency in the Stochastic Frontier Production Function Model. *Journal of Econometrics* 19: 233–38. [CrossRef]

Karpoff, Jonathan. 1987. The Relation between Price Changes and Trading Volume: A Survey. *Journal of Financial and Quantitative Analysis* 22: 109–26. [CrossRef]

Lee, Charles M. C., and Bhaskaran Swaminathan. 2000. Price Momentum and Trading Volume. *Journal of Finance* 55: 2017–69. [CrossRef]

Muravyev, Dmitriy, Neil D. Pearson, and John Paul Broussard. 2013. Is There Price Discovery in Equity Options? *Journal of Financial Economics* 107: 259–83. [CrossRef]

Ni, Xiaoyan. 2007. Stock Option Returns: A Puzzle. Available online: https://ssrn.com/abstract=959024 (accessed on 15 January 2013).

Ofek, Eli, Matthew Richardson, and Robert F. Whitelaw. 2004. Limited Arbitrage and Short Sales Restrictions: Evidence from the Options Markets. *Journal of Financial Economics* 74: 305–42. [CrossRef]

Pan, Jun, and Allen M. Poteshman. 2006. The Information in Option Volume for Future Stock Prices. *Review of Financial Studies* 19: 871–908. [CrossRef]

Roll, Richard, Eduardo Schwartz, and Avanidhar Subrahmanyam. 2010. O/S: The Relative Trading Activity in Options and Stock. *Journal of Financial Economics* 96: 1–17. [CrossRef]

Scheinkman, José A., and Wei Xiong. 2003. Overconfidence and Speculative Bubbles. *Journal of Political Economy* 111: 1183–220. [CrossRef]

Statman, Meir, Steven Thorley, and Keith Vorkink. 2006. Investor Overconfidence and Trading Volume. *Review of Financial Studies* 19: 1531–65. [CrossRef]

Stephen, Jens A., and Robert E. Whaley. 1990. Intraday Price Change and Trading Volume Relations in the Stock and Stock Option Markets. *Journal of Finance* 45: 191–220. [CrossRef]

Ülkü, Numan, Fahad Ali, Saidgozi Saydumarov, and Deniz Ikizlerli. 2023. COVID Caused a Negative Bubble. Who Profited? Who Lost? How Stock Markets Changed? *Pacific-Basin Finance Journal* 79: 102044. [CrossRef]

Vijh, Anand M. 1990. Liquidity of the CBOE Equity Options. *Journal of Finance* 45: 1157–79. [CrossRef]

Xing, Yuhang, Xiaoyan Zhang, and Rui Zhao. 2010. What Does the Individual Option Volatility Smirk Tell Us about Future Equity Returns? *Journal of Financial and Quantitative Analysis* 45: 641–62. [CrossRef]

MDPI