econometrics

Recent Advances in Theory and Methods for the Analysis of High Dimensional and High Frequency Financial Data

Edited by
Norman R. Swanson and Xiye Yang

Printed Edition of the Special Issue Published in *Econometrcis*

www.mdpi.com/journal/econometrics

MDPI

Recent Advances in Theory and Methods for the Analysis of High Dimensional and High Frequency Financial Data

Recent Advances in Theory and Methods for the Analysis of High Dimensional and High Frequency Financial Data

Editors

Norman R. Swanson
Xiye Yang

MDPI • Basel • Beijing • Wuhan • Barcelona • Belgrade • Manchester • Tokyo • Cluj • Tianjin

Editors

Norman R. Swanson
Rutgers University
USA

Xiye Yang
Rutgers University
USA

Editorial Office
MDPI
St. Alban-Anlage 66
4052 Basel, Switzerland

This is a reprint of articles from the Special Issue published online in the open access journal *Econometrics* (ISSN 2225-1146) (available at: https://www.mdpi.com/journal/econometrics/special_issues/High_Dimensional_Frequency_Data).

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

LastName, A.A.; LastName, B.B.; LastName, C.C. Article Title. *Journal Name* **Year**, *Volume Number*, Page Range.

ISBN 978-3-0365-0852-8 (Hbk)
ISBN 978-3-0365-0853-5 (PDF)

Contents

About the Editors

Norman R. Swanson was educated at the University of Waterloo and the University of California, San Diego. He is currently a Professor in the Economics Department at Rutgers University. He has held previous positions at Pennsylvania State University (1994–1999), Texas A&M University (1999–2001), and Purdue University (2001–2002). His primary research interests include econometric theory, financial and macro econometrics, time series analysis, and forecasting. He is a fellow of the *Journal of Econometrics*, which is the top field journal in econometrics, and he is currently an associate editor of the *Journal of Business and Economic Statistics*, the *International Journal of Forecasting*, and the *European Journal of Pure and Applied Mathematics*. He is, or has served, as guest editor for journals ranging from *Journal of Econometrics* to *Journal of Business and Economic Statistics*, and has previously acted as an associate editor to a number of journals, including *Studies in Nonlinear Dynamics and Econometrics* and *Empirical Economics*. He is a member of various professional organizations, including the Econometric Society, the American Statistical Association, the American Economic Association, and the Canadian Economic Association. He has recent scholarly publications in leading economics and statistics journals including *Econometrica*, *Journal of Econometrics*, *Review of Economics and Statistics*, *Journal of Business and Economic Statistics*, and the *Journal of the American Statistical Association*, among others; and is, or has acted as, ongoing visiting scholar to various central banks, including The Federal Reserve Bank of Philadelphia and the Bank of Canada.

Xiye Yang an Assistant Professor in the Department of Economics at Rutgers University. Prior to joining Rutgers, Xiye received his PhD from the University of Amsterdam (2015) and MPhil from Tinbergen Institute (2011). In 2009, Xiye graduated with a BA in Economics and a BSc in Mathematics and Applied Mathematics from Peking University, China. His research interests lie in the areas of econometric theory, financial econometrics, asset pricing, and empirical finance.

Preface to "Recent Advances in Theory and Methods for the Analysis of High Dimensional and High Frequency Financial Data"

Recently, considerable attention has been placed on the development and application of tools useful for the analysis of the high-dimensional and/or high-frequency datasets that now dominate the landscape. The purpose of this Special Issue is to collect both methodological and empirical papers that develop and utilize state-of-the-art econometric techniques for the analysis of such data.

Norman R. Swanson, Xiye Yang
Editors

econometrics

MDPI

Article

Using the Entire Yield Curve in Forecasting Output and Inflation

Eric Hillebrand [1], Huiyu Huang [2], Tae-Hwy Lee [3,*] and Canlin Li [4]

1 Aarhus University and CREATES, Fuglesangs Allé 4, 8210 Aarhus V, Denmark; ehillebrand@creates.au.dk
2 ICBC Credit Suisse Asset Management, Beijing 100033, China; huang.huiyu@icbccs.com.cn
3 Department of Economics, University of California, Riverside, CA 92521, USA
4 Monetary and Financial Market Analysis Section, Division of Monetary Affairs, Federal Reserve Board, Washington, DC 20551, USA; canlin.li@frb.gov
* Correspondence: taelee@ucr.edu

Received: 17 June 2018; Accepted: 21 August 2018; Published: 29 August 2018

Abstract: In forecasting a variable (forecast target) using many predictors, a factor model with principal components (PC) is often used. When the predictors are the yield curve (a set of many yields), the Nelson–Siegel (NS) factor model is used in place of the PC factors. These PC or NS factors are combining information (CI) in the predictors (yields). However, these CI factors are not "supervised" for a specific forecast target in that they are constructed by using only the predictors but not using a particular forecast target. In order to "supervise" factors for a forecast target, we follow Chan et al. (1999) and Stock and Watson (2004) to compute PC or NS factors of many forecasts (not of the predictors), with each of the many forecasts being computed using one predictor at a time. These PC or NS factors of forecasts are combining forecasts (CF). The CF factors are supervised for a specific forecast target. We demonstrate the advantage of the supervised CF factor models over the unsupervised CI factor models via simple numerical examples and Monte Carlo simulation. In out-of-sample forecasting of monthly US output growth and inflation, it is found that the CF factor models outperform the CI factor models especially at longer forecast horizons.

Keywords: level, slope, and curvature of the yield curve; Nelson-Siegel factors; supervised factor models; combining forecasts; principal components

JEL Classification: C5; E4; G1

1. Introduction

The predictive power of the yield curve for macroeconomic variables has been documented in the literature for a long time. Many different points on the yield curve have been used and various methodologies have been examined. For example, Stock and Watson (1989) find that two interest rate spreads, the difference between the six-month commercial paper rate and the six-month Treasury bill rate, and the difference between the ten-year and one-year Treasury bond rates, are good predictors of real activity, thus contributing to their index of leading indicators. Bernanke (1990), Friedman and Kuttner (1993), Estrella and Hardouvelis (1991), and Kozicki (1997), among many others, have investigated a variety of yields and yield spreads individually on their ability to forecast macroeconomic variables. Hamilton and Kim (2002) as well as Diebold et al. (2005) provide a brief summary of this line of research and the link between the yield curve and macroeconomic variables.

Various macroeconomic models for exploring the yield curve information for real activity prediction are proposed. Ang and Piazzesi (2003) and Piazzesi (2005) study the role of macroeconomic variables in an arbitrage-free affine yield curve model. Estrella (2005) constructs an analytical rational expectations model to investigate the reasons for the success of the slope of the yield curve (the spread

between long-term and short-term government bond rates) in predicting real economic activity and inflation. The model in Ang et al. (2006), Piazzesi and Wei is an arbitrage-free dynamic model (using lags of GDP growth and yields as regressors) that characterizes expectations of GDP growth. Rudebusch and Wu (2008) provide an example of a macro-finance specification that employs more macroeconomic structure and includes both rational expectations and inertial elements.

Stock and Watson (1999, 2002) investigate forecasts of output growth and inflation using over a hundred of economic indicators, including many interest rates and yield spreads. Stock and Watson (2002, 2012) advocate methods that aim at solving the large-N predictor problem, particularly those using principal components (PC). Ang et al. (2006) suggest the use of the short rate, the five-year to three-month yield spread, and lagged GDP growth in forecasting GDP growth out-of-sample. The choice of these two yield curve characteristics, as they argue, is because they have almost one-to-one correspondence with the first two principal components of the short rate and five yield spreads that account for 99.7% of quarterly yield curve variation.

Alternatively to the PC factor approach on the large-N predictor information set, Diebold and Li (2006) propose the Nelson and Siegel (1987) (NS) factors for the large-N yields. They use a modified three-factor NS model to capture the dynamics of the yield curve and show that the three NS factors may be interpreted as level, slope, and curvature. Diebold et al. (2006) examine the correlations between NS yield factors and macroeconomic variables. They find that the level factor is highly correlated with inflation and that the slope factor is highly correlated with real activity. For more on the yield curve background and the three characteristics of the yield curve, see Litterman and Scheinkman (1991) and Diebold and Li (2006).

In this paper, we utilize the yield curve information for prediction of macro-economic variables. Using a large number of yield curve points with different maturities yields a large-N problem in the predictive regression. The PC factors or the NS factors of the yield curve may be used to reduce the large dimension of the predictors. However, the PC and NS factors of the yield curve are not supervised for a specific variable to forecast. These factors simply combine information (CI) of many predictors (yields) without having to look at a forecast target. Hence, the conventional CI factor models (using factors of the predictors) are unsupervised for any forecast target.

Our goal in this paper is to consider factor models where the factors are computed with a particular forecast target in mind. Specifically, we consider the PC or NS factors of forecasts (not of predictors), with each of the forecasts formed using one predictor at a time. (It could be generalized to make each forecast from using more than one predictor, e.g., a subset of the N predictors, in which case there can be as many as 2^N forecasts to combine.) These factors will combine the forecasts (CF). The PC factors of forecasts are combined forecasts using the combining weights that solves a singular value problem for a set of forecasts, while the NS factors of forecasts are combined forecasts using the combining weights obtained from orthogornal polynomials that emulate the shape of a yield curve (in level, slope, and curvature). The PC or NS factors of the many forecasts are supervised for a forecasting target. The main idea of the CF-factor model is to focus on the space spanned by *forecasts* rather than the space spanned by *predictors*. The factorization of forecasts (CF-factor model) can substantially improve forecasting performance compared to the factorization of predictors (CI-factor model). This is because the CF-factor model takes the forecast target into the factorization, while the conventional CI-factor model is blind to the forecast target because the factorization uses only information on predictors.

For both CI and CF schemes, the NS factor model can be relevant only when the yield curve is used as predictors while the PC factor model can be used in general. The NS factors are specific to the yield curve factors such as level, slope, and curvature factors. When the predictors are from the points on the yield curve, the NS factor models proposed here is nearly the same as the PC factors. Given the similarity of NS and PC and the generality of PC, we begin the paper with the PC models to understand the mechanism of the supervision in CF-factor models. We demonstrate how the supervised CF factor models outperform the unsupervised CI factor model, under the presence of many predictors (50 points on the yield curve at each time). The empirical work shows that there are

potentially big gains in the CF-factor models. In out-of-sample forecasting of U.S. monthly output growth and inflation, it is found that the CF factor models (CF-NS and CF-PC) are substantially better than the conventional CI factors models (CI-NS and CI-PC). The advantage of supervised factors is even greater for longer forecast horizons.

The paper is organized as follows: in Section 2, we describe the CI and CF frameworks and principal component approaches for their estimation, present theoretical results about supervision, and an example to provide intuition. Section 3 provides simulations of supervision under different noise, predictor correlation, and predictor persistence conditions. In Section 4, we introduce the NS component approaches for the CI and CF frameworks. In Section 5, we show the out-of-sample performance of the proposed methods in forecasting U.S. monthly output growth and inflation. Section 6 presents the conclusions.

2. Supervising Factors

2.1. Factor Models

Let y_{t+h} denote the variable to be forecast (output growth or inflation) using yield curve information stamped at time t, where h denotes the forecast horizon. The predictor vector x_t contains information about the yield curve at various maturities: $x_t := (x_{1t},\ x_{2t},\ \ldots,\ x_{Nt})'$, where $x_{it} := x_t(\tau_i)$ denotes the yield at time t with maturity τ_i $(i = 1, 2, \ldots, N)$.

Consider the CI model when N is large

$$y_{t+h} = (1\ x_t')\alpha + \varepsilon_{t+h}, \qquad (t = 1, 2, \ldots, T) \tag{1}$$

for which the forecast at time T is

$$\hat{y}_{T+h}^{\text{CI-OLS}} = (1\ x_T')\hat{\alpha}, \tag{2}$$

with $\hat{\alpha}$ estimated by OLS using the information up to time T. A problem is that here the mean-squared forecast error (MSFE) is of order $O\left(\frac{N}{T}\right)$ increasing with N.[1] A solution to this problem is to reduce the dimension either by selecting a subset of the N predictors, e.g., by Lasso type regression (Tibshirani 1996) or by using factor models of, e.g., Stock and Watson (2002). In this paper, we focus on using the factor model rather than selecting a subset of the N predictors.[2]

2.1.1. CI-Factor Model

The conventional factor model is the CI factor model for x_t of the form

$$x_t = \Lambda_{\text{CI}} f_{\text{CI},t} + v_{\text{CI},t}, \qquad (t = 1, \ldots, T), \tag{3}$$

where Λ_{CI} is $N \times k_{\text{CI}}$ and $f_{\text{CI},t}$ is $k_{\text{CI}} \times 1$. The estimated factor loadings $\hat{\Lambda}_{\text{CI}}$ are obtained either by following Stock and Watson (2002) and Bai (2003), or by following Nelson and Siegel (1987) and Diebold and Li (2006). The latter approach is discussed in Section 4. The factors are then estimated by

$$\hat{f}_{\text{CI},t} = \hat{\Lambda}_{\text{CI}}' x_t. \tag{4}$$

As this model computes the factors from all N predictors of x_t directly, it will be called "CI-factor". The forecast $\hat{y}_{T+h} = (1\ \hat{f}_{\text{CI},T}')\hat{\alpha}_{\text{CI}}$ can be formed using $\hat{\alpha}_{\text{CI}}$ estimated at time T from the regression

$$y_t = (1\ \hat{f}_{\text{CI},t-h}')\alpha_{\text{CI}} + u_{\text{CI},t}, \qquad (t = h + 1, \ldots, T). \tag{5}$$

---

[1] This is explained in Bai and Ng 2008; Huang and Lee 2010; Stock and Watson 2002.

[2] Bai and Ng (2008) consider CI factor models with a selected subset (targeted predictors).

In matrix form, we write the factor model (3) and (5) for the vector of forecast target observations y and for the $T \times N$ matrix of predictors X as follows:[3]

$$X = F_{\text{CI}}\Lambda_{\text{CI}}' + v_{\text{CI}}, \tag{6}$$
$$y = F_{\text{CI}}\alpha_{\text{CI}} + u_{\text{CI}}, \tag{7}$$

where y is the $T \times 1$ vector of observations, F_{CI} is a $T \times k_{\text{CI}}$ matrix of factors, Λ_{CI} is an $N \times k_{\text{CI}}$ matrix of factor loadings, α_{CI} is a $k_{\text{CI}} \times 1$ parameter vector, v_{CI} is a $T \times N$ random matrix, and u_{CI} is a $T \times 1$ vector of random errors.

Remark 1. ***(No supervision in CI-factor model):*** *Consider the joint density of* (y_{t+h}, x_t)

$$D(y_{t+h}, x_t; \theta) = D_1(y_{t+h}|x_t; \theta)D_2(x_t; \theta), \tag{8}$$

where D_1 *is the conditional density of* y_{t+h} *given* x_t*, and* D_2 *is the marginal density of* x_t*. The CI-factor model assumes a situation where the joint density operates a "cut" in the terminology of Barndorff-Nielsen (1978) and Engle et al. (1983), such that*

$$D(y_{t+h}, x_t; \theta) = D_1(y_{t+h}|x_t; \theta_1)D_2(x_t; \theta_2), \tag{9}$$

where $\theta = (\theta_1\ \theta_2')'$*, and* $\theta_1 = \alpha$*,* $\theta_2 = (F, \Lambda)'$ *are "variation-free". Under this situation, the forecasting equation in (5) is obtained from the conditional model* D_1 *and the factor equation in (3) is solely obtained from the marginal model* D_2 *of the predictors. The computation of the factors is entirely from the marginal model* D_2 *that is blind to the forecast target* y_{t+h}*.*

While the CI factor analysis of a large predictor matrix X solves the dimensionality problem, it computes the factors using information in X only, without accounting for the variable y to be forecast, and therefore the factors are not supervised for the forecast target. Our goal in this paper is to improve this approach by accounting for the forecast target in the computation of the factors. The procedure will be called *supervision*.

There are some attempts in the literature to supervise factor computation for a given forecast target. For example, Bair et al. (2006) and Bai and Ng (2008) consider factors of selected predictors that are informative for a specified forecast target; Zou et al. (2006) consider sparse loadings of principal components; De Jong (1993) and Groen and Kapetanios (2016) consider partial least squares regression; De Jong and Kiers (1992) consider principal covariate regression; Armah and Swanson (2010) select variables for factor proxies that have the maximum predictive power for the variable being forecast; and some weighted principal components have been used to downweight noisier series.

In this paper, we consider the CF-factor model that computes factors from forecasts rather than from predictors. This approach has been proposed in Chan et al. (1999) and in Stock and Watson (2004), there labeled "principal component forecast combination". We will refer to this approach as CF-PC (combining forecasts principal components). The details are as follows.

---

[3] The suppressed time stamp of y and X captures the h-lag relation for the forecast horizon and we treat the data centered so that we do not include a constant term explicitly in the regression for notational simplicity.

2.1.2. CF-Factor Model

The forecasts from a CF-factor model are computed in two steps. The first step is to estimate the factors of the individual forecasts. Let the individual forecasts be formed by regressing the forecast target y_{t+h} using the ith individual predictor x_{it}:

$$\hat{y}_{T+h}^{(i)} := a_{i,T} + b_{i,T} x_{iT} \qquad (i = 1, 2, \ldots, N). \tag{10}$$

Stack the N individual forecasts into a vector $\hat{\mathbf{y}}_{t+h} := (\hat{y}_{t+h}^{(1)}, \hat{y}_{t+h}^{(2)}, \ldots, \hat{y}_{t+h}^{(N)})'$ and consider a factor model of $\hat{\mathbf{y}}_{t+h}$:

$$\hat{\mathbf{y}}_{t+h} = \Lambda_{\mathrm{CF}} f_{\mathrm{CF},t+h} + v_{\mathrm{CF},t+h}. \tag{11}$$

The CF-factor is estimated from

$$\hat{f}_{\mathrm{CF},t+h} := \hat{\Lambda}_{\mathrm{CF}}' \hat{\mathbf{y}}_{t+h}. \tag{12}$$

The second step is to estimate the forecasting equation (for which the estimated CF-factors from the first step are used as regressors)[4]

$$y_{t+h} = \hat{f}_{\mathrm{CF},t+h}' \alpha_{\mathrm{CF}} + u_{\mathrm{CF},t+h}. \tag{13}$$

Then, the CF-factor forecast at time T is

$$\hat{y}_{T+h}^{\mathrm{CF}} = \hat{f}_{\mathrm{CF},T+h}' \hat{\alpha}_{\mathrm{CF}}, \tag{14}$$

where $\hat{\alpha}_{\mathrm{CF}}$ is estimated. See (Chan et al. 1999; Huang and Lee 2010; Stock and Watson 2004).

To write the CF-factor model in matrix form, we assume for notational simplicity that the data has been centered so that we do not include a constant term. We regress y on the columns x_i of X, $i = 1, \ldots, N$, one at a time, and write the fitted values in (10) as

$$\hat{y}^{(i)} = x_i (x_i' x_i)^{-1} x_i' y =: x_i b_i. \tag{15}$$

Collect the fitted values in the matrix

$$\hat{Y} = [\hat{y}^{(1)}\ \hat{y}^{(2)}\ \cdots\ \hat{y}^{(N)}] := XB \in \mathbb{R}^{T \times N}, \tag{16}$$

where $B = \mathrm{diag}(b_1, \ldots, b_N) \in \mathbb{R}^{N \times N}$ is a diagonal matrix containing the regression coefficients. We call B the *supervision matrix*. Then, the CF-factor model is

$$\hat{Y} = F_{\mathrm{CF}} \Lambda_{\mathrm{CF}}' + v_{\mathrm{CF}}, \tag{17}$$

$$y = F_{\mathrm{CF}} \alpha_{\mathrm{CF}} + u_{\mathrm{CF}}, \tag{18}$$

where F_{CF} is a $T \times k_{\mathrm{CF}}$ matrix of factors of $\hat{Y} = XB$, Λ_{CF} is an $N \times k_{\mathrm{CF}}$ matrix of factor loadings, α_{CF} is an $k_{\mathrm{CF}} \times 1$ parameter vector, v_{CF} is a $T \times N$ random matrix, and u_{CF} is a $T \times 1$ vector of random errors. In the rest of the paper, the subscripts CI and CF may be omitted for simplicity.

We use principal components (PC) as discussed in Stock and Watson (2002), Bai (2003), and Bai and Ng (2006). For the specific case of yield curve data, we use NS components as discussed in Nelson and Siegel (1987) and Diebold and Li (2006). We use both CF and CI approaches together with PC factors and NS factors. Our goal is to show that forecasts using supervised factor models (CF-PC and CF-NS) are better than forecasts from conventional unsupervised factor models (CI-PC and CI-NS).

---

[4] Given the dependent nature of macroeconomic and financial time series, the forecasting equation can be extended to allow the supervision to be based on the relation between y_t and some predictors after controlling for lagged dependent variables and to allow the dynamic factor structure, which we leave for future work.

We show analytically and in simulations how supervision works to improve factor computation with respect to a specified forecast target. In Section 5, we present empirical evidence.

Remark 2. ***(Estimation of** B**):*** *The CF-factor model in (17) and (18) with $B = I_N$ (identity matrix) is a special case when there is no supervision. In this case, the CF-factor model collapses to the CI-factor model. If B were consistently estimated by minimizing the forecast error loss, then the CF-factor model with the "optimal" B would outperform the CI-factor model. However, as the dimension of the supervision matrix B grows with N^2, B is an "incidental parameter" matrix and can not be estimated consistently. See Neyman and Scott (1948) and Lancaster (2000). Any estimation error in B translates into forecast error in the CF-factor model. Whether there is any virtue in considering Bayesian methods of estimating B, while still avoiding this problem, is left for future research. Instead, in this paper, we circumvent this difficulty by imposing that $B = diag(b_1, \ldots, b_N)$ be a diagonal matrix and by estimating the diagonal elements b_i's from the ordinary least squares regression in (10) or (15) with one predictor x_i at a time. The supervision matrix B can be non-diagonal in general. As imposing the diagonality on B may be restrictive, it would be an interesting empirical question to examine if the CF-factor forecast with this restriction and the estimation strategy of B can still outperform the CI-factor forecast with $B = I_N$. Our empirical results in Section 5 (Table 1) support this simple estimation strategy for the diagonal matrix B, in favor of the CF-factor model.*

Remark 3. ***(Combining forecasts with many predictors):*** *It is generally believed that it is difficult to estimate the forecast combination weights when N is large. Therefore, the equal weights $\left(\frac{1}{N}\right)$ have been widely used instead of estimating weights.*[5] *It is often found in the literature that equally-weighted combined forecasts are often the best. Stock and Watson (2004) call this the "forecast combination puzzle". See also Timmermann (2006). Smith and Wallis (2009) explore a possible explanation of the forecast combination puzzle and conclude that it is due to estimation error of the combining weights.*

Now, we note that, in the CF-factor model described above, we can consistently *estimate the combining weights. From the CF-factor forecast (14) and the estimated factor (12),*

$$\hat{y}_{T+h} = \hat{f}'_{CF,T+h}\hat{\alpha}_{CF} = \left(\hat{\mathbf{y}}'_{T+h}\hat{\Lambda}_{CF}\right)\hat{\alpha}_{CF} := \hat{\mathbf{y}}'_{T+h}\hat{w}, \tag{19}$$

where

$$\hat{w} := \hat{\Lambda}_{CF}\hat{\alpha}_{CF} \tag{20}$$

is estimated consistently as long as $\hat{\Lambda}_{CF}$ and $\hat{\alpha}_{CF}$ are estimated consistently.

2.2. Singular Value Decomposition

In this section, we formalize the concept of supervision and explain how it improves factor extraction. We compare the two different approaches CI-PC (Combining Information—Principal Components) and CF-PC (Combining Forecasts—Principal Components) in a linear forecast problem of the time series y given predictor data X. We explain the advantage of the CF-PC approach over CI-PC in Section 2.3 and give some examples in Section 2.4. We explore the advantage of supervision in simulations in Section 3.2. As an alternative to PC factors, we propose the use of NS factors in Section 4.

Principal components of predictors X (CI-PC): Let $X \in \mathbb{R}^{T\times N}$ be a matrix of regressors and let

$$X = R\Sigma W' \in \mathbb{R}^{T\times N} \tag{21}$$

[5] An exception is Wright (2009), who uses Bayesian model averaging (BMA) for pseudo out-of-sample prediction of U.S. inflation, and finds that it generally gives more accurate forecasts than simple equal-weighted averaging. He uses $N = 107$ predictors.

be the singular value decomposition of X, with $\Sigma \in \mathbb{R}^{T\times N}$ diagonal rectangular, that is, diagonal square matrix padded with zero rows below the square if $\min(T,N) = N$ or padded with zero columns next to the square if $\min(T,N) = T$, $R \in \mathbb{R}^{T\times T}$, and $W \in \mathbb{R}^{N\times N}$ is unitary. Write

$$X'X = W\Sigma' R' R\Sigma W' = W\Sigma'\Sigma W', \tag{22}$$

where $\Sigma'\Sigma := \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$ is diagonal and square. Therefore, W contains the eigenvectors of $X'X$. For a matrix $A \in \mathbb{R}^{T\times N}$, denote by $A_k \in \mathbb{R}^{T\times k}$ the matrix consisting of the first $k \leq N$ columns of A. Then, W_k is the matrix containing the singular vectors corresponding to the $k = k_{\text{CI}}$ largest singular values $(\sigma_1, \ldots, \sigma_k)$. The first k principal components are given by

$$F_{\text{CI}} := XW_k = R\Sigma W'W_k = R\Sigma \begin{bmatrix} I_k \\ \mathbf{0} \end{bmatrix} = R\Sigma_k = R_k\Sigma_{kk}, \tag{23}$$

where I_k is the $k \times k$ identity matrix, $\mathbf{0}$ is an $(N-k)\times k$ matrix of zeros, and Σ_{kk} is the $k \times k$ upper-left diagonal block of Σ. Note that the first k principal components F_{CI} of X are constant multiples of columns of R_k as Σ_{kk} is diagonal. The projection (forecast) of y onto F_{CI} is given by

$$\begin{aligned} \hat{y}_{\text{CI-PC}} &:= F_{\text{CI}}(F_{\text{CI}}'F_{\text{CI}})^{-1}F_{\text{CI}}'y = XW_k(W_k'X'XW_k)^{-1}W_k'X'y \\ &= R_k\Sigma_{kk}(\Sigma_{kk}'R_k'R_k\Sigma_{kk})^{-1}\Sigma_{kk}'R_k'y = R_k(R_k'R_k)^{-1}R_k'y = R_kR_k'y, \end{aligned} \tag{24}$$

as $R_k'R_k = I_k$. Therefore, the CI forecast, $\hat{y}_{\text{CI-PC}}$, is the projection of y onto R_k. The CI forecast error and the CI sum of squared error (SSE) are

$$y - \hat{y}_{\text{CI-PC}} = y - R_kR_k'y = (I_T - R_kR_k')y, \tag{25}$$

$$SSE_{\text{CI-PC}} = ||y - \hat{y}_{\text{CI-PC}}||^2 = y'(I_T - R_kR_k')y, \tag{26}$$

as $(I_T - R_kR_k')$ is symmetric idempotent.

Bai (2003) shows that, under general assumptions on the factor and error structure, F_{CI} is a consistent and asymptotically normal estimator of $F_{\text{CI}}H$, where H is an invertible $k \times k$ matrix.[6] This identification problem is also clear from Equation (24), and it conveniently allows us to identify the principal components $F_{\text{CI}} = R_k\Sigma_{kk}$ as $F_{\text{CI}} = R_k$ since Σ_{kk} is diagonal. The principal components are scalar multiples of the first k columns of R. Bai's result shows that principal components can be estimated consistently only up to linear combinations. Bai and Ng (2006) show that the parameter vector α in the forecast equation can be estimated consistently for $\alpha'H^{-1}$ with an asymptotically normal distribution.

Principal components of forecasts $\hat{Y}$ (CF-PC): To generate forecasts in a CF-factor scheme, we regress y on the columns x_i of X, $i = 1, \ldots, N$, one at a time, and calculate the fitted values of (15). Collect the fitted values in the matrix as in (16), with $B = \text{diag}(b_1, \ldots, b_N)$ containing the regression coefficients in its diagonal. Compute the singular value decomposition of $\hat{Y}$:

$$\hat{Y} = S\Theta V', \tag{27}$$

[6] In order for the objects in Bai's (2003) analysis to converge, he introduces scaling such that the singular values are the eigenvalues of the matrix $X'X/T$. Then, the singular vectors are multiplied by $\sqrt{T}$. In our notation, the singular value decomposition becomes $X = \sqrt{T}R\frac{\Sigma}{\sqrt{T}}W'$.

with $\Theta \in \mathbb{R}^{T\times N}$ is diagonal rectangular, and $S \in \mathbb{R}^{T\times T}$, $V \in \mathbb{R}^{N\times N}$ unitary. Pick the first $k = k_{\text{CF}}$ principal components of $\hat{Y}$,

$$F_{\text{CF}} := \hat{Y}V_k = S\Theta V'V_k = S\Theta \begin{bmatrix} I_k \\ \mathbf{0} \end{bmatrix} = S\Theta_k = S_k\Theta_{kk}, \tag{28}$$

where V_k is the $N \times k$ matrix of the singular vectors corresponding to the k largest singular values $(\theta_1, \ldots, \theta_k)$ and Θ_{kk} is the $k \times k$ upper-left diagonal block of Θ. Again, we can identify the estimated k principal components of $\hat{Y}$ with $F_{\text{CF}} = S_k$, where F_{CF} is the $T \times k_{\text{CF}}$ matrix of factors of $\hat{Y}$. The projection (forecast) of y onto F_{CF} is given by:

$$\begin{aligned} \hat{y}_{\text{CF-PC}} &:= F_{\text{CF}}(F'_{\text{CF}}F_{\text{CF}})^{-1}F'_{\text{CF}}y = \hat{Y}V_k(V'_k\hat{Y}'\hat{Y}V_k)^{-1}V'_k\hat{Y}'y \\ &= S_k\Theta_{kk}(\Theta'_{kk}S'_kS_k\Theta_{kk})^{-1}\Theta'_{kk}S'_ky = S_k(S'_kS_k)^{-1}S'_ky = S_kS'_ky \end{aligned} \tag{29}$$

as $S'_kS_k = I_k$. The CF forecast, $\hat{y}_{\text{CF-PC}}$, is the projection of y onto S_k. The CF forecast error and the CF SSE are

$$y - \hat{y}_{\text{CF-PC}} = y - S_kS'_ky = (I_T - S_kS'_k)y, \tag{30}$$

$$SSE_{\text{CF-PC}} = ||y - \hat{y}_{\text{CF-PC}}||^2 = y'(I_T - S_kS'_k)y, \tag{31}$$

as $(I_T - S_kS'_k)$ is symmetric idempotent.

2.3. Supervision

In this sub-section, we explain the advantage of CF-PC over CI-PC in factor computation. We call the advantage "supervision", which is defined as follows:

Definition 1. *(Supervision). The advantage of CF-PC over CI-PC, called* supervision*, is the selection of principal components according to their contribution to variation in y, as opposed to selection of principal components according to their contribution to variation in the columns of X. This is achieved by selecting principal components from a matrix of forecasts of y.*

We use the following measures of supervision of CF-PC in comparison with CI-PC.

Definition 2. *(Absolute Supervision). Absolute supervision is the difference of the sums of squared errors (SSE) of CI-PC and CF-PC:*

$$s_{abs}(X, y, k_{\text{CI}}, k_{CF}) := ||y - \hat{y}_{\text{CI-PC}}||^2 - ||y - \hat{y}_{\text{CF-PC}}||^2 = y'(S_{k_{CF}}S'_{k_{CF}} - R_{k_{CI}}R'_{k_{CI}})y. \tag{32}$$

Definition 3. *(Relative Supervision). Relative supervision is the ratio of the sums of squared errors of CI-PC over CF-PC:*

$$s_{rel}(X, y, k_{\text{CI}}, k_{CF}) := \frac{||y - \hat{y}_{\text{CI-PC}}||^2}{||y - \hat{y}_{\text{CF-PC}}||^2} = \frac{y'(I_T - R_{k_{CI}}R'_{k_{CI}})y}{y'(I_T - S_{k_{CF}}S'_{k_{CF}})y}. \tag{33}$$

Remark 4. *When $k_{CI} = k_{CF} = N$, there is no room for supervision*

$$s_{abs}(X, y, N, N) = y'(SS' - RR')y = y'(I_T - I_T)y = 0 \tag{34}$$

because $SS' = RR' = I_T$. Relative supervision is defined only for $k_{CF} < N$.

For the sake of simplifying the notation and presentation, we consider the same number of factors in CI and CF factor models with $k_{\text{CI}} = k_{\text{CF}} = k$ for the rest of the paper.

Remark 5. *S_k is a block of a basis change matrix that in the expression $y'S_k$ returns the first k coordinates of y with respect to the new basis. This new basis is the one with respect to which the mapping $\hat{Y}\hat{Y}' = XBBX' = S\Theta\Theta'S'$ becomes diagonal, with singular values in descending order such that the first k columns of S correspond to the k largest singular values. Therefore, $y'S_kS_k'y$ is the sum of the squares of these coordinates. Broadly speaking, the S_k are the k largest components of y in the sense of $\hat{Y}$ and its construction from the single regression coefficients. Thus, $y'S_kS_k'y$ is the sum of the squares of the k coefficients in y that contributes most to the variation in the columns of $\hat{Y}$.*

Analogously, R_k is a block of a basis change matrix that for $y'R_k$ returns the first k coordinates of y with respect to the basis that diagonalizes the mapping $XX' = R\Sigma\Sigma'R'$. Therefore, $y'R_kR_k'y$ is the sum of squares of the k coordinates of y selected according to their contribution to variation in the columns of X.

We emphasize *the factors that explain most of the variation of the columns of X, i.e., the eigenvectors associated with the largest eigenvalues of XX', which are selected in the principal component analysis of X, may have little to do with the factors that explain most of the variation of y, however.* The relation between X and y in the data-generating process can, at worst, completely reverse the order of principal components in the columns of X and in y. We demonstrate this in the following Example 1.

2.4. Example 1

In this subsection, we give a small example to facilitate intuition for the supervision mechanics of CF-PC. Example 1 illustrates how the supervision of factor computation defined in Definition 1 operates. In Example 2 in the next section, we add randomness to Example 1 to explore the effect of stochasticity in a well-understood problem.

Let

$$X = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/4 \\ 0 & 0 & 0 & 1/5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{35}$$

with $T = 6$ and $N = 5$. The singular value decomposition of $X = R\Sigma W$ is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{3} & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{5} \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}. \tag{36}$$

Let

$$y = (1, 2, 3, 4, 5, 0)'. \tag{37}$$

Then, the diagonal matrix B that contains the coefficients of y w.r.t. each column of X is

$$B = \text{diag}(4, 9, 1, 25, 16), \tag{38}$$

and

$$\hat{Y} := XB = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 2 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \tag{39}$$

The singular value decomposition of $\hat{Y} = XB = S\Theta V$ is

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 5 & 0 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}. \tag{40}$$

We set $k_{\text{CI}} = k_{\text{CF}} = k$ and compare CI-PC and CF-PC with the same number of principal components. Recall from (23) that $F_{CI} = R\Sigma_k$ and from (28) that $F_{CF} = S\Theta_k$. The absolute supervision and relative supervision, defined in (32) and (33), are computed for each k :

	$s_{abs}(X, y, k_{\text{CI}}, k_{\text{CF}})$	$s_{rel}(X, y, k_{\text{CI}}, k_{\text{CF}})$
$k = 1$	24	1.8
$k = 2$	36	3.6
$k = 3$	36	8.2
$k = 4$	24	25.0
$k = 5$	0	N/A

See Appendix A for the calculation. The absolute supervision is all positive and the relative supervision is larger than 1 for all $k < N$.

As noted in Remarks 1 and 5, the relation between X and y is crucial. In this example, the magnitude of the components in y is reversed from the order in X. For X, the ordering of the columns of X with respect to the largest eigenvalues of XX' is $\{3, 1, 2, 5, 4\}$. For y, the ordering of the columns of X with respect to the largest eigenvalues of $\hat{Y}\hat{Y}'$ is $\{4, 5, 2, 1, 3\}$. For example, consider the case $k = 2$, i.e., we choose two out of five factors in the principal component analysis. CI-PC, the analysis of X, will pick the columns 3 and 1 of X, that is, the vectors $(1, 0, 0, 0, 0, 0)'$ and $(0, 1/2, 0, 0, 0, 0)'$. These correspond to the two largest singular values 1 and $1/2$ of X. CF-PC, the analysis of $\hat{Y}$, will pick columns 4 and 5 of X, that is, the vectors $(0, 0, 0, 0, 1/5, 0)'$ and $(0, 0, 0, 1/4, 0, 0)'$. These correspond to the two largest singular values 5 and 4 of $\hat{Y}$. The regression coefficients in $B = \text{diag}(4, 9, 1, 25, 16)$ de-emphasize columns 3 and 1 of X and emphasize columns 4 and 5 of X.

3. Monte Carlo

There are several simplifications in the construction of Example 1, which we relax by the following extensions:

(a) Adding randomness makes the estimation of the regression coefficients in B a statistical problem. The sampling errors influence the selection of the components of $\hat{Y}$. (b) Adding correlation among regressors (columns of X) introduces correlation among individual forecasts (columns of $\hat{Y}$), increasing the effect of sampling error in the selection of the components of $\hat{Y}$. (c) Increasing N to realistic magnitudes, in particular in the presence of highly correlated regressors, will increase estimation error in the principal components due to collinearity.

We address the first extension (a) in Example 2. All three extensions (a), (b), (c) are addressed in Example 3 of Section 3.2.

3.1. Example 2

Consider adding some noise to X, y in Example 1. Let v be a $T \times N$ matrix of independent random numbers, each entry distributed as $N(0, \sigma_v^2)$, and u be a vector of independent random numbers, each distributed as $N(0, \sigma_u^2)$. In this example, the new regressor matrix X is the sum of X in Example 1 and the noise term v, and the new y is the sum of y in Example 1 and the noise term u. For simplicity,

we set $\sigma_v = \sigma_u$ in the simulations and let both range from 0.01 to 3. This covers a substantial range of randomness given the magnitude of the numbers in X and y. For each scenario of $\sigma_v = \sigma_u$, we generate 1000 random matrices v and random vectors u and calculate the Monte Carlo average of the sums of squared errors (SSE).

Figure 1 plots the Monte Carlo average of the SSEs for selection of $k = 1$ to $k = 4$ components. For standard deviations $\sigma_v = \sigma_u$ close to zero, the sum of squared errors are as calculated in Example 1. As the noise increases, the advantage of CF over CI decreases but remains substantial, in particular for smaller numbers of principal components. For $k = 5$ estimated components (not shown), the SSEs of CI-PC and CF-PC coincide because $k = N$.

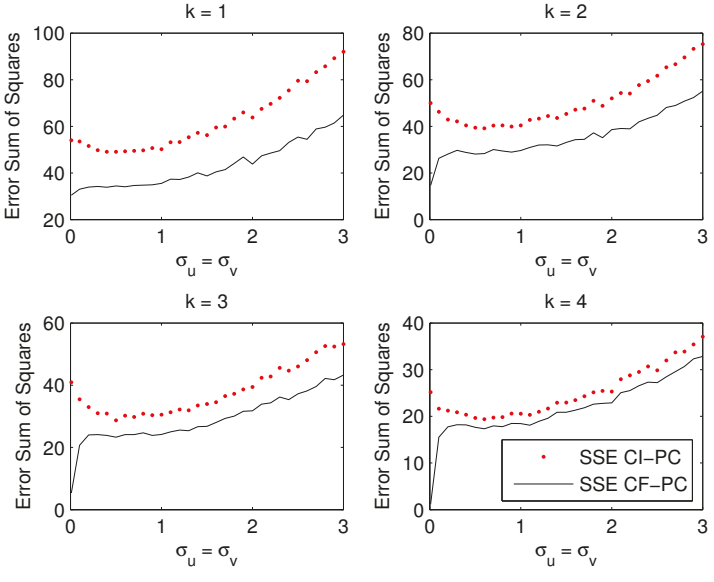


Figure 1. For Example 2. Monte Carlo averages of the sum of squared errors (SSE) against a grid of standard deviations $\sigma_u = \sigma_v$ ranging from 0.01 to 3 in factor and forecast equations, for a selection of $k = 1$ to $k = 4$ components. When the standard deviation is close to zero, the SSE are close to the ones reported in Example 1. With increasing noise, the advantage of CF over CI decreases but remains substantial, in particular for few components. For $k = 5 = N$ (not shown), the SSE of CI-PC and CF-PC coincide, as shown in Remark 4.

3.2. Example 3

We consider the data-generating process (DGP)

$$X = F\Lambda' + v, \tag{41}$$

$$y = F\alpha + u, \tag{42}$$

where y is the $T \times 1$ vector of observations, F is a $T \times r$ matrix of factors, Λ is an $N \times r$ matrix of factor loadings, α is an $r \times 1$ parameter vector, v is a $T \times N$ random matrix, and u is a $T \times 1$ vector of random errors. We set $T = 200$, $N = 50$ and consider $r = 3$ data-generating factors.

Note that, under this DGP, the CI-PC model in Equations (6) and (7) is correctly specified if the correct number of factors is identified, i.e., $k_{\text{CI}} = r$. Even under this DGP, however, an insufficient

number of factors, $k_{\text{CI}} < r$, can still result in an advantage of the CF-PC model over the CI-PC model. We will explore this question in this section.

Factors and persistence: For each run in the simulation, we generate the r factors in F as independent AR(1) processes with zero mean and a normally distributed error with mean zero and variance one:

$$F_{t,i} = \phi F_{t-1,i} + \varepsilon_{t,i},\ t = 2, \ldots, T,\ i = 1, \ldots, r. \tag{43}$$

We consider a grid of 19 different AR(1) coefficients ϕ, equidistant between 0 and 0.90. We consider $r = 3$ data-generating factors and $k \in \{1, 2, 3, 4\}$ estimated factors.

Contemporaneous factor correlation: Given a correlation coefficient ρ for adjacent regressors, the $N \times r$ matrix Λ of factor loadings is obtained from the first r columns of an upper triangular matrix from a Cholesky decomposition of

$$\begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{N-1} \\ \rho & 1 & \rho & \cdots & \rho^{N-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{N-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{N-1} & \rho^{N-2} & \rho^{N-3} & \cdots & 1 \end{bmatrix}. \tag{44}$$

We consider a grid of 19 different values for ρ, equidistant between the points -0.998 and 0.998. In this setup, the 10th value is very close to $\rho = 0$. Then, the covariance matrix of the regressors is given by

$$\mathbb{E}X'X = \mathbb{E}[(\Lambda F' + v')(F\Lambda' + v)] = \Omega_F + \Omega_v, \tag{45}$$

where $\Omega_F = \Lambda\Lambda'$ and $\Omega_v = \mathbb{E}v'v$ is given by the identity matrix in our simulations. The relation $\mathbb{E}F'F = I$ is due to the independence of the factors, but may be subject to substantial finite sample error, in particular for ϕ close to one, for well-known reasons.

Relation of X and y: The $r \times 1$ parameter vector α is drawn randomly from a standard normal distribution for each run in the simulation. This allows α to randomly shuffle which factors are important for y.

Noise level: We set $\sigma_u = \sigma_v$ and let it range between 0.1 and 3 in steps of 0.1. We add the case of 0.01 that essentially corresponds to a deterministic factor model.

For a given number $r = 3$ of data-generating factors, the simulation setup varies along the dimensions ϕ (19 points), k (4 points), ρ (19 points), $\sigma_u = \sigma_v$ (31 points). For every single scenario, we run 1000 simulations and calculate the SSEs of CI-PC and CF-PC, and the relative supervision $s_{rel}(X, y, k, k)$. Then, we take the Monte Carlo average of the SSEs and $s_{rel}(X, y, k, k)$ over the 1000 simulations.[7]

The Monte Carlo results are presented in Figures 2–4. Each figure contains four panels that plot the situation for $k = 1, 2, 3, 4$ estimated number of factors. The main findings from the figures can be summarized as follows:

1. Figure 2: If the number of estimated factors k is below the true number $r = 3$, as shown in top panels, the supervision becomes smaller with increasing noise. If the correct number of factors or

[7] In relation to the empirical application using the yield data in Section 5, we could have calibrated the simulation design to make the Monte Carlo more realistic for the empirical application in Section 5. Nevertheless, our Monte Carlo design covers wide ranges of the parameter values for the noise levels, correlation structures (ρ and ϕ) in the yield data. Figure 2 shows that the supervision is smaller with larger noise levels, which may be rather obvious intuitively. Figure 4 shows that the advantage of supervision when the factors are persistence, which depends on the number of factors k relative to the true number of factors r. Particularly interesting is Figure 3 which shows that the advantage of supervision is smaller when the contemporaneous correlation ρ between predictors is larger, which may be relevant for the yield data because the yields with different maturities may be moderately contemporaneously correlated. We thank a referee for pointing this out.

more are estimated $(k \geq r)$, as in bottom panels, the advantage of supervision increases with the noise level $\sigma_u = \sigma_v$, Even in this case when the CI-PC is the correct model $(k \geq r)$, supervision becomes larger as the noise increases.

2. Figure 3: The advantage of supervision is greatest when the contemporaneous correlation ρ between predictors is minimal. For almost perfect correlation, the advantage of supervision disappears. This is true regardless of whether the correct number of factors is estimated or not. Intuitively, for near-perfect factor correlation, the difference between those factors that explain variation in the columns of X and those that explain variation in $\hat{Y}$ vanishes, and so supervision becomes meaningless.
3. Figure 4: If the correct number of factors or more are estimated $(k \geq r)$, the advantage of supervision decreases with factor persistence ϕ. High persistence induces spurious contemporaneous correlation, and in this sense the situation is related to the result in No. 2. If the number of estimated factors is below the true number of factors $(k < r)$, however, the advantage of supervision increases with factor persistence.

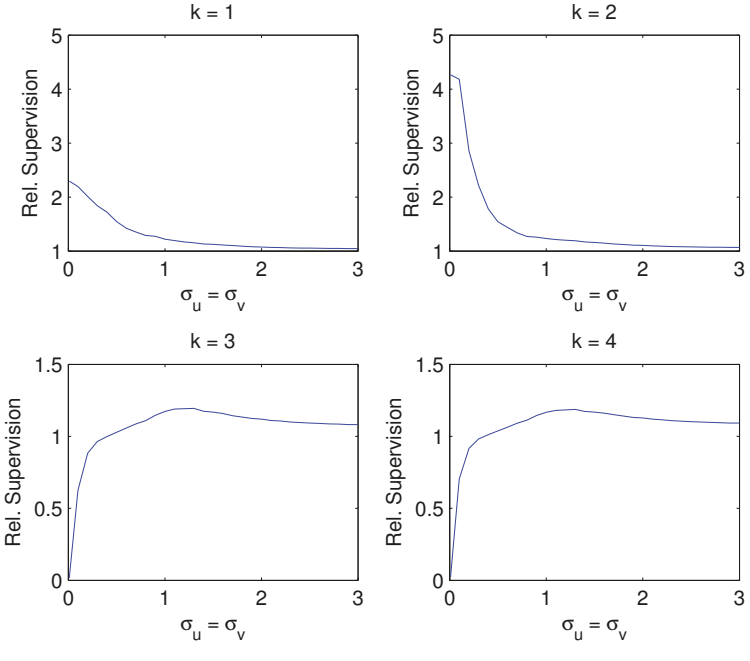


Figure 2. Supervision dependent on noise. Relative supervision against a grid of standard deviations in factor and forecast equation $\sigma_u = \sigma_v$, ranging from 0.01 to 3, while the factor serial correlation is fixed at $\phi = 0$ and the contemporaneous factor correlation is $\rho = 0$.

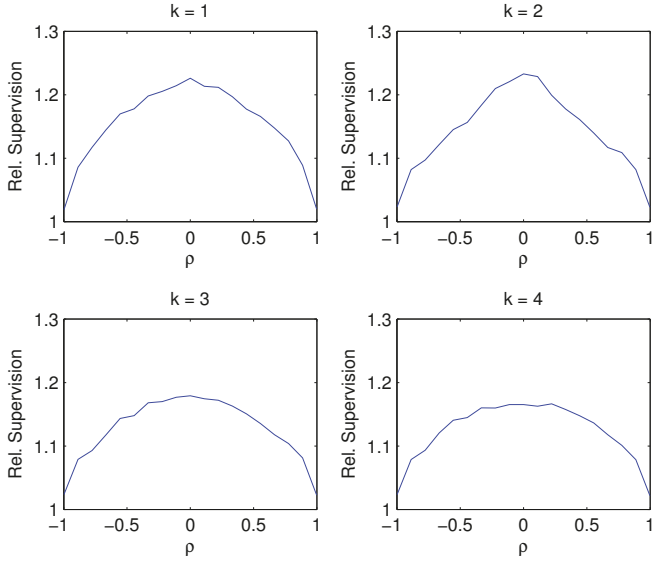


Figure 3. Supervision dependent on contemporaneous factor correlation ρ. Relative supervision against a grid of contemporaneous correlation coefficients ρ ranging from -0.998 to 0.998, while the factor serial correlation ϕ is fixed at zero and the noise level is fixed at $\sigma_u = \sigma_v = 1$.

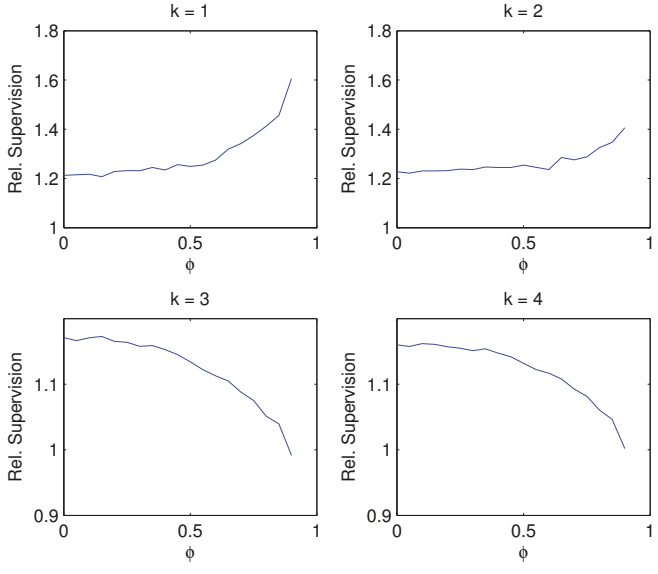


Figure 4. Supervision dependent on factor persistence ϕ. Relative supervision against a grid of AR(1) coefficients ϕ ranging from 0 to 0.9, while the noise level is fixed at $\sigma_u = \sigma_v = 1$ and the contemporaneous regressor correlation is $\rho = 0$.

4. Supervising Nelson–Siegel Factors

In the previous section, we have examined the factor model based on principal components. When the predictors are points on the yield curve, an alternative factor model can be constructed based on Nelson–Siegel (NS) components. We introduce two new factor models, CF-NS and CI-NS, by replacing principal components with NS components in CF-PC and CI-PC models. Like CI-PC, CI-NS is unsupervised. Like CF-PC, CF-NS is supervised for the particular forecast target of interest.

4.1. Nelson–Siegel Components of the Yield Curve

As an alternative to using principal components in the factor model, one can apply the modified Nelson–Siegel (NS) three-factor framework of Diebold and Li (2006) to factorize the yield curve. Nelson and Siegel (1987) propose Laguerre polynomials $L_n(z) = \frac{e^z}{n!}\frac{d^n}{dz^n}(z^n e^{-z})$ with weight function $w(z) = e^{-z}$ to model the instantaneous nominal forward rate (forward rate curve)

$$\begin{aligned} f_t(\tau) &= \beta_1 + (\beta_2 + \beta_3)\left(L_0(z)e^{-\theta\tau}\right) - \beta_3\left(L_1(z)e^{-\theta\tau}\right) \\ &= \beta_1 + (\beta_2 + \beta_3)e^{-\theta\tau} - \beta_3(1-\theta\tau)e^{-\theta\tau} \\ &= \beta_1 + \beta_2 e^{-\theta\tau} + \beta_3\theta\tau e^{-\theta\tau}, \end{aligned} \tag{46}$$

where $z = \theta\tau$, $L_0(z) = 1$, $L_1(z) = 1 - \theta\tau$, and $\beta_j \in \mathbb{R}$ for all j. The decay parameter θ may change over time, but we fixed $\theta = 0.0609$ for all t following Diebold and Li (2006).[8]

Then, the continuously compounded zero-coupon nominal yield $x_t(\tau)$ of the bond with maturity τ months at time t is

$$x_t(\tau) = \frac{1}{\tau}\int_0^\tau f_t(s)ds = \beta_1 + \beta_2\left(\frac{1-e^{-\theta\tau}}{\theta\tau}\right) + \beta_3\left(\frac{1-e^{-\theta\tau}}{\theta\tau} - e^{-\theta\tau}\right). \tag{47}$$

Allowing the β_j's to change over time and adding the approximation error v_{it}, we obtain the following approximate NS factor model for the yield curve for $i = 1, \ldots, N$:

$$\begin{aligned} x_t(\tau_i) &= \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\theta\tau_i}}{\theta\tau_i}\right) + \beta_{3t}\left(\frac{1-e^{-\theta\tau_i}}{\theta\tau_i} - e^{-\theta\tau_i}\right) + v_{it} \\ &= \begin{bmatrix} 1 & \left(\frac{1-e^{-\theta\tau_i}}{\theta\tau_i}\right) & \left(\frac{1-e^{-\theta\tau_i}}{\theta\tau_i} - e^{-\theta\tau_i}\right) \end{bmatrix} \begin{bmatrix} \beta_{1t} \\ \beta_{2t} \\ \beta_{3t} \end{bmatrix} + v_{it} \\ &= \lambda_i' f_t + v_{it}, \end{aligned} \tag{48}$$

where $f_t = (\beta_{1t}, \beta_{2t}, \beta_{3t})'$ are the three NS factors and $\lambda_i' = \begin{bmatrix} 1 & \left(\frac{1-e^{-\theta\tau_i}}{\theta\tau_i}\right) & \left(\frac{1-e^{-\theta\tau_i}}{\theta\tau_i} - e^{-\theta\tau_i}\right) \end{bmatrix}$ are the factor loadings. Because $x_t(\infty) = \beta_{1t}$, $x_t(\infty) - x_t(0) = -\beta_{2t}$, and $[x_t(0) + x_t(\infty)] - 2x_t(\tau_m)$ with $\tau_m = 24$ (say) is proportional to $-\beta_{3t}$, the three NS factors $(\beta_{1t}, \beta_{2t}, \beta_{3t})'$ are associated with level, slope, and curvature of the yield curve.

---

[8] Diebold and Li (2006) show that fixing Nelson–Siegel decay parameter at $\theta = 0.0609$ maximizes the curvature loading at the two-year bond maturity and allows better identifications of the three NS factors. They also show that allowing the θ to be a free parameter does not improve the forecasting performance. Therefore, following their advice, we fix $\theta = 0.0609$ and did not estimate it. A small θ (for a slow decaying curve) fits the curve for long maturities better and a large θ (for a fast decaying curve) fits the curve for short maturities better.

4.2. CI-NS and CF-NS

4.2.1. NS Components of Predictors X (CI-NS)

We have N predictors of yields $x_t = (x_{1t}, x_{2t}, \ldots, x_{Nt})'$ where $x_{it} = x_t(\tau_i)$ denotes the yield to maturity τ_i months at time t, $(i = 1, 2, \ldots, N)$. Stacking x_{it} for $i = 1, 2, \ldots, N$, (48) can be written as

$$x_t = \Lambda_{\text{CI}} f_{\text{CI},t} + v_{\text{CI},t}, \tag{49}$$

or

$$x_{it} = \lambda'_{\text{CI},i} f_{\text{CI},t} + v_{\text{CI},it}, \tag{50}$$

where λ_i denotes the i-th row of

$$\Lambda_{\text{CI}} = \begin{pmatrix} 1 & \frac{1-e^{-\theta\tau_1}}{\theta\tau_1} & (\frac{1-e^{-\theta\tau_1}}{\theta\tau_1} - e^{-\theta\tau_1}) \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\theta\tau_N}}{\theta\tau_N} & (\frac{1-e^{-\theta\tau_N}}{\theta\tau_N} - e^{-\theta\tau_N}) \end{pmatrix}, \tag{51}$$

which is the $N \times 3$ matrix of *known* factor loadings because we fix $\theta = 0.0609$ following Diebold and Li (2006). The NS factors $\hat{f}_{\text{CI},t} = (\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t})'$ are estimated from regressing x_{it} on $\lambda'_{\text{CI},i}$ (over $i = 1, \ldots, N$) by fitting the yield curve period by period for each t.

Then, we consider a linear forecast equation

$$y_t = (1 \; \hat{f}'_{\text{CI},t-h})\alpha_{\text{CI}} + u_{\text{CI},t}, \qquad t = h+1, \ldots, T, \tag{52}$$

in order to forecast y_{t+h} (such as output growth or inflation). We first estimate $\hat{\alpha}_{\text{CI}}$ using the information up to time T and then form the forecast we call CI-NS by

$$\hat{y}^{\text{CI-NS}}_{T+h} = (1 \; \hat{f}'_{\text{CI},T})\hat{\alpha}_{\text{CI}}. \tag{53}$$

This method is comparable to CI-PC with number of factors fixed at $k = 3$. It differs from CI-PC, however, in that the three NS factors $(\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t})$ have intuitive interpretations as level, slope and curvature of the yield curve, while the first three principal components may not have a clear interpretation. In the empirical section, we also consider two alternative CI-NS forecasts by including only the level factor $\hat{\beta}_{1t}$ (denoted CI-NS ($k = 1$)), and only the level and slope factors $(\hat{\beta}_{1t}, \hat{\beta}_{2t})$ (denoted CI-NS ($k = 2$)) to see whether the level factor or the combination of level and slope factors have dominant contribution in forecasting output growth and inflation.

4.2.2. NS Components of Forecasts $\hat{Y}$ (CF-NS)

While CI-NS solves the large-N dimensionality problem by reducing the N yields to three factors $\hat{f}_{\text{CI},t} = (\hat{\beta}_{1t}, \hat{\beta}_{2t}, \hat{\beta}_{3t})'$, it computes the factors entirely from yield curve information x_t only, without accounting for the variable y_{t+h} to be forecast. Similar in spirit to CF-PC, here we can improve CI-NS by supervising the factor computation, which we term as CF-NS.

The CF-NS forecast is based on the NS factors of $\hat{\mathbf{y}}_{t+h} := (\hat{y}^{(1)}_{t+h}, \hat{y}^{(2)}_{t+h}, \ldots, \hat{y}^{(N)}_{t+h})'$, a vector of the N individual forecasts as in (10) and (11),

$$\hat{\mathbf{y}}_{t+h} = \Lambda_{\text{CF}} f_{\text{CF},t+h} + v_{\text{CF},t+h}, \tag{54}$$

with $\Lambda_{\text{CF}} = \Lambda_{\text{CI}}$ in (51). Hence, $\Lambda_{\text{CI}} = \Lambda_{\text{CF}} = \Lambda$ for the NS factor models. Note that, when the NS factors loadings are normalized to sum up to one, the three CF-NS factors

$$\begin{aligned}\hat{f}_{\text{CF},t+h} &= \Lambda' \hat{y}_{t+h} \\ &= \left(\begin{array}{ccc} \frac{1}{s_1}\sum_{i=1}^{N} \hat{y}_{T+h}^{(i)} & \frac{1}{s_2}\sum_{i=1}^{N}\left(\frac{1-e^{-\theta\tau_i}}{\theta\tau_i}\right)\hat{y}_{T+h}^{(i)} & \frac{1}{s_3}\sum_{i=1}^{N}\left(\frac{1-e^{-\theta\tau_i}}{\theta\tau_i} - e^{-\theta\tau_i}\right)\hat{y}_{T+h}^{(i)} \end{array} \right)'\end{aligned} \tag{55}$$

are weighted individual forecasts with the three *normalized* NS loadings, with $s_1 = N$, $s_2 = \sum_{i=1}^{N}\left(\frac{1-e^{-\theta\tau_i}}{\theta\tau_i}\right)$, and $s_3 = \sum_{i=1}^{N}\left(\frac{1-e^{-\theta\tau_i}}{\theta\tau_i} - e^{-\theta\tau_i}\right)$. The CF-NS forecast can be obtained from the forecasting equation

$$\begin{aligned} y_{t+h} &= \hat{f}'_{\text{CF},t+h}\alpha_{\text{CF}} + u_{\text{CF},t+h}, \\ \hat{y}_{T+h}^{\text{CF-NS}} &= \hat{f}'_{\text{CF},T+h}\hat{\alpha}_{\text{CF}}, \end{aligned} \tag{56}$$

which is denoted CF-NS($k = 3$). The parameter vector $\hat{\alpha}_T$ is estimated using information up to time T. Using only the first factor or the first two factors, one can obtain the forecasts CF-NS($k = 1$) and CF-NS($k = 2$).

Note that, while the CF-PC method can be used for data of many kinds, the CF-NS method we propose is tailored to forecasting using the yield curve. It uses fixed factor loadings in Λ that are the NS exponential factor loadings for yield curve modeling, and hence avoids the estimation of factor loadings. In contrast, CF-PC needs to estimate Λ.

Also note that, by construction, CF-NS($k = 1$) is the equally weighted combined forecast $\frac{1}{N}\sum_{i=1}^{N}\hat{y}_{T+h}^{(i)}$.

5. Forecasting Output Growth and Inflation

This section presents the empirical analysis where we describe the data, implement forecasting methods introduced in the previous sections on forecasting output growth and inflation, and analyze out-of-sample forecasting performances. This allows us to analyze the differences between output growth and inflation forecasting using the same yield curve information and to compare the strengths of different methods.

5.1. Data

Let y_{t+h} denote the variable to be forecast (output growth or inflation) using yield information up to time t, where h denotes the forecast horizon. The predictor vector $\mathbf{x}_t = (x_t(\tau_1),\ x_t(\tau_2),\ \ldots,\ x_t(\tau_N))'$ contains the information about the yield curve at various maturities: $x_t(\tau_i)$ denotes the zero coupon yield of maturity τ_i months at time t $(i = 1, 2, \ldots, N)$.

Two forecast targets, output growth and inflation, are constructed respectively as monthly growth rate of Personal Income (PI, seasonally adjusted annual rate) and monthly change in CPI (Consumer Price Index for all urban consumers: all items, seasonally adjusted) from 1970:01 to 2010:01. PI and CPI data are obtained from the web site of the Federal Reserve Bank of St. Louis (FRED2).

We apply the following data transformations. For the monthly growth rate of PI, we set $y_{t+h} = 1200[(1/h)\ln(\text{PI}_{t+h}/\text{PI}_t)]$ as the forecast target (as used in Ang et al. (2006)). For the consumer price index (CPI), we set $y_{t+h} = 1200[(1/h)\ln(\text{CPI}_{t+h}/\text{CPI}_t)]$ as the forecast target (as used in Stock and Watson (2007)).[9]

[9] $y_{t+h} = 1200[(1/h)\ln(\text{CPI}_{t+h}/\text{CPI}_t) - \ln(\text{CPI}_t/\text{CPI}_{t-1})]$ is used in Bai and Ng (2008).

Our yield curve data consist of U.S. government bond prices, coupon rates, and coupon structures, as well as issue and redemption dates from 1970:01 to 2009:12.[10] We calculate zero-coupon bond yields using the unsmoothed Fama and Bliss (1987) approach. We measure bond yields on the second day of each month. We also apply several data filters designed to enhance data quality and focus attention on maturities with good liquidity. First, we exclude floating rate bonds, callable bonds and bonds extended beyond the original redemption date. Second, we exclude outlying bond prices less than 50 or greater than 130 because their price discounts/premium are too high and imply thin trading, and we exclude yields that differ greatly from yields at nearby maturities. Finally, we use only bonds with maturity greater than one month and less than fifteen years because other bonds are not actively traded. Indeed, to simplify our subsequent estimation, using linear interpolation we pool the bond yields into fixed maturities of 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 33, 36, 39, 42, 45, 48, 51, 54, 57, 60, 63, 66, 72, 78, 84, 90, 96, 102, 108, and 120 months, where a month is defined as 30.4375 days.[11]

We examine some descriptive statistics (not reported for space) of the two forecast targets and yield curve level, slope, and curvature (empirical measures), over the full sample from 1970:01 to 2009:12 and the out-of-sample evaluation period from 1995:02 to 2010:01. We observe that both PI growth and CPI inflation become more moderate and less volatile from around the mid-1980s. This has become a stylized fact known as the "Great Moderation". In particular, there is a substantial drop in persistency of CPI inflation. The volatility and persistency of the yield curve slope and curvature do not change much. The yield curve level, however, decreases and stabilizes.

In predicting macroeconomic variables using the term structure, yield spreads between yields with various maturities and the short rate are commonly used in the literature. One possible reason for this practice is that yield levels are treated as I(1) processes, so yield spreads will likely be I(0). Similarly, macroeconomic variables are typically assumed to be I(1) and transformed properly into I(0), so that, in using yield spreads to forecast macro targets, issues such as spurious regression are avoided. In this paper, however, we use yield levels (not spreads) to predict PI growth and CPI inflation (not change in inflation), for the following reasons. First, whether yields and inflation are I(1) or I(0) is still arguable. Stock and Watson (1999, 2012) use yield spreads and treat inflation as I(1), so they forecast change in inflation. Inoue and Kilian (2008), however, treat inflation as I(0). Since our target is forecasting inflation, not change in inflation, we will treat CPI inflation as well as yields as I(0) in our empirical analysis. Second, we emphasize real-time, out-of-sample forecasting performance more than in-sample concerns. As long as out-of-sample forecast performance is unaltered or even improved, we think the choice of treating the variables as I(1) or I(0) variables does not matter much.[12] Third, using yield levels will allow us to provide clearer interpretations for questions such as what part of the yield curve contributes the most towards predicting PI growth or CPI inflation, and how the different parts of the yield curve interact in the prediction, etc.

5.2. Out-of-Sample Forecasting

All forecasting models are estimated in a rolling window scheme with window size $R = 300$ months ending at month t (starting at $t - R + 1$). In the evaluation period from $t =$ 1995:02

[10] As a robust check, we apply our method to the original yield data of Diebold and Li (2006) and also to the sub-samples in our data set. The results are essentially the same as those summarized at the end of Section 5.

[11] It may be interesting to explore whether different maturity yields might have different effects on the forecast outcome. However, the present paper is focused on the comparison between CF and CI, rather than a detailed CI-only analysis, e.g., to find the best maturity yield for the forecast outcome. Nevertheless, our CI-NS model has reflected such effects as the three NS factors (level, slope, and curvature) are different combinations of bond maturities as shown in Equation (55). The different coefficients on the NS factors suggest that different bond maturities have different effects on the forecast outcome, as Gogas et al. (2015) has found.

[12] While not reported for space, we tried forecasting change in inflation and found forecasting inflation directly using all yield levels improves out-of-sample performances of most forecasting methods by a large margin.

to t = 2010:01 (180 months), the first rolling sample to estimate models begins at 1970:02 and ends at 1995:01, the second rolling sample is for 1970:03–1995:02, the third 1970:04–1995:03, and so on. The out-of-sample evaluation period is from 1995:02 to 2010:01 (hence out-of-sample size $P = 180$).[13] In all NS-related methods (CI and CF), we set θ, the parameter that governs the exponential decay rate, at 0.0609 for reasons discussed in Diebold and Li (2006).[14] We compare h-months-ahead out-of-sample forecasting results of those methods introduced so far for $h = 1, 3, 6, 12, 18, 24, 30, 36$ months ahead.

Figure 5 illustrates what economic contents these factors in CF-PC may bear. It shows that the first PC assigns about equal weights to all $N = 50$ individual forecasts that use yields at various maturities (in months) so that it may be interpreted as the factor that captures the level of the yield curve; the second PC assigns roughly increasing weights so that it may be interpreted as the factor capturing the slope; and the third PC assigns roughly first decreasing then increasing weights, so that it may be interpreted as factor capturing curvature.

Tables 1 and 2 present the root mean squared forecast errors (RMSFE) of PC methods with $k = 1, 2, 3, 4, 5$, and of NS methods with $k = 1, 2, 3$, for PI growth (Table 1A) and for CPI inflation (Table 2A) forecasts using all 50 yield levels.[15] In Panel A of Tables 1 and 2, we report the Root Mean Squared Forecast Errors (RMSFE, which is the squared root of the MSFE of a model).[16] In Panel B of Tables 1 and 2, we report Relative Supervision of CI-PC vs. CF-PC and Relative Supervision of CI-NS vs. CF-NS, according to Definition 3, which is the ratio of the MSFEs of two CI and CF models. The relative supervision in Panel B can be obtained from RMSFEs in Panel A. For simplicity of presentation in Panel B, we present the relative supervision only with the same number of factors ($k_{\mathrm{CI}} = k_{\mathrm{CF}}$ and $k_{\mathrm{NS}} = k_{\mathrm{NS}}$).

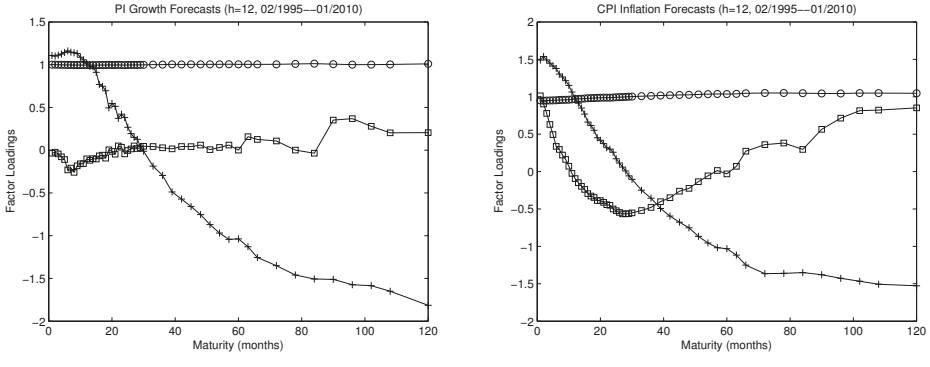


Figure 5. *Cont.*

[13] As a robust check, we have also tried with different sample splits for the estimation and prediction periods, i.e., the number of in-sample regression observations and the out-of-sample evaluation observations. We find that the results are similar.

[14] For different values of θ, the performances of CI-NS and CF-NS change only marginally.

[15] While we report the results for $k = 4, 5$ for CF-PC, we do not report for $k = 4, 5$ for CF-NS. Svennsson (1995) and Christensen et al. (2009) (CDR 2009) extend the three factor NS model to four or five factor NS models. CDR's dynamic generalized NS model has five factors with one level factor, two slope factors and two curvature factors. The Svensson and CDR extensions are useful to fit the yield curve at longer maturities (>10 years). Because we only used yields with maturities ≤10 years, the second curvature factor loadings will look similar to the slope factor loadings and we will have collinearity problem. CDR use yields up to 30 years. The 4th and 5th factors have no clear economic intrepretations and are hard to explain. For these reasons, we report results for $k = 1, 2, 3$ for the CF-NS model.

[16] For the statistical significance of the loss-difference (see Definition 2), the asymptotic p-values of the Diebold–Mariano statistics are all very close to zero especially for larger values of the forecast horizon h.

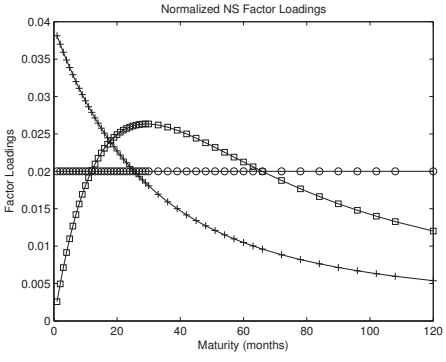


Figure 5. Factor loadings of principal components and Nelson–Siegel factors. The first two panels: factor loadings of the first three principal components in CF-PC ($k = 3$) averaged over the out-of-sample period (02/1995–01/2010), for both PI growth (first panel) and CPI inflation (second panel). The abscissa refers to the 50 individual forecasts that use yields at the 50 maturities (in months). The loading of the first principal component has the circle-symbol, the second the cross-symbol, and the third the square-symbol. The third panel: three normalized Nelson–Siegel (NS) exponential loadings in CF-NS that correspond to the three NS factors, respectively. The abscissa refers to the 50 individual forecasts that use yields at the 50 maturities (in months). The circled line denotes the first normalized NS factor loading $1/N$, the crossed line denotes the second normalized NS factor loading $(1 - e^{-\theta\tau})/(\theta\tau)$, divided by the sum, and the squared line denotes the third normalized NS factor loading $(1 - e^{-\theta\tau})/(\theta\tau) - e^{-\theta\tau}$, divided by the sum, where τ denotes maturity and θ is fixed at 0.0609.

We find that, in general, supervised factorization performs better. The CF schemes (CF-PC and CF-NS) perform substantially better than the CI schemes (CI-PC and CI-NS). Within the same CF or CI schemes, two alternative factorizations work similarly: CF-PC and CF-NS are about the same, and CI-PC and CI-NS are about the same. We summarize our findings from Figure 5 and Tables 1 and 2 as follows.

1. *Supervision is similar for CF-PC and CF-NS*. The factor loadings for CF-NS and for CF-PC are similar as shown in Figure 5. Panel (c) of the figure plots three normalized NS exponential loadings in CF-NS that correspond respectively to the three NS factors. Note that the factor loadings in CF-NS are pre-specified while those in CF-PC are estimated from the N individual forecasts. Nevertheless, their shapes in panel (a) look very similar to those of the CF-PC loadings in panels (a) and (b) (apart from the signs). Accordingly, out-of-sample forecasting performance of CF-PC and CF-NS are very similar as shown in Panel A of Tables 1 and 2.
2. *Supervision is substantial*. Supervised factor models perform better than unsupervised factor models in forecasting. Both CF-PC and CF-NS are much better than CI-PC and CI-NS models as shown in Panel B of Tables 1 and 2.
3. *Supervision is generally stronger for a longer forecast horizon h*. The advantage of CF-PC over CI-PC generally increases with forecast horizon h, as shown in Panel B of Tables 1 and 2.[17]

[17] We conducted a Monte Carlo (not reported), which are consistent with the empirical results that the supervision is stronger for a longer forecast horizon h.

4. *We often get the best supervised predictions with a single factor* ($k = 1$) *with the CF-factor models.*[18] Since CF-NS($k = 1$) is the equally weighted combined forecast as noted in Section 4.2.2, this is another case of the forecast combination puzzle discussed in Remark 3 that the equal-weighted forecast combination is hard to beat. Since CF-PC($k = 1$) is numerically identical to CF-NS($k = 1$) as shown in Figure 5, CF-PC($k = 1$) is also effectively equally weighted forecast averaging.[19]

Table 1. Out-of-sample forecasting of personal income growth.

Panel A. Root Mean Squared Forecast Errors								
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
CI-PC($k = 1$)	5.64	3.56	2.99	2.78	2.61	2.50	2.46	2.42
CI-PC($k = 2$)	5.67	3.64	3.12	3.00	2.81	2.66	2.55	2.45
CI-PC($k = 3$)	5.71	3.69	3.19	3.08	2.92	2.77	2.63	2.49
CI-PC($k = 4$)	5.72	3.76	3.23	3.12	2.93	2.77	2.58	2.36
CI-PC($k = 5$)	5.74	3.78	3.26	3.15	2.98	2.81	2.61	2.38
CI-NS($k = 1$)	5.84	3.84	3.28	3.06	2.86	2.69	2.53	2.41
CI-NS($k = 2$)	5.71	3.71	3.20	3.11	2.93	2.77	2.62	2.48
CI-NS($k = 3$)	5.72	3.69	3.19	3.09	2.93	2.78	2.63	2.47
CF-PC($k = 1$)	5.60	3.45	2.83	2.54	2.24	1.95	1.75	1.58
CF-PC($k = 2$)	5.56	3.43	2.83	2.62	2.31	1.93	1.76	1.61
CF-PC($k = 3$)	5.60	3.44	2.94	2.78	2.47	2.02	1.65	1.48
CF-PC($k = 4$)	5.63	3.60	3.08	2.83	2.39	1.97	1.67	1.45
CF-PC($k = 5$)	5.63	3.60	3.05	2.87	2.41	2.05	1.69	1.51
CF-NS($k = 1$)	5.60	3.45	2.83	2.54	2.24	1.95	1.75	1.58
CF-NS($k = 2$)	5.56	3.43	2.84	2.62	2.30	1.95	1.76	1.62
CF-NS($k = 3$)	5.59	3.44	2.94	2.79	2.47	2.02	1.64	1.48

[18] Figlewski and Urich (1983) talked about various constrained models in forming a combination of forecasts and examined when we need more than the simple averaging combined forecast. They discussed a sufficient condition when the simple average of forecasts is the optimal forecast combination: "Under the most extensive set of constraints, forecast errors are assumed to have zero mean and to be independent and identically distributed. In this case the optimal forecast is the simple average." This corresponds to CF-PC($k = 1$) and CF-NS($k = 1$) when the first factor ($k = 1$) in PC or NS is sufficient for the CF factor model. It is clearly the case in CF-NS as shown in Equation (55). One can show that the first PC (corresponding to the largest singular value) would also be the simple average. Hence, in terms of the CF-factor model, the forecast combination puzzle amounts to the fact that we often do not need the second PC factor. Interestingly, (Figlewski and Urich 1983, p. 696) continued to note the cases when the simple average is not optimal: "However, the hypothesis of independence among forecast errors is overwhelmingly rejected for our data-errors are highly positively correlated with one another." On the other hand, they also noted other reasons why the simple average may still be preferred, as they wrote, "Because the estimated error structure was not completely stable over time, the models which adjusted for correlation did not achieve lower mean squared forecast error than the simple average in out-of-sample tests. Even so, we find...that forecasts from these models, while less accurate than the simple mean, do contain information which is not fully reflected in prices in the money market, and is therefore economically valuable." We thank a referee for letting us know on this from Figlewski and Urich (1983).

[19] While the simple equally weighted forecast combination can be implemented without the use of PCA or without making reference to the NS model, it is important to note that the simple average combined forecast indeed corresponds the first CF-PC factor (CF-PC($k = 1$)) or the first CF-NS factor (CF-NS($k = 1$)). In view of Figlewski and Urich (1983), it will be useful to know when the first factor ($k = 1$) is enough so that the simple average is good or when the higher order factors ($k > 1$) may be necessary as they contain more information in addition to the first CF-factor. This is important in understanding the forecast combination puzzle. The forecast combination puzzle is about whether to include only the first CF factor or more.

Table 1. *Cont.*

Panel B. Relative Supervision $s_{rel}(X, y, k_{CI}, k_{CF})$								
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
CI-PC($k = 1$) vs. CF-PC($k = 1$)	1.01	1.06	1.12	1.20	1.36	1.64	1.98	2.35
CI-PC($k = 2$) vs. CF-PC($k = 2$)	1.04	1.13	1.22	1.31	1.48	1.90	2.10	2.32
CI-PC($k = 3$) vs. CF-PC($k = 3$)	1.04	1.15	1.18	1.23	1.40	1.88	2.54	2.83
CI-PC($k = 4$) vs. CF-PC($k = 4$)	1.03	1.09	1.10	1.22	1.50	1.98	2.39	2.65
CI-PC($k = 5$) vs. CF-PC($k = 5$)	1.04	1.10	1.14	1.20	1.53	1.88	2.39	2.48
CI-NS($k = 1$) vs. CF-NS($k = 1$)	1.09	1.24	1.34	1.45	1.63	1.90	2.09	2.33
CI-NS($k = 2$) vs. CF-NS($k = 2$)	1.05	1.17	1.27	1.41	1.62	2.02	2.22	2.34
CI-NS($k = 3$) vs. CF-NS($k = 3$)	1.05	1.15	1.18	1.23	1.41	1.89	2.57	2.79

The forecast target is Output Growth $y_{t+h} = 1200 \times \log(PI_{t+h}/PI_t) \div h$. Out-of-sample forecasting period is 02/1995–01/2010. In Panel A, reported are the Root Mean Squared Forecast Errors (which is the squared root of the MSFE of a model). In Panel B, reported are Relative Supervision of CI-PC vs. CF-PC and Relative Supervision of CI-NS vs. CF-NS, according to Definition 3, which is the ratio of the MSFEs of the two models. For simplicity of presentation, we present the relative supervision in Panel B only with the same number of factors ($k_{CI} = k_{CF} = k$ and $k_{NS} = k_{NS} = k$).

Table 2. Out-of-sample forecasting of CPI inflation.

Panel A. Root Mean Squared Forecast Errors								
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
CI-PC($k = 1$)	3.77	2.86	2.25	1.92	1.94	2.16	2.47	2.75
CI-PC($k = 2$)	4.21	3.45	2.96	2.76	2.77	2.84	2.96	3.08
CI-PC($k = 3$)	4.24	3.50	3.00	2.82	2.88	2.98	3.10	3.19
CI-PC($k = 4$)	4.31	3.57	3.05	2.87	2.91	3.00	3.12	3.18
CI-PC($k = 5$)	4.30	3.58	3.07	2.93	3.00	3.10	3.20	3.23
CI-NS($k = 1$)	3.95	3.12	2.62	2.48	2.60	2.79	2.97	3.10
CI-NS($k = 2$)	4.22	3.46	2.98	2.82	2.88	2.98	3.09	3.18
CI-NS($k = 3$)	4.24	3.50	3.01	2.83	2.89	2.99	3.11	3.20
CF-PC($k = 1$)	3.65	2.67	1.91	1.31	1.01	0.90	0.96	1.08
CF-PC($k = 2$)	3.66	2.70	1.93	1.35	1.10	1.05	1.11	1.19
CF-PC($k = 3$)	3.68	2.72	1.97	1.47	1.29	1.19	1.19	1.20
CF-PC($k = 4$)	3.74	2.80	2.01	1.47	1.22	1.14	1.15	1.17
CF-PC($k = 5$)	3.74	2.79	1.98	1.45	1.20	1.12	1.18	1.20
CF-NS($k = 1$)	3.65	2.68	1.91	1.31	1.02	0.90	0.96	1.08
CF-NS($k = 2$)	3.66	2.70	1.93	1.35	1.10	1.05	1.10	1.19
CF-NS($k = 3$)	3.68	2.73	1.97	1.47	1.29	1.20	1.19	1.20
Panel B. Relative Supervision $s_{rel}(X, y, k_{CI}, k_{CF})$								
	$h = 1$	$h = 3$	$h = 6$	$h = 12$	$h = 18$	$h = 24$	$h = 30$	$h = 36$
CI-PC($k = 1$) vs. CF-PC($k = 1$)	1.07	1.15	1.39	2.15	3.69	5.76	6.62	6.48
CI-PC($k = 2$) vs. CF-PC($k = 2$)	1.32	1.63	2.35	4.18	6.34	7.32	7.11	6.70
CI-PC($k = 3$) vs. CF-PC($k = 3$)	1.33	1.66	2.32	3.68	4.98	6.27	6.79	7.07
CI-PC($k = 4$) vs. CF-PC($k = 4$)	1.33	1.63	2.30	3.81	5.69	6.93	7.36	7.39
CI-PC($k = 5$) vs. CF-PC($k = 5$)	1.32	1.65	2.40	4.08	6.25	7.66	7.35	7.25
CI-NS($k = 1$) vs. CF-NS($k = 1$)	1.17	1.36	1.88	3.58	6.50	9.61	9.57	8.24
CI-NS($k = 2$) vs. CF-NS($k = 2$)	1.33	1.64	2.38	4.36	6.85	8.05	7.89	7.14
CI-NS($k = 3$) vs. CF-NS($k = 3$)	1.33	1.64	2.33	3.71	5.02	6.21	6.83	7.11

The forecast target is Inflation $y_{t+h} = 1200 \times \log(CPI_{t+h}/CPI_t) \div h$. Out-of-sample forecasting period is 02/1995–01/2010. In Panel A, reported are the Root Mean Squared Forecast Errors (which is the squared root of the MSFE of a model). In Panel B, reported are Relative Supervision of CI-PC vs. CF-PC and Relative Supervision of CI-NS vs. CF-NS, according to Definition 3, which is the ratio of the MSFEs of the two models. For simplicity of presentation, we present the relative supervision in Panel B only with the same number of factors ($k_{CI} = k_{CF} = k$ and $k_{NS} = k_{NS} = k$).

6. Conclusions

For forecasting in the presence of many predictors, it is often useful to reduce the dimension by a factor model (in a dense case) or by variable selection (in a sparse case). In this paper, we consider a factor model. In particular, we examine the supervised principal component analysis of Chan et al. (1999). The model is called CF-PC, as the principal components of many forecasts are the combined forecasts.

The CF-PC extracts factors from the space spanned by forecasts rather than from the space spanned by predictors. This factorization of the forecasts improves forecast performance compared to factor analysis of the predictors. We extend the CF-PC to CF-NS, which uses the NS factor model in place of the PC factor model, for the application where the predictors are the yield curve. While the yield curve is a functional data consisting of many different maturity points on a curve at each time, the NS factors can parsimoniously capture the shapes of the curve.

We have applied the CF-PC and CF-NS models in forecasting output growth and inflation using a large number of bond yields to examine if the supervised factorization improves forecast performance. In general, we have found that CF-PC and CF-NS perform substantially better than CI-PC and CI-NS, that the advantage of supervised factor models is even larger for longer forecast horizons, and that the two alternative factor models based on PC and NS factors are similar and perform similarly.

Author Contributions: All authors contributed equally to the paper.

Acknowledgments: We would like to thank the two referees for helpful comments. We also thank Jonathan Wright and seminar participants at FRB of San Francisco, FRB of St Louis, Federal Reserve Board (Washington DC), Bank of Korea, SETA meeting, Stanford Institute of Theoretical Economics (SITE), University of Cambridge, NCSU, UC Davis, UCR, UCSB, UCSD, USC, Purdue, LSU, Indiana, Drexel, OCC, WMU, and SNU, for useful discussions and comments. All errors are our own. E.H. acknowledges support from the Danish National Research Foundation. The views presented in this paper are solely those of the authors and do not necessarily represent those of ICBCCS, the Federal Reserve Board or their staff.

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

Appendix A. Calculation of Absolute and Relative Supervision in Example 1

Using R and Σ_k obtained from the SVD for CI in (36), and S and Θ_k obtained from the SVD for CF in (40), we calculate the absolute supervision and relative supervision for each k. The CI factors are $F_{CI} = R\Sigma_k$ from (23), and the CF factors $F_{CF} = S\Theta_k$ from (28).

For $k = 1$,

$$F_{\text{CI}} = \begin{bmatrix} 1&0&0&0&0&0 \\ 0&1&0&0&0&0 \\ 0&0&1&0&0&0 \\ 0&0&0&1&0&0 \\ 0&0&0&0&1&0 \\ 0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad F_{\text{CF}} = \begin{bmatrix} 0&0&0&0&1&0 \\ 0&0&0&1&0&0 \\ 0&0&1&0&0&0 \\ 0&1&0&0&0&0 \\ 1&0&0&0&0&0 \\ 0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} 5 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 5 \\ 0 \end{bmatrix}, \tag{A1}$$

$$\hat{y}_{\text{CI-PC}} = R_1 R_1' y = (1,0,0,0,0,0)', \tag{A2}$$

$$\hat{y}_{\text{CF-PC}} = S_1 S_1' y = (0,0,0,0,5,0)', \tag{A3}$$

$$||y - \hat{y}_{\text{CI-PC}}||^2 = ||(1,\ 2,\ 3,\ 4,\ 5,\ 0)' - (1,0,0,0,0,0)'||^2 = 54, \tag{A4}$$

$$||y - \hat{y}_{\text{CF-PC}}||^2 = ||(1,\ 2,\ 3,\ 4,\ 5,\ 0)' - (0,0,0,0,5,0)'||^2 = 30. \tag{A5}$$

Hence, $s_{abs}(X, y, 1, 1) = ||y - \hat{y}_{\text{CI-PC}}||^2 - ||y - \hat{y}_{\text{CF-PC}}||^2 = 54 - 30 = 24$, and $s_{rel}(X, y, 1, 1) = ||y - \hat{y}_{\text{CI-PC}}||^2 / ||y - \hat{y}_{\text{CF-PC}}||^2 = 54/30 = 1.8$.

For $k = 2$,

$$F_{\text{CI}} = \begin{bmatrix} 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} 1&0\\ 0&\frac{1}{2}\\ 0&0\\ 0&0\\ 0&0\\ 0&0 \end{bmatrix} = \begin{bmatrix} 1&0\\ 0&\frac{1}{2}\\ 0&0\\ 0&0\\ 0&0\\ 0&0 \end{bmatrix}, \quad F_{\text{CF}} = \begin{bmatrix} 0&0&0&0&1&0\\ 0&0&0&1&0&0\\ 0&0&1&0&0&0\\ 0&1&0&0&0&0\\ 1&0&0&0&0&0\\ 0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} 5&0\\ 0&4\\ 0&0\\ 0&0\\ 0&0\\ 0&0 \end{bmatrix} = \begin{bmatrix} 0&0\\ 0&0\\ 0&0\\ 0&4\\ 5&0\\ 0&0 \end{bmatrix}, \quad \text{(A6)}$$

$$\hat{y}_{\text{CI-PC}} = R_2 R_2' y = (1,2,0,0,0,0)', \quad \text{(A7)}$$
$$\hat{y}_{\text{CF-PC}} = S_2 S_2' y = (0,0,0,4,5,0)', \quad \text{(A8)}$$

$$||y - \hat{y}_{\text{CI-PC}}||^2 = ||(1,\,2,\,3,\,4,\,5,\,0)' - (1,2,0,0,0,0)'||^2 = 50, \quad \text{(A9)}$$
$$||y - \hat{y}_{\text{CF-PC}}||^2 = ||(1,\,2,\,3,\,4,\,5,\,0)' - (0,0,0,4,5,0)'||^2 = 14. \quad \text{(A10)}$$

Hence, $s_{abs}(X,y,2,2) = ||y - \hat{y}_{\text{CI-PC}}||^2 - ||y - \hat{y}_{\text{CF-PC}}||^2 = 50 - 14 = 36$, and $s_{rel}(X,y,2,2) = ||y - \hat{y}_{\text{CI-PC}}||^2 / ||y - \hat{y}_{\text{CF-PC}}||^2 = 50/14 = 3.6$.

For $k = 3$,

$$F_{\text{CI}} = \begin{bmatrix} 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} 1&0&0\\ 0&\frac{1}{2}&0\\ 0&0&\frac{1}{3}\\ 0&0&0\\ 0&0&0\\ 0&0&0 \end{bmatrix} = \begin{bmatrix} 1&0&0\\ 0&\frac{1}{2}&0\\ 0&0&\frac{1}{3}\\ 0&0&0\\ 0&0&0\\ 0&0&0 \end{bmatrix}, \quad F_{\text{CF}} = \begin{bmatrix} 0&0&0&0&1&0\\ 0&0&0&1&0&0\\ 0&0&1&0&0&0\\ 0&1&0&0&0&0\\ 1&0&0&0&0&0\\ 0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} 5&0&0\\ 0&4&0\\ 0&0&3\\ 0&0&0\\ 0&0&0\\ 0&0&0 \end{bmatrix} = \begin{bmatrix} 0&0&0\\ 0&0&0\\ 0&0&3\\ 0&4&0\\ 5&0&0\\ 0&0&0 \end{bmatrix}, \quad \text{(A11)}$$

$$\hat{y}_{\text{CI-PC}} = R_3 R_3' y = (1,2,3,0,0,0)', \quad \text{(A12)}$$
$$\hat{y}_{\text{CF-PC}} = S_3 S_3' y = (0,0,3,4,5,0)', \quad \text{(A13)}$$

$$||y - \hat{y}_{\text{CI-PC}}||^2 = ||(1,\,2,\,3,\,4,\,5,\,0)' - (1,2,3,0,0,0)'||^2 = 41, \quad \text{(A14)}$$
$$||y - \hat{y}_{\text{CF-PC}}||^2 = ||(1,\,2,\,3,\,4,\,5,\,0)' - (0,0,3,4,5,0)'||^2 = 5. \quad \text{(A15)}$$

Hence, $s_{abs}(X,y,3,3) = ||y - \hat{y}_{\text{CI-PC}}||^2 - ||y - \hat{y}_{\text{CF-PC}}||^2 = 41 - 5 = 36$, and $s_{rel}(X,y,3,3) = ||y - \hat{y}_{\text{CI-PC}}||^2 / ||y - \hat{y}_{\text{CF-PC}}||^2 = 41/5 = 8.2$.

For $k = 4$,

$$F_{\text{CI}} = \begin{bmatrix} 1&0&0&0&0&0\\ 0&1&0&0&0&0\\ 0&0&1&0&0&0\\ 0&0&0&1&0&0\\ 0&0&0&0&1&0\\ 0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} 1&0&0&0\\ 0&\frac{1}{2}&0&0\\ 0&0&\frac{1}{3}&0\\ 0&0&0&\frac{1}{4}\\ 0&0&0&0\\ 0&0&0&0 \end{bmatrix} = \begin{bmatrix} 1&0&0&0\\ 0&\frac{1}{2}&0&0\\ 0&0&\frac{1}{3}&0\\ 0&0&0&\frac{1}{4}\\ 0&0&0&0\\ 0&0&0&0 \end{bmatrix}, \quad F_{\text{CF}} = \begin{bmatrix} 0&0&0&0&1&0\\ 0&0&0&1&0&0\\ 0&0&1&0&0&0\\ 0&1&0&0&0&0\\ 1&0&0&0&0&0\\ 0&0&0&0&0&1 \end{bmatrix} \begin{bmatrix} 5&0&0&0\\ 0&4&0&0\\ 0&0&3&0\\ 0&0&0&2\\ 0&0&0&0\\ 0&0&0&0 \end{bmatrix} = \begin{bmatrix} 0&0&0&0\\ 0&0&0&2\\ 0&0&3&0\\ 0&4&0&0\\ 5&0&0&0\\ 0&0&0&0 \end{bmatrix}, \quad \text{(A16)}$$

$$\hat{y}_{\text{CI-PC}} = R_4 R_4' y = (1,2,3,4,0,0)', \quad \text{(A17)}$$
$$\hat{y}_{\text{CF-PC}} = S_4 S_4' y = (0,2,3,4,5,0)', \quad \text{(A18)}$$

$$||y - \hat{y}_{\text{CI-PC}}||^2 = ||(1,\ 2,\ 3,\ 4,\ 5,\ 0)' - (1,2,3,4,0,0)'||^2 = 25, \tag{A19}$$

$$||y - \hat{y}_{\text{CF-PC}}||^2 = ||(1,\ 2,\ 3,\ 4,\ 5,\ 0)' - (0,2,3,4,5,0)'||^2 = 1. \tag{A20}$$

Hence, $s_{abs}(X, y, 4, 4) = ||y - \hat{y}_{\text{CI-PC}}||^2 - ||y - \hat{y}_{\text{CF-PC}}||^2 = 25 - 1 = 24$, and $s_{rel}(X, y, 4, 4) = ||y - \hat{y}_{\text{CI-PC}}||^2 / ||y - \hat{y}_{\text{CF-PC}}||^2 = 25/1 = 25$.

For $k = 5$, $s_{abs}(X, y, 5, 5) = y'(SS' - RR')y = 0$ because

$$||y - \hat{y}_{\text{CI-PC}}||^2 = ||(1,\ 2,\ 3,\ 4,\ 5,\ 0)' - (1,2,3,4,5,0)'||^2 = 0, \tag{A21}$$

$$||y - \hat{y}_{\text{CF-PC}}||^2 = ||(1,\ 2,\ 3,\ 4,\ 5,\ 0)' - (1,2,3,4,5,0)'||^2 = 0. \tag{A22}$$

Hence, as noted in Remark 4, $s_{abs}(X, y, 5, 5) = ||y - \hat{y}_{\text{CI-PC}}||^2 - ||y - \hat{y}_{\text{CF-PC}}||^2 = 0 - 0 = 0$, and $s_{rel}(X, y, 5, 5)$ is not defined for $k = N = 5$.

References

Ang, Andrew, and Monika Piazzesi. 2003. A No-Arbitrage Vector Autoregression of Term Structure Dynamics with Macroeconomic and Latent Variables. *Journal of Monetary Economics* 50: 745–87. [CrossRef]

Ang, Andrew, Monika Piazzesi, and Min Wei. 2006. What Does the Yield Curve Tell Us about GDP Growth? *Journal of Econometrics* 131: 359–403. [CrossRef]

Armah, Nii Ayi, and Norman R. Swanson. 2010. Seeing Inside the Black Box: Using Diffusion Index Methodology to Construct Factor Proxies in Large Scale Macroeconomic Time Series Environments. *Econometric Reviews* 29: 476–510. [CrossRef]

Bai, Jushan. 2003. Inferential Theory for Factor Models of Large Dimensions. *Econometrica* 71: 135–71. [CrossRef]

Bai, Jushan and Serena Ng. 2006. Confidence Intervals for Diffusion Index Forecasts and Inference for Factor-Augmented Regressions. *Econometrica* 74: 1133–50. [CrossRef]

Bai, Jushan and Serena Ng. 2008. Forecasting Economic Time Series Using Targeted Predictors. *Journal of Econometrics* 146: 304–17. [CrossRef]

Bair, Eric, Trevor Hastie, Debashis Paul, and Robert Tibshirani. 2006. Prediction by Supervised Principal Components. *Journal of the American Statistical Association* 101: 119–37. [CrossRef]

Barndorff-Nielsen, Ole. 1978. *Information and Exponential Families in Statistical Theory*. New York: Wiley.

Bernanke, Ben. 1990. On the Predictive Power of Interest Rates and Interest Rate Spreads, Federal Reserve Bank of Boston. *New England Economic Review* November/December: 51–68.

Chan, Lewis, James Stock, and Mark Watson. 1999. A Dynamic Factor Model Framework for Forecast Combination. *Spanish Economic Review* 1: 91–121. [CrossRef]

Christensen, Jens, Francis Diebold, and Glenn Rudebusch. 2009. An Arbitrage-free Generalized Nelson–Siegel Term Structure Model. *Econometrics Journal* 12: C33–C64. [CrossRef]

De Jong, Sijmen. 1993. SIMPLS: An Alternative Approach to Partial Least Squares Regression. *Chemometrics and Intelligent Laboratory Systems* 18: 251–61. [CrossRef]

De Jong, Sijmen and Henk Kiers. 1992. Principal Covariate Regression: Part I. Theory. *Chemometrics and Intelligent Laboratory Systems* 14: 155–64. [CrossRef]

Diebold, Francis, and Canlin Li. 2006. Forecasting the Term Structure of Government Bond Yields. *Journal of Econometrics* 130: 337–64. [CrossRef]

Diebold, Francis, Monika Piazzesi, and Glenn Rudebusch. 2005. Modeling Bond Yields in Finance and Macroeconomics. *American Economic Review* 95: 415–20. [CrossRef]

Diebold, Francis, Glenn Rudebusch, and Boragan Aruoba. 2006. The Macroeconomy and the Yield Curve: A Dynamic Latent Factor Approach. *Journal of Econometrics* 131: 309–38. [CrossRef]

Engle, Robert, David Hendry, and Jean-Francois Richard. 1983. Exogeneity. *Econometrica* 51: 277–304. [CrossRef]

Estrella, Arturo. 2005. Why Does the Yield Curve Predict Output and Inflation? *The Economic Journal* 115: 722–44. [CrossRef]

Estrella, Arturo, and Gikas Hardouvelis. 1991. The Term Structure as a Predictor of Real Economic Activity. *Journal of Finance* 46: 555–76. [CrossRef]

Fama, Eugene, and Robert Bliss. 1987. The Information in Long-maturity Forward Rates. *American Economic Review* 77: 680–92.

Figlewski, Stephen, and Thomas Urich. 1983. Optimal Aggregation of Money Supply Forecasts: Accuracy, Profitability and Market Efficiency. *Journal of Finance* 38: 695–710. [CrossRef]

Friedman, Benjamin, and Kenneth Kuttner. 1993. Why Does the Paper-Bill Spread Predict Real Economic Activity? In *New Research on Business Cycles, Indicators and Forecasting*. Edited by James Stock and Mark Watson. Chicago: University of Chicago Press, pp. 213–54.

Gogas, Periklis, Theophilos Papadimitriou, and Efthymia Chrysanthidou. 2015. Yield Curve Point Triplets in Recession Forecasting. *International Finance* 18: 207–26. [CrossRef]

Groen, Jan, and George Kapetanios. 2016. Revisiting Useful Approaches to Data-Rich Macroeconomic Forecasting. *Computational Statistics & Data Analysis* 100: 221–39.

Hamilton, James, and Dong Heon Kim. 2002. A Reexamination of the Predictability of Economic Activity Using the Yield Spread. *Journal of Money, Credit, and Banking* 34: 340–60. [CrossRef]

Huang, Huiyu, and Tae-Hwy Lee. 2010. To Combine Forecasts or To Combine Information? *Econometric Reviews* 29: 534–70. [CrossRef]

Inoue, Atsushi, and Lutz Kilian. 2008. How Useful is Bagging in Forecasting Economic Time Series? A Case Study of U.S. CPI Inflation. *Journal of the American Statistical Association* 103: 511–22. [CrossRef]

Kozicki, Sharon. 1997. Predicting Real Growth and Inflation with the Yield Spread, Federal Reserve Bank of Kansas City. *Economic Review* 82: 39–57.

Lancaster, Tony. 2000. The incidental parameter problem since 1948. *Journal of Econometrics* 95: 391–413. [CrossRef]

Litterman, Robert, and Jose Scheinkman. 1991. Common Factors Affecting Bond Returns. *Journal of Fixed Income* 1: 54–61. [CrossRef]

Nelson, Charles, and Andrew Siegel. 1987. Parsimonious Modeling of Yield Curves. *Journal of Business* 60: 473–89. [CrossRef]

Neyman, Jerzy, and Elizabeth Scott. 1948. Consistent Estimation from Partially Consistent Observations. *Econometrica* 16: 1–32. [CrossRef]

Piazzesi, Monika. 2005. Bond Yields and the Federal Reserve. *Journal of Political Economy* 113: 311–44. [CrossRef]

Rudebusch, Glenn, and Tao Wu. 2008. A Macro-Finance Model of the Term Structure, Monetary Policy, and the Economy. *Economic Journal* 118: 906–26. [CrossRef]

Smith, Jeremy, and Kenneth Wallis. 2009. A Simple Explanation of the Forecast Combination Puzzle. *Oxford Bulletin of Economics and Statistics* 71: 331–55. [CrossRef]

Stock, James, and Mark Watson. 1989. New Indexes of Coincident and Leading Indicators. In *NBER Macroeconomic Annual*. Edited by Olivier Blanchard and Stanley Fischer. Cambridge: MIT Press, vol. 4.

Stock, James, and Mark Watson. 1999. Forecasting Inflation. *Journal of Monetary Economics* 44: 293–335. [CrossRef]

Stock, James, and Mark Watson. 2002. Forecasting Using Principal Components from a Large Number of Predictors. *Journal of the American Statistical Association* 97: 1167–79. [CrossRef]

Stock, James, and Mark Watson. 2004. Combination Forecasts of Output Growth in a Seven-country Data Set. *Journal of Forecasting* 23: 405–30. [CrossRef]

Stock, James, and Mark Watson. 2007. Has Inflation Become Harder to Forecast? *Journal of Money, Credit, and Banking* 39: 3–34. [CrossRef]

Stock, James, and Mark Watson. 2012. Generalized Shrinkage Methods for Forecasting Using Many Predictors. *Journal of Business and Economic Statistics* 30: 481–93. [CrossRef]

Svensson, Lars. 1995. Estimating Forward Interest Rates with the Extended Nelson–Siegel Method. *Quarterly Review* 3: 13–26.

Tibshirani, Robert. 1996. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society B* 58: 267–88.

Timmermann, Alan. 2006. Forecast Combinations. In *Handbook of Economic Forecasting*. Edited by Graham Elliott, Clive Granger and Alan Timmermann. Amsterdam: North-Holland, vol. 1, chp. 4.

Wright, Jonathan. 2009. Forecasting US Inflation by Bayesian Model Averaging. *Journal of Forecasting* 28: 131–44. [CrossRef]

Zou, Hui, Trevor Hastie, and Robert Tibshirani. 2006. Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics* 15: 262–86. [CrossRef]

econometrics

MDPI

Article

Covariance Prediction in Large Portfolio Allocation

Carlos Trucíos [1,*], Mauricio Zevallos [2], Luiz K. Hotta [2] and André A. P. Santos [3,4]

[1] São Paulo School of Economics, FGV, São Paulo 01332-000, Brazil
[2] Department of Statistics, University of Campinas, Campinas 13083-859, Brazil; amadeus@ime.unicamp.br (M.Z.); hotta@ime.unicamp.br (L.K.H.)
[3] UC3M-Santander Big Data Institute, Universidad Carlos III de Madrid, Getafe 28903, Spain; andre.portela@ufsc.br
[4] Department of Economics, Universidade Federal de Santa Catarina, Florianópolis 88040-970, Brazil
* Correspondence: ctrucios@gmail.com

Received: 12 November 2018; Accepted: 2 May 2019; Published: 9 May 2019

Abstract: Many financial decisions, such as portfolio allocation, risk management, option pricing and hedge strategies, are based on forecasts of the conditional variances, covariances and correlations of financial returns. The paper shows an empirical comparison of several methods to predict one-step-ahead conditional covariance matrices. These matrices are used as inputs to obtain out-of-sample minimum variance portfolios based on stocks belonging to the S&P500 index from 2000 to 2017 and sub-periods. The analysis is done through several metrics, including standard deviation, turnover, net average return, information ratio and Sortino's ratio. We find that no method is the best in all scenarios and the performance depends on the criterion, the period of analysis and the rebalancing strategy.

Keywords: Minimum variance portfolio; risk; shrinkage; S&P 500

JEL Classification: C13; C53; C58; G11

1. Introduction

Forecasting returns, volatilities and conditional correlations has attracted the interest of researchers and practitioners in finance since these factors are crucial, for example, in portfolio allocation, risk management, option pricing and hedging strategies; see, for instance, Engle (2009), Hlouskova et al. (2009) and Boudt et al. (2013) for some references.

A well-known stylised fact in multivariate time series of financial returns is that not only conditional variances but also conditional covariances and correlations evolve over time. To describe this evolution, several methods have been proposed in the literature. In general, these methods involve different ways to circumvent the issue of dimensionality. The treatment of this problem is vital for the estimation of large portfolios (composed of hundreds or thousands of assets). As noted by Engle et al. (2017), when dealing with portfolios composed of a thousand time series, many multivariate GARCH models present unsatisfactory performance or computational problems in their estimation. For some multivariate GARCH models, estimation problems arise even for smaller dimensions; see, for instance, Laurent et al. (2012), Caporin and McAleer (2014), Caporin and Paruolo (2015) and de Almeida et al. (2018).

Our empirical application is based on an investor who adopts the minimum variance criterion in order to decide on portfolio allocations. A very large body of literature in portfolio optimization considers this particular policy; see, for instance, Clarke et al. (2011 2006) for extensive practitioner-oriented studies on the performance and composition of minimum variance portfolios. This policy can be seen as a particular case of the traditional mean-variance optimisation.

The mean-variance problem, however, is known to be very sensitive to estimation of the mean returns (Frahm 2010; Jagannathan and Ma 2003).[1] Very often, the estimation error in the mean returns degrades the overall portfolio performance and introduces an undesirable level of portfolio turnover. In fact, existing evidence suggests that the performance of optimal portfolios that do not rely on estimated mean returns is usually better, see DeMiguel et al. (2009).

To obtain the minimum variance portfolio, the key input is the estimate of the conditional covariance matrix. As far as we known, there are few works in the literature comparing the estimation of this matrix for large portfolios, with Creal et al. (2011), Hafner and Reznikova (2012), Engle et al. (2017), Nakagawa et al. (2018) and Moura and Santos (2018) being especially relevant. Given the myriad of models and methods in the literature to estimate the covariance matrix, empirical studies about the comparison of estimates in large portfolios are most welcome.

The paper is intended to assess the performance of several methods to predict one-step-ahead conditional covariance matrices in large portfolios. This is done empirically, by comparing the out-of-sample performance of minimum variance portfolios based on S&P500 stocks traded from 2 January 2000 to 30 November 2017, using measures such as average (AV), standard deviation (SD), information ratio (IR), Sortino's ratio (SR) (Sortino and van der Meer 1991), turnover (TO) and average portfolio net of transaction cost (AV^{net}). Since not all stocks of the index were traded during the whole period, we consider portfolios of dimension $N = 174$ stocks. To assess the robustness of the results, we also the analyse three sub-periods: the pre-crisis period (January 2004 to December 2007), the subprime crisis period (January 2008 to June 2009), and the post-crisis period (July 2009 to November 2017).

We consider several attractive methods and models including recent proposals used by practitioners and academics to predict one-step-ahead conditional covariance matrices. They are selected mainly because they use different approaches to overcome the issue of dimensionality problem. Specifically, the paper compares the DCC model as used in Engle et al. (2017), the DECO model of Engle and Kelly (2012), the OGARCH model of Alexander and Chibumba (1996), the RiskMetrics 1994 and the RiskMetrics 2006 (Zumbach 2007) methods, the generalised principal volatility components analysis (GPVC) proposed by Li et al. (2016) as a generalisation of the procedure of Hu and Tsay (2014), and we also apply the robust version of the GPVC method proposed by Trucíos et al. (2019). DCC models are estimated using composite likelihood, as advocated in Pakel et al. (2014). In addition, the linear shrinkage (LS) and non-linear shrinkage (NLS) of Ledoit and Wolf (2004a) and Ledoit and Wolf (2012), respectively, are applied on all the previous methods. Therefore, compared to Engle et al. (2017), Hafner and Reznikova (2012) and Nakagawa et al. (2018), the set of competing methods is much bigger and the device of shrinkage is assessed in all the compared methods. We consider a total of 47 methods, including the equal-weighted portfolio strategy. This constitutes the main contribution of the paper.

The rest of the paper is organised as follows: Section 2 presents the methods and models used to predict the one-step-ahead volatility covariance matrix. It also presents the composite likelihood used to estimate the DCC model and the shrinkage method as presented in Pakel et al. (2014). The empirical application is given in Section 3. Section 4 concludes and the list of the estimation methods is in the Appendix A.

2. The Forecast Methods

Denote by $r_{i,t}$, $i = 1, \ldots, N$, $t = 1, \ldots, T$ the return of the i-th asset at time t, where N is the number of assets under consideration to construct the portfolio and T denotes the sample size. For simplicity, consider that $E(r_{i,t}|\mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_{t-1}$ denotes the information available at time $(t-1)$. Let $\mathbf{r}_t = (r_{1,t}, \ldots, r_{N,t})'$; the conditional covariance matrix is defined as $\mathbf{H}_t = \mathrm{Cov}(\mathbf{r}_t|\mathcal{F}_{t-1})$

[1] See Wied et al. (2013) for a test for the presence of structural breaks in minimum variance portfolios

with elements $h_{i,j,t} = \mathrm{Cov}(\mathrm{r_{i,t}}, \mathrm{r_{j,t}}|\mathcal{F}_{\mathrm{t}-1})$. At time $(t-1)$, we are interested in estimating $\mathbf{H}_t$ in order to select a portfolio for the period $(t-1, t]$. In the following we present some methods to estimate it.

2.1. The RiskMetrics Methods

One of the most popular methods used in risk analysis is the RiskMetrics method developed by the RiskMetrics Group at JP Morgan. We call this the RiskMetrics 1994 (RM1994) method. The main feature of the RiskMetrics method is that the predicted volatility is a linear function of the present and past squared returns. Although it has being widely used, it has some problems. In order to overcome some of these problems, the same group developed the RM2006 method. Like the RM1994 method, the RM2006 method is also data-oriented, in the sense that it was calibrated and tested to have good performance with the majority of the target empirical data, and was developed to take into account some of the stylised facts and weaknesses detected in the RM1994 method. We can summarize the main modifications in three types. In the first type, considering that the volatility has a long memory feature, the weights decay logarithmically instead of exponentially, as happens in the RM1994 method. The second is that the weights depend on the forecast horizon. The third is that the conditional distribution of the return is not multivariate Gaussian; the distribution is based on the estimated devolatilised residuals and it can be roughly defined as a Student-t distribution with scale correction. Finally, the return levels are modelled considering the lagged correlation between returns.

2.2. The CCC Model

The constant conditional correlation model (Bollerslev 1990) is one of the simplest MGARCH models to estimate, since basically the variances are modelled independently and the covariances are obtained using the conditional standard deviation and a constant conditional correlation matrix. The conditional covariance matrix $\mathbf{H}_t$ evolves according to:

$$\begin{aligned}
\mathbf{H}_t &= \mathbf{D}_t\mathbf{R}\mathbf{D}_t, && (1)\\
\mathbf{D}_t &= Diag(d_{1,t}, \ldots, d_{N,t}), && (2)\\
\mathbf{R} &= Diag(\mathbf{H})^{-1/2}\mathbf{H}Diag(\mathbf{H})^{-1/2}, && (3)\\
\mathbf{H} &= \mathrm{Cov}(\mathrm{r_t}), && (4)
\end{aligned}$$

with $d_{i,t}^2 = Var(r_{i,t}|\mathcal{F}_{t-1})$ (marginal univariate conditional variances). The advantage of the CCC model is its easy estimation, although, the main disadvantage is the strong assumption that conditional correlations are time-invariant. Engle (2002) extended this idea in a dynamic conditional correlation way, as detailed in the next section.

2.3. The DCC Model

In this section, we describe the scalar DCC model of Engle (2002) as used in Pakel et al. (2014) and Engle et al. (2017), and the composite likelihood. The non-linear shrinkage method, which is also used to estimate the DCC model, is presented in Section 2.8. In the DCC model, the marginal univariate conditional variances $d_{i,t}^2 = Var(r_{i,t}|\mathcal{F}_{t-1})$ are modelled first. Define the devolatilised residuals as $\mathbf{s}_t = (r_{1,t}/d_{1,t}, \ldots, r_{N,t}/d_{N,t})'$. We use the DCC model with correlation targeting as in Engle et al. (2017). The conditional covariance matrix $\mathbf{H}_t$ evolves according to:

$$\begin{aligned}
\mathbf{H}_t &= \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t, && (5)\\
\mathbf{R}_t &= Diag(\mathbf{Q}_t)^{-1/2}\mathbf{Q}_t Diag(\mathbf{Q}_t)^{-1/2}, && (6)\\
\mathbf{Q}_t &= (1-\alpha-\beta)\mathbf{C} + \alpha\mathbf{s}_{t-1}\mathbf{s}_{t-1}^{'} + \beta\mathbf{Q}_{t-1}, && (7)
\end{aligned}$$

where $\mathbf{D}_t$ is a diagonal matrix with the i-th element of the diagonal equal to $d_{i,t}^2$, $C = Corr(\mathbf{r}_t) = Cov(\mathbf{s}_t)$ is the unconditional correlation matrix, and $R_t = Corr(\mathbf{r}_t|\mathcal{F}_{t-1}) = Cov(\mathbf{s}_t|\mathcal{F}_{t-1})$ is the conditional correlation matrix at time t. The parameters α and β are non-negative with $\alpha + \beta < 1$. We have

$$\mathbf{r}_t|\mathcal{F}_{t-1} \sim WS(0, \mathbf{H}_t), \tag{8}$$

where $WS(0, \mathbf{H}_t)$ means a multivariate distribution with mean zero and covariance matrix $\mathbf{H}_t$.

The model is usually estimated in three stages. In each stage, the estimation is conditional on the estimates found in previous stages. The stages are: (1) estimate $\mathbf{D}_t$ usually assuming a GARCH(1,1) model for each $t = 1, \ldots, T$, and evaluate the devolatilised residuals; (2) select an estimator of the correlation target matrix C using the devolatilised residuals; and (3) estimate the parameters α and β. We will comment on stage one in the application section and on stage 2 in Section 2.8. In the third stage, even with only two parameters, one may face estimation problems with a large number of assets because it is necessary to invert the conditional covariance matrix $\mathbf{H}_t$ (for each $t = 1, \ldots, T$). One way to overcome this problem is through the use of the composite (log-)likelihood[2] to compute it. This method was proposed in the 2008 version of Pakel et al. (2014). In the 2014 version, they showed that the estimators of α and β, given by maximizing the composite likelihood, are consistent although not efficient. They evaluate the composite likelihood by summing the likelihood of all contiguous pairs. Thus, there are only $(N-1)$ bivariate terms and for any contiguous pair it is only necessary to invert a matrix of order two. For instance, let $\mathbf{r}^{(i)} = (r_{i,1}, \ldots, r_{i,T})'$, $i = 1, \ldots, N$, i.e., the series of returns of the ith asset, and denote by $l_i(\alpha, \beta; \mathbf{r}^{(i)}, \mathbf{r}^{(i+1)})$ the likelihood of the pair $(\mathbf{r}^{(i)}, \mathbf{r}^{(i+1)})$, $i = 1, \ldots, N-1$, assuming that each pair comes from a bivariate DCC model, defined similarly as the model given by Equations (5–7). Then, the composite likelihood is given by:

$$CL(\alpha, \beta; \mathbf{r}^{(i)}, i = 1, \ldots, N) = \sum_{i=1}^{N-1} l_i(\alpha, \beta; \mathbf{r}^{(i)}, \mathbf{r}^{(i+1)}). \tag{9}$$

Engle et al. (2017) argue that the estimator of the conditional covariance matrix given by the DCC model using composite likelihood in stage three with the estimation of the unconditional correlation matrix using non-linear shrinkage in stage two is robust against model misspecification in large dimensions (N).

2.4. The DECO Model

Engle and Kelly (2012) propose a dynamic equicorrelation (DECO) model as a trade-off between a model which imposes many restrictions in the covariance matrix and a less structured model. They contend that imposing too much structure can lead to an efficient estimation when the restrictions are correct, but can suffer from breakdown in the presence of misspecification. On the other hand, the lack of restrictions may lead to the issue of dimensionality. Considering this trade-off, they propose a model where the cross-correlations between any pair of returns are equal on the same day, but it can vary over time. In addition, as in the CCC and DCC models, the DECO model also assumes that the marginals are modelled by a univariate volatility model. Using the same notation, we have $d_{i,t}^2 = Var(r_{i,t}|\mathcal{F}_{t-1})$, and the covariance matrix is written as $\mathbf{H}_t = \mathbf{D}_t\mathbf{R}_t\mathbf{D}_t$ as in Equation (5). The equicorrelation matrix is given by:

$$\mathbf{R}_t = (1 - \rho_t)\mathbf{I}_N + \rho_t\mathbf{J}_N, \tag{10}$$

where ρ_t is the equicorrelation, $\mathbf{I}_N$ denotes the N-dimensional identity matrix and $\mathbf{J}_N$ is the $N \times N$ matrix of ones. According to Engle and Kelly (2012), $\mathbf{R}_t^{-1}$ exist if and only if $\rho_t \neq 1$ and

[2] From now on we just call the log-likelihood likelihood.

$\rho_t \neq -1/(N-1)$, and $\mathbf{R}_t$ is positive definite if and only if $\rho_t \in (-1/(N-1), 1)$. The evaluation of the likelihood is easy because we have closed forms for $\mathbf{R}_t^{-1}$ and $\det(\mathbf{R}_t)$, given by:

$$\mathbf{R}_t^{-1} = \frac{1}{1-\rho_t}\mathbf{I}_N - \frac{\rho_t}{(1-\rho_t)(1+[N-1]\rho_t)}\mathbf{J}_N, \tag{11}$$

and

$$\det(\mathbf{R}_t) = (1-\rho_t)^{N-1}\left[1+(N-1)\rho_t\right], \tag{12}$$

respectively. This description of the DECO model corresponds to a single block. The DECO model can also be used considering many blocks, as described in Engle and Kelly (2012).

2.5. The OGARCH Model

Alexander and Chibumba (1996) propose the Orthogonal GARCH (OGARCH) model, a dimension reduction technique to model the conditional covariance matrix. The model intends to simplify the problem of modelling an N-dimensional system into modelling a system of K-dimension orthogonal components where those components are obtained through principal component analysis ($K \leq N$). Since the components are orthogonal, the conditional covariance matrix of the whole system can be obtained as:

$$\mathbf{H}_t = \mathbf{A}\mathbf{D}_t\mathbf{A}' + \mathbf{V}_\epsilon, \tag{13}$$

where $\mathbf{A}$ is an $N \times k$ matrix whose columns are the normalised eigenvectors associated with the unconditional covariance matrix, $\mathbf{D}_t$ is a diagonal matrix whose elements are the conditional variances of the k principal orthogonal components associated with the k largest eigenvalues, and $\mathbf{V}_\epsilon$ is the covariance matrix of the errors that can be ignored. The conditional variances of each component can be modelled by a GARCH-type model.

Alexander and Chibumba (1996) and Alexander (2002) emphasise the importance of using a number of components k much smaller than N. However, Bauwens et al. (2006) and Becker et al. (2015) suggest using $k = N$ to avoid problems related with the inverse of $\mathbf{H}_t$. The OGARCH model with $k = N$ is a particular case of the GO-GARCH model (Van der Weide 2002).

2.6. The Generalised Principal Volatility Components Model

The generalised principal volatility components (GPVC) procedure is a dimension reduction technique recently proposed by Li et al. (2016), which decomposes a series into two groups of volatility components. The first group corresponds to a small number of components with volatility evolving over time while the second one corresponds to components whose volatility is constant over time. The GPVC procedure considers an orthogonal matrix $\mathbf{M} = [\mathbf{A} : \mathbf{B}]$ and decomposes an N-dimensional vector $\mathbf{y}_t = (y_{1t}, ..., y_{Nt})'$ with $E(\mathbf{y}_t|\mathcal{F}_{t-1}) = 0$ into:

$$\mathbf{y}_t = \mathbf{M}\mathbf{M}'\mathbf{y}_t = (\mathbf{A}\mathbf{A}' + \mathbf{B}\mathbf{B}')\mathbf{y}_t = \mathbf{A}\mathbf{f}_t + \mathbf{ffl}_t, \tag{14}$$

with $\mathbf{f}_t = \mathbf{A}'\mathbf{y}_t$ and $\mathbf{ffl}_t = \mathbf{B}\mathbf{B}'\mathbf{y}_t$. The matrix $\mathbf{M}$ is obtained through the decomposition $\mathbf{G}\mathbf{M} = \mathbf{\Lambda}\mathbf{M}$, where $\mathbf{\Lambda}$ is a diagonal matrix with elements given by the eigenvalues in decreasing order and $\mathbf{M}$ is the associated matrix of normalised eigenvectors. The columns of matrices $\mathbf{A}$ and $\mathbf{B}$ are the eigenvectors associated with the non-zero and zero eigenvalues, respectively, which are obtained from the eigenvalue decomposition of the matrix $\mathbf{G}$. In practice, $\mathbf{G}$ is given by:

$$\mathbf{G} = \sum_{k=1}^{g}\sum_{t=1}^{T}\omega(\mathbf{y}_t)E^2\left[\left(\mathbf{y}_t\mathbf{y}_t' - \mathbf{\Sigma}\right)I(\|\mathbf{y}_{t-k}\| \leq \|\mathbf{y}_t\|)\right], \tag{15}$$

where g is a positive integer that gives the maximum lag order considered, $\omega(\cdot)$ is a weight function, $\mathbf{\Sigma}$ is the unconditional covariance matrix and $\|\cdot\|$ is the L_1 norm. Then, after some calculations, the conditional covariance matrix can be obtained by:

$$\mathbf{H}_t = \mathbf{A}\mathbf{H}_t^f\mathbf{A}' + \mathbf{A}\mathbf{A}'\mathbf{\Sigma}\mathbf{B}\mathbf{B}' + \mathbf{B}\mathbf{B}'\mathbf{\Sigma}, \tag{16}$$

where $\mathbf{H}_t^f$ is the conditional covariance matrix of the volatility components with volatility evolving over time and the remaining are terms as defined previously[3]. The matrix $\mathbf{G}$ is estimated as:

$$\hat{\mathbf{G}} = \sum_{k=1}^{g}\sum_{\tau=1}^{T}\omega(\mathbf{y}_\tau)\left[\frac{1}{T-k}\sum_{t=k+1}^{T}\left[\left(\mathbf{y}_t\mathbf{y}_t' - \hat{\mathbf{\Sigma}}\right)I(\|\mathbf{y}_{t-k}\| \leq \|\mathbf{y}_\tau\|)\right]\right]^2. \tag{17}$$

The estimated version of Equation (16) is obtained by replacing the true values with the estimated ones.

2.7. The Robust GPVC Model

Trucíos et al. (2019) show the non-robustness of the GPVC procedure of Li et al. (2016) and propose an alternative procedure to obtain volatility components that is robust to outliers. This procedure is based on a robust estimator of the unconditional covariance matrix, a weighted estimator of $E\left[\left(\mathbf{y}_t\mathbf{y}_t' - \mathbf{\Sigma}\right)I(\|\mathbf{y}_{t-k}\| \leq \|\mathbf{y}_t\|)\right]$, and robustified filters. The matrix (17) is replaced by a less sensitive matrix, defined as:

$$\hat{\mathbf{G}}^R = \sum_{k=1}^{g}\sum_{\tau=1}^{T}\omega(\mathbf{y}_\tau)\left[\sum_{t=k+1}^{T}\delta^*(d_t^2)\left\{\left(\mathbf{y}_t\mathbf{y}_t' - \hat{\mathbf{\Sigma}}^R\right)I(\|\mathbf{y}_{t-k}\| \leq \|\mathbf{y}_\tau\|)\right\}\right]^2, \tag{18}$$

where d_t^2 is the robust squared Mahalanobis distance given by $d_t^2 = (\mathbf{y}_t - \hat{}^R)'\hat{\mathbf{\Sigma}}_t^{-1}(\mathbf{y}_t - \hat{}^R)$ with $\hat{\mathbf{\Sigma}}_t = 0.01\rho(\mathbf{y}_{t-1}'\mathbf{y}_{t-1}) + 0.99\hat{\mathbf{\Sigma}}_{t-1}$, $\hat{\mathbf{\Sigma}}_1 = \hat{\mathbf{\Sigma}}^R$ and $\hat{}^R$, $\hat{\mathbf{\Sigma}}^R$ being robust estimates of the unconditional mean and covariance matrix. Trucíos et al. (2019) use the minimum covariance determinant (MCD) estimator of Rousseeuw (1984), implemented by the algorithm of Hubert et al. (2012). The robust filters, $\rho(\cdot)$ and $\delta(\cdot)$ are given by $\rho(x_t) = x_t$ if $d_t^2 \leq c$, $\rho(x_t) = \hat{\Sigma}^R$ if $d_t^2 > c$; $\delta(x) = 1$ if $x \leq c$, $\delta(x) = 1/x$ if $x > c$ and $\delta^*(\cdot) = \delta(\cdot)/||\delta(\cdot)||$, where $\|\cdot\|$ is the L_1 norm. For details, see Trucíos et al. (2019).

To avoid returns corresponding to periods with high volatility being considered as possible outliers, the robust procedure incorporates in the squared Mahalanobis distance a covariance matrix evolving over time, which can be seen as a robust RM1994 method with $\lambda = 0.99$.

Finally, the conditional covariance matrix $\mathbf{H}_t$ is obtained as in Equation (16).

2.8. Linear and Non-Linear Shrinkage

Besides the estimation of the covariance matrix ($\mathbf{H}_t$), in some of the aforementioned models, we have to estimate the unconditional covariance or correlation matrix; for instance, the matrix $\mathbf{C}$ in Equation (7) of the DCC model. Generally, the estimation of the unconditional correlation (covariance) matrix is done using the sample correlation (covariance) matrix. However, this is inefficient in the large dimensional case because we could end up with a number of parameters with the same order of magnitude as the dataset, or even larger (see, for instance, the simulation study in the Appendix of Engle et al. (2017)). In general, comparing the eigenvalues of the true correlation matrix with the eigenvalues of the sample correlation matrix, there is a tendency to underestimate the smaller eigenvalues and overestimate the larger ones. A natural way to reduce this bias is to increase the smaller eigenvalues and decrease the larger sample eigenvalues and then reconstruct the estimate of the

[3] Note that when $\mathbf{\Sigma} = \mathbf{I}$, $\mathbf{H}_t = \mathbf{A}\mathbf{H}_t^f\mathbf{A}' + \mathbf{B}\mathbf{B}'\mathbf{\Sigma} = \mathbf{A}\mathbf{H}_t^f\mathbf{A}' + \mathbf{\Sigma}_{\mathbf{ffl}}$ as presented in Li et al. (2016).

correlation matrix. This is the main idea behind the shrinkage method. Engle et al. (2017) analyse the use of three types the shrinkage: linear shrinkage of Ledoit and Wolf (2004b) with shrinkage target given by (a multiple of) the identity matrix; linear shrinkage of Ledoit and Wolf (2004a) with shrinkage target given by the equicorrelation matrix; and the non-linear shrinkage of Ledoit and Wolf (2012) for the estimation of the unconditional correlation matrix in Equation (7). Using simulation, they conclude that the three types of shrinkage have better performance than the use of the sample correlation matrix in the estimation of $\mathbf{H}_t$, and the best performance is obtained from the non-linear shrinkage. They conclude that the application of non-linear shrinkage improves the estimation, and the improvement generally increases for a larger number of assets. In the application, they also apply the non-linear shrinkage to the estimated one-step-ahead conditional covariance matrix, which is not done in the simulation study. In the empirical application, they construct portfolios of global minimum variance with portfolio sizes 100, 500 and 1000 and updated monthly. As in the simulation study, they construct portfolios with H_t modelled by DCC and CCC models and the RiskMetrics 2006 method. However, besides applying the linear and non-linear shrinkage to the target correlation matrix, they also apply the shrinkages to the one-step-ahead prediction of the volatility matrix. The best performance is achieved by the DCC model with the non-linear shrinkage applied only to the estimation of the intercept matrix, followed by the non-linear shrinkage applied both to the intercept matrix and to the one-step-ahead prediction matrix. We use the linear shrinkage towards the equicorrelation matrix, because in Engle et al. (2017) it presented slightly better performance than the shrinkage towards the identity matrix, although the estimator does not belong to the class of rotation-equivariant estimators.

For a light introduction to the main idea behind shrinkage, suppose we want to estimate the covariance matrix $\mathbf{\Sigma}$ and we have an estimate $\hat{\mathbf{C}}$ based on a sample of size T. For instance, $\hat{\mathbf{C}}$ could be the sample covariance matrix and $\mathbf{\Sigma}$, the population matrix (unconditional covariance matrix). This is the case of the estimation of the DCC, where $\mathbf{\Sigma}$ is the intercept matrix. When the ratio N/T, called concentration ratio, becomes large, we have in-sample overfitting due to the excessive number of parameters, introducing a bias in the estimation of the eigenvalues. One way to correct this problem is through the shrinkage method.

For the linear shrinkage towards the equicorrelation matrix, denote by $\hat{c}_{ij}$ the element of the estimate $\hat{C}$. The mean of the estimated correlations is given by:

$$\bar{r} = \frac{2}{(N-1)N} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \frac{\hat{c}_{i,j}}{\sqrt{\hat{c}_{i,i}\hat{c}_{j,j}}}, \tag{19}$$

such that for the target matrix F we have $f_{i,i} = \hat{c}_{i,i}$ and $f_{i,j} = \bar{r}\sqrt{\hat{c}_{i,i}\hat{c}_{j,j}}$. The shrinkage estimate is given by:

$$\hat{\mathbf{\Sigma}}_{Shrink} = \delta F + (1-\delta)\hat{C}, \tag{20}$$

where the shrinkage intensity, δ, is such that it minimizes the expected quadratic loss as in Ledoit and Wolf (2004a). For the shrinkage intensity δ, define the quadratic loss function

$$L(\delta) = ||\delta F + (1-\delta)\hat{C} - \Sigma||^2.$$

Ledoit and Wolf (2004a) propose to use the shrinkage intensity, which minimizes the risk function $R(\delta) = E(L(\delta))$. The formulae and the derivation of the estimated shrinkage intensity can be found in the Appendix B of Ledoit and Wolf (2004a).

Regarding the non-linear shrinkage, let $\hat{C}$ having dimension $(N \times N)$, $(\hat{\lambda}_1, \ldots, \hat{\lambda}_N)$, sorted in descending order, be the set of eigenvalues, and $(\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_N)$ the corresponding eigenvectors, such that:

$$\hat{C} = \sum_{i=1}^{N} \hat{\lambda}_i \hat{\mathbf{u}}_i \hat{\mathbf{u}}_i'. \tag{21}$$

For an investor holding a portfolio with weights ω, the estimated variance is given by $\omega'\hat{C}\omega$. The non-linear shrinkage of Ledoit and Wolf (2004b) is a transformation from $(\hat{\lambda}_1, \ldots, \hat{\lambda}_N)$ to $\tilde{\lambda} = (\tilde{\lambda}_1, \ldots, \tilde{\lambda}_N)$, such that substituting $\hat{\lambda}_i$ for $\tilde{\lambda}_i$ in Equation (21) gives a consistent estimator of the out-of-sample variance $\omega'\Sigma\omega'$. Denote by $\lambda = (\lambda_1, \ldots, \lambda_N)$ the set of eigenvalues of Σ in descending order. Ledoit and Wolf (2004b) define QuEST functions $(q_1(\lambda), \ldots, q_N(\lambda)$, such that $\tilde{\lambda}$ minimizes the Euclidean distance between the QuEST functions and the sample eigenvalues, i.e., given by:

$$\tilde{\lambda} = \arg \min_{\lambda \in [0,\infty)^N} \sum_{i=1}^{N} [q_i(\lambda) - \hat{\lambda}_i]^2. \tag{22}$$

A definition of the QuEST functions and a rigorous exposition of non-linear shrinkage can be found in Ledoit and Wolf (2012), while a lighter presentation can be found in the Supplementary Material of Engle et al. (2017).

3. Empirical Application

3.1. Data and Methods

In this section, we implement the procedures described in Section 2 and use the predicted one-step-ahead conditional covariance matrix to construct the minimum variance portfolio (MVP) of the stocks used in the composition of the S&P 500 index, traded from 2 January 2000 to 30 November 2017. Because not all stocks of the index were traded during the whole period, we ended up with $N = 174$ stocks.

To evaluate the out-of-sample portfolio performance, we consider a rolling window scheme. The out-of-sample portfolio performance is evaluated in four different periods, namely: pre-crisis period (January 2004 to December 2007, 1008 days), subprime crisis period (January 2008 to June 2009, 378 days), post-crisis period (July 2009 to November 2017, 2218 days), and full period (January 2004 to November 2017, 3503 days). In each window, the one-step-ahead covariance matrix is estimated and the MVP values with and without short-sale constraints are obtained. The weights in the MVP portfolio are rebalanced with both daily and monthly frequencies. In the latter case, we follow Engle et al. (2017), that is, we obtain the portfolio returns daily but update the weights monthly (following the common convention we use 21 consecutive trading days as a month). Monthly updating is common in practice to reduce transaction costs.

The procedures described in Section 2 are combined with the linear and non-linear shrinkage estimator described in Subsection 2.8. The linear and non-linear shrinkage are applied at the beginning and/or at the end of the estimation procedure. A detailed description of each combination of the estimation procedures is given in the Appendix A. In addition, for the sake of comparison, we also implement the naive equal-weighted portfolio. In the line of Engle et al. (2017), Gambacciani and Paolella (2017), Trucíos et al. (2018) among others, we consider the following annualised out-of-sample performance measures. Denote by $R_p = \{r_{p,1}, \ldots, r_{p,k}\}$ the observed out-of-sample returns from a given method where k in the length of the out-of-sample period. The measures considered in this paper: the annualised average portfolio return (AV), standard deviation portfolio return (SD), information ratio (IR), Sortino's ratio (SR) and average turnover (TO) are computed as follows:

AV: equal to $252 \times \bar{R}_p$, where $\bar{R}_p$ is the average of the elements of R_p.
SD: equal to $\sqrt{252} \times S_p$, where S_p is the standard deviation of the elements of R_p.
IR: AV/SD.
SR: AV$/\sqrt{252 \times S^{*2}}$, where S^{*2} is the mean of $r^*_{p,i}, i = 1, \ldots, k$, with $r^*_{p,i} = r^2_{p,i}$ if $r_{p,i}$ less than the minimal acceptable return, which is taken to zero, and zero otherwise.
TO: $k^{-1} \sum_{t=2}^{k} \sum_{j=1}^{N} |\omega_{j,t} - \omega_{j,t-1}|$ where $\omega_{j,t}$ is the portfolio weight at time t for the j-th asset, and k is the number of the out-of-sample portfolio returns.

As pointed out by Kirby and Ostdiek (2012), Santos and Ferreira (2017), Olivares-Nadal and DeMiguel (2018), among others, transaction costs (c) can have an impact on the portfolio's performance. In order to take into account those costs, we also compute the portfolio returns net of transaction cost. For a given c, the portfolio return net of transaction costs at time t is given by $r_{p,t}^{net} = (1 - c \times turnover_t)(1 + r_{p,t}) - 1$ and then the annualised average portfolio return net of transaction costs is $\text{AV}^{net} = 252 \times \bar{R}_p^{net}$ where $\bar{R}_p^{net}$ is the average of the portfolio return net of transaction costs $r_{p,1}^{net}, \ldots, r_{p,k}^{net}$. We consider $c = 20bp$ (intermediate) and $c = 50bp$ (high level) transaction costs where a basis point (bp) is a unit of measure commonly used in finance and is equivalent to 0.01%. The annualised average portfolio return net of transation costs considering $c = 20bp$ and $c = 50bp$ are denoted by AV_{20bp}^{net} and AV_{50bp}^{net}, respectively.

3.2. Results

Tables 1–8 report annualised out-of-sample performance measures for MVP with performance for the pre-crisis, crisis, post-crisis and full periods. Tables 1–4 report the results for daily rebalanced portfolios whereas Tables 5–8 report the results for monthly rebalanced portfolios. We also have results for MVP with no short-sale constraints. However, in this paper we focus on the results for MVP with short-sale constraints and give a short summary of the main findings for the case without short-sale constraints. A detailed analysis of the case without short-sale constraints is given in the Supplementary Material.

In Tables 1–8 we report (in parentheses) the rank of the methods according to the SD criterion in the second column. Moreover, for each criterion, the best five methods are highlighted in shadowed cells. The equal-weighted portfolio strategy is represented by $1/N$.

Taking into account the fact that portfolios are chosen in order to have the minimum variance, the analysis is first done according to the SD criterion. For portfolios rebalanced daily or monthly, the largest SD is reported by the equal-weight portfolio strategy. For portfolios rebalanced daily (Tables 1–4), the five smallest SDs are obtained by the DCC based-methods, except in the crisis period, in which case the five smallest SDs are spread among the DCC, OGARCH and GPVC based-methods. In the crisis-period, the smallest SD is obtained by the GPVC procedure with the non-linear shrinkage applied to the one-step-ahead conditional covariance matrix. For portfolios rebalanced monthly (Tables 5–8), the smallest SDs are obtained by the RM2006-LS[4], NLS-DCC, NLS-GPVC and RM2006-LS procedures for the full, pre-crisis, crisis and post-crisis periods, respectively.

The best performance in terms of the AV criterion differs depending on the period and rebalance strategy. For instance, for daily rebalancing the best performance in the full period is achieved by the RPVC followed by the RPVC with non-linear shrinkage applied to the one-step-ahead conditional covariance matrix. However, for the pre-crisis, crises and post-crisis periods, the best performance is achieved by the OGARCH with non-linear shrinkage applied to the unconditional covariance matrix (NLS-OGARCH), RPVC with linear shrinkage applied to the one-step-ahead conditional covariance matrix (RPVC-LS) and RiskMetrics method with linear shrinkage applied to the one-step-ahead conditional covariance matrix (RM1994-LS), respectively. For monthly rebalancing, the best performances in the full, pre-crisis, crisis and post-crisis periods are achieved by the RPVC, OGARCH-NLS, GPVC-LS and equal-weight portfolio strategy, respectively.

In terms of average turnover, the five smallest average turnovers are in the OGARCH and GPVC groups, with the best performance being achieved by the OGARCH with non-linear shrinkage applied to the one-step-ahead conditional covariance matrix in almost all cases. The only two exceptions are observed in the crisis period, in which case the best performance is achieved by the GPVC procedure with non-linear shrinkage applied to the one-step-ahead conditional covariance matrix.

[4] The acronyms are described in the Appendix A.

Additionally, note that regardless of whether portfolio is rebalanced daily or monthly, the average turnover reported by all dimension reduction techniques is smaller than reported by the non-dimension reduction procedures.

Table 1. Annualised performance measures: AV, SD, IR, SR and TO stand for the average, standard deviation, information ratio, Sortino's ratio and turnover of the out-of-sample MVP returns. AV^{net}_{20bp} and AV^{net}_{50bp} stand for the average out-of-sample MVP return net of transaction costs considering 20 and 50 basis-points, respectively. Period January 2004 to November 2017. The shaded cells denote the top five for each criterion. Weights are rebalanced on a daily basis considering short-selling constraints.

	AV	**SD**	**IR**	**SR**	**TO**	$\mathbf{AV^{net}_{20bp}}$	$\mathbf{AV^{net}_{50bp}}$
1/N	8.302	20.058 (47)	0.414	0.570	-	-	-
CCC	7.706	11.839 (12)	0.651	0.890	0.297	7.509	7.279
CCC LS	7.004	11.881 (14)	0.590	0.807	0.307	6.815	6.578
CCC NLS	7.876	11.932 (17)	0.660	0.905	0.277	7.685	7.470
LS CCC	7.506	11.816 (11)	0.635	0.868	0.302	7.311	7.078
NLS CCC	7.345	11.809 (10)	0.622	0.848	0.298	7.153	6.923
LS CCC LS	6.628	11.918 (16)	0.556	0.759	0.305	6.439	6.205
NLS CCC NLS	7.522	11.910 (15)	0.632	0.865	0.303	7.327	7.091
DCC	7.737	11.613 (2)	0.666	0.908	0.308	7.532	7.296
DCC LS	6.941	11.689 (5)	0.594	0.810	0.314	6.749	6.508
DCC NLS	7.711	11.695 (6)	0.659	0.905	0.285	7.513	7.292
LS DCC	7.707	11.613 (1)	0.664	0.904	0.308	7.502	7.266
NLS DCC	7.629	11.616 (3)	0.657	0.894	0.307	7.424	7.188
LS DCC LS	6.907	11.688 (4)	0.591	0.806	0.314	6.715	6.474
NLS DCC NLS	7.645	11.699 (7)	0.653	0.896	0.283	7.447	7.227
RM2006	8.649	11.809 (9)	0.732	0.995	0.271	8.446	8.234
RM2006 LS	8.746	11.724 (8)	0.746	1.017	0.282	8.564	8.343
RM2006 NLS	8.734	11.865 (13)	0.736	1.011	0.268	8.537	8.327
RM1994	8.502	12.220 (22)	0.696	0.947	0.283	8.289	8.069
RM1994 LS	8.391	12.012 (18)	0.699	0.953	0.277	8.196	7.979
RM1994 NLS	8.763	12.151 (19)	0.721	0.990	0.225	8.581	8.405
DECO	5.980	12.258 (25)	0.488	0.660	0.297	5.797	5.568
DECO NLS	6.103	12.485 (41)	0.489	0.669	0.360	5.884	5.604
LS DECO	5.980	12.257 (24)	0.488	0.660	0.297	5.797	5.568
NLS DECO	5.981	12.257 (23)	0.488	0.660	0.297	5.798	5.569
NLS DECO NLS	6.103	12.485 (42)	0.489	0.669	0.360	5.884	5.604
OGARCH	8.363	12.341 (27)	0.678	0.936	0.095	8.271	8.196
OGARCH LS	7.052	12.544 (43)	0.562	0.773	0.103	6.974	6.893
OGARCH NLS	8.126	12.154 (20)	0.669	0.928	0.072	8.052	7.996
LS OGARCH	7.951	12.477 (39)	0.637	0.877	0.095	7.860	7.786
NLS OGARCH	8.365	12.341 (27)	0.678	0.936	0.095	8.273	8.198
LS OGARCH LS	6.880	12.710 (44)	0.541	0.743	0.101	6.802	6.723
NLS OGARCH NLS	8.126	12.154 (20)	0.669	0.928	0.072	8.051	7.996
GPVC	7.825	12.467 (38)	0.628	0.861	0.132	7.700	7.598
GPVC LS	7.438	12.274 (26)	0.606	0.834	0.106	7.341	7.259
GPVC NLS	6.727	12.369 (31)	0.544	0.749	0.113	6.621	6.533
LS GPVC	7.994	12.452 (36)	0.642	0.891	0.117	7.872	7.781
NLS GPVC	7.672	12.433 (33)	0.617	0.845	0.130	7.547	7.447
LS GPVC LS	7.470	12.429 (32)	0.601	0.826	0.161	7.359	7.238
NLS GPVC NLS	6.725	12.365 (30)	0.544	0.749	0.113	6.619	6.533
RPVC	9.657	12.785 (45)	0.755	1.047	0.222	9.479	9.310
RPCV LS	7.989	12.439 (34)	0.642	0.889	0.180	7.861	7.724
RPVC NLS	9.186	12.485 (40)	0.736	1.026	0.184	9.035	8.893
LS RPVC	8.543	12.347 (29)	0.692	0.953	0.201	8.387	8.235
NLS RPVC	8.064	13.142 (46)	0.614	0.850	0.191	7.904	7.755
LS RPCV LS	7.493	12.439 (35)	0.602	0.828	0.167	7.378	7.252
NLS RPVC NLS	7.658	12.460 (37)	0.615	0.850	0.172	7.509	7.376

Table 2. Annualised performance measures: AV, SD, IR, SR and TO stand for the average, standard deviation, information ratio, Sortino's ratio and turnover of the out-of-sample MVP returns. AV^{net}_{20bp} and AV^{net}_{50bp} stand for the average out-of-sample MVP return net of transaction costs considering 20 and 50 basis-points, respectively. Period January 2004 to December 2007. The shaded cells denote the top five for each criterion. Weights are rebalanced on a daily basis considering short-selling constraints.

	AV	**SD**	**IR**	**SR**	**TO**	$\mathbf{AV^{net}_{20bp}}$	$\mathbf{AV^{net}_{50bp}}$
1/N	12.732	12.755 (47)	0.998	1.418	-	-	-
CCC	11.425	8.381 (6)	1.363	1.963	0.256	11.137	10.934
CCC LS	9.818	8.495 (14)	1.156	1.655	0.264	9.569	9.362
CCC NLS	11.157	8.404 (11)	1.328	1.907	0.247	10.863	10.668
LS CCC	11.305	8.394 (9)	1.347	1.940	0.258	11.030	10.826
NLS CCC	11.461	8.399 (10)	1.365	1.966	0.251	11.195	10.997
LS CCC LS	9.632	8.628 (17)	1.116	1.596	0.258	9.386	9.183
NLS CCC NLS	11.172	8.426 (12)	1.326	1.910	0.263	10.901	10.692
DCC	11.144	8.203 (3)	1.359	1.947	0.263	10.843	10.636
DCC LS	9.450	8.394 (8)	1.126	1.605	0.268	9.201	8.992
DCC NLS	10.919	8.234 (5)	1.326	1.898	0.253	10.609	10.410
LS DCC	11.103	8.199 (2)	1.354	1.941	0.263	10.802	10.596
NLS DCC	11.035	8.196 (1)	1.346	1.929	0.262	10.733	10.527
LS DCC LS	9.423	8.391 (7)	1.123	1.601	0.268	9.174	8.965
NLS DCC NLS	10.829	8.226 (4)	1.316	1.884	0.252	10.519	10.321
RM2006	11.983	8.553 (15)	1.401	2.045	0.258	11.630	11.426
RM2006 LS	10.988	8.435 (13)	1.303	1.887	0.268	10.728	10.516
RM2006 NLS	9.852	8.686 (19)	1.134	1.619	0.259	9.520	9.318
RM1994	9.496	9.148 (29)	1.038	1.503	0.282	9.121	8.902
RM1994 LS	8.498	8.866 (23)	0.959	1.374	0.275	8.182	7.967
RM1994 NLS	10.080	9.112 (28)	1.106	1.584	0.220	9.742	9.571
DECO	9.282	9.062 (25)	1.024	1.457	0.253	9.040	8.840
DECO NLS	8.998	9.197 (32)	0.978	1.388	0.302	8.725	8.487
LS DECO	9.280	9.063 (26)	1.024	1.456	0.253	9.039	8.838
NLS DECO	9.271	9.064 (27)	1.023	1.455	0.254	9.030	8.829
NLS DECO NLS	8.998	9.197 (33)	0.978	1.388	0.302	8.725	8.487
OGARCH	13.356	9.188 (31)	1.454	2.097	0.083	13.165	13.100
OGARCH LS	11.565	10.105 (45)	1.144	1.602	0.088	11.435	11.367
OGARCH NLS	12.805	9.203 (34)	1.391	1.998	0.071	12.638	12.582
LS OGARCH	13.068	9.257 (36)	1.412	2.030	0.081	12.885	12.821
NLS OGARCH	13.362	9.188 (30)	1.454	2.098	0.083	13.172	13.106
LS OGARCH LS	11.305	10.326 (46)	1.095	1.528	0.082	11.175	11.110
NLS OGARCH NLS	12.804	9.203 (34)	1.391	1.997	0.071	12.637	12.582
GPVC	11.497	9.268 (37)	1.241	1.757	0.109	11.246	11.163
GPVC LS	11.024	9.282 (39)	1.188	1.680	0.082	10.835	10.772
GPVC NLS	11.210	9.320 (43)	1.203	1.690	0.099	10.993	10.918
LS GPVC	12.213	9.294 (41)	1.314	1.868	0.094	11.953	11.881
NLS GPVC	11.274	9.348 (44)	1.206	1.703	0.108	11.020	10.938
LS GPVC LS	10.325	9.288 (40)	1.112	1.559	0.129	10.153	10.052
NLS GPVC NLS	11.165	9.318 (42)	1.198	1.683	0.097	10.949	10.876
RPVC	12.966	8.680 (18)	1.494	2.169	0.193	12.642	12.492
RPCV LS	10.423	9.000 (24)	1.158	1.646	0.152	10.218	10.100
RPVC NLS	12.233	8.697 (20)	1.407	2.018	0.171	11.951	11.818
LS RPVC	11.635	8.577 (16)	1.357	1.944	0.175	11.354	11.218
NLS RPVC	10.878	8.829 (22)	1.232	1.760	0.171	10.579	10.447
LS RPCV LS	10.304	9.271 (38)	1.111	1.558	0.139	10.125	10.016
NLS RPVC NLS	10.628	8.760 (21)	1.213	1.723	0.158	10.336	10.215

Table 3. Annualised performance measures: AV, SD, IR, SR and TO stand for the average, standard deviation, information ratio, Sortino's ratio and turnover of the out-of-sample MVP returns. AV^{net}_{20bp} and AV^{net}_{50bp} stand for the average out-of-sample MVP return net of transaction costs considering 20 and 50 basis-points, respectively. Period January 2008 to June 2009. The shaded cells denote the top five for each criterion. Weights are rebalanced on a daily basis considering short-selling constraints.

	AV	**SD**	**IR**	**SR**	**TO**	$\mathbf{AV^{net}_{20bp}}$	$\mathbf{AV^{net}_{50bp}}$
1/N	−30.668	43.046 (47)	−0.713	−0.960	-	-	-
CCC	−25.407	22.009 (20)	−1.154	−1.464	0.362	−25.564	−25.799
CCC LS	−25.522	22.003 (19)	−1.160	−1.471	0.365	−25.680	−25.917
CCC NLS	−23.682	22.613 (27)	−1.047	−1.344	0.300	−23.820	−24.026
LS CCC	−26.288	21.934 (13)	−1.199	−1.516	0.369	−26.448	−26.686
NLS CCC	−27.144	21.965 (16)	−1.236	−1.558	0.365	−27.301	−27.537
LS CCC LS	−27.052	21.967 (17)	−1.232	−1.553	0.368	−27.211	−27.449
NLS CCC NLS	−25.372	22.346 (25)	−1.135	−1.446	0.326	−25.521	−25.743
DCC	−26.520	21.580 (5)	−1.229	−1.554	0.389	−26.683	−26.928
DCC LS	−26.702	21.596 (7)	−1.236	−1.563	0.391	−26.866	−27.112
DCC NLS	−24.636	21.926 (12)	−1.124	−1.446	0.312	−24.777	−24.989
LS DCC	−26.639	21.582 (6)	−1.234	−1.561	0.390	−26.802	−27.047
NLS DCC	−27.020	21.596 (7)	−1.251	−1.581	0.392	−27.184	−27.429
LS DCC LS	−26.833	21.599 (9)	−1.242	−1.570	0.392	−26.997	−27.243
NLS DCC NLS	−24.899	21.952 (14)	−1.134	−1.460	0.311	−25.039	−25.249
RM2006	−22.728	21.862 (11)	−1.040	−1.326	0.281	−22.858	−23.054
RM2006 LS	−22.912	21.815 (10)	−1.050	−1.338	0.279	−23.041	−23.235
RM2006 NLS	−21.267	21.958 (15)	−0.969	−1.264	0.216	−21.372	−21.529
RM1994	−20.793	22.108 (22)	−0.941	−1.205	0.260	−20.914	−21.096
RM1994 LS	−21.234	22.053 (21)	−0.963	−1.232	0.259	−21.355	−21.537
RM1994 NLS	−20.974	22.161 (23)	−0.946	−1.236	0.178	−21.060	−21.188
DECO	−31.859	22.706 (33)	−1.403	−1.742	0.408	−32.030	−32.288
DECO NLS	−29.187	22.618 (28)	−1.291	−1.633	0.386	−29.358	−29.615
LS DECO	−31.854	22.706 (32)	−1.403	−1.742	0.408	−32.026	−32.284
NLS DECO	−31.829	22.702 (31)	−1.402	−1.741	0.408	−32.001	−32.258
NLS DECO NLS	−29.188	22.618 (29)	−1.291	−1.633	0.386	−29.359	−29.615
OGARCH	−21.671	23.390 (36)	−0.927	−1.218	0.107	−21.722	−21.799
OGARCH LS	−21.745	23.360 (35)	−0.931	−1.223	0.108	−21.796	−21.873
OGARCH NLS	−20.118	21.541 (3)	−0.934	−1.223	0.071	−20.153	−20.205
LS OGARCH	−23.677	24.009 (45)	−0.986	−1.291	0.109	−23.728	−23.804
NLS OGARCH	−21.671	23.390 (36)	−0.927	−1.218	0.107	−21.722	−21.799
LS OGARCH LS	−23.571	23.957 (41)	−0.984	−1.288	0.109	−23.622	−23.699
NLS OGARCH NLS	−20.118	21.541 (3)	−0.934	−1.223	0.071	−20.153	−20.205
GPVC	−19.789	22.287 (24)	−0.888	−1.151	0.105	−19.831	−19.894
GPVC LS	−16.841	22.700 (30)	−0.742	−0.973	0.113	−16.890	−16.964
GPVC NLS	−23.692	21.444 (1)	−1.105	−1.434	0.050	−23.711	−23.740
LS GPVC	−18.380	22.823 (34)	−0.805	−1.079	0.112	−18.429	−18.503
NLS GPVC	−20.574	21.983 (18)	−0.936	−1.207	0.102	−20.614	−20.674
LS GPVC LS	−21.137	23.982 (43)	−0.881	−1.144	0.193	−21.208	−21.315
NLS GPVC NLS	−23.716	21.451 (2)	−1.106	−1.435	0.050	−23.735	−23.764
RPVC	−17.369	23.870 (40)	−0.728	−0.962	0.188	−17.446	−17.561
RPCV LS	−15.911	23.839 (39)	−0.667	−0.888	0.189	−15.990	−16.109
RPVC NLS	−22.229	22.432 (26)	−0.991	−1.296	0.114	−22.277	−22.350
LS RPVC	−21.004	23.672 (38)	−0.887	−1.153	0.195	−21.076	−21.183
NLS RPVC	−25.119	27.169 (46)	−0.925	−1.231	0.156	−25.192	−25.302
LS RPCV LS	−21.164	23.982 (44)	−0.883	−1.145	0.193	−21.235	−21.342
NLS RPVC NLS	−25.492	23.964 (42)	−1.064	−1.389	0.115	−25.543	−25.620

Table 4. Annualised performance measures: AV, SD, IR, SR and TO stand for the average, standard deviation, information ratio, Sortino's ratio and turnover of the out-of-sample MVP returns. AV^{net}_{20bp} and AV^{net}_{50bp} stand for the average out-of-sample MVP return net of transaction costs considering 20 and 50 basis-points, respectively. Period July 2009 to November 2017. The shaded cells denote the top five for each criterion. Weights are rebalanced on a daily basis considering short-selling constraints.

	AV	**SD**	**IR**	**SR**	**TO**	$\mathbf{AV^{net}_{20bp}}$	$\mathbf{AV^{net}_{50bp}}$
1/N	13.130	16.057 (47)	0.818	1.148	-	-	-
CCC	11.830	10.561 (15)	1.120	1.606	0.306	11.669	11.427
CCC LS	11.455	10.599 (18)	1.081	1.554	0.318	11.288	11.037
CCC NLS	11.932	10.502 (10)	1.136	1.628	0.288	11.781	11.554
LS CCC	11.713	10.540 (13)	1.111	1.595	0.310	11.550	11.304
NLS CCC	11.525	10.512 (11)	1.096	1.571	0.308	11.362	11.119
LS CCC LS	11.193	10.629 (19)	1.053	1.514	0.315	11.027	10.778
NLS CCC NLS	11.640	10.552 (14)	1.103	1.583	0.317	11.473	11.222
DCC	12.213	10.366 (1)	1.178	1.681	0.315	12.047	11.797
DCC LS	11.736	10.429 (8)	1.125	1.612	0.322	11.566	11.311
DCC NLS	11.942	10.383 (5)	1.150	1.644	0.295	11.786	11.553
LS DCC	12.204	10.366 (1)	1.177	1.679	0.314	12.038	11.788
NLS DCC	12.175	10.367 (3)	1.174	1.674	0.314	12.009	11.761
LS DCC LS	11.715	10.427 (7)	1.124	1.609	0.322	11.545	11.290
NLS DCC NLS	11.922	10.383 (4)	1.148	1.641	0.293	11.768	11.536
RM2006	12.648	10.498 (9)	1.205	1.686	0.275	12.502	12.283
RM2006 LS	13.314	10.403 (6)	1.280	1.812	0.289	13.160	12.930
RM2006 NLS	13.542	10.518 (12)	1.288	1.820	0.281	13.393	13.169
RM1994	13.243	10.941 (35)	1.210	1.691	0.287	13.091	12.863
RM1994 LS	13.613	10.686 (25)	1.274	1.799	0.281	13.463	13.239
RM1994 NLS	13.430	10.808 (32)	1.243	1.747	0.235	13.305	13.117
DECO	11.144	10.806 (29)	1.031	1.478	0.298	10.986	10.749
DECO NLS	11.007	11.214 (39)	0.982	1.410	0.383	10.805	10.501
LS DECO	11.145	10.806 (31)	1.031	1.478	0.298	10.987	10.750
NLS DECO	11.145	10.806 (29)	1.031	1.478	0.298	10.987	10.750
NLS DECO NLS	11.007	11.214 (39)	0.982	1.410	0.383	10.805	10.501
OGARCH	11.333	10.671 (22)	1.062	1.508	0.098	11.280	11.201
OGARCH LS	10.030	10.684 (24)	0.939	1.334	0.109	9.972	9.885
OGARCH NLS	10.927	11.000 (36)	0.993	1.422	0.072	10.889	10.833
LS OGARCH	11.145	10.658 (20)	1.046	1.485	0.099	11.092	11.012
NLS OGARCH	11.333	10.671 (22)	1.062	1.508	0.098	11.280	11.201
LS OGARCH LS	10.194	10.669 (21)	0.956	1.360	0.108	10.136	10.050
NLS OGARCH NLS	10.926	11.000 (37)	0.993	1.421	0.072	10.889	10.833
GPVC	10.992	11.289 (41)	0.974	1.377	0.148	10.913	10.795
GPVC LS	10.052	10.781 (27)	0.932	1.324	0.116	9.991	9.898
GPVC NLS	10.008	11.374 (44)	0.880	1.251	0.132	9.939	9.835
LS GPVC	10.681	11.061 (38)	0.966	1.366	0.128	10.612	10.508
NLS GPVC	10.985	11.300 (42)	0.972	1.375	0.146	10.907	10.790
LS GPVC LS	11.203	10.569 (16)	1.060	1.532	0.170	11.114	10.981
NLS GPVC NLS	10.030	11.364 (43)	0.883	1.256	0.131	9.962	9.858
RPVC	12.892	11.524 (46)	1.119	1.592	0.241	12.766	12.578
RPCV LS	11.084	10.772 (26)	1.029	1.466	0.192	10.985	10.836
RPVC NLS	13.327	11.476 (45)	1.161	1.682	0.202	13.221	13.062
LS RPVC	12.331	10.816 (33)	1.140	1.636	0.214	12.219	12.051
NLS RPVC	12.630	10.801 (28)	1.169	1.677	0.207	12.521	12.358
LS RPCV LS	11.256	10.596 (17)	1.062	1.535	0.177	11.164	11.026
NLS RPVC NLS	12.145	10.837 (34)	1.121	1.619	0.189	12.047	11.898

Table 5. Annualised performance measures: AV, SD, IR, SR and TO stand for the average, standard deviation, information ratio, Sortino's ratio and turnover of the out-of-sample MVP returns. AV^{net}_{20bp} and AV^{net}_{50bp} stand for the average out-of-sample MVP return net of transaction costs considering 20 and 50 basis-points, respectively. Period January 2004 to November 2017. The shaded cells denote the top five for each criterion. Weights are rebalanced on a monthly basis considering short-selling constraints.

	AV	**SD**	**IR**	**SR**	**TO**	$\mathbf{AV^{net}_{20bp}}$	$\mathbf{AV^{net}_{50bp}}$
1/N	8.302	20.058 (47)	0.414	0.570	-	-	-
CCC	7.946	12.262 (10)	0.648	0.902	0.319	7.938	7.925
CCC LS	6.832	12.263 (11)	0.557	0.775	0.329	6.823	6.810
CCC NLS	7.725	12.388 (16)	0.624	0.867	0.296	7.717	7.704
LS CCC	7.731	12.261 (9)	0.631	0.878	0.323	7.723	7.709
NLS CCC	7.588	12.278 (14)	0.618	0.859	0.319	7.579	7.566
LS CCC LS	6.471	12.359 (15)	0.524	0.728	0.325	6.462	6.449
NLS CCC NLS	7.758	12.439 (20)	0.624	0.869	0.321	7.749	7.736
DCC	7.425	12.182 (3)	0.610	0.845	0.325	7.416	7.403
DCC LS	6.567	12.200 (6)	0.538	0.747	0.334	6.558	6.544
DCC NLS	6.901	12.247 (8)	0.563	0.780	0.302	6.892	6.879
LS DCC	7.386	12.184 (4)	0.606	0.840	0.325	7.377	7.364
NLS DCC	7.296	12.193 (5)	0.598	0.829	0.325	7.287	7.274
LS DCC LS	6.518	12.203 (7)	0.534	0.741	0.334	6.509	6.495
NLS DCC NLS	6.781	12.266 (12)	0.553	0.764	0.300	6.772	6.760
RM2006	7.350	12.012 (2)	0.612	0.843	0.287	7.342	7.329
RM2006 LS	7.442	11.870 (1)	0.627	0.867	0.294	7.434	7.421
RM2006 NLS	7.101	12.274 (13)	0.579	0.798	0.296	7.093	7.081
RM1994	7.777	12.644 (29)	0.615	0.848	0.296	7.769	7.756
RM1994 LS	7.157	12.391 (17)	0.578	0.796	0.292	7.149	7.136
RM1994 NLS	7.906	12.606 (27)	0.627	0.865	0.254	7.899	7.888
DECO	5.631	12.899 (43)	0.437	0.608	0.317	5.622	5.609
DECO NLS	5.641	13.162 (44)	0.429	0.599	0.386	5.630	5.614
LS DECO	5.631	12.899 (42)	0.437	0.608	0.317	5.622	5.609
NLS DECO	5.631	12.899 (41)	0.437	0.608	0.317	5.622	5.609
NLS DECO NLS	5.640	13.162 (45)	0.429	0.599	0.386	5.630	5.614
OGARCH	7.819	12.556 (24)	0.623	0.859	0.101	7.816	7.812
OGARCH LS	6.848	12.687 (32)	0.540	0.744	0.113	6.845	6.840
OGARCH NLS	7.985	12.451 (22)	0.641	0.891	0.078	7.984	7.981
LS OGARCH	7.581	12.716 (37)	0.596	0.821	0.103	7.579	7.575
NLS OGARCH	7.821	12.555 (23)	0.623	0.859	0.101	7.818	7.814
LS OGARCH LS	7.029	12.893 (40)	0.545	0.751	0.111	7.026	7.021
NLS OGARCH NLS	7.993	12.451 (21)	0.642	0.891	0.078	7.991	7.988
GPVC	7.282	12.707 (34)	0.573	0.789	0.155	7.277	7.271
GPVC LS	7.225	12.435 (19)	0.581	0.801	0.120	7.222	7.218
GPVC NLS	6.560	12.672 (31)	0.518	0.712	0.132	6.557	6.552
LS GPVC	7.200	12.713 (36)	0.566	0.783	0.138	7.196	7.190
NLS GPVC	7.223	12.697 (33)	0.569	0.782	0.153	7.219	7.212
LS GPVC LS	6.521	12.568 (25)	0.519	0.718	0.172	6.516	6.509
NLS GPVC NLS	6.568	12.665 (30)	0.519	0.713	0.130	6.565	6.559
RPVC	8.453	12.712 (35)	0.665	0.920	0.248	8.446	8.436
RPCV LS	7.355	12.415 (18)	0.592	0.822	0.193	7.350	7.342
RPVC NLS	8.011	12.816 (39)	0.625	0.863	0.201	8.005	7.997
LS RPVC	7.000	12.615 (28)	0.555	0.765	0.227	6.994	6.985
NLS RPVC	6.488	13.243 (46)	0.490	0.676	0.203	6.482	6.474
LS RPCV LS	6.535	12.588 (26)	0.519	0.718	0.180	6.530	6.523
NLS RPVC NLS	6.874	12.741 (38)	0.540	0.743	0.182	6.869	6.862

Table 6. Annualised performance measures: AV, SD, IR, SR and TO stand for the average, standard deviation, information ratio, Sortino's ratio and turnover of the out-of-sample MVP returns. AV^{net}_{20bp} and AV^{net}_{50bp} stand for the average out-of-sample MVP return net of transaction costs considering 20 and 50 basis-points, respectively. Period January 2004 to December 2007. The shaded cells denote the top five for each criterion. Weights are rebalanced on a monthly basis considering short-selling constraints.

	AV	**SD**	**IR**	**SR**	**TO**	$\mathbf{AV^{net}_{20bp}}$	$\mathbf{AV^{net}_{50bp}}$
1/N	12.732	12.755 (47)	0.998	1.418	-	-	-
CCC	10.636	8.697 (7)	1.223	1.750	0.265	10.629	10.618
CCC LS	7.605	8.868 (19)	0.858	1.208	0.284	7.597	7.585
CCC NLS	10.356	8.732 (8)	1.186	1.694	0.254	10.349	10.338
LS CCC	10.158	8.738 (9)	1.163	1.659	0.269	10.150	10.139
NLS CCC	10.186	8.758 (11)	1.163	1.659	0.263	10.179	10.168
LS CCC LS	7.319	9.045 (23)	0.809	1.137	0.281	7.312	7.300
NLS CCC NLS	9.802	8.809 (16)	1.113	1.581	0.277	9.795	9.784
DCC	10.939	8.612 (3)	1.270	1.825	0.271	10.932	10.920
DCC LS	7.691	8.796 (15)	0.874	1.233	0.286	7.683	7.671
DCC NLS	10.763	8.661 (6)	1.243	1.782	0.260	10.756	10.745
LS DCC	10.923	8.608 (2)	1.269	1.823	0.271	10.915	10.904
NLS DCC	10.889	8.599 (1)	1.266	1.819	0.269	10.882	10.871
LS DCC LS	7.672	8.795 (14)	0.872	1.230	0.284	7.664	7.653
NLS DCC NLS	10.725	8.649 (5)	1.240	1.778	0.258	10.718	10.707
RM2006	10.378	8.765 (13)	1.184	1.706	0.292	10.369	10.357
RM2006 LS	9.295	8.629 (4)	1.077	1.540	0.300	9.287	9.275
RM2006 NLS	9.578	8.884 (20)	1.078	1.527	0.313	9.569	9.556
RM1994	8.112	9.545 (37)	0.850	1.209	0.323	8.103	8.089
RM1994 LS	6.813	9.279 (24)	0.734	1.033	0.317	6.804	6.791
RM1994 NLS	9.912	9.282 (25)	1.068	1.520	0.265	9.904	9.892
DECO	6.883	9.577 (39)	0.719	1.009	0.277	6.875	6.864
DECO NLS	6.257	9.784 (43)	0.640	0.887	0.340	6.247	6.233
LS DECO	6.882	9.577 (39)	0.719	1.008	0.277	6.875	6.863
NLS DECO	6.873	9.577 (41)	0.718	1.007	0.277	6.865	6.854
NLS DECO NLS	6.257	9.784 (44)	0.640	0.887	0.340	6.247	6.233
OGARCH	12.682	9.305 (26)	1.363	1.958	0.088	12.680	12.676
OGARCH LS	11.229	10.166 (45)	1.105	1.556	0.097	11.226	11.222
OGARCH NLS	12.878	9.376 (29)	1.374	1.971	0.063	12.877	12.874
LS OGARCH	12.588	9.346 (28)	1.347	1.928	0.088	12.586	12.582
NLS OGARCH	12.682	9.305 (26)	1.363	1.958	0.088	12.680	12.676
LS OGARCH LS	11.414	10.359 (46)	1.102	1.548	0.090	11.411	11.408
NLS OGARCH NLS	12.878	9.376 (29)	1.374	1.971	0.063	12.877	12.874
GPVC	11.014	9.504 (36)	1.159	1.636	0.145	11.010	11.004
GPVC LS	11.064	9.438 (31)	1.172	1.657	0.105	11.061	11.057
GPVC NLS	10.637	9.478 (33)	1.122	1.569	0.134	10.634	10.628
LS GPVC	11.235	9.595 (42)	1.171	1.652	0.120	11.232	11.226
NLS GPVC	10.939	9.576 (38)	1.142	1.611	0.145	10.935	10.929
LS GPVC LS	9.183	9.503 (35)	0.966	1.345	0.139	9.179	9.174
NLS GPVC NLS	10.656	9.473 (32)	1.125	1.572	0.132	10.652	10.647
RPVC	11.558	8.741 (10)	1.322	1.896	0.216	11.552	11.544
RPCV LS	10.172	9.038 (22)	1.126	1.594	0.174	10.168	10.161
RPVC NLS	11.023	8.761 (12)	1.258	1.791	0.193	11.018	11.010
LS RPVC	9.859	8.845 (18)	1.115	1.566	0.202	9.854	9.846
NLS RPVC	9.802	8.925 (21)	1.098	1.558	0.193	9.797	9.789
LS RPCV LS	9.188	9.490 (34)	0.968	1.346	0.151	9.184	9.178
NLS RPVC NLS	9.995	8.828 (17)	1.132	1.600	0.183	9.990	9.982

Table 7. Annualised performance measures: AV, SD, IR, SR and TO stand for the average, standard deviation, information ratio, Sortino's ratio and turnover of the out-of-sample MVP returns. AV^{net}_{20bp} and AV^{net}_{50bp} stand for the average out-of-sample MVP return net of transaction costs considering 20 and 50 basis-points, respectively. Period January 2008 to June 2009. The shaded cells denote the top five for each criterion. Weights are rebalanced on a monthly basis considering short-selling constraints.

	AV	**SD**	**IR**	**SR**	**TO**	$\mathbf{AV^{net}_{20bp}}$	$\mathbf{AV^{net}_{50bp}}$
1/N	−30.668	43.046 (47)	−0.713	−0.960	-	-	-
CCC	−25.344	22.796 (13)	−1.112	−1.460	0.381	−25.355	−25.371
CCC LS	−25.390	22.796 (14)	−1.114	−1.462	0.383	−25.401	−25.418
CCC NLS	−23.540	23.614 (33)	−0.997	−1.318	0.318	−23.550	−23.566
LS CCC	−25.295	22.748 (12)	−1.112	−1.463	0.385	−25.306	−25.323
NLS CCC	−26.760	22.916 (21)	−1.168	−1.532	0.373	−26.771	−26.787
LS CCC LS	−26.818	22.925 (22)	−1.170	−1.535	0.375	−26.828	−26.844
NLS CCC NLS	−25.222	23.536 (26)	−1.072	−1.414	0.335	−25.233	−25.248
DCC	−26.146	22.840 (16)	−1.145	−1.508	0.419	−26.158	−26.175
DCC LS	−26.248	22.850 (17)	−1.149	−1.513	0.419	−26.259	−26.276
DCC NLS	−23.982	23.354 (24)	−1.027	−1.359	0.333	−23.993	−24.008
LS DCC	−26.384	22.858 (18)	−1.154	−1.520	0.419	−26.395	−26.412
NLS DCC	−27.056	22.905 (20)	−1.181	−1.554	0.419	−27.067	−27.084
LS DCC LS	−26.495	22.865 (19)	−1.159	−1.526	0.421	−26.506	−26.523
NLS DCC NLS	−24.849	23.458 (25)	−1.059	−1.401	0.331	−24.859	−24.875
RM2006	−22.356	22.084 (4)	−1.012	−1.327	0.356	−22.367	−22.383
RM2006 LS	−23.045	22.006 (2)	−1.047	−1.370	0.356	−23.055	−23.071
RM2006 NLS	−21.109	23.116 (23)	−0.913	−1.206	0.274	−21.116	−21.126
RM1994	−22.685	22.716 (11)	−0.999	−1.307	0.337	−22.695	−22.711
RM1994 LS	−23.388	22.619 (10)	−1.034	−1.350	0.335	−23.398	−23.413
RM1994 NLS	−21.739	23.572 (28)	−0.922	−1.215	0.235	−21.745	−21.755
DECO	−28.184	24.101 (42)	−1.169	−1.550	0.404	−28.197	−28.215
DECO NLS	−27.588	23.858 (35)	−1.156	−1.533	0.367	−27.599	−27.617
LS DECO	−28.182	24.100 (41)	−1.169	−1.550	0.404	−28.195	−28.213
NLS DECO	−28.166	24.098 (40)	−1.169	−1.549	0.404	−28.178	−28.197
NLS DECO NLS	−27.591	23.859 (36)	−1.156	−1.533	0.367	−27.602	−27.620
OGARCH	−20.677	23.592 (30)	−0.877	−1.145	0.124	−20.680	−20.683
OGARCH LS	−20.855	23.577 (29)	−0.885	−1.155	0.126	−20.857	−20.860
OGARCH NLS	−19.608	22.343 (7)	−0.878	−1.156	0.063	−19.610	−19.613
LS OGARCH	−20.516	24.433 (44)	−0.840	−1.098	0.130	−20.518	−20.522
NLS OGARCH	−20.677	23.592 (30)	−0.877	−1.145	0.124	−20.680	−20.683
LS OGARCH LS	−20.564	24.390 (43)	−0.843	−1.103	0.132	−20.567	−20.570
NLS OGARCH NLS	−19.608	22.343 (7)	−0.878	−1.156	0.061	−19.610	−19.613
GPVC	−14.454	22.017 (3)	−0.657	−0.868	0.138	−14.457	−14.462
GPVC LS	−14.100	22.418 (9)	−0.629	−0.831	0.136	−14.103	−14.107
GPVC NLS	−20.436	22.235 (5)	−0.919	−1.209	0.048	−20.438	−20.440
LS GPVC	−15.361	22.807 (15)	−0.674	−0.902	0.165	−15.364	−15.368
NLS GPVC	−14.829	21.853 (1)	−0.679	−0.892	0.134	−14.832	−14.837
LS GPVC LS	−17.991	24.031 (38)	−0.749	−0.995	0.226	−17.996	−18.004
NLS GPVC NLS	−20.471	22.244 (6)	−0.920	−1.210	0.048	−20.472	−20.474
RPVC	−15.076	23.561 (27)	−0.640	−0.849	0.203	−15.080	−15.086
RPCV LS	−14.841	23.612 (32)	−0.629	−0.837	0.201	−14.844	−14.850
RPVC NLS	−23.341	23.711 (34)	−0.984	−1.289	0.134	−23.344	−23.349
LS RPVC	−18.340	23.935 (37)	−0.766	−1.017	0.226	−18.345	−18.353
NLS RPVC	−26.862	26.877 (46)	−0.999	−1.331	0.178	−26.868	−26.876
LS RPCV LS	−17.991	24.031 (38)	−0.749	−0.995	0.226	−17.996	−18.004
NLS RPVC NLS	−25.379	24.937 (45)	−1.018	−1.338	0.140	−25.383	−25.388

Table 8. Annualised performance measures: AV, SD, IR, SR and TO stand for the average, standard deviation, information ratio, Sortino's ratio and turnover of the out-of-sample MVP returns. AV^{net}_{20bp} and AV^{net}_{50bp} stand for the average out-of-sample MVP return net of transaction costs considering 20 and 50 basis-points, respectively. Period July 2009 to November 2017. The shaded cells denote the top five for each criterion. Weights are rebalanced on a monthly basis considering short-selling constraints.

	AV	**SD**	**IR**	**SR**	**TO**	$\mathbf{AV^{net}_{20bp}}$	$\mathbf{AV^{net}_{50bp}}$
1/N	13.130	16.057 (47)	0.818	1.148	-	-	-
CCC	12.592	10.935 (21)	1.152	1.666	0.333	12.583	12.569
CCC LS	12.200	10.873 (17)	1.122	1.628	0.342	12.191	12.178
CCC NLS	12.038	10.852 (15)	1.109	1.600	0.310	12.029	12.017
LS CCC	12.455	10.936 (22)	1.139	1.648	0.340	12.446	12.432
NLS CCC	12.466	10.894 (18)	1.144	1.656	0.333	12.457	12.443
LS CCC LS	11.992	10.932 (20)	1.097	1.592	0.336	11.983	11.970
NLS CCC NLS	12.656	10.945 (23)	1.156	1.679	0.340	12.646	12.633
DCC	11.729	10.801 (11)	1.086	1.554	0.336	11.720	11.706
DCC LS	11.873	10.761 (7)	1.103	1.590	0.340	11.864	11.850
DCC NLS	10.560	10.716 (3)	0.985	1.402	0.315	10.551	10.538
LS DCC	11.714	10.800 (10)	1.085	1.551	0.333	11.705	11.691
NLS DCC	11.701	10.801 (12)	1.083	1.549	0.333	11.692	11.678
LS DCC LS	11.845	10.761 (6)	1.101	1.585	0.340	11.835	11.821
NLS DCC NLS	10.534	10.714 (2)	0.983	1.397	0.312	10.525	10.512
RM2006	11.197	10.716 (4)	1.045	1.476	0.271	11.189	11.177
RM2006 LS	11.987	10.531 (1)	1.138	1.627	0.279	11.978	11.966
RM2006 NLS	10.943	10.776 (8)	1.016	1.442	0.291	10.935	10.923
RM1994	13.040	11.344 (37)	1.150	1.630	0.275	13.032	13.021
RM1994 LS	12.758	11.017 (27)	1.158	1.653	0.273	12.750	12.738
RM1994 NLS	12.229	11.068 (28)	1.105	1.566	0.252	12.222	12.211
DECO	11.054	11.293 (34)	0.979	1.415	0.321	11.045	11.032
DECO NLS	11.262	11.791 (45)	0.955	1.392	0.409	11.251	11.235
LS DECO	11.054	11.293 (34)	0.979	1.415	0.321	11.045	11.032
NLS DECO	11.055	11.293 (36)	0.979	1.415	0.321	11.046	11.033
NLS DECO NLS	11.262	11.791 (45)	0.955	1.392	0.409	11.251	11.235
OGARCH	10.576	10.959 (25)	0.965	1.377	0.103	10.573	10.569
OGARCH LS	9.694	10.852 (16)	0.893	1.281	0.120	9.691	9.686
OGARCH NLS	10.568	11.197 (33)	0.944	1.348	0.088	10.566	10.563
LS OGARCH	10.200	10.921 (19)	0.934	1.332	0.103	10.197	10.192
NLS OGARCH	10.580	10.958 (24)	0.966	1.377	0.103	10.577	10.573
LS OGARCH LS	9.853	10.847 (14)	0.908	1.304	0.115	9.850	9.845
NLS OGARCH NLS	10.581	11.197 (32)	0.945	1.350	0.088	10.579	10.576
GPVC	9.374	11.733 (43)	0.799	1.121	0.161	9.370	9.363
GPVC LS	9.194	11.126 (31)	0.826	1.168	0.124	9.191	9.186
GPVC NLS	9.425	11.598 (41)	0.813	1.146	0.145	9.421	9.416
LS GPVC	9.295	11.436 (38)	0.813	1.143	0.141	9.291	9.285
NLS GPVC	9.379	11.742 (44)	0.799	1.122	0.159	9.375	9.368
LS GPVC LS	9.617	10.737 (5)	0.896	1.283	0.180	9.612	9.605
NLS GPVC NLS	9.435	11.585 (40)	0.815	1.149	0.145	9.432	9.426
RPVC	11.163	11.486 (39)	0.972	1.376	0.268	11.155	11.144
RPCV LS	9.965	10.803 (13)	0.922	1.319	0.201	9.959	9.951
RPVC NLS	12.158	11.600 (42)	1.048	1.497	0.218	12.152	12.143
LS RPVC	10.150	11.125 (30)	0.912	1.291	0.239	10.143	10.133
NLS RPVC	10.847	11.091 (29)	0.978	1.388	0.212	10.841	10.833
LS RPCV LS	9.637	10.782 (9)	0.894	1.279	0.187	9.632	9.624
NLS RPVC NLS	11.130	10.964 (26)	1.015	1.451	0.189	11.125	11.118

As for the annualised average portfolio returns taking into account transaction costs, the procedures with the five largest values of AV^{net}_{20bp} and AV^{net}_{50bp} are the same procedures with the largest AV, except in some cases in the pre-crisis period, where one of five largest AV^{net}_{50bp} is obtained by the NLS-OGARCH-NLS procedure.

For each period, the five best methods in terms of information criteria are the same (except in Table 8, where four methods are the same). We omit the analysis in the crisis period because these criteria values are negative. Overall, for daily rebalancing, RiskMetrics based methods are among the best in the full and post-crisis periods, RPVC and RPVC-NLS are among the best in the full and pre-crisis periods, and NLS-OGARCH and LS-OGARCH are among the best in the pre-crisis period. For monthly rebalancing, some OGARCH-based methods are among the best in the pre-crisis and full periods, some CCC-based methods are among the best in the post-crisis and full periods, RM1994-LS is among the best for the post-crisis period, and RPVC is among the best for the full period.

The analysis of Tables 1–8 reveals that none of the methods is the best in all scenarios and the performance depends on the criterion, the period and the rebalancing strategy. In this sense, the analysis will focus on the full period (Tables 1 and 5) in order to account for periods with different volatility levels. When portfolios are rebalanced on a daily basis, we find that DCC-based methods are the best in terms of SD; RM2006-LS, RM2006-NL, RPVC and RPVC-NLS are the best in terms of {AV, AV^{net}_{20bp}, AV^{net}_{50bp}} and {IR, SR}, and some OGARCH-based are the best regarding TO. For monthly rebalanced portfolios, the best methods in terms of SD are DCC, LS-DCC, NLS-DCC, RM2006 and RM2006-LS, whereas the best performances in terms of {AV, AV^{net}_{20bp}, AV^{net}_{50bp}} and {IR, SR} are given by (RPVC, RPVC-NLS), (OGARCH-NLS,NLS-OGARCH-NLS) and CCC. In addition, the equal-weighted strategy is the second best in terms of AV, but the worst regarding SD, IR and SR criteria.

To show when the shrinkage method improves performance in terms of SD, the analysis is again focused on the full period (Tables 1 and 5). For daily and monthly portfolio rebalancing : shrinkage always improves the performance of the RM2004 and GPVC methods (except LS-GPVC for monthly rebalancing) whereas it always worsens the DCC method; linear shrinkage at the end improves RM2006; just linear/non-linear shrinkage at the beginning improves DECO; OGARCH-NLS and NLS-OGARCH-NLS improves OGARCH; LS-CCC improves CCC (as well as NLS-DCC for daily rebalancing). Additionally, for daily rebalancing, shrinkage always improves the performance of RPVC (except LS-GPVC), whereas for monthly rebalancing, linear shrinkage applied at the beginning and/or end improves RPVC. Nakagawa et al. (2018) also reports that in some cases the use of non-linear shrinkage on the unconditional covariance matrix of the devolatilised returns in the DCC model increases the standard deviation of the out-of-sample portfolio returns.

We now discuss the effect of shrinkage in terms of AV^{net}_{50bp}. For daily rebalancing, shrinkage improves the performance of the RM2006 and DECO methods, and worsens the performance of the DCC and RPVC methods. In addition, CCC-NLS is better than CCC, RM1994-NLS is better than RM1994, and LS-GPVC is better than GPVC. For monthly rebalancing, shrinkage does not improve the performance of the CCC, DCC, GPVC and RPVC methods. In addition, RM2006-LS is better than RM2006, RM1994-NLS is better than RM1994, DECO-NLS and NLS-DECO-NLS are better than DECO, and OGARCH-NLS and NLS-OGARCH-NLS are better than OGARCH.

Finally, we list next the main findings when short-selling is allowed for optimisation of the portfolio variance. A detailed analysis of these cases is given in the Supplementary Material. First, none of the methods is the best in all scenarios and the performance depends on the criterion, the sample period and the portfolio rebalancing scheme. Second, the analysis of the full period reveals that for daily rebalancing, DCC methods are the best regarding SD and are among the best in terms of IR and SR. RM1994-LS and RM2006-LS are the best according to AV, AV^{net}_{20bp}, AV^{net}_{50bp}, IR and SR. For monthly rebalancing, DCC-LS and LS-DCC-LS are among the best in terms of SD, RM2006-NLS is the best in terms of SD and is among the best regarding IR and SR. RM 1994 and RM1994-LS are the first and second best in terms of AV, AV^{net}_{20bp}, AV^{net}_{50bp} but are among the worst in terms of SD. Third, the analysis of the turnover and average net returns in the no short-sale constraints case must be carefully done. This is because since no limits are imposed on the weights of the portfolio, large turnover values can be obtained and consequently we can have a large loss (average return) but huge net gain (average net portfolio return taking into account transaction costs). Fourth, in many cases shrinkage improves the performance of the methods in terms of SD, and this improvement can be substantial. Fifth, the top-five

models in terms of SD are the same in both restricted and unrestricted minimum variance portfolios for daily rebalancing, except in the crisis period.

4. Conclusions

The main conclusion of the paper is that none of the methods is the best in all scenarios and the performance depends on the criterion, the sample period, the portfolio rebalancing scheme and whether or not short-selling constraints are included in the portfolio optimisation process.

When short-selling constraints are included in the portfolio optimisation process, the main results can be summarised as follows. First, none of the methods is the best in all scenarios and the performance depends on the criterion, the sample period and the portfolio rebalancing scheme. Second, when considering the SD criterion, the five smallest SDs are obtained by the DCC based-methods, except in the crisis period, in which case, the five smallest SDs are spread among the DCC, OGARCH and GPVC based-methods. In the crisis-period, the smallest SDs are obtained by the GPVC procedure with the non-linear shrinkage applied to the one-step-ahead conditional covariance matrix. For portfolios rebalanced monthly, the smallest SDs are obtained by the RM2006-LS, NLS-DCC, NLS-GPVC and RM2006-LS procedures for the full, pre-crisis, crisis and post-crisis periods, respectively. Third, unlike Engle et al. (2017) and Nakagawa et al. (2018), we do not find that applying non-linear shrinkage to the unconditional correlation matrix of the devolatilised returns improves the performance of the portfolio in terms of SD when the DCC model is used, and this also happens when applied in other methods. It is important to point out that Engle et al. (2017) use portfolio of 1000 assets, Nakagawa et al. (2018) use portfolios of 100, 500 and 1000 assets and we use a portfolio with 174 assets.

When short-selling is allowed for optimisation of the portfolio variance, the main conclusions are: none of the methods is the best in all scenarios and the performance depends on the criterion, the sample period and the portfolio rebalancing scheme; in many cases shrinkage improves the performance of the methods in terms of SD and this improvement can be substantial; for daily rebalancing the top-five models in terms of SD are the same of those when short-selling constraints are imposed, except in the crisis period cases. Finally, focusing on the analysis of the full period cases we can say that overall the DCC and Riskmetrics-based methods are the best; and the analysis of the turnover and average net returns in the no short-selling constraints case should be carefully done.

Supplementary Materials: The following are available online at http://www.mdpi.com/2225-1146/7/2/19/s1, File: Covariance Prediction in Large Portfolio Allocation: Supplementary Material.

Author Contributions: This paper has been a collaborative effort, with all authors contributing equally to this work. This includes conceptualization and investigation of the main ideas in the manuscript, methodology proposals, and formal analysis, as well as all aspects of the writing process.

Funding: The first three authors acknowledge financial support from the São Paulo Research Foundation (FAPESP), grants 2016/18599-4, 2018/03012-3, 2013/00506-1 and 2018/04654-9. The fourth author is grateful to the National Council for Scientific and Technological Development (CNPq) for grant 303688/2016-5. The third author is also grateful to CNPq for grant 313035/2017-2.

Acknowledgments: The first three authors acknowledge support of the Centre for Applied Research on Econometrics, Finance and Statistics (CAREFS) and the Centre of Quantitative Studies in Economics and Finance (CEQEF). The authors are also grateful to two anonymous referees and the academic editor for providing useful comments and suggestions on earlier version of the paper.

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

Appendix A. Estimation Methods

Here we present the detailed list of the estimation methods implemented in the paper. The marginal variances in the CCC, DCC and DECO models were modelled by the GJR-(1,1) model (Glosten et al. 1993) and the parameters were estimated by quasi-maximum likelihood assuming a Student-*t* distribution. The volatility components in the GPVC and RPVC procedures were modelled by the GJR(1,1)-cDCC(1,1) model and its robust version proposed by Boudt et al. (2013) and

Laurent et al. (2016), respectively. The univariate variances in the OGARCH model were also modelled by the GJR-(1,1).

In the GPVC and RPVC procedures, the number of selected volatility components was estimated using criteria of Ahn and Horenstein (2013), Bai and Ng (2002) and Kaiser-Guttman Guttman (1954), and using the ratio estimator proposed by Lam and Yao (2012). Following these criteria and the suggestions in Trucíos et al. (2019), we use one volatility component in the GPVC procedure and four volatility components in the RPVC procedure.

The CCC, DCC, DECO, RM1994 and RM2006 procedures were implemented using the MFE Matlab Toolbox of Kevin Sheppard. The OGARCH, GPVC and RPVC procedures were implemented in R (R Core Team 2017) using the R packages *rugarch* of Ghalanos (2017), *Rcpp* of Eddelbuettel and François (2011) and *covRobust* of Wang et al. (2017). For the shrinkage procedures, we used the R packages *RiskPortfolios* (Ardia et al. 2018) and *nlshrink* (Ramprasad 2016) for the linear and non-linear shrinkage, respectively, coupled with the MATLAB toolbox QuEST (Ledoit and Wolf 2017) for the non-linear shrinkage and the MATLAB function *covCor*[5]. Whenever a program presented other options, we used the default options.

CCC based-methods

- CCC: Estimated by quasi-maximum likelihood.
- LS-CCC: Estimated as in CCC, but with the unconditional covariance matrix (Equation (4)) estimated using linear shrinkage.
- NLS-CCC: Estimated as in LS-CCC, but replacing linear by the non-linear shrinkage.
- CCC-LS: Estimated as in CCC, with the application of the linear shrinkage to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- CCC-NLS: Estimated as in CCC-LS, but replacing linear by non-linear shrinkage.
- LS-CCC-LS: Estimated as in LS-CCC, with the application of non-linear shrinkage to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- NLS-CCC-NLS: Estimated as in NLS-CCC, with the application of non-linear shrinkage to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.

DCC based-methods

- DCC: Estimated by composite likelihood (Pakel et al. 2014) using consecutive pairs.
- LS-DCC: Estimated as in DCC, but with the unconditional covariance matrix of the devolatilised returns ($\mathbf{C}$ in Equation (7)) estimated using linear shrinkage.
- NLS-DCC: Estimated as in LS-DCC, but replacing linear by non-linear shrinkage.
- DCC-LS: Estimated as in DCC, with the application of linear shrinkage to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- DCC-NLS: Estimated as in DCC-LS, but replacing linear by non-linear shrinkage.
- LS-DCC-LS: Estimated as in LS-DCC, with the application of linear shrinkage to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- NLS-DCC-NLS: Estimated as in NLS-DCC, with the application of non-linear shrinkage to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.

DECO based-methods

- DECO: Estimated using a single block.
- LS-DECO: Estimated as in DECO, but the unconditional covariance matrix of the devolatilised returns is estimated using linear shrinkage.

[5] Available at www.econ.uzh.ch/en/people/faculty/wolf/publications.

- NLS-DECO: Estimated as in LS-DECO, but replacing linear by non-linear shrinkage.
- DECO-NLS: Estimated as in DECO-LS, but non-linear shrinkage is applied to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- NLS-DECO-NLS: Estimated as in NLS-DECO model, but with non-linear shrinkage applied to the $\mathbf{H}_{T+1}$ and linear shrinkage towards the equicorrelation matrix

Because in the DECO model the estimated unconditional covariance matrix and $\mathbf{H}_{T+1}$ are already equicorrelated there is no sense in using linear shrinkage towards the equicorrelation matrix, since it has no effect.

RiskMetrics based-methods

- RM1994: RM1994 method.
- RM1994-LS: Estimated as in RM1994 with linear shrinkage applied to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- RM1994-NLS: Estimated as in RM1994-LS but replacing linear by non-linear shrinkage.
- RM2006[6]: RM2006 method (Zumbach 2007).
- RM2006-LS: Estimated as in RM2006 with linear shrinkage applied to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- RM2006-NLS: Estimated as in RM2006-LS but replacing linear by non-linear shrinkage.

OGARCH based-methods

- OGARCH: The OGARCH model considers $k = N$ components.
- LS-OGARCH: Estimated as in OGARCH, but the unconditional covariance matrix used in the spectral decomposition is estimated using linear shrinkage.
- NLS-OGARCH: Estimated as in LS-OGARCH, but replacing linear by non-linear shrinkage.
- OGARCH-LS: Estimated as in OGARCH with the linear shrinkage applied to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- OGARCH-NLS: Estimated as in OGARCH-LS, but replacing linear by non-linear shrinkage.
- LS-OGARCH-LS: Estimated as in LS-OGARCH, but linear shrinkage is applied to the predicted one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- NLS-OGARCH-NLS: Estimated as in NLS-OGARCH, but non-linear shrinkage is applied to the predicted one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.

GPVC based-methods

- GPVC: The GPVC procedure considers $k = 1$ volatility component, as explained later. We use $g = 5$ as in Li et al. (2016).
- LS-GPVC: Estimated as in the GPVC model with the unconditional covariance matrix $\hat{\mathbf{\Sigma}}$ in Equation (17) estimated using linear shrinkage.
- NLS-GPVC: Estimated as in LS-GPVC, but replacing linear by non-linear shrinkage.
- GPVC-LS: Estimated as in GPVC with linear shrinkage applied to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- GPVC-NLS: Estimated as in GPVC-LS, but replacing linear by non-linear shrinkage.
- LS-GPVC-LS: Estimated as in LS-GPVC with linear shrinkage applied to the predicted one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.

[6] This method was implemented using the MFE Matlab Toolbox of Kevin Sheppard with the default options. An R implementation of the same procedure can be found in Trucios (2017).

- NLS-GPVC-NLS: Estimated as in NLS-GPVC with non-linear shrinkage applied to the predicted one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.

RPVC based-methods

- RPVC: The RPVC procedure considers $k = 4$ volatility components, as explained later. We use $g = 5$ as in Li et al. (2016) and c as in Trucíos et al. (2019).
- LS-RPVC: Estimated as in RPVC, but linear shrinkage is applied to the robust unconditional covariance matrix $\hat{\mathbf{\Sigma}}^R$ used in Equation (18).
- NLS-RPVC: Estimated as in LS-RPVC, but replacing linear by non-linear shrinkage.
- RPVC-LS: Estimated as in RPVC with linear shrinkage applied to the one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- RPVC-NLS: Estimated as in RPVC-LS, but replacing linear by non-linear shrinkage.
- LS-RPVC-LS: Estimated as in LS-RPVC with the linear shrinkage applied to the predicted one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.
- NLS-RPVC-NLS: Estimated as in NLS-RPVC with non-linear shrinkage applied to the predicted one-step-ahead conditional covariance matrix $\mathbf{H}_{T+1}$.

References

Ahn, Seung C, and Alex R. Horenstein. 2013. Eigenvalue ratio test for the number of factors. *Econometrica* 81: 1203–27.

Alexander, Carol. 2002. Principal component models for generating large GARCH covariance matrices. *Economic Notes: Review of Banking, Finance and Monetary Economics* 31: 337–59. [CrossRef]

Alexander, Carol O., and Aubrey Muyeke Chibumba. 1996. *Multivariate Orthogonal Factor GARCH*. Sussex: University of Sussex Discussion Papers in Mathematics.

Ardia, David, Kris Boudt, and Jean-Philippe Gagnon-Fleury. 2018. *RiskPortfolios: Computation of Risk-Based Portfolios*. R package version 2.1.2. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2911021 (accessed on 28 June 2018).

Bai, Jushan, and Serena Ng. 2002. Determining the number of factors in approximate factor models. *Econometrica* 70: 191–221. [CrossRef]

Bauwens, Luc, Sébastien Laurent, and Jeroen V.K. Rombouts. 2006. Multivariate GARCH models: A survey. *Journal of Applied Econometrics* 21: 79–109. [CrossRef]

Becker, Ralf, Adam E. Clements, Mark B. Doolan, and A. Stan Hurn. 2015. Selecting volatility forecasting models for portfolio allocation purposes. *International Journal of Forecasting* 31: 849–61. [CrossRef]

Bollerslev, Tim. 1990. Modelling the coherence in short-run nominal exchange rates: A multivariate generalized ARCH model. *Review of Economics and Statistics* 72: 498–505. [CrossRef]

Boudt, Kris, Jon Danielsson, and Sébastien Laurent. 2013. Robust forecasting of dynamic conditional correlation GARCH models. *International Journal of Forecasting* 29: 244–57. [CrossRef]

Caporin, Massimiliano, and Michael McAleer. 2014. Robust ranking of multivariate GARCH models by problem dimension. *Computational Statistics & Data Analysis* 76: 172–85.

Caporin, Massimiliano, and Paolo Paruolo. 2015. Proximity-structured multivariate volatility models. *Econometric Reviews* 34: 559–93. [CrossRef]

Clarke, Roger, Harindra De Silva, and Steven Thorley. 2011. Minimum-variance portfolio composition. *Journal of Portfolio Management* 37: 31. [CrossRef]

Clarke, Roger G., Harindra De Silva, and Steven Thorley. 2006. Minimum-variance portfolios in the US equity market. *The Journal of Portfolio Management* 33: 10–24. [CrossRef]

Creal, Drew, Siem Jan Koopman, and André Lucas. 2011. A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics* 29: 552–563.

de Almeida, Daniel, Luiz K. Hotta, and Esther Ruiz. 2018. MGARCH models: Trade-off between feasibility and flexibility. *International Journal of Forecasting 34*(1), 45–63. [CrossRef]

DeMiguel, Victor, Lorenzo Garlappi, and Raman Uppal. 2009. Optimal versus naive diversification: How inefficient is the 1/n portfolio strategy? *Review of Financial Studies* 22: 1915–53. [CrossRef]

Eddelbuettel, Dirk, and Romain François. 2011. Rcpp: Seamless R and C++ integration. *Journal of Statistical Software* 40: 1–18. doi:10.18637/jss.v040.i08. [CrossRef]

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. 2009. *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton: Princeton University Press.

Engle, Robert, and Bryan Kelly. 2012. Dynamic equicorrelation. *Journal of Business & Economic Statistics* 30: 212–28.

Engle, Robert F., Olivier Ledoit, and Michael Wolf. 2017. Large dynamic covariance matrices. *Jounal of Business and Economic Statistics* (doi:10.1080/07350015.2017.1345683). [CrossRef]

Frahm, Gabriel. 2010. Linear statistical inference for global and local minimum variance portfolios. *Statistical Papers* 51: 789–812. [CrossRef]

Gambacciani, Marco, and Marc S. Paolella. 2017. Robust normal mixtures for financial portfolio allocation. *Econometrics and Statistics* 3: 91–111. [CrossRef]

Ghalanos, Alexios. 2017. *Rugarch: Univariate GARCH Models*. R package version 1.3-8. Available online: https://cran.r-project.org/web/packages/rugarch/ (accessed on 15 October 2017)

Glosten, Lawrence R., Ravi Jagannathan, and David E. Runkle. 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance* 48: 1779–801. [CrossRef]

Guttman, Louis. 1954. Some necessary conditions for common factor analysis. *Psychometrika* 19: 149–61. [CrossRef]

Hafner, Christian M and Olga Reznikova. 2012. On the estimation of dynamic conditional correlation models. *Computational Statistics & Data Analysis* 56: 3533–3545.

Hlouskova, Jaroslava, Kurt Schmidheiny, and Martin Wagner. 2009. Multistage predictions for multivariate GARCH models: Closed form solution and the value for portfolio management. *Journal of Empirical Finance* 16: 330–6. [CrossRef]

Hu, Yu-Pin, and Ruey S. Tsay. 2014. Principal volatility component analysis. *Journal of Business & Economic Statistics* 32: 153–164.

Hubert, Mia, Peter J. Rousseeuw, and Tim Verdonck. 2012. A deterministic algorithm for robust location and scatter. *Journal of Computational and Graphical Statistics* 21: 618–37. [CrossRef]

Jagannathan, Ravi, and Tongshu Ma. 2003. Risk reduction in large portfolios: Why imposing the wrong constraints helps. *The Journal of Finance* 58: 1651–84. [CrossRef]

Kirby, Chris, and Barbara Ostdiek. 2012. It's all in the timing: simple active portfolio strategies that outperform naive diversification. *Journal of Financial and Quantitative Analysis* 47: 437–67. [CrossRef]

Lam, Clifford, and Qiwei Yao. 2012. Factor modeling for high-dimensional time series: Inference for the number of factors. *The Annals of Statistics* 40: 694–726. [CrossRef]

Laurent, Sébastien, Christelle Lecourt, and Franz C. Palm. 2016. Testing for jumps in conditionally Gaussian ARMA–GARCH models, a robust approach. *Computational Statistics & Data Analysis* 100: 383–400.

Laurent, Sébastien, Jeroen V. K. Rombouts, and Francesco Violante. 2012. On the forecasting accuracy of multivariate garch models. *Journal of Applied Econometrics* 27: 934–55. [CrossRef]

Ledoit, Olivier, and Michael Wolf. 2004a. Honey, I shrunk the sample covariance matrix. *The Journal of Portfolio Management* 30: 110–9. [CrossRef]

Ledoit, Olivier, and Michael Wolf. 2004b. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis* 88: 365–411. [CrossRef]

Ledoit, Olivier, and Michael Wolf. 2012. Nonlinear shrinkage estimation of large-dimensional covariance matrices. *The Annals of Statistics* 40: 1024–60. [CrossRef]

Ledoit, Olivier, and Michael Wolf. 2017. Numerical implementation of the quest function. *Computational Statistics & Data Analysis* 115: 199–223.

Li, Weiming, Jing Gao, Kunpeng Li, and Qiwei Yao. 2016. Modeling multivariate volatilities via latent common factors. *Journal of Business & Economic Statistics* 34: 564–73.

Moura, Guilherme Valle, and Andre A. P Santos. 2018. Forecasting Large Stochastic Covariance Matrices. Working Paper. Available online: https://ssrn.com/abstract=3222808 (accessed on 15 Septembe 2018).

Nakagawa, Kei, Mitsuyoshi Imamura, and Kenichi Yoshida. 2018. Risk-based portfolios with large dynamic covariance matrices. *International Journal of Financial Studies* 6: 52. [CrossRef]

Olivares-Nadal, Alba V., and Victor DeMiguel. 2018. A robust perspective on transaction costs in portfolio optimization. *Operations Research* 66: 733–9. [CrossRef]

Pakel, Cavit, Neil Shephard, Kevin Sheppard, and Robert Engle. 2017. *Fitting Vast Dimensional Time-Varying Covariance Models*. NYU Working Paper No. FIN-08-009. Available online: https://ssrn.com/abstract=1354497 (accessed on 3 March 2018).

R Core Team. 2017. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Ramprasad, Pratik. 2016. *nlshrink: Non-Linear Shrinkage Estimation of Population Eigenvalues and Covariance Matrices*. R package version 1.0.1. Available online: https://cran.r-project.org/web/packages/nlshrink/ (accessed on 27 March 2018)

Rousseeuw, Peter J. 1984. Least median of squares regression. *Journal of the American Statistical Association* 79: 871–880. [CrossRef]

Santos, André Alves Portela and Alexandre R. Ferreira. 2017. On the choice of covariance specifications for portfolio selection problems. *Brazilian Review of Econometrics* 37: 89–122. [CrossRef]

Sortino, Frank A., and Robert van der Meer. 1991. Downside risk. *Journal of Portfolio Management* 17: 27–31. [CrossRef]

Trucios, Carlos. 2017. *RM2006: RiskMetricsss 2006 Methodology*. R package version 0.1.0. Available online: https://cran.r-project.org/web/packages/RM2006/ (accessed on 9 May 2018)

Trucíos, Carlos, Luiz K. Hotta, and Pedro L. Valls Pereira. 2019. On the robustness of the principal volatility components. *Journal of Empirical Finance* 52: 201–219. [CrossRef]

Trucíos, Carlos, Luiz K. Hotta, and Esther Ruiz. 2018. Robust bootstrap densities for dynamic conditional correlations: Implications for portfolio selection and value-at-risk. *Journal of Statistical Computation and Simulation* 88: 1976–2000. [CrossRef]

Van der Weide, Roy. 2002. GO-GARCH: A multivariate generalized orthogonal GARCH model. *Journal of Applied Econometrics* 17: 549–64. [CrossRef]

Wang, Naisyin, Adrian Raftery, and Chris Fraley. 2017. *covRobust: Robust Covariance Estimation via Nearest Neighbor Cleaning*. R package version 1.1-3. Available online: https://cran.r-project.org/web/packages/covRobust/ (accessed on 27 March 2018)

Wied, Dominik, Daniel Ziggel, and Tobias Berens. 2013. On the application of new tests for structural changes on global minimum-variance portfolios. *Statistical Papers* 54: 955–75. [CrossRef]

Zumbach, Gilles O. 2007. A Gentle Introduction to the RM2006 Methodology. Available online: https://ssrn.com/abstract=1420183 (accessed on 12 March 2018).

econometrics

MDPI

Article

Bivariate Volatility Modeling with High-Frequency Data

Marius Matei [1,2,3,*], Xari Rovira [4] and Núria Agell [4]

[1] Department of Economics, Macquarie Business School, Macquarie University, Sydney, NSW 2109, Australia
[2] Systemic Risk Monitoring Division, Financial Stability Department, National Bank of Romania, Bucharest 030031, Romania
[3] Centre for Macroeconomic Modelling, National Institute of Economic Research 'Costin C. Kiriţescu', Romanian Academy, Bucharest 050711, Romania
[4] Department of Operations, Innovation and Data Sciences, ESADE Business School, Ramon Llull University, E-08172 Sant Cugat, Spain; xari.rovira@esade.edu (X.R.); nuria.agell@esade.edu (N.A.)
* Correspondence: mateimar@gmail.com; Tel.: +61-466-484-775 or +40-744-380-012

Received: 6 August 2018; Accepted: 6 September 2019; Published: 15 September 2019

Abstract: We propose a methodology to include night volatility estimates in the day volatility modeling problem with high-frequency data in a realized generalized autoregressive conditional heteroskedasticity (GARCH) framework, which takes advantage of the natural relationship between the realized measure and the conditional variance. This improves volatility modeling by adding, in a two-factor structure, information on latent processes that occur while markets are closed but captures the leverage effect and maintains a mathematical structure that facilitates volatility estimation. A class of bivariate models that includes intraday, day, and night volatility estimates is proposed and was empirically tested to confirm whether using night volatility information improves the day volatility estimation. The results indicate a forecasting improvement using bivariate models over those that do not include night volatility estimates.

Keywords: high-frequency; volatility; forecasting; realized measures; bivariate GARCH

JEL Classification: C32; C53; C58

1. Introduction

We aim to improve volatility modeling by adding information that exists on latent volatility processes while the markets are closed and no transactions occur. We build upon the observation that the price at market closing usually differs from the price at market opening, despite no transactions occurring between the two recordings. Models previously proposed usually estimate volatility by including information on past day and intraday volatility, estimated from day-recorded prices and sampled at various time intervals. Some papers have proposed methods to address overnight returns. The latent volatility component apparent in periods when markets are closed, highlighted by the difference between the two prices, may be the effect of events that occurred during the market closing, both domestic or international, or may be due to other latent factors that usually influence the financial markets, and may prove useful in volatility modeling. We propose an estimation of this night latent volatility and suggest a new model that uses day, intraday, and night volatility information to model day volatility. What distinguishes our contribution from other papers published on similar topics is that we propose a two-factor structure in a realized generalized autoregressive conditional heteroskedasticity (GARCH) setting that takes advantage of the natural relationship between the realized measure and the conditional (day and night) variance. The mathematical structure is thus elegant, facilitates volatility estimation, and allows the inclusion of return-volatility dependence. We call the structure bivariate

because it uses both day and night volatility information, as opposed to the univariate ones that only use day information. To strengthen the robustness of our empirical research, we further extended this idea to a number of realized GARCH models that use day and intraday volatility information, creating an equivalent set of bivariate models that additionally use night volatility information. We obtained a class of realized GARCH models that incorporate day, night, and intraday volatility measures; they were assessed against their counterparts that did not include night volatility information using an extended set of 10 stock prices. Empirical results of the forecasting performance assessment show a degree of improvement of the newly proposed models over those that do not include night volatility measures. This finding suggests the potential of our method for volatility forecasting problems for financial assets and other assets with night latent volatility information.

Financial volatility modeling has benefited significantly from the availability of high-frequency data. The main interest in modeling using frequently sampled information and integrating it into models built to estimate day conditional variance was initiated by Andersen and Bollerslev (1998), who used realized volatility estimates extracted from intraday data (realized variance) as better estimates of conditional volatility than squared returns. They proved that by adding up squared intraday returns, the forecasted volatility would correlate closely to the future latent volatility factor.

Engle (2002) was among the first econometricians who extended the standard GARCH model to include an exogenous realized measure (the realized variance) in the conditional variance (GARCH) equation. In this model, the realized measures' variation is not explained; thus, such models (GARCH-*X*) are considered incomplete. Engle and Gallo (2006) proposed the multiplicative error model (MEM), which was the first attempt to contain a separate GARCH structure equation for the realized measure. A similar complete model nested in a MEM setting is the high frequency based volatility (HEAVY) model of Shephard and Sheppard (2010). Both MEM and HEAVY models are difficult to use as they work with multiple latent processes—for every realized measure used, there is a corresponding latent volatility process. The Realized GARCH model proposed by Hansen et al. (2012) combines a GARCH structure for returns with realized measures of volatility. Compared with MEM and HEAVY models, the Realized GARCH model takes advantage of the natural relationship between the realized measure and the conditional variance. Instead of introducing additional latent factors, it proposes a single measurement equation in which the realized measure is a consistent estimator of the integrated variance. Besides its elegant mathematical structure, the Realized GARCH model is easy to estimate, captures the return-volatility dependence (leverage effect), and has been empirically shown to outperform conventional GARCH. A more robust version of the Realized GARCH model was introduced by Banulescu-Radu et al. (2019), suggesting a variant that is less sensitive to outliers and minimizes the impact on volatility of days with extreme negative volatility shocks. A realized exponential GARCH model that can use multiple realized volatility measures for the modeling of a return series, using a similar framework, has also been proposed (Hansen and Huang 2016). Finding that the Realized GARCH model was insufficient for capturing the long memory of underlying volatility, Huang et al. (2016) developed a parsimonious variant of the Realized GARCH model by introducing Corsi's (2009) heterogeneous autoregressive (HAR) specification in the volatility dynamics. A multivariate GARCH model that incorporates realized measures of variances and covariances was also introduced by Hansen et al. (2014), but it did not suggest the introduction of night volatility information. Bollerslev et al. (2018) proposed asymmetric multivariate volatility models that exploit estimates of variances and covariances based on the signs of high-frequency returns to allow for more nuanced responses to positive and negative return shocks than the threshold leverage effect. Hansen et al. (2019) proposed a multivariate GARCH model that incorporates realized measures for the covariance matrix of returns.

Overnight (close-to-open) volatility is usually higher than the five-minute realized volatility estimated during trading hours, and the close-to-open price differential may trigger a distorting effect on the realized volatility. Thus, the inclusion of overnight returns when constructing the realized conditional covariance matrix of the daily returns has been empirically documented to reduce

information loss and consequently improve volatility forecasting. A common approach to account for volatility during the market's closing hours has been to calculate a close-to-open return from the price change recorded between the trading day closing and the next trading day opening, and then add its squared value to the sum of intraday returns (Bollerslev et al. 2009; Martens 2002; Blair et al. 2001). Hansen and Lunde (2005) compounded optimal weights corresponding to overnight returns and to the sum of intraday returns, and Fleming and Kirby (2011) and Fuertes and Olmo (2013) further applied it. De Pooter et al. (2008) and Fleming et al. (2003) computed it in matrix form by incorporating the cross-product of the vector of overnight returns in the summation of the matrix that provided the covariance matrix of the daily returns, acknowledging that the outer product of the vector of overnight returns is an inaccurate estimator of the integrated covariance matrix for the period when markets were closed (Fleming et al. 2003). Koopman et al. (2005); Martens (2002); and Angelidis and Degiannakis (2008) excluded the noisy overnight returns to compute an estimate of volatility during trading hours, instead of daily volatility; then, they scaled up the sum of intraday returns to cover the whole 24-h day. The literature has not yet reached a consensus on the best method of accounting for overnight returns; however, Ahoniemi and Lanne (2013) suggested that the weighted sum of the squared overnight return and the sum of intraday squared returns was the most accurate measure of realized volatility for the Standard&Poor's' S&P 500 index.

This paper suggests a method of capturing and incorporating night volatility into the day conditional volatility equation of one low-frequency as well as a number of high-frequency GARCH models. We propose a two-factor structure of the conditional variance, one for night and one for day variance, in a realized GARCH setting that takes advantage of the natural relationship between the realized measure and the conditional (day and night) variance. The mathematical structure is thus elegant, facilitates volatility estimation, and allows the inclusion of the return-volatility dependence. A general framework is formulated; based on it, a set of GARCH models is adapted such that it uses the estimation of night latent volatility to model day conditional volatility. This approach enabled us to document, in an empirical context, whether the introduction of the night volatility component, in the two-factor structure and realized GARCH setting we propose, improved the volatility modeling for each of the models discussed. The new models are called bivariate as they use both night and day volatility information and are defined to work in typical financial settings, such as volatility modeling of stock and commodity prices. We assessed the performance of the bivariate models by comparing the error functions of the forecasts of the bivariate models with those obtained when the simple versions of the models, which do not use night volatility information, were used. We call the latter models univariate models. The scope of this study was thus to analyze whether the use of night volatility information in the forms proposed improves the modeling of day volatility.

The paper proceeds as follows. Section 2 proposes the new set of bivariate realized models. Section 3 describes the data and methodology, and Section 4 summarizes the results. The paper concludes with Section 5, where final remarks are presented, and some future lines of research are proposed.

2. Bivariate Realized Models

2.1. Base Model

Existing high-frequency GARCH models estimate day conditional variance using day and intraday volatility information. We developed a class of realized models that allow constructing day volatility estimates with day, intraday, and night volatility information. Models previously proposed use return and volatility information estimated from trades that occurred during the trading day to estimate next-day volatility. However, latent volatility existing between the trading periods (called night volatility) has scarcely been considered in the day volatility estimation problem. The idea emerged from an observation on financial stock time series; prices at market closing differ from those at market opening the following trading day, although during the night the market is closed and thus no

transactions occur, so no intranight information exists. Despite the lack of night trades, latent (night) volatility still occurs, causing a price mismatch. We examined whether this latent night volatility can be modeled and whether, if incorporated into the conditional volatility modeling, it would help to provide better estimates of day volatility. Compared to other researchers that also modeled overnight returns, we proposed a two-factor structure in a realized GARCH setting with a GARCH equation that links day/night volatility to returns, night/day volatility, and intraday volatility of the previous day. This allowed us to retain the benefits of the Realized GARCH model of Hansen et al. (2012), namely, to take advantage of the natural relationship between the realized measure and the conditional day (and night for the models we proposed in the current paper) variance in an elegant structure that facilitates volatility estimation, allowed us to capture the return-volatility dependence, and was previously proved to outperform traditional GARCH. Below, we presented a method to capture this volatility and to insert it into the day conditional volatility equation.

The starting model is a reduced form Bivariate Realized GARCH model, which is a Realized GARCH model with night volatility information and exogenous realized measures, defined as follows:

$$r_t = r_t^{\bullet} + r_t^{\circ},\ z_t^{\circ} = \frac{r_t^{\circ} - \mu^{\circ}}{\sqrt{h_t^{\circ}}},\ z_t^{\bullet} = \frac{r_t^{\bullet} - \mu^{\bullet}}{\sqrt{h_t^{\bullet}}}, \tag{1}$$

$$\log h_t^{\circ} = \omega^{\circ} + \tau^{(\circ 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\circ 2)}\left(z_{t-1}^{\circ}\right) + \beta^{\circ}\ \log h_{t-1}^{\circ} + \gamma^{\circ}\ \log x_{t-1}, \tag{2}$$

$$\log h_t^{\bullet} = \omega^{\bullet} + \tau^{(\bullet 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\bullet 2)}\left(z_{t-1}^{\circ}\right) + \beta^{\bullet}\ \log h_{t-1}^{\bullet} + \gamma^{\bullet}\ \log x_{t-1}, \tag{3}$$

where $\bullet$ denotes the night information, $^{\circ}$ denotes the day information of the vector, r_t is the return, $z_t \sim iid(0,1)$, $u_t \sim iid\left(0, \sigma_u^2\right)$, $h_t^{\circ} = var(r_t^{\circ}|\mathcal{F}_{t-1})$, $h_t^{\bullet} = var\left(r_t^{\bullet}|\mathcal{F}_{t-1}\right)\ \mathcal{F}_t = \sigma(r_t,\ x_t,\ r_{t-1},\ x_{t-1}, \ldots)$, $r_t^{\circ} = 100 \times \left(\log(price)_{t_{close}} - \log(price)_{t_{open}}\right)$, and $r_t^{\bullet} = 100 \times \left(\log(price)_{t_{open}} - \log(price)_{t-1_{close}}\right)$. As such, r_t is the sum between night $r_t^{\bullet}$ and day r_t° returns, z_t° represents the standardized day returns, and $z_t^{\bullet}$ represents the standardized night returns, whereas μ° is the means of day returns and $\mu^{\bullet}$ is the means of night returns. All τ's are coefficients of the standardized returns that follow to be estimated through the maximum log-likelihood function (MLE). If marked by $^{\circ}$, τ represents the coefficients of the standardized returns in the equation of conditional day volatility, and if marked by $\bullet$, τ represents the coefficients of the standardized returns in the equation of conditional night volatility. The numbers next to $^{\circ}$ or $\bullet$ are for indexing purposes: For example, $\tau^{(\circ 1)}$ and $\tau^{(\circ 2)}$ are two coefficients of the standardized returns in the equation of conditional day volatility that follow to be estimated through MLE.

Thus, the base model is formed of three equations: The return equation, which is the sum between day (open-to-close) returns and night (close-to-open) returns, and two conditional volatility equations, as follows: The first expresses day volatility as a function of previous day (z_{t-1}°) and night ($z_{t-1}^{\bullet}$; standardized) returns, conditional day variance (h_{t-1}°), and a realized measure of volatility (x_{t-1}; realized kernel, high–low, realized variance, etc.). The second defines night volatility as a function of previous day (z_{t-1}°) and night ($z_{t-1}^{\bullet}$; standardized) returns, conditional night variance ($h_{t-1}^{\bullet}$), and a realized measure of volatility (x_{t-1}). Notably, in this model (called reduced form for this reason), the realized measure is not endogenized nor linked to the day volatility measure through a measurement equation, but rather is treated as an exogenous variable. We added this equation to the complete form of the model that was documented in the next section. The realized measure was compounded from intraday prices recorded throughout the day.

2.2. Extended Models

We used the base model structure and extended its idea to a class of best-known GARCH-type models. We used this approach as all models used share the same structure and thus similar properties, which enabled us to set up a similar bivariate configuration. The aim was to construct a group of models that takes advantage of night volatility estimation, and also defines the existing natural relationship

between the realized measures and the conditional day and night variance. As such, we proposed four new realized models and one non-realized model: Bivariate Realized GARCH (1,1), with an endogenous component of realized measure and therefore a separate measurement equation, which we will call a complete version model; Bivariate Exponential GARCH-X (Bivariate EGARCH-X), that is a bivariate exponential generalized autoregressive conditional heteroskedastic model with an exogenous realized measure; Bivariate Realized EGARCH (1,1); Bivariate Realized GARCH (2,2); and Bivariate EGARCH (1,1). The detailed specifications of the bivariate models we propose are provided in Table 1.

Table 1. Summary of the bivariate realized generalized autoregressive conditional heteroskedasticity (GARCH) models proposed.

Model	Return Equations	GARCH Equations	Measurement Equation
Bivariate EGARCH (1,1)	$r_t = r_t^{\bullet} + r_t^{\circ}$ $z_t^{\circ} = \frac{r_t^{\circ} - \mu^{\circ}}{\sqrt{h_t^{\circ}}}$ $z_t^{\bullet} = \frac{r_t^{\bullet} - \mu^{\bullet}}{\sqrt{h_t^{\bullet}}}$	$\log h_t^{\circ} = \omega^{\circ} + \varepsilon^{(\circ 1)}\left[\tau^{(\circ 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\circ 2)}\left(z_{t-1}^{\circ}\right)\right] + \varepsilon^{(\circ 2)}\left\{\left[\tau^{(\circ 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\circ 2)}\left(z_{t-1}^{\circ}\right)\right]^2 - 1\right\} + \beta^{\circ} \log h_{t-1}^{\circ}$ $\log h_t^{\bullet} = \omega^{\bullet} + \varepsilon^{(\bullet 1)}\left[\tau^{(\bullet 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\bullet 2)}\left(z_{t-1}^{\circ}\right)\right] + \varepsilon^{(\bullet 2)}\left\{\left[\tau^{(\bullet 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\bullet 2)}\left(z_{t-1}^{\circ}\right)\right]^2 - 1\right\} + \beta^{\bullet} \log h_{t-1}^{\bullet}$	
Bivariate EGARCH-X	$r_t = r_t^{\bullet} + r_t^{\circ}$ $z_t^{\circ} = \frac{r_t^{\circ} - \mu^{\circ}}{\sqrt{h_t^{\circ}}}$ $z_t^{\bullet} = \frac{r_t^{\bullet} - \mu^{\bullet}}{\sqrt{h_t^{\bullet}}}$	$\log h_t^{\circ} = \omega^{\circ} + \varepsilon^{(\circ 1)}\left[\tau^{(\circ 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\circ 2)}\left(z_{t-1}^{\circ}\right)\right] + \varepsilon^{(\circ 2)}\left\{\left[\tau^{(\circ 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\circ 2)}\left(z_{t-1}^{\circ}\right)\right]^2 - 1\right\} + \beta^{\circ} \log h_{t-1}^{\circ} + \gamma^{\circ} \log x_{t-1}$ $\log h_t^{\bullet} = \omega^{\bullet} + \varepsilon^{(\bullet 1)}\left[\tau^{(\bullet 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\bullet 2)}\left(z_{t-1}^{\circ}\right)\right] + \varepsilon^{(\bullet 2)}\left\{\left[\tau^{(\bullet 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\bullet 2)}\left(z_{t-1}^{\circ}\right)\right]^2 - 1\right\} + \beta^{\bullet} \log h_{t-1}^{\bullet} + \gamma^{\bullet} \log x_{t-1}$	
Bivariate Realized EGARCH (1,1)	$r_t = r_t^{\bullet} + r_t^{\circ}$ $z_t^{\circ} = \frac{r_t^{\circ} - \mu^{\circ}}{\sqrt{h_t^{\circ}}}$ $z_t^{\bullet} = \frac{r_t^{\bullet} - \mu^{\bullet}}{\sqrt{h_t^{\bullet}}}$	$\log h_t^{\circ} = \omega^{\circ} + \varepsilon^{(\circ 1)}\left[\tau^{(\circ 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\circ 2)}\left(z_{t-1}^{\circ}\right)\right] + \varepsilon^{(\circ 2)}\left\{\left[\tau^{(\circ 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\circ 2)}\left(z_{t-1}^{\circ}\right)\right]^2 - 1\right\} + \beta^{\circ} \log h_{t-1}^{\circ} + \gamma^{\circ} \log x_{t-1}$ $\log h_t^{\bullet} = \omega^{\bullet} + \varepsilon^{(\bullet 1)}\left[\tau^{(\bullet 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\bullet 2)}\left(z_{t-1}^{\circ}\right)\right] + \varepsilon^{(\bullet 2)}\left\{\left[\tau^{(\bullet 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\bullet 2)}\left(z_{t-1}^{\circ}\right)\right]^2 - 1\right\} + \beta^{\bullet} \log h_{t-1}^{\bullet} + \gamma^{\bullet} \log x_{t-1}$	$\log x_t = \xi + \varphi \log h_t^{\circ} + \vartheta \log h_t^{\bullet} + \delta^{(\circ 1)} z_t^{\bullet} + \delta^{(\circ 2)} z_t^{\circ} + u_t$
Bivariate Realized GARCH (1,1) complete form	$r_t = r_t^{\bullet} + r_t^{\circ}$ $z_t^{\circ} = \frac{r_t^{\circ} - \mu^{\circ}}{\sqrt{h_t^{\circ}}}$ $z_t^{\bullet} = \frac{r_t^{\bullet} - \mu^{\bullet}}{\sqrt{h_t^{\bullet}}}$	$\log h_t^{\circ} = \omega^{\circ} + \tau^{(\circ 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\circ 2)}\left(z_{t-1}^{\circ}\right) + \beta^{\circ} \log h_{t-1}^{\circ} + \gamma^{\circ} \log x_{t-1}$ $\log h_t^{\bullet} = \omega^{\bullet} + \tau^{(\bullet 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\bullet 2)}\left(z_{t-1}^{\circ}\right) + \beta^{\bullet} \log h_{t-1}^{\bullet} + \gamma^{\bullet} \log x_{t-1}$	$\log x_t = \xi + \varphi \log h_t^{\circ} + \vartheta \log h_t^{\bullet} + \delta^{(\circ 1)} z_t^{\bullet} + \delta^{(\circ 2)} z_t^{\circ} + u_t$
Bivariate Realized GARCH (2,2)	$r_t = r_t^{\bullet} + r_t^{\circ}$ $z_t^{\circ} = \frac{r_t^{\circ} - \mu^{\circ}}{\sqrt{h_t^{\circ}}}$ $z_t^{\bullet} = \frac{r_t^{\bullet} - \mu^{\bullet}}{\sqrt{h_t^{\bullet}}}$	$\log h_t^{\circ} = \omega^{\circ} + \tau^{(\circ 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\circ 2)}\left(z_{t-1}^{\circ}\right) + \alpha^{\circ} \log\left(max\left(r_{t-1}^{\circ 2}, 10^{-20}\right)\right) + \beta^{(\circ 1)} \log h_{t-1}^{\circ} + \beta^{(\circ 2)} \log h_{t-2}^{\circ} + \gamma^{(\circ 1)} \log x_{t-1} + \gamma^{(\circ 2)} \log x_{t-2}$ $\log h_t^{\bullet} = \omega^{\bullet} + \tau^{(\bullet 1)}\left(z_{t-1}^{\bullet}\right) + \tau^{(\bullet 2)}\left(z_{t-1}^{\circ}\right) + \alpha^{\bullet} \log\left(max\left(r_{t-1}^{\bullet 2}, 10^{-20}\right)\right) + \beta^{(\bullet 1)} \log h_{t-1}^{\bullet} + \beta^{(\bullet 2)} \log h_{t-2}^{\bullet} + \gamma^{(\bullet 1)} \log x_{t-1} + \gamma^{(\bullet 2)} \log x_{t-2}$	$\log x_t = \xi + \varphi \log h_t^{\circ} + \vartheta \log h_t^{\bullet} + \varepsilon_1\left[\delta^{(\circ 1)} z_t^{\bullet} + \delta^{(\circ 2)} z_t^{\circ}\right] + \varepsilon_2\left\{\left[\delta^{(\circ 1)} z_t^{\bullet} + \delta^{(\circ 2)} z_t^{\circ}\right]^2 - 1\right\} + u_t$

Next, we summarized the main features of each model. All share similar return equations as in the case of the base model—the daily return r_t is the sum between open-to-close return (day return) r_t° and close-to-open return (night return) $r_t^{\bullet}$. The GARCH equations share distinct properties but they have unique features as well. All define the day (open-to-close) volatility h_t° as a function of day z_t° and night $z_t^{\bullet}$ standardized returns as defined above, and also as a function of the previous day (open-to-close) volatility. Except for the Bivariate EGARCH (1,1) and the reduced form Bivariate Realized GARCH models, all other models also include the relationship between day volatility h_t° and intraday volatility x_{t-1} in the GARCH equation. Since Bivariate EGARCH (1,1) is not a realized model, it does not contain intraday information. In our Bivariate EGARCH-X model, intraday volatility x_{t-1} is treated as an exogenous variable and is thus not linked to any other variable. However, all other realized models incorporate a third equation, the measurement equation, which defines the joint dependence between

r_t and x_t. x_t is thus "endogenized" by being formulated as a function of day (open-to-close) volatility, night (close-to-open) volatility, and day and night standardized returns (z_t° and $z_t^{\bullet}$, respectively).

3. Data and Estimation Methodology

We used tick data sampled along 3537 trading days during the period of 30 August 2004–31 December 2018, corresponding to 10 stocks: AIG (American International Group, Inc.), AXP (American Express Company), BAC (Bank of America Corporation), CSCO (Cisco Systems, Inc.), F (Ford Motor Company (F)), GE (General Electric Company), INTC (Intel Corporation), JPM (JPMorgan Chase & Co.), MSFT (Microsoft Corporation), and T (AT&T Inc.). To avoid the outliers that would result from quiet days, the half trading days around the Christmas and Thanksgiving holidays were removed.

We opted for estimating intraday volatility by compounding realized kernels instead of the more widely used realized variance, as it is generally acknowledged that squared daily returns provide a poor estimation of actual intraday volatility. Realized kernels are robust for microstructure errors or frictions, which are known to cause endogenous and dependent noise terms. They are used to estimate the quadratic variation in an efficient price process when the time stamps in every day do not match (non-synchronous, with irregularly spaced observations) and when the high-frequency time series described by the prices are noisy with many microstructure effects. We compounded the realized kernels as measures of intraday volatility (x_t) using the methodology of Barndorff-Nielsen et al. (2009, 2011). The framework is given by Y, a variable that is the sum of a Brownian semi-martingale and a jump process, as follows:

$$Y_t = \int_0^t a_u du + \int_0^t \sigma_u dW_u + J_t. \tag{4}$$

For the purpose of our exercise, we need to find the quadratic variation of Y, $[Y] = \int_0^T \sigma_u^2 du + \sum_{i=1}^{N_T} C_i^2$. Barndorff-Nielsen et al. (2009, 2011) estimated it from the noisy discrete observations X_{τ_j} of Y_{τ_j}, $0 = \tau_0 < \tau_1 < \ldots < \tau_n = T$, where $X_{\tau_j} = Y_{\tau_j} + U_{\tau_j}$ and U_{τ_j} represents the market microstructure effects (noise). Barndorff-Nielsen et al. (2009, 2011) estimated this quadratic variation by proposing realized kernels, a non-negative estimator that is constructed as follows.

The first challenge with the tick data is the non-synchronicity. Non-synchronous trading occurs when the trades or quotes appear at irregularly spaced times across stocks, which is usually the case in stock markets, especially those with low liquidity or stale prices. Barndorff-Nielsen et al. (2011) solved this by suggesting a refresh time when all the stocks are traded. We implemented the same method by recording the prices only when (and immediately after) all of them were traded.

To eliminate start and end effects and their associated errors, which are averaged through this procedure, we proceeded to jittering (averaging) the first and last two prices, as also suggested by Barndorff-Nielsen et al. (2011)Barndorff-Nielsen et al. Having synchronized and constructed the time series by jittering at the initial and final time points, we defined the semi-definite realized kernels, as follows, according to Barndorff-Nielsen et al. (2009, 2011):

$$K(X) = \sum_{h=-H}^{H} k\left(\frac{h}{H+1}\right)\gamma_h, \text{ where } \gamma_h = \sum_{j=|h|+1}^{n} x_j x_{j-|h|}, \tag{5}$$

where $k(x)$ is a kernel weight function that has the $k(0) = 1$, $k'(0) = 0$ property, and k is twice differentiable with continuous derivatives.

Barndorff-Nielsen et al. (2009) used a Parzen kernel as it satisfies the smoothness conditions through $k'(0) = k'(1) = 0$, and its estimates are positive. We made the same choice, and used the same Parzen kernel function:

$$k(x) = \begin{cases} 1 - 6x^2 + 6x^3, & 0 \le x \le 1/2 \\ 2(1-x)^3, & 1/2 \le x \le 1 \\ 0, & x > 1 \end{cases}. \tag{6}$$

The optimal choice of bandwidth, according to Barndorff-Nielsen et al. (2009), which we chose to use, is $H^* = c^* \xi^{4/5} n^{3/5}$, with $c^* = \left\{ \frac{k''(0)^2}{k^{0,0}_{\bullet}} \right\}^{1/5}$ and $\xi^2 = \frac{\omega^2}{\sqrt{T \int_0^T \sigma_u^4 du}}$, where $c^* = \left((12)^2\right)^{1/5} = 3.5134$ for the Parzen kernel. $\int_0^T \sigma_u^4 du$ is called the integrated quarticity, and, in our empirical exercise, it equals RV_{sparse}. This denotes a subsampled realized variance based on 20-min returns. By calculating 1200 realized variances by shifting the first observation recorded time in 1-s increments, we obtained a number of realized variance estimators. We averaged them and obtained RV_{sparse}. ω^2 was estimated by calculating the realized variance using every ith trade. We varied the starting point, and thereby produced i realized variances, namely RV^i_{dense}. Thus, our ω^2 estimator was calculated as:

$$\hat{\omega}^2_{(j)} = \frac{RV^{(j)}_{dense}}{2n_{(j)}}, \; j = 1, \ldots, i, \tag{7}$$

where $n_{(j)}$ is the number of non-zero returns used to estimate $RV^{(j)}_{dense}$. The estimate of ω^2 is then the average of the j estimates,

$$\hat{\omega}^2 = \frac{1}{i} \sum_{j=1}^{i} \hat{\omega}^2_{(j)}. \tag{8}$$

By design, the realized kernel is positive semi-definite and the rate of convergence is $n^{1/5}$.

We estimated the in-sample and out-of-sample (3000th day in the sample, 24 November 2016, the cutoff point) in both the univariate and bivariate models with respect to each of the 10 stocks. The univariate models considered are the standard realized versions of the GARCH model (Realized GARCH, Realized EGARCH, EGARCH-X, and Realized GARCH (2,2)), as well as the EGARCH model. The estimated bivariate models are those mentioned in Section 2 (Bivariate EGARCH, reduced and complete forms of Bivariate Realized GARCH, Bivariate Realized EGARCH, Bivariate EGARCH-X, and Bivariate Realized GARCH (2,2)).

The estimation was performed by maximizing the total log-likelihood functions (MLE), namely the sum of partial log-likelihood functions for the returns and for the intraday measures; the ranking criterion with respect to the MLE was the partial log-likelihood function for returns solely. We used MLE to estimate both the proposed bivariate models and a number of univariate models that do not include night volatility information.

The log-likelihood function used in the estimation of the above models takes the form $l\left(r_t^{\bullet}, r_t^{\circ}, x_t\right) = L_1$ for Bivariate EGARCH and Bivariate EGARCH-X, or $l\left(r_t^{\bullet}, r_t^{\circ}, x_t\right) = L_1 + L_2$ for Bivariate Realized GARCH complete version, Bivariate Realized EGARCH (1,1), and Bivariate Realized GARCH (2,2) (Appendix A), where $L_1 = -\frac{1}{2} \sum_{t=1}^{n} \left\{ 2\log(2\pi) + \log\left(1-\rho^2\right) + \log h_t^{\bullet} + \log h_t^{\circ} + \frac{(r_t^{\bullet} - \mu^{\bullet})^2 / h_t^{\bullet} + (r_t^{\circ} - \mu^{\circ})^2 / h_t^{\circ}}{(1-\rho^2)} - \frac{2\rho}{(1-\rho^2)} \frac{(r_t^{\bullet} - \mu^{\bullet})(r_t^{\circ} - \mu^{\circ})}{\sqrt{h_t^{\bullet} h_t^{\circ}}} \right\}$ and $L_2 = -\frac{1}{2} \sum_{t=1}^{n} \left\{ \log(2\pi) + \log\left(\sigma_u^2\right) + u_t^2 / \sigma_u^2 \right\}$.

To evaluate whether introducing night volatility estimations in models' equations improves the day volatility estimation, we calculated two loss functions, root mean squared error (RMSE) and mean absolute error (MAE). Based on these, we documented the number of models for each in-sample and out-of-sample estimation for each of the 10 stocks, at which MAE and RMSE were smaller. This allowed us to draw conclusions about the better performance of the bivariate or univariate models. Based on the size of the loss functions obtained at each estimation, we analyzed the performance of the new models that included night volatility estimates. This contributed to our objective by documenting whether or not night volatility information improves the estimation of day volatility with respect to the main GARCH-type of models proposed in the literature.

The maximized log-likelihood functions in univariate and bivariate estimations are provided in Tables A1 and A2 in Appendix B. As the log-likelihood functions of the bivariate models differ from those of the univariate versions (for the bivariate estimation, we maximized a

bi-dimensional vector $\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}$ with a non-null correlation factor (ρ) between its subvectors), it makes little sense to compare the values of the MLEs across the univariate and bivariate models to document an improvement or loss of performance when introducing night volatility estimates. Specifically, the log-likelihood function for the bivariate models is: $\log l(r_t^{\bullet}, r_t^{\circ}) = -\frac{1}{2}\sum_{t=1}^{n}\left\{2\log(2\pi) + \log\left(1-\rho^2\right) + \log(h_t^{\bullet}) + \log(h_t^{\circ}) + \frac{r_t^{\bullet 2}/h_t^{\bullet} + r_t^{\circ 2}/h_t^{\circ}}{(1-\rho^2)} - \frac{2\rho}{(1-\rho^2)}\frac{r_t^{\bullet} r_t^{\circ}}{\sqrt{h_t^{\bullet} h_t^{\circ}}}\right\}$, where $\rho = corr\left(r_t^{\circ}, r_t^{\bullet}\right)$. In the univariate models' case, the log-likelihood function is $\log l(r_t) = -\frac{1}{2}\sum_{t=1}^{n}\left[\log(2\pi) + \log\left(h_t\right) + \frac{(r_t-\mu)^2}{h_t}\right]$ for EGARCH and EGARCH-X, and $\log l(r_t) = -\frac{1}{2}\sum_{t=1}^{n}\left[2\log(2\pi) + \log\left(h_t\right) + \frac{(r_t-\mu)^2}{h_t} + \log\left(\sigma_u^2\right) + \frac{u_t^2}{\sigma_u^2}\right]$ for Realized EGARCH, Realized GARCH, and Realized GARCH (2,2). As such, we could not use this method to evaluate the performance of the bivariate models, as we would be comparing the values of estimations of different functions.

Thus, for the purpose of documenting the gain or loss in accuracy, we used the standard method in econometrics for evaluating the models' performance—that of calculating two loss functions (RMSE and MAE)—which would better assess whether adding night volatility information with a two-factor structure in a realized GARCH setting improves estimations of next-day volatility.

4. Results

The standard method used in econometrics to evaluate models' performance is to calculate the size of the loss functions, among which RMSE and MAE are the most common and reliable. We calculated them for both in-sample and out-of-sample estimations, and our results indicate an improvement when night volatility estimations were included in the equations of the day conditional volatility in almost every case.

We worked with a number of models that have different features and for which adding an estimation of night volatility may contribute to the volatility estimation. For example, by inspecting the results for RMSE (in-sample estimation) in Table 2, the improvement was evident for 55 out of 60 cases (1 loss function result × 6 models evaluated × 10 stocks). The cases in which the improvement could not be documented are marked with red (for RMSE) or green (for MAE) numbers in Table 2. In the five cases in which this was not evident, four of them were for Realized GARCH (2,2). This means that Realized GARCH (2,2) only shows some features that did not work better when the night volatility estimates were considered given the way in which the model was designed. This may be because, compared to the other models that model next-day volatility by only using information from the previous day and night, Realized GARCH (2,2) uses information on the previous night volatility as well as information on returns and volatility of the previous two days. We thought that this might be the problem with this model, but it would need to be proven empirically; we left this question for future work.

Table 2. Loss functions in univariate and bivariate estimations; in-sample.

Stock		EGARCH		EGARCH-X		Realized EGARCH		Realized GARCH				Realized GARCH (2,2)	
		Univ	Biv	Univ	Biv	Univ	Biv	Univ	Biv (com)	Univ	Biv (red)	Univ	Biv
AIG	RMSE	203.3	188.6	203.7	195.1	189.8	195.8	254.0	218.2	254.0	219.7	190.6	250.2
	MAE	18.0	15.1	20.1	16.5	17.2	15.2	22.9	21.0	22.9	21.1	17.4	21.4
AXP	RMSE	6.9	6.3	6.7	6.2	6.8	5.5	6.8	5.5	7.0	5.2	7.0	7.1
	MAE	3.1	2.3	3.1	2.6	3.1	2.2	3.2	2.0	3.3	1.9	3.0	2.3
BAC	RMSE	16.7	15.9	16.3	15.4	15.9	15.1	16.2	15.7	16.4	15.3	15.9	15.1
	MAE	4.5	3.5	4.0	3.2	4.1	2.8	4.3	3.0	3.8	3.1	4.2	2.7
CSCO	RMSE	6.5	5.6	6.6	6.4	6.8	5.7	6.4	5.5	6.8	5.9	6.8	6.6
	MAE	3.1	2.2	3.1	2.4	3.3	2.5	3.0	2.2	3.0	1.9	3.2	2.3
F	RMSE	16.9	16.3	16.4	15.1	16.3	15.5	16.1	15.7	16.5	15.3	16.2	14.8
	MAE	4.0	3.2	3.8	3.3	4.2	3.3	4.3	3.3	4.2	2.9	4.0	3.0
GE	RMSE	6.8	6.0	6.9	6.5	6.6	5.8	6.9	5.5	6.7	5.5	6.4	7.0
	MAE	3.3	1.9	2.7	2.2	3.1	1.8	3.2	2.2	3.2	1.8	2.8	2.4
INTC	RMSE	16.6	15.9	16.3	15.4	16.4	15.3	16.5	15.0	16.5	15.1	16.1	15.2
	MAE	4.4	3.1	4.0	3.5	4.3	3.1	4.2	3.0	3.7	3.2	3.9	2.7
JPM	RMSE	11.4	10.4	11.1	9.7	10.7	10.6	11.1	9.9	10.8	10.1	11.1	9.9
	MAE	3.9	2.6	3.5	2.7	3.5	2.7	3.3	3.0	3.6	2.2	3.6	2.7
MSFT	RMSE	6.8	6.4	6.6	6.1	6.8	5.6	6.9	5.6	6.8	5.5	6.6	7.1
	MAE	3.3	1.8	2.8	2.5	3.1	2.1	3.3	2.0	3.4	2.1	3.4	2.9
T	RMSE	16.8	15.9	16.5	15.2	16.7	15.2	16.4	15.4	16.6	15.7	16.2	14.9
	MAE	4.3	3.5	4.2	3.4	4.0	3.0	4.1	3.0	4.5	3.5	4.5	3.1

This conclusion was strengthened by examining the MAE results. When considering MAE as an evaluation tool, the bivariate models produced superior forecasting ability in 59 out of 60 cases, indicating an improvement for the models that included night volatility estimation in the day volatility modeling. However, in only one case out of 60 was the improvement not evident, for the same Realized GARCH (2,2) model. As such, the model itself appears to be problematic, not the evaluation we performed. As mentioned above, we thought that the problem with this model was that it models conditional day volatility by including in the model information on day volatility and returns from the previous two days, instead of one day only as we did for the other models. In Bivariate Realized GARCH (2,2), we considered only one-night volatility information instead of considering the night volatility estimation from the previous two nights.

Univ and *Biv* stand for *Univariate* and *Bivariate*, respectively, while *com* and *red* stand for *complete* and *reduced*, respectively. Red and green numbers indicate the stances in which bivariate models perform worse than the univariate ones (when evaluated according to RMSE or MAE, respectively).

When examining the results for the out-of-sample estimations in Table 3, we found that of 60 evaluations with RMSE, 53 showed forecasting improvement when night volatility information was used. In the seven cases in which the improvement was not evident, three were recorded for the same Realized GARCH (2,2) model. The remaining four belonged to various other models, one for each. However, we observed another pattern. Most of the failures in documenting an improvement were for the same stock: AIG. This suggests that the results were sensitive not only to the model (as we explained earlier with the way in which Realized GARCH (2,2) was built), but were also sensitive to the stock choice. Since AIG persistently failed in showing an improvement when using night volatility information, AIG price recordings should be more carefully examined to understand what makes it less sensitive to this modeling suggestion, including examining the amount of the stock price differential (the difference between the market closing and the market opening prices), and also understanding the roots of the volatility transmission for this stock in particular. Again, we left this as exploratory work for the future paper. When ranked according to MAE, 58 results out of 60 indicated improvement,

whereas only two cases (among them, one for Realized GARCH (2,2)) did not. Again, both estimations indicated strong evidence in favor of including night volatility estimation in the modeling problem of day volatility.

Table 3. Loss functions in univariate and bivariate estimations; out-of-sample.

		EGARCH		**EGARCH-X**		**Realized EGARCH**		**Realized GARCH**				**Realized GARCH (2,2)**	
Stock		**Univ**	**Biv**	**Univ**	**Biv**	**Univ**	**Biv**	**Univ**	**Biv (com)**	**Univ**	**Biv (red)**	**Univ**	**Biv**
AIG	RMSE	565.5	574.3	552.0	532.9	544.0	566.4	552.9	572.4	552.9	573.3	538.9	584.7
	MAE	109.0	100.4	106.9	104.1	103.8	102.5	122.0	103.1	122.1	103.0	104.6	121.1
AXP	RMSE	14.3	14.0	14.5	13.3	14.2	12.8	14.0	13.0	14.0	12.6	13.9	12.8
	MAE	8.8	8.1	9.0	9.1	8.6	7.8	8.7	7.7	8.7	8.1	9.0	7.7
BAC	RMSE	44.0	43.1	43.2	42.1	44.5	42.6	43.5	42.1	43.3	42.6	43.3	43.1
	MAE	19.4	17.7	18.4	17.6	18.0	17.2	18.3	17.8	18.6	17.2	18.0	17.4
CSCO	RMSE	14.2	14.2	14.4	13.1	14.3	13.2	13.9	12.8	14.2	13.1	13.8	12.7
	MAE	8.9	8.0	9.0	8.9	8.7	7.8	9.0	8.1	9.0	8.0	9.2	8.1
F	RMSE	44.0	43.0	43.0	42.4	44.3	42.3	43.4	42.6	43.5	42.0	43.3	43.4
	MAE	19.1	18.1	18.6	17.4	18.6	16.7	18.5	17.9	18.3	17.0	18.6	17.0
GE	RMSE	14.7	14.4	14.3	13.3	14.0	13.3	13.9	12.7	14.3	12.6	13.9	13.0
	MAE	9.5	7.6	8.8	8.7	8.6	7.8	8.8	8.2	8.4	8.0	9.4	7.8
INTC	RMSE	44.4	43.3	43.2	42.2	44.4	42.0	43.7	42.2	43.4	41.9	43.4	43.3
	MAE	19.6	18.1	18.5	17.1	18.2	17.0	18.4	18.0	18.3	17.1	18.3	17.3
JPM	RMSE	26.7	26.1	26.1	25.3	25.4	24.0	25.2	24.8	25.5	24.6	25.4	25.1
	MAE	12.5	11.9	12.9	11.6	12.1	10.9	12.1	11.7	12.4	11.9	12.3	11.3
MSFT	RMSE	14.6	14.3	14.0	13.5	13.8	13.2	13.7	12.7	14.3	12.7	14.0	12.3
	MAE	8.7	8.0	9.2	9.0	8.6	7.6	8.5	7.6	8.8	8.1	9.0	7.7
T	RMSE	43.8	42.9	43.1	42.3	43.8	42.5	43.3	42.1	43.4	42.0	43.2	43.4
	MAE	19.6	17.9	18.4	17.5	18.0	16.7	18.3	17.6	18.6	17.0	18.3	17.6

Counting the number of cases that fail to show improvement is valuable for two reasons: (1) It is the best tool when comparing models evaluated through MLE given that the log-likelihood functions were not similar for looking at the size of the MLE values; and (2) the cases in which we failed to see improvement indicated some consistency for a specific model and a specific stock. This opens the opportunity for future work in which we might try to understand why the Realized GARCH (2,2) model and AIG stock persistently indicated less evidence compared with other models and stocks, where by adding night volatility information, we produced improved volatility estimation.

Red and green numbers indicate the stances in which bivariate models perform worse than the univariate ones (when evaluated according to RMSE or MAE, respectively).

Thus, we concluded that the proposed bivariate models improved the forecasting performance compared with the univariate models; as such, adding night volatility estimations according to the methodology suggested improves next-day volatility estimates.

5. Conclusions

This paper provided a methodology that captures and integrates night volatility into the modeling of day volatility. In univariate context, this method led to formulating four bivariate realized GARCH models (Bivariate EGARCH-X, Bivariate Realized GARCH, Bivariate Realized GARCH (2,2), and Bivariate Realized EGARCH) and one bivariate non-realized model (Bivariate EGARCH). The novelty of this method is the incorporation of a night measure of volatility into the models, computed from price changes between the closing and opening of the trading market with a two-factor structure of the conditional variance in a realized GARCH setting that takes advantage of the natural relationship

between the realized measure and the conditional variance. This captures the leverage effect and maintains an elegant mathematical structure that facilitates the estimation of volatility.

With respect to assessing forecasting performance, the first finding was that rankings were sensitive to the stock and model choice but displayed little sensitivity to the ranking criterion and estimation methodology. However, the bivariate models were proved to perform better in most instances, compared with the univariate models. As such, we concluded that by adding night volatility estimates in the volatility models according to the methodology described, better estimates of next-day volatility could be obtained. This represents a step further from including high-frequency data in the modeling problem of the GARCH models in that estimates of night volatility are added into the equation of the day conditional variance according to the novel methodology we suggest.

The assessment to multivariate assets (e.g., portfolios of stocks) could be extended in future work by documenting a method of forecasting volatility of assets using the principal component (PC) analysis or other statistical procedures that use the orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables, taking advantage of the autoregressive conditional heteroskedastic models we proposed that use estimates of day, intraday, and night volatility. We might refer to these models as PC Bivariate Realized GARCH models and these might be used to formulate the general form of one multivariate asset's conditional variance–covariance matrix expressed in terms of conditional variances of the compounding assets and of their principal components. This would allow the estimation of the volatility of one multivariate asset through estimations of the volatility of principal components using day, intraday, and night volatility information. Then, by reducing the n-multivariate to a $n - k$ stock dimension (n and k positive integers), we could estimate the new models and assess their one-day-ahead forecasting performance. Constructing models that use volatility information from the previous two days and two nights may further improve the modeling of volatility, as we noted by inspecting the results for the current bivariate form of Realized GARCH (2,2). Disseminating among the stocks according to their underlying volatility features may provide a better method of more consistently modeling their volatility patterns.

Integration of volatility estimates of highly interlinked markets that are open during the closing time of the reference market is another suggestion for further research. For example, proposing models for the U.S. market that estimate day volatility using night volatility estimates from the Asian markets open during the non-trading times of the U.S. market would allow for integration in such models of systemic risk and financial contagion related elements, with likely benefits for volatility estimation and forecasting.

Author Contributions: Individual contributions to the current paper were as follows: Conceptualization, M.M., X.R., and N.A.; Methodology, M.M.; Software, M.M.; Validation, M.M.; Formal Analysis, M.M.; Investigation, M.M.; Resources, M.M., X.R., and N.A.; Data Curation, M.M.; Writing—Original Draft Preparation, M.M.; Writing—Review & Editing, M.M.; Visualization, M.M.; Supervision, M.M., X.R., and N.A.; Project Administration, M.M., X.R., and N.A.; Funding Acquisition, M.M., X.R., and N.A.

Funding: M.M. acknowledges support from the Agency for Management of University and Research Grants (AGAUR) of the Government of Catalonia (Resolution IUE/2681/2008 of 8 August, DOGC no. 5208 of 03.09.08), ESADE Business School (Ramon Llull University), as well as from the University of Tasmania (ARC DP130100168).

Acknowledgments: We acknowledge the contribution of Peter Reinhard Hansen, Zhuo (Albert) Huang and Howard Howan Shek to the definition of the reduced form Bivariate Realized GARCH model described by Equations (1) to (3). We are grateful for the comments from Mardi Dungey and participants in the European Economic Association and Econometric Society Meeting 2012, Econometric Society Australasian Meeting 2013, and FIRN 2012 conference.

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

Disclaimer: The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bank of Romania.

Appendix A. Log-Likelihood Function for the Bivariate Models

The data are bivariate vectors compounded of two univariate vectors that refer to uncorrelated sets of information (we considered first that night volatility was uncorrelated with day volatility): $\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}|F_{t-1} \sim N(0, \begin{pmatrix} h_t^{\bullet} & 0 \\ 0 & h_t^{\circ} \end{pmatrix})$. Accordingly, the random vector $\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}$ depends solely on the information set available at time $t-1$, and has a normal distribution with $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$ mean and a variance equal to the variance–covariance matrix $\begin{pmatrix} h_t^{\bullet} & 0 \\ 0 & h_t^{\circ} \end{pmatrix}$. The latter is equivalent to $var\left(r_t^{\circ}\right) = \sigma_t^{\circ}$, $var\left(r_t^{\bullet}\right) = h_t^{\bullet}$ and $cov\left(r_t^{\circ}, r_t^{\bullet}\right) = 0$. The total volatility is given as $r_t = r_t^{\circ} + r_t^{\bullet}$. Theory states that when a random vector (such as $\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}$) is normally distributed, then its components are also normal. $\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}|F_{t-1} \sim N(0, \begin{pmatrix} h_t^{\bullet} & 0 \\ 0 & h_t^{\circ} \end{pmatrix})$ shows that $r_t^{\bullet}|F_{t-1}, \sim N\left(0, h_t^{\bullet}\right)$ and $r_t^{\circ}|F_{t-1} \sim N\left(0, h_t^{\circ}\right)$. Since a sum of two normal variables is a normal variable with the average equal to the arithmetic sum of the two component averages, $r_t|F_{t-1} \sim N\left(0,\ h_t^{\bullet} + h_t^{\circ}\right)$, then the density function of $r_t|F_{t-1}$ has the form of a normal variable, that is, $f(r_t) = \frac{1}{\sqrt{\sigma_t^*}\sqrt{2\pi}} e^{\frac{r_t^2}{2h_t^*}}$, where $h_t^* = h_t^{\bullet} + h_t^{\circ}$ is the variance of r_t. Since n observations of $t = 1, \ldots, n$ are made, the likelihood function is the $\begin{pmatrix} r_1 \\ \ldots \\ r_n \end{pmatrix}$ vector's density, and $r_1, \ldots, r_n$ are independent of each other, so the likelihood function is $l(r_t) = \prod_{t=1}^{n} f(r_t) = \prod_{t=1}^{n} \frac{1}{\sqrt{h_t^*}\sqrt{2\pi}} e^{\frac{r_t^2}{2h_t^*}} = \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\prod_{t=1}^{n} \frac{1}{\sqrt{h_t^*}}\right)^{-\frac{1}{2}\sum_{t=1}^{n} \frac{r_t^2}{h_t^*}}$. Taking the log of this expression and using the logarithm properties, the log-likelihood function of the total returns r_t will become $\log l(r_t) = \log\left(\prod_{t=1}^{n}\left(\frac{1}{\sqrt{2\pi}}\right)^n\right) + \log\left(\left(\prod_{t=1}^{n} \frac{1}{\sqrt{h_t^*}}\right)\right) - \frac{1}{2}\sum_{t=1}^{n} \frac{r_t^2}{h_t^*} = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{n} \log\left(h_t^*\right) - \frac{1}{2}\sum_{t=1}^{n} \frac{r_t^2}{h_t^*} = -\frac{1}{2}\sum_{t=1}^{n}\left[\log(2\pi) + \log\left(h_t^*\right) + \frac{r_t^2}{h_t^*}\right]$.

If we considered a more complete model with a non-null correlation between r_t° and $r_t^{\bullet}$ (meaning that night volatility influences day volatility), that is, $corr\left(r_t^{\circ}, r_t^{\bullet}\right) = \rho \neq 0$, the formulation of the log-likelihood function slightly changes. Observe first that ρ does not depend on t, that is, the correlation is not time dependent. Then, the covariance will be $\left(r_t^{\circ}, r_t^{\bullet}\right) = corr\left(r_t^{\circ}, r_t^{\bullet}\right)\sqrt{var\left(r_t^{\bullet}\right)var\left(r_t^{\circ}\right)} = \rho\sqrt{h_t^{\bullet}h_t^{\circ}}$. This means that in the new model (with a non-null correlation), the variance–covariance matrix takes the form $\begin{pmatrix} h_t^{\bullet} & \rho\sqrt{h_t^{\bullet}h_t^{\circ}} \\ \rho\sqrt{h_t^{\bullet}h_t^{\circ}} & h_t^{\circ} \end{pmatrix}$, having the variances of r_t° and $r_t^{\bullet}$ on the first diagonal, and the covariance between r_t° and $r_t^{\bullet}$ on the second diagonal, that is $cov\left(r_t^{\circ}, r_t^{\bullet}\right)$ (since $cov\left(r_t^{\circ}, r_t^{\bullet}\right) = cov\left(r_t^{\bullet}, r_t^{\circ}\right)$). As such, the $|F_{t-1}$ conditioned distribution of the $\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}$ vector is $\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}|F_{t-1} \sim N(0, \begin{pmatrix} h_t^{\bullet} & \rho\sqrt{h_t^{\bullet}h_t^{\circ}} \\ \rho\sqrt{h_t^{\bullet}h_t^{\circ}} & h_t^{\circ} \end{pmatrix})$. The conditional variance of r_t, $var(r_t|F_{t-1})$, is $var(r_t|F_{t-1}) = var\left(r_t^{\bullet}|F_{t-1}\right) + var\left(r_t^{\circ}|F_{t-1}\right) + 2cov\left(r_t^{\bullet}|F_{t-1},\ r_t^{\circ}|F_{t-1}\right) = h_t^{\bullet} + h_t^{\circ} + 2\rho\sqrt{h_t^{\bullet}h_t^{\circ}}$, that is, $h_t^* = h_t^{\bullet} + h_t^{\circ} + 2\rho\sqrt{h_t^{\bullet}h_t^{\circ}}$. The log-likelihood function of $r_t = r_t^{\bullet} + r_t^{\circ}$ will be the same as the one iterated for the null correlation case, the only difference being that the variance encloses the correlation term $h_t^* = h_t^{\bullet} + h_t^{\circ} + 2\rho\sqrt{h_t^{\bullet}h_t^{\circ}}$.

However, we want to consider the log-likelihood function of the bivariate vector $\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}$ and not that of the univariate vector $r_t = r_t^{\bullet} + r_t^{\circ}$. As such, to define the new log-likelihood function, we considered the density function of the bi-dimensional normal $\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}$. The general form of a p-dimensional normal vector $N_p(\mu, \Sigma)$ (a matrix with μ vector average and Σ variance–covariance matrix) takes the form $f(x) = \frac{1}{(\sqrt{2\pi})^p} \frac{1}{\sqrt{det(\Sigma)}} e^{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)}$, where x is any vector for which the density function has been calculated with p arguments, $det(\Sigma)$ is the determinant of the variance–covariance matrix Σ, and $(x-\mu)'\Sigma^{-1}(x-\mu)$ is the matrix product between the transpose of the $(x-\mu)$ vector, the inverse of matrix Σ, and the $(x-\mu)$ vector. As such, with $p = 2$ for the particular case of a bi-dimensional vector $\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}$, the density function is $f\left(r_t^{\bullet}, r_t^{\circ}\right) = \frac{1}{(\sqrt{2\pi})^2} \frac{1}{\sqrt{det(\Sigma)}} e^{-\frac{1}{2}(r_t^{\bullet}, r_t^{\circ})\Sigma^{-1}\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}}$ in which $\mu = 0$ and $\Sigma = \begin{pmatrix} h_t^{\bullet} & \rho\sqrt{h_t^{\bullet} h_t^{\circ}} \\ \rho\sqrt{h_t^{\bullet} h_t^{\circ}} & h_t^{\circ} \end{pmatrix}$. Since $det(\Sigma) = h_t^{\bullet} h_t^{\circ} - \rho^2 h_t^{\bullet} h_t^{\circ} = h_t^{\bullet} h_t^{\circ}\left(1-\rho^2\right)$, then its log form is $\log(det(\Sigma)) = \log(h_t^{\bullet}) + \log(h_t^{\circ}) + \log\left(1-\rho^2\right)$. The inverse matrix of the variance–covariance matrix is $\Sigma^{-1} = \frac{1}{h_t^{\bullet} h_t^{\circ}(1-\rho^2)} \begin{pmatrix} h_t^{\bullet} & -\rho\sqrt{h_t^{\bullet} h_t^{\circ}} \\ -\rho\sqrt{h_t^{\bullet} h_t^{\circ}} & h_t^{\circ} \end{pmatrix}$. As such, the product $-\frac{1}{2}\left(r_t^{\bullet}, r_t^{\circ}\right)\Sigma^{-1}\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix}$ becomes $-\frac{1}{2}\left(r_t^{\bullet}, r_t^{\circ}\right)\Sigma^{-1}\begin{pmatrix} r_t^{\bullet} \\ r_t^{\circ} \end{pmatrix} = -\frac{1}{2}\frac{r_t^{\bullet 2} h_t^{\circ} + r_t^{\circ 2} h_t^{\bullet} - 2 r_t^{\bullet} r_t^{\circ} \rho \sqrt{h_t^{\bullet} h_t^{\circ}}}{h_t^{\bullet} h_t^{\circ}(1-\rho^2)}$. Thus, the log-likelihood function $\log l(r_t^{\bullet}, r_t^{\circ})$ is obtained by multiplying the functions $f\left(r_t^{\bullet}, r_t^{\circ}\right)$ for the $t = 1, \ldots, n$, and by taking the log of the resulting product $\log l(r_t^{\bullet}, r_t^{\circ}) = -\frac{1}{2}\sum_{t=1}^{n}\left\{2\log(2\pi) + \log\left(1-\rho^2\right) + \log(h_t^{\bullet}) + \log(h_t^{\circ}) + \frac{r_t^{\bullet 2} h_t^{\circ} + r_t^{\circ 2} h_t^{\bullet} - 2 r_t^{\bullet} r_t^{\circ} \rho \sqrt{h_t^{\bullet} h_t^{\circ}}}{h_t^{\bullet} h_t^{\circ}(1-\rho^2)}\right\}$.

By performing some simple iterations in the expression above, we obtained the final form of the bivariate log-likelihood function as $\log l(r_t^{\bullet}, r_t^{\circ}) = -\frac{1}{2}\sum_{t=1}^{n}\left\{2\log(2\pi) + \log\left(1-\rho^2\right) + \log(h_t^{\bullet}) + \log(h_t^{\circ}) + \frac{r_t^{\bullet 2}/h_t^{\bullet} + r_t^{\circ 2}/h_t^{\circ}}{(1-\rho^2)} - \frac{2\rho}{(1-\rho^2)}\frac{r_t^{\bullet} r_t^{\circ}}{\sqrt{h_t^{\bullet} h_t^{\circ}}}\right\}$.

Appendix B

Table A1. Maximized log-likelihood functions in univariate and bivariate estimations; in-sample.

Stock	EGARCH		EGARCH-X		Realized EGARCH		Realized GARCH			Realized GARCH (2,2)	
	Univ	Biv	Univ	Biv	Univ	Biv	Univ	Biv (com)	Biv (red)	Univ	Biv
AIG	−1721.9	−2900.0	−1710.1	−2821.5	−1711.8	−2845.7	−1709.1	−2874.4	−2875.4	−1701.3	−2849.9
AXP	−1668.6	−2742.6	−1637.8	−2790.4	−1645.5	−2842.3	−1642.7	−2855.3	−2857.5	−1638.3	−2904.8
BAC	−1506.6	−2499.9	−1473.0	−2438.8	−1475.5	−2437.7	−1478.5	−2443.0	−2439.7	−1471.5	−2437.1
CSCO	−1722.4	−2886.2	−1709.7	−2820.9	−1712.9	−2841.1	−1711.7	−2876.9	−2876.1	−1702.1	−2845.1
F	−1673.5	−2746.3	−1644.2	−2791.8	−1645.9	−2841.5	−1644.1	−2853.8	−2855.2	−1642.8	−2898.9
GE	−1504.9	−2498.1	−1474.2	−2433.2	−1475.8	−2442.8	−1477.9	−2446.1	−2440.1	−1467.1	−2440.7
INTC	−1505.4	−2497.5	−1471.4	−2434.7	−1478.1	−2439.8	−1475.5	−2445.9	−2439.8	−1468.1	−2437.7
JPM	−1658.1	−2750.6	−1616.4	−2699.0	−1619.6	−2702.0	−1625.9	−2714.3	−2703.1	−1615.4	−2683.7
MSFT	−1668.0	−2743.3	−1639.9	−2792.0	−1639.3	−2840.6	−1642.8	−2851.0	−2855.3	−1639.9	−2903.1
T	−1507.1	−2497.5	−1470.9	−2434.4	−1477.5	−2438.6	−1478.2	−2442.1	−2440.0	−1471.3	−2439.3

Table A2. Maximized log-likelihood functions in univariate and bivariate estimations; out-of-sample.

Stock	EGARCH		EGARCH-X		Realized EGARCH		Realized GARCH			Realized GARCH (2,2)	
	Univ	Biv	Univ	Biv	Univ	Biv	Univ	Biv (com)	Biv (red)	Univ	Biv
AIG	−399.5	−1032.1	−394.2	−795.4	−387.7	−749.6	−372.5	−777.9	−774.4	−383.6	−736.9
AXP	−313.3	−607.7	−309.6	−565.7	−310.7	−576.5	−305.9	−561.3	−561.8	−308.4	−573.1
BAC	−344.6	−654.0	−341.9	−687.9	−352.0	−671.4	−337.0	−673.7	−676.2	−337.6	−670.0
CSCO	−407.3	−1033.4	−386.0	−790.5	−392.5	−752.5	−376.0	−770.1	−777.9	−372.5	−732.2
F	−308.9	−602.1	−308.0	−560.3	−307.6	−566.1	−305.6	−573.0	−563.7	−307.7	−570.1
GE	−348.8	−657.6	−339.0	−687.8	−353.7	−666.6	−348.5	−678.1	−672.0	−339.1	−672.9
INTC	−347.5	−659.5	−345.9	−681.7	−351.3	−678.7	−336.2	−674.3	−674.8	−340.5	−676.8
JPM	−330.9	−607.9	−326.1	−589.3	−324.1	−582.2	−316.0	−579.1	−573.1	−323.2	−584.8
MSFT	−403.5	−1030.5	−393.2	−787.9	−389.4	−745.4	−368.9	−772.7	−780.4	−386.3	−733.3
T	−315.0	−603.8	−305.2	−568.1	−303.8	−569.8	−301.4	−570.6	−572.1	−304.2	−570.6

References

Ahoniemi, Katja, and Markku Lanne. 2013. Overnight Stock Returns and Realized Volatility. *International Journal of Forecasting* 29: 592–604. [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]

Angelidis, Timotheos, and Stavros Degiannakis. 2008. Volatility Forecasting: Intra-day versus Inter-day Models. *Journal of International Financial Markets, Institutions and Money* 18: 449–65. [CrossRef]

Banulescu-Radu, Denisa, Peter Reinhard Hansen, Zhuo Huang, and Marius Matei. 2019. Volatility During the Financial Crisis Through the Lens of High Frequency Data: A Realized GARCH Approach. Available online: https://sites.google.com/site/peterreinhardhansen/research-papers/volatilityduringthefinancialcrisisthroughthelensofhighfrequencydataarealizedgarchapproach (accessed on 5 May 2019).

Barndorff-Nielsen, Ole Eiler, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. 2009. Realized Kernels in Practice: Trades and Quotes. *The Econometrics Journal* 12: C1–C32. [CrossRef]

Barndorff-Nielsen, Ole Eiler, Peter Reinhard Hansen, Asger Lunde, and Neil Shephard. 2011. Multivariate Realised Kernels: Consistent Positive Semi-Definite Estimators of the Covariation of Equity Prices with Noise and Non-Synchronous Trading. *Journal of Econometrics* 162: 149–69. [CrossRef]

Blair, Bevan J., Ser-Huang Poon, and Stephen J. Taylor. 2001. Forecasting S&P 100 Volatility: The Incremental Information Content of Implied Volatilities and High-Frequency Index Returns. *Journal of Econometrics* 105: 5–26.

Bollerslev, Tim, George Tauchen, and Hao Zhou. 2009. Expected Stock Returns and Variance Risk Premia. *Review of Financial Studies* 22: 4463–92. [CrossRef]

Bollerslev, Tim, Andrew J. Patton, and Rogier Quaedvlieg. 2018. Multivariate Leverage Effects and Realized Semicovariance GARCH Models. Available online: https://ssrn.com/abstract=3164361 (accessed on 10 February 2019).

Corsi, Fulvio. 2009. A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics* 7: 174–96. [CrossRef]

De Pooter, Michiel, Martin Martens, and Dick van Dijk. 2008. Predicting the Daily Covariance Matrix for S&P 100 Stocks Using Intraday Data—But Which Frequency to Use? *Econometric Reviews* 27: 199–229.

Engle, Robert F. 2002. New Frontiers of ARCH Models. *Journal of Applied Econometrics* 17: 425–46. [CrossRef]

Engle, Robert F., and Giampiero M. Gallo. 2006. A Multiple Indicators Model for Volatility Using Intra-daily Data. *Journal of Econometrics* 131: 3–27. [CrossRef]

Fleming, Jeff, and Chris Kirby. 2011. Long Memory in Volatility and Trading Volume. *Journal of Banking and Finance* 35: 1714–26. [CrossRef]

Fleming, Jeff, Chris Kirby, and Barbara Ostdiek. 2003. The Economic Value of Volatility Timing Using "Realized" Volatility. *Journal of Financial Economics* 67: 473–509. [CrossRef]

Fuertes, Ana-Maria, and Jose Olmo. 2013. Optimally Harnessing Inter-day and Intra-day Information for Daily Value-at-Risk Prediction. *International Journal of Forecasting* 29: 28–42. [CrossRef]

Hansen, Peter Reinhard, and Zhuo Huang. 2016. Exponential GARCH Modeling with Realized Measures of Volatility. *Journal of Business & Economic Statistics* 34: 269–87.

Hansen, Peter Reinhard, and Asger Lunde. 2005. A Realized Variance for the Whole Day Based on Intermittent High-frequency Data. *Journal of Financial Econometrics* 3: 525–54. [CrossRef]

Hansen, Peter Reinhard, Zhuo Huang, and Howard Howan Shek. 2012. Realized GARCH: A Joint Model for Returns and Realized Measures of Volatility. *Journal of Applied Econometrics* 27: 877–906. [CrossRef]

Hansen, Peter Reinhard, Asger Lunde, and Valeri Voev. 2014. Realized Beta GARCH: A Multivariate GARCH Model with Realized Measures of Volatility. *Journal of Applied Econometrics* 29: 774–99. [CrossRef]

Hansen, Peter Reinhard, Pawel Janus, and Siem Jan Koopman. 2019. Realized Wishart-GARCH: A Score-driven Multi-Asset Volatility Model. *Journal of Financial Econometrics* 17: 1–32. [CrossRef]

Huang, Zhuo, Hao Liu, and Tianyi Wang. 2016. Modeling Long Memory Volatility Using Realized Measures of Volatility: A Realized HAR GARCH Model. *Economic Modelling* 52: 812–21. [CrossRef]

Koopman, Siem Jan, Borus Jungbacker, and Eugenie Hol. 2005. Forecasting Daily Variability of the S&P 100 Stock Index Using Historical, Realised, and Implied Volatility Measurements. *Journal of Empirical Finance* 12: 445–75.

Martens, Martin. 2002. Measuring and Forecasting S&P 500 Index-Futures Volatility Using High-frequency Data. *Journal of Futures Markets* 22: 497–518.

Shephard, Neil, and Kevin Sheppard. 2010. Realising the Future: Forecasting with High-frequency-based Volatility (HEAVY) Models. *Journal of Applied Econometrics* 25: 197–231. [CrossRef]

econometrics

MDPI

Article

Representation of Japanese Candlesticks by Oriented Fuzzy Numbers

Krzysztof Piasecki [1,*] and Anna Łyczkowska-Hanćkowiak [2]

[1] Department of Investment and Real Estate, Poznań University of Economics and Business, 61-875 Poznań, Poland
[2] Institute of Finance, WSB University in Poznań, 61-895 Poznań, Poland; anna.lyczkowska-hanckowiak@wsb.poznan.pl
* Correspondence: krzysztof.piasecki@ue.poznan.pl; Tel.: +48-618-543-531

Received: 1 September 2019; Accepted: 11 December 2019; Published: 18 December 2019

Abstract: The Japanese candlesticks' technique is one of the well-known graphic methods of dynamic analysis of securities. If we apply Japanese candlesticks for the analysis of high-frequency financial data, then we need a numerical representation of any Japanese candlestick. Kacprzak et al. have proposed to represent Japanese candlesticks by ordered fuzzy numbers introduced by Kosiński and his cooperators. For some formal reasons, Kosiński's theory of ordered fuzzy numbers has been revised. The main goal of our paper is to propose a universal method of representation of Japanese candlesticks by revised ordered fuzzy numbers. The discussion also justifies the need for such revision of a numerical model of the Japanese candlesticks. There are considered the following main kinds of Japanese candlestick: White Candle (White Spinning), Black Candle (Black Spinning), Doji Star, Dragonfly Doji, Gravestone Doji, and Four Price Doji. For example, we apply numerical model of Japanese candlesticks for financial portfolio analysis.

Keywords: Japanese candlestick; ordered fuzzy number; Kosiński's number; oriented fuzzy number; dynamic analysis of securities

JEL Classification: C02; C43; G19

1. Introduction

The notion of ordered fuzzy number (OFN) is introduced by Kosiński (Kosiński and Słysz 1993; Kosiński et al. 2002, 2003) as an extension of the concept of fuzzy number (FN), which is widely interpreted as an imprecise approximation of a real number. They intuitively determine OFN as FN equipped with information about the location of the approximated number. This additional information is given as orientation of OFN. For this reason, we can interpret OFN as an imprecise approximation of a real number which may change in the direction determined by orientation. The monograph (Prokopowicz et al. 2017) is a competent source of information about the contemporary state of knowledge on OFN defined by Kosiński. On the other hand, Kosiński (2006) has shown that there exist improper OFNs which cannot be represented by a pair of FNs and its orientation. Then, we cannot to apply any knowledge about fuzzy sets to solve practical problems described by improper OFN. Therefore, any considerations which use improper OFN may be not fruitful. For formal reasons, Kosiński's theory was revised (Piasecki 2018) in such a way that the revised OFN definition fully corresponds to the intuitive Kosiński's definition of OFN. In this paper, we recall OFN defined by Kosiński as "Kosiński's number" which is in line with suggestions given by other researchers (Prokopowicz 2016; Prokopowicz and Pedrycz 2015). Moreover, we propose the OFNs defined in a revised way to be called oriented fuzzy numbers. This proposal has been thoroughly justified in

Piasecki (2019). In this way, in the concept of OFNs, we distinguished two of its types: Kosiński's numbers and oriented fuzzy numbers.

A Japanese candlestick (JC) is a style of financial chart used to describe variability of financial assets' exchange quotations. Candlestick charts were developed in the 18th century by a Japanese rice trader, Munehisa Homma (Morris 2006). The JC charting techniques were introduced to the Western world by Nison (1991). Among other things, the market quotation changes are described in this way: that any White Candle describes a rise in quotations and any Black Candle describes a fall in quotations. According to the traditional convention, any classical JC is represented by a four-element set of its real numbers: open price, close price, low price, and high price. Some applications of JCs are presented, for example in (Detollenaere and Mazza 2014; Fock et al. 2005; Jasemi et al. 2011; Marshall et al. 2006; Kamo and Dagli 2009; Tsung-Hsun et al. 2012, 2015).

Any original JC charting technique is a graphic tool for recording quotations volatility. Any JC is a very convenient tool for synthetic recording of high-frequency time series of financial data. If we use JCs for the analysis of high-frequency financial data, then then we need their linguistic or numerical representation.

Lee et al. (2006) proposed to describe some JC attributes by linguistic variables evaluated by linguistic labels which have a meaning depending on the applied pragmatics of the natural language. By its very nature of things, each such description is imprecise information. For this reason, Zadeh (1975a, 1975b, 1975c) proposed to describe each linguistic variable by its values defined as a fuzzy subset in the predefined space. Then, these linguistic variables may be transformed with the use of fuzzy set theory. Lee et al. (2006) assessed JC attributes by linguistic labels represented by trapezoidal fuzzy numbers. In an analogous way, JCs are described in (Kamo and Dagli 2009; Naranjo et al. 2018; Naranjo and Santos 2019). After (Herrera and Herrera-Viedma 2000), we can say that an application of imprecise linguistic assessments for decision analysis is very beneficial because it introduces a more flexible framework which allows us to represent the information in a more direct and adequate way when we are unable to express it precisely. Then, all the authors mentioned in this paragraph use linguistic labels representing JC as an input signal for the different systems supporting financial decision-making. The output signals received in this way are very useful.

On the other hand, linguistically described JCs cannot be used for calculating portfolio JC as a sum of their components' JCs (Łyczkowska-Hanćkowiak and Piasecki 2018b). This task requires the JCs numerical representation given as their quantified models.

For any security, each quantified JC may be used as such an approximation of its market price that it is additionally equipped with information about this price trend. For these reasons, Kacprzak et al. (2013) have described JC by means of Kosiński's numbers. Kacprzak's description has a significant disadvantage. There, each Black Candle is described by improper OFN. So, Kacprzak's approach is not convenient for financial analysis with JCs use.

The above conclusion justifies remodeling JCs with the use of the revised theory of OFN. The main aim of our paper is to describe JCs by means of oriented fuzzy numbers.

Our approach to represent JCs by OFNs differs to the approach represented by Marszałek and Burczyński (2013a, 2013b, 2014), who define the membership function of candles as some density function. In our opinion, such an approach is not compatible with the essence of the JCs.

Our paper is organized as follows. In Section 2, we present two kinds of trapezoidal OFNs: Kosiński's numbers and oriented fuzzy numbers. At the end of this section, we justify the postulated restricting applications of the oriented fuzzy numbers. Section 3 contains the main information about JCs. In Section 4, we briefly discuss Marszałek's approach and Kacprzak's approach to modeling JCs by OFNs. Section 5 is the main part of our paper. In this section we propose our method of representing JCs by oriented fuzzy numbers. In Section 6, we determine the imprecise expected return rate by means of JCs models. Next, we show the possibility of application of this return rate for fuzzy portfolio analysis. This section is based on (Piasecki 2017). Section 7 presents a simple case study in which

the introduced numerical model of JCs is applied for financial portfolio analysis. Section 8 contains final conclusions.

2. Fuzzy Numbers

The symbol $\mathcal{F}(\mathbb{R})$ denotes the family of all fuzzy sets in the real line $\mathbb{R}$. A commonly accepted model of imprecise number is the fuzzy number (FN), generally defined by Dubois and Prade (1978) as some kind of fuzzy subset $\mathcal{L} \in \mathcal{F}(\mathbb{R})$. Thanks to the results obtained in (Goetschel and Voxman 1986), any FN can be equivalently defined as follows:

Theorem 1 (Delgado et al. 1998). *For any FN $\mathcal{L}$ there exists such a non-decreasing sequence $(a,b,c,d) \subset \mathbb{R}$ that $L(a,b,c,d,L_L,R_L) = \mathcal{L} \in \mathcal{F}(\mathbb{R})$ is determined by its membership function $\mu_L(\cdot|a,b,c,d,L_L,R_L) \in [0,1]^{\mathbb{R}}$ described by the identity:*

$$\mu_L(x|a,b,c,d,L_L,R_L) = \begin{cases} 0, & x \notin [a,d], \\ L_L(x), & x \in [a,b], \\ 1, & x \in [b,c], \\ R_L(x), & x \in [c,d], \end{cases} \tag{1}$$

where the left reference function $L_L \in [0,1]^{[a,b]}$ and the right reference function $R_L \in [0,1]^{[c,d]}$ are upper semi-continuous monotonic ones meeting the conditions:

$$L_L(b) = R_L(c) = 1, \tag{2}$$

$$\forall_{x \in]a,d[} : \ \mu_{\mathcal{L}}(x|a,b,c,d,L_L,R_L) > 0. \tag{3}$$

Let us note that identity (1) additionally describes the extended notation of numerical intervals, which is used in this work. The family of all FNs is denoted by the symbol $\mathbb{F}$. For any $z \in [b,c]$, a FN $\mathcal{L}(a,b,c,d,L_L,R_L)$ is a formal model of linguistic variable "about z". Understanding the phrase "about z" depends on the applied pragmatics of the natural language. In (Dubois and Prade 1979), arithmetic operations on FNs are introduced in such a way that they are coherent with the Zadeh's extension principle. In our paper, we do not apply this arithmetic. Therefore, a description of arithmetic operations on FNs is omitted here.

2.1. Kosiński's Number

The concept of ordered fuzzy numbers (OFN) was introduced by Kosiński and his co-writers in the series of papers (Kosiński and Słysz 1993; Kosiński et al. 2002, 2003) as an extension of the concept of FN. Thus, any OFN should be determined with the use of any fuzzy subset in real line $\mathbb{R}$. On the other side, Kosiński has defined OFN as an ordered pair of functions from the unit interval $[0,1]$ into $\mathbb{R}$. This pair is not similar to any fuzzy subset in $\mathbb{R}$. It means that the Kosiński's proposal of OFN notion is not an extension of FN. Thus, we can not to accept Kosiński's original terminology. For this reason, we are agreeing with other scientists (Prokopowicz 2016; Prokopowicz and Pedrycz 2015) that the OFN defined by Kosiński should be called the Kosiński's number (KN). Let the symbol $\langle f \rangle^{\triangleleft}$ denote the pseudo-inversion of monotonic continuous surjection $f \in [r,s]^{[t,u]}$. KNs are originally defined as follows:

Definition 1. *For any sequence $(a,b,c,d) \subset \mathbb{R}$, the KN $\overset{\leftrightarrow}{S}(a,b,c,d,\langle f_s \rangle^{\triangleleft}, \langle g_s \rangle^{\triangleleft})$ is defined as an ordered pair (f_S, g_S) of monotonic continuous surjections $f_S : [0,1] \to UP_S = [a,b]$ and $g_S : [0,1] \to DOWN_S = [c,d]$ fulfilling the condition:*

$$f_s(0) = a \text{ and } f_s(1) = b \quad \text{and } g_s(1) = c \text{ and } \quad g_s(0) = d. \tag{4}$$

For any KN $\overleftrightarrow{S}(a,b,c,d,\langle f_s\rangle^{\triangleleft},\langle g_s\rangle^{\triangleleft})$, the function $f_S \in UP_s^{[0,1]}$ is called the up-function. Then, the function $g_S \in DOWN_s^{[0,1]}$ is called the down-function. Moreover, for KN $\overleftrightarrow{S}(a,b,c,d,\langle f_s\rangle^{\triangleleft},\langle g_s\rangle^{\triangleleft})$, the number $a \in \mathbb{R}$ is called the starting point and the number $d \in \mathbb{R}$ is called the ending point. The space of all KNs is denoted by the symbol $\overleftrightarrow{\mathbb{K}}$. Any sequence $(a,b,c,d) \subset \mathbb{R}$ satisfies exactly one of the following conditions:

$$b < c \ \| \ (b = c \text{ and } a < d), \tag{5}$$

$$b > c \ \| \ (b = c \text{ and } a > d), \tag{6}$$

$$b = c \text{ and } a = d. \tag{7}$$

If the condition (5) is fulfilled, then the KN $\overleftrightarrow{S}(a,b,c,d,\langle f_s\rangle^{\triangleleft},\langle g_s\rangle^{\triangleleft})$ has a positive orientation. For this case, an example of graphs of KN is presented in Figure 1a. Any positively oriented K'N is interpreted as an imprecise number, which may increase. If the condition (6) is fulfilled, then KN $\overleftrightarrow{S}(a,b,c,d,\langle f_s\rangle^{\triangleleft},\langle g_s\rangle^{\triangleleft})$ is negatively oriented. Then, it is interpreted as an imprecise number, which may decrease. For the case (7), KN $\overleftrightarrow{S}(a,b,c,d,\langle f_s\rangle^{\triangleleft},\langle g_s\rangle^{\triangleleft})$ represents the interval $[a,b] \subset \mathbb{R}$.

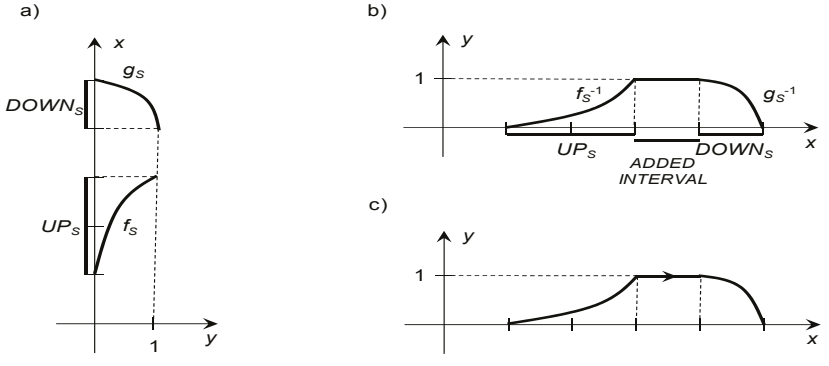


Figure 1. (**a**) Positively oriented Kosiński's pair, (**b**) membership function of FN determined by positive oriented KN, (**c**) arrow denotes the positive orientation of KN. Source: (Kosiński et al. 2002).

In (Piasecki 2018) it is shown that we have:

Theorem 2. *For any sequence $(a,b,c,d) \subset \mathbb{R}$, the KN $\overleftrightarrow{S}(a,b,c,d,\langle f_s\rangle^{\triangleleft},\langle g_s\rangle^{\triangleleft})$ is explicitly determined by its membership relation $\mu_s\left(\cdot\middle|a,b,c,d,\langle f_s\rangle^{\triangleleft},\langle g_s\rangle^{\triangleleft}\right) \subset \mathbb{R}\times[0,1]$ given by the identity:*

$$\mu_L(x|a,b,c,d,\langle f_s\rangle^{\triangleleft},\langle g_s\rangle^{\triangleleft}) = \begin{cases} 0, & x \notin [a,d] \equiv [d,a], \\ \langle f_s\rangle^{\triangleleft}(x), & x \in [a,b] \equiv [b,a], \\ 1, & x \in [b,c] \equiv [c,b], \\ \langle g_s\rangle^{\triangleleft}(x), & x \in [c,d] \equiv [d,c]. \end{cases} \tag{8}$$

The graph of membership function of KN $\overleftrightarrow{S}\left(a,b,c,d,f_S^{\triangleleft},g_S^{\triangleleft}\right)$ is presented in Figure 1b. A membership relation of any KN may be represented by the graph which has an extra arrow from the starting point to the ending one. This arrow denotes the KN orientation, which shows supplementary information about possible changes of the approximated number. The graph of KN membership function with positive orientation is shown in Figure 1c. Next, examples of such graphs are presented in Figure 2.

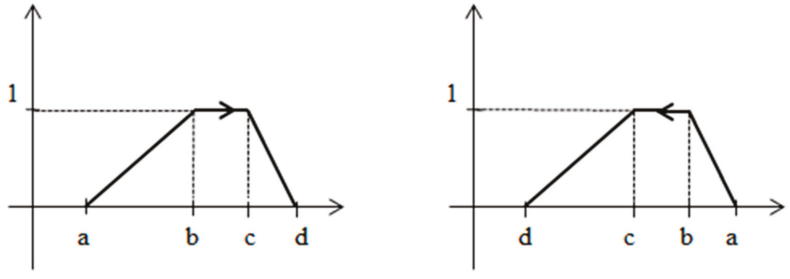


Figure 2. The membership relation of trapezoidal KN (TrKN) $\overleftrightarrow{Tr}(a,b,c,d)$ with: (**a**) positive orientation and (**b**) negative orientation. Source: Own elaboration.

If the sequence $(a,b,c,d) \subset \mathbb{R}$ is not monotonic, then the membership relation $\mu_s(x|a,b,c,d,\langle f_s\rangle^{\lhd},\langle g_s\rangle^{\lhd}) \subset \mathbb{R}\times[0,1]$ is not a function. Then, this membership relation cannot be considered as a membership function of any fuzzy set. Therefore, for any non-monotonic sequence $(a,b,c,d) \subset \mathbb{R}$, the KN $\mathcal{S}\left(a,b,c,d,f_S^{\lhd},g_S^{\lhd}\right)$ is called an improper one (Kosiński 2006). An example of a membership relation of a negatively oriented improper KN is presented in Figure 3. The remaining KNs are called proper ones. Some examples of proper KNs were presented in Figure 2.

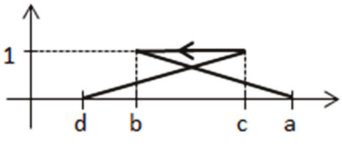


Figure 3. Membership relation of negatively oriented improper TrKN $\overleftrightarrow{Tr}(a,b,c,d)$. Source: Own elaboration.

Kosiński et al. (2002, 2003) determined arithmetic operators for KNs as an extension of results obtained for FNs in (Goetschel and Voxman 1986). In our paper, we do not apply this arithmetic. Therefore, a description of arithmetic operations on FNs is omitted here.

Because in this paper we restrict our main considerations to the case of trapezoidal numbers, we will take into account the following definition:

Definition 2. *For any sequence* $(a,b,c,d) \subset \mathbb{R}$*, the trapezoidal KN (TrKN)* $\overleftrightarrow{Tr}(a,b,c,d)$ *is defined as KN determined by its membership relation* $\mu_{Tr}(\cdot|a,b,c,d) \subset \mathbb{R}\times[0,1]$ *given by the identity:*

$$\mu_T(x) = \mu_{Tr}(x|a,b,c,d) = \begin{cases} 0, & x \notin [a,d] \equiv [d,a], \\ \frac{x-a}{b-a}, & x \in [a,\ b] \equiv [b,a] \\ 1, & x \in [b,c] \equiv [c,b] \\ \frac{x-d}{c-d}, & x \in [c,d] \equiv [d,c]. \end{cases} \tag{9}$$

The symbol $\overleftrightarrow{\mathbb{K}}_{Tr}$ denotes the space of all TrKNs. Some examples of membership relations of TrKN are presented in Figures 2 and 3.

Kosiński et al. (2002, 2003) determine arithmetic operators for TrKNs in the following way:

- The "dot multiplication" $\boxdot$ defined for any pair $\left(\beta, \overleftrightarrow{Tr}(a,b,c,d)\right) \in \mathbb{R}\times\overleftrightarrow{\mathbb{K}}_{Tr}$ by the identity:

$$\beta \boxdot \overleftrightarrow{Tr}(a,b,c,d) = \overleftrightarrow{Tr}(\beta{\cdot}a, \beta{\cdot}b, \beta{\cdot}c, \beta{\cdot}d), \tag{10}$$

- The "Kosiński's addition" $\boxplus_K$ defined for any pair $\left(\overset{\leftrightarrow}{Tr}(a,b,c,d), \overset{\leftrightarrow}{Tr}(e,f,g,h)\right) \in \left(\check{\mathbb{K}}_{Tr}\right)^2$ by the identity:

$$\overset{\leftrightarrow}{Tr}(a,b,c,d) \boxplus_K \overset{\leftrightarrow}{Tr}(e,f,g,h) = \overset{\leftrightarrow}{Tr}(a+e, b+f, c+g, d+h). \tag{11}$$

The determining "dot multiplication" is coherent with the Zadeh's extension principle. On the other hand, "Kosiński's addition" is not coherent with the Zadeh's extension principle. In (Kosiński 2006) it is shown that there exist such pairs of proper TrKNs that their Kosiński's sum is an improper one. This is the main reason why Kosiński's theory of OFNs was revised in (Piasecki 2018).

2.2. Oriented Fuzzy Numbers

Only in the case of any monotonic sequence $(a,b,c,d) \subset \mathbb{R}$ may the membership relation $\mu_s\left(\cdot|a,b,c,d,\langle f_s\rangle^{\triangleleft}, \langle g_s\rangle^{\triangleleft}\right) \subset \mathbb{R} \times [0,1]$ be interpreted as a membership function $\mu_s\left(\cdot|a,b,c,d,\langle f_s\rangle^{\triangleleft}, \langle g_s\rangle^{\triangleleft}\right) \in [0,1]^{\mathbb{R}}$. Thus, we distinguish the following kind of proper OFN.

Definition 3. (Piasecki 2018) *For any monotonic sequence $(a,b,c,d) \subset \mathbb{R}$, the oriented fuzzy number (OFN) $\overset{\leftrightarrow}{L}(a,b,c,d,S_L,E_L) = \overset{\leftrightarrow}{\mathcal{L}}$ is the pair of orientation $\overrightarrow{a,d} = (a,d)$ and fuzzy set $\mathcal{L} \in \mathcal{F}(\mathbb{R})$ described by membership function $\mu_L(\cdot|a,b,c,d,S_L,E_L) \in [0,1]^{\mathbb{R}}$ given by the identity:*

$$\mu_L(x|a,b,c,d,S_L,E_L) = \begin{cases} 0, & x \notin [a,d] \equiv [d,a], \\ S_L(x), & x \in [a,b] \equiv [b,a], \\ 1, & x \in [b,c] \equiv [c,b], \\ E_L(x), & x \in [c,d] \equiv [d,c]. \end{cases} \tag{12}$$

where the starting function $S_L \in [0,1]^{[a,b]}$ and the ending function $E_L \in [0,1]^{[c,d]}$ are upper semi-continuous monotonic ones meeting the conditions (3) and

$$S_L(b) = E_L(c) = 1. \tag{13}$$

The symbol $\mathbb{K}$ denotes the space of all OFNs. Any OFN is a proper KN and any proper KN is OFN. Therefore, we have $\mathbb{K} \subset \check{\mathbb{K}}$. The positive and negative orientations of OFN are determined in the same way as for the case of KN. The interpretation of a OFN orientation is identical to the interpretation of a KN orientation. For the family of all positively oriented OFNs and the family of all negatively oriented OFN, we respectively denote by the symbols $\mathbb{K}^+$ and $\mathbb{K}^-$. For any OFN $\overset{\leftrightarrow}{L}(a,b,c,d,S_L,E_L)$, the condition (7) implies that:

$$a = b = c = d. \tag{14}$$

Then, the OFN $\overset{\leftrightarrow}{L}(a,a,a,a,S_L,E_L) = [\![a]\!]$ describes the real number $a \in \mathbb{R}$ which is not oriented. Summing up, we see that:

$$\mathbb{K} = \mathbb{K}^+ \cup \mathbb{R} \cup \mathbb{K}^-. \tag{15}$$

We restrict our main considerations to the case of trapezoidal OFN, defined as follows:

Definition 4. *For any monotonic sequence $(a,b,c,d) \subset \mathbb{R}$, the trapezoidal OFN (TrOFN) $\overset{\leftrightarrow}{Tr}(a,b,c,d)$ is defined as OFN determined with use its membership function $\mu_{Tr}(\cdot|a,b,c,d) \in [0,1]^{\mathbb{R}}$ given by identity (9).*

The symbol $\mathbb{K}_{Tr}$ denotes the space of all TrOFNs. In our considerations, we will only use the arithmetic operators determined for TrOFN. Then, the"Kosiński addition" $\boxplus_K$ is replaced by "addition" $\boxplus$ defined for any pair $\left(\overleftrightarrow{Tr}(a,b,c,d), \overleftrightarrow{Tr}(p-a,q-b,\ r-c,s-d)\right) \in (\mathbb{K}_{Tr})^2$ by the identity:

$$\begin{aligned} &\overleftrightarrow{Tr}(a,b,\ c,d) \boxplus \overleftrightarrow{Tr}(p-a,q-b,\ r-c,s-d) \\ &= \begin{cases} \overleftrightarrow{Tr}(\min\{p,q\},q,\ r,\max\{r,s\}), & (q<r) \vee (q=r \wedge p \leq s), \\ \overleftrightarrow{Tr}(\max\{p,q\},q,\ r,\min\{r,s\}), & (q>r) \vee (q=r \wedge p > s). \end{cases} \end{aligned} \tag{16}$$

The space $\mathbb{K}_{Tr}$ is closed under the addition $\boxplus$ (Piasecki 2018). Moreover, sums obtained by using the addition $\boxplus$ are the best approximation of sums obtained by means of the Kosiński's addition $\boxplus_K$ (Piasecki 2018). The restrictions imposed in Definition 4 allow that TrOFNs may be analyzed with use of the fuzzy set theory. On the other hand, these methods of analysis are not applicable for TrKNs. Therefore, if we intend to apply trapezoidal OFNs for any additive model, then we should restrict our considerations to the use of the additive semigroup $\langle \mathbb{K}_{Tr},\ \boxplus \rangle$. This principle is shortly called the postulate restricting applications to the use of TrOFNs. It is very important restriction because we have many additive models of real-world problems. For example, financial portfolio analysis is an additive model.

Furthermore, the addition $\boxplus$ is not commutative. It implies that any multiple sum determined by $\boxplus$ may be dependent on its summands' ordering (Piasecki 2018). For this reason, the assumed order of performing multiple additions may require additional justifications.

For any OFN, the "dot multiplication" operator is determined by the identity (10). As we know, the "dot multiplication" is coherent to the Zadeh's extension principle. Therefore, using this principle, we can extend the "dot multiplication" operator to any unary operator linked to monotonic function $G : \mathbb{R} \supset \mathbb{A} \to \mathbb{R}$. This extension is determined in the following way:

If TrOFN $\overleftrightarrow{Tr}(a,b,c,d)$ is represented by its membership function $\mu_{Tr}(\cdot|a,b,c,d) \in [0,1]^{\mathbb{R}}$, then the membership function $\mu_G \in [0,1]^{\mathbb{R}}$ of OFN $\overleftrightarrow{\mathcal{G}} = G\left(\overleftrightarrow{Tr}(a,b,c,d)\right)$ is given by the identity:

$$\mu_G(x) = \mu_{Tr}\left(G^{-1}(x)\middle|G(a),G(b),G(c),G(d)\right). \tag{17}$$

While discussing the results obtained, we will also refer to the disorientation map $\overline{\overline{\Psi}} : \mathbb{K} \to \mathbb{F}$ determined in (Piasecki 2019) as follows:

$$\overline{\overline{\Psi}}\left(\overleftrightarrow{\mathcal{L}}(a,b,c,d,S_L,E_L)\right) = \begin{cases} L(a,b,c,d,S_L,E_L) & \overleftrightarrow{\mathcal{L}}(a,b,c,d,S_L,E_L) \in \mathbb{K}^+ \cup\ R, \\ L(d,c,b,a,E_L,S_L) & \overleftrightarrow{\mathcal{L}}(a,b,c,d,S_L,E_L) \in \mathbb{K}^-. \end{cases} \tag{18}$$

3. Japanese Candlesticks

We understand the term "security" as an authorization to receive a future financial revenue, payable to a certain maturity. Japanese Candlestick (JC) is a kind of financial chart used to describe the volatility of a fixed security quotation. The JC charting techniques are perfectly described by Nison (1991). The basic elements of JC charting techniques are the descriptions of a single JC.

Let fixed security be given as $\mathcal{Y}$. For the assumed time period $[0,T]$, the volatility of the security $\mathcal{Y}$ quotations are described by the function $Q : [0,T] \to \mathbb{R}^+$. Then, JC represents four important pieces of information about $\mathcal{Y}$ quotations:

- The open price:

$$P_o = Q(0), \tag{19}$$

- The close price:

$$P_c = Q(T), \tag{20}$$

- The high price:
$$P_h = \max\{Q(t) : t \in [0, T]\}, \tag{21}$$
- The low price:
$$P_l = \min\{Q(t) : t \in [0, T]\}. \tag{22}$$

The high price and the low price together are called extreme prices. In brokerage house reports, each JC is usually described by the data set $\{P_o, P_c, P_l, P_h\}$ of reported prices. In general, JCs are optional composed of:

- The body determined as a rectangle between the open price and the closed price.
- The upper wick determined as the line between the body and the high price.
- The lower wick determined as the line between the body and the low price.

The body illustrates the opening and closing trades. We can note the following cases here:

$$P_o < P_c, \tag{23}$$

$$P_o > P_c, \tag{24}$$

$$P_o = P_c, \tag{25}$$

If the condition (23) is met, then the candle body is white. Then, the JC is called White Candle or White Spinning. Any Black Candle may be considered as a bullish signal, i.e., a forecast of an increase in quotations. If the condition (24) is satisfied, then the candle body is black. Then, the JC is called Black Candle or Black Spinning. Any Black Candle may be considered as a bearish signal, i.e., a forecast of a decline in quotations. The general case of a White Candle and a Black Candle are presented in the Figure 4.

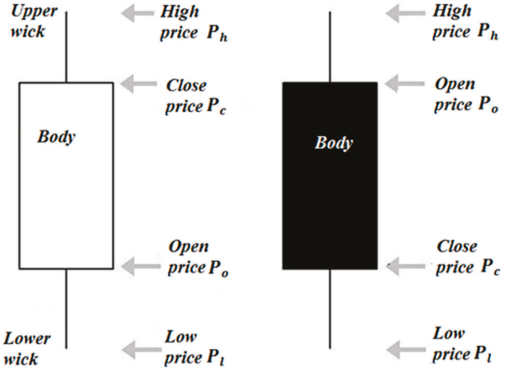


Figure 4. Japanese candlesticks. Source: Own elaboration.

Any JC need not have either a body or a wick. If a JC fulfills the condition (25) then it does not have a body. Such a JC is called a Doji. We distinguish the following main kinds of Doji:

- Doji Star is a JC fulfilling the condition:
$$P_l < P_o = P_c < P_h, \tag{26}$$
- Dragonfly Doji is a JC fulfilling the condition:
$$P_l < P_o = P_c = P_h, \tag{27}$$

- Gravestone Doji is a JC fulfilling the condition:

$$P_l = P_o = P_c < P_h, \tag{28}$$

- Four Price Doji is a JC fulfilling the condition:

$$P_l = P_o = P_c = P_h, \tag{29}$$

These kinds of Doji are presented in Figure 5.

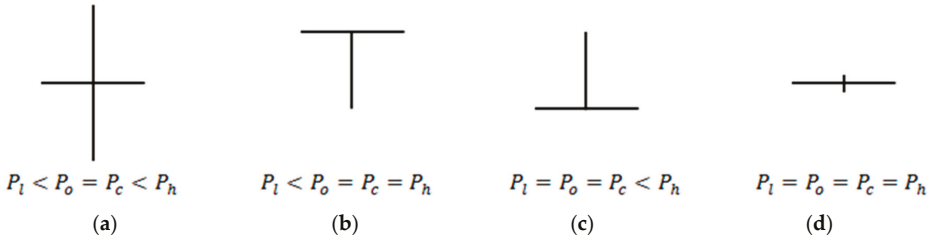


Figure 5. The main kinds of Doji: (**a**) Doji Star, (**b**)Dragonfly Doji, (**c**) Gravestone Doji, and (**d**) Four Price Doji. Source: Own elaboration

Example 1. *During the session on the Warsaw Stock Exchange (WSE) on July 23, 2019, we observed quotations of the following companies: Assecopol (ACP), CYFRPLSAT (CPS), ENERGA (ENG), JSW (JSW), KGHM (KGH), LOTOS (LTS), ORANGEPL (OPL), PGE(PGE), PKOBP (PKO). For each observed company, the results of our observations are recorded in Table 1 as reported price sets.*

Table 1. Selected quotations on the Warsaw Stock Exchange (WSE) on 23 July, 2019. Source: (Bankier.pl n.d.) and own elaboration.

Company	Open Price P_o	Close Price P_c	Low Price P_l	High Price P_h	Kind of JC
ACP	56.00	56.55	56.00	57.15	White Candle
CPS	30.14	30.00	29.74	30.24	Black Candle
ENG	7.60	7.60	7.58 *	7.71 **	Doji Star
JSW	42.60	41.92	41.80	42.68	Black Candle
KGH	97.50	96.62	95.94	97.96	Black Candle
LTS	88.70	87.48	86.82	88.70	Black Candle
OPL	6.18	6.38	6.17	6.47	White Candle
PGE	9.48	9.32	9.25	9.57	Black Candle
PKO	42.60	42.60	42.49 **	42.79 *	Doji Star

* Earlier extreme price P_{xe} ** later extreme price P_{xl}.

Each reported price set may be graphically presented by the JC mentioned in the last column of Table 1. Furthermore, we see that the quotations of ACP are assigned with a White Candle without a lover wick. In addition, all extreme prices are marked here with stars. The notions of earlier and later extreme prices are explained in Section 5.

4. Modeling of Japanese Candlesticks by Kosiński's Numbers

Let us take into account two facts:

- Any OFN may be interpreted as an imprecise approximation of a real number, which may change in direction determined by orientation.
- Each JC may be used as such an approximation of market price that it is additionally equipped with information about this price trend.

For this reason, any JC may be represented by means of OFN. In this section, we focus our attention on the use of KN for modeling candles. In the literature we find two such models.

4.1. The Marszałek's Approach

Let fixed security be given as $\mathcal{Y}$ quoted in the assumed time period $[0, T]$, Marszałek and Burczyński (2013a, 2013b, 2014) consider the observed quotation trend $Q : [0, T] \to \mathbb{R}^+$ as a realization of random variable. Considered security, $\mathcal{Y}$, can be represented by the following parameters: the open price P_o, the close price P_c, the high price P_h, and the low price P_l.

At the beginning, the authors select two real numbers, $S_1,\ S_2 \in \mathbb{R}$, fulfilling the condition:

$$P_l < S_1 \leq S_2 < P_h. \tag{30}$$

Then, they calculate conditional probabilities:

$$f_q(x) = P(Q(t) > x \big| Q(t) > S_2), \tag{31}$$

$$g_q(x) = P(Q(t) < x \big| Q(t) < S_1). \tag{32}$$

The above probabilities determine the functions $f_q : [S_2, P_h] \to [0, 1]$ and $g_q : [P_l, S_1] \to [0, 1]$. Different methods of determining numbers $S_1,\ S_2$ and estimating the probabilities (31) and (32) are discussed in (Marszałek and Burczyński 2013a, 2013b, 2014). However, in all the examples given there we observe that:

$$P_o < S_1 \leq S_2 < P_c. \tag{33}$$

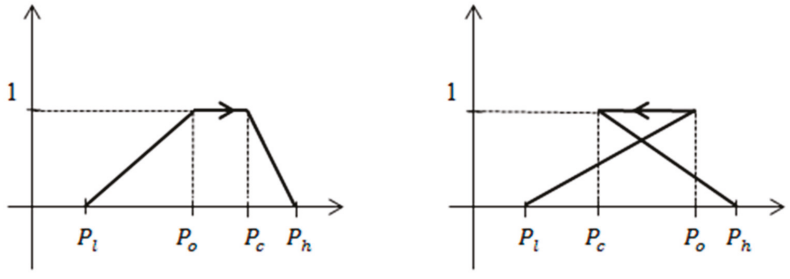
This restriction results from the requirements of probability estimation methods. Finally, authors save the observed quotation trend $Q : [0, T] \to \mathbb{R}^+$ as KN $\overleftrightarrow{Q} \in \mathbb{K}$, determined in the following way:

$$\overleftrightarrow{Q} = \overleftrightarrow{S}(a, b, c, d, \langle f_s \rangle^{\triangleleft}, \langle g_s \rangle^{\triangleleft}) = \begin{cases} \overleftrightarrow{S}\left(P_l, S_1, S_2, P_h, \left\langle g_q \right\rangle^{\triangleleft}, \left\langle f_q \right\rangle^{\triangleleft}\right) & P_o < P_c \\ \overleftrightarrow{S}\left(P_h, S_2, S_1, P_l, \left\langle f_q \right\rangle^{\triangleleft}, \left\langle g_q \right\rangle^{\triangleleft}\right) & P_o > P_c \end{cases}. \tag{34}$$

It is easy to see that the essence of the above KN is similar to the JC essence. For this reason, we agree with Marszałek and Burczyński that the above KN $\overleftrightarrow{Q}$ may be called oriented fuzzy candlestick (OFC). In our opinion, OFCs are an excellent short record of quotation volatility. In (Marszałek and Burczyński 2013a, 2013b, 2014), it is shown that OFC is very useful for financial decision making. Unfortunately, OFC does not save information about the low price P_l and high price P_h. Therefore, any OFC is not a quantitative model of JC.

4.2. The Kacprzak's Approach

Kacprzak et al. (2013) propose ex cathedra the universal way of presenting JC as TrKN $\overleftrightarrow{Tr}(P_l, P_o, P_c, P_h)$. Then, any White Candle is represented by positively oriented TrKN, which is a proper one. The membership function of this White Candle is presented in Figure 6a. In line with the Kacprzak's approach, any Black Candle is described by negatively oriented TrKN, which is an improper one. In Figure 6b, we can see the graph of a membership relation describing a Black Candle. Any Doji is represented by proper TrKN.



(**a**) (**b**)

Figure 6. Membership relation representing Japanese Candlesticks (JCs): (**a**) White Candle, (**b**) Black Candle. Source: Own elaboration.

Example 2. *We describe each Japanese candle observed in Example 1 by TrKN determined by means of the Kacprzak' method. Obtained representations are presented in Table 2. All improper TrKNs are indicated in red.*

Table 2. Representation of Japanese candles observed in Example 1 in the WSE on 23 July, 2019. Black: proper KN; Red: improper KN; Green: positively oriented FN; Blue: negatively oriented FN.

Company	Kacprzak's Method of Representation by TrKN	Proposed Method of Representation by TrOFN
ACP	$\overleftrightarrow{Tr}(56.00,\ 56.00,\ 56.55,\ 57.15)$	$\overleftrightarrow{\mathcal{JC}}_{ACP} = \overleftrightarrow{Tr}(56.00,\ 56.00,\ 56.55,\ 57.15)$
CPS	$\overleftrightarrow{Tr}(29.74,\ 30.14,\ 30.00,\ 30.24)$	$\overleftrightarrow{\mathcal{JC}}_{CPS} = \overleftrightarrow{Tr}(30.24,\ 30.14,\ 30.00,\ 29.74)$
ENG	$\overleftrightarrow{Tr}(7.58,\ 7.60,\ 7.60,\ 7.71)$	$\overleftrightarrow{\mathcal{JC}}_{ENG} = \overleftrightarrow{Tr}(7.58,\ 7.60,\ 7.60,\ 7.71)$
JSW	$\overleftrightarrow{Tr}(41.80,\ 42.60,\ 41.92,\ 42.68)$	$\overleftrightarrow{\mathcal{JC}}_{JSW} = \overleftrightarrow{Tr}(42.68,\ 42.60,\ 41.92,\ 41.80)$
KGH	$\overleftrightarrow{Tr}(95.94,\ 97.50,\ 96.62,\ 97.96)$	$\overleftrightarrow{\mathcal{JC}}_{KGH} = \overleftrightarrow{Tr}(97.96,\ 97.50,\ 96.62,\ 95.94)$
LTS	$\overleftrightarrow{Tr}(86.82,\ 88.70,\ 87.48,\ 88.70)$	$\overleftrightarrow{\mathcal{JC}}_{LTS} = \overleftrightarrow{Tr}(88.70,\ 88.70,\ 87.48,\ 86.82)$
OPL	$\overleftrightarrow{Tr}(6.17,\ 6.18,\ 6.38,\ 6.47)$	$\overleftrightarrow{\mathcal{JC}}_{OPL} = \overleftrightarrow{Tr}(6.17,\ 6.18,\ 6.38,\ 6.47)$
PGE	$\overleftrightarrow{Tr}(9.25,\ 9.48,\ 9.32,\ 9.57)$	$\overleftrightarrow{\mathcal{JC}}_{PGE} = \overleftrightarrow{Tr}(9.57,\ 9.48,\ 9.32,\ 9.25)$
PKO	$\overleftrightarrow{Tr}(42.49,\ 42.60, 42.60,\ 42.79)$	$\overleftrightarrow{\mathcal{JC}}_{PKO} = \overleftrightarrow{Tr}(42.79,\ 42.60, 42.60,\ 42.42)$

To our best knowledge, for the moment, the described Kacprzak's representation of JCs have not found an application in finance theory or practice. We suppose that it results from the fact that Kacprzak's approach to JCs' modeling is not coherent with the postulate restricting applications to the use of TrOFNs. Thus, in the next section, we will propose a universal model of representing JCs by TrOFN.

5. Representation of Japanese Candlesticks by Trapezoidal Oriented Fuzzy Numbers (TrOFN)

In above section, we conclude that any JC should be represented by TrOFN. That is why we are reconsidering the fixed JC, $\widetilde{\mathcal{JC}}$. Let $\widetilde{\mathcal{JC}}$ be reported by the data set $\{P_o, P_c, P_l, P_h\}$ of reported prices. Then, $\widetilde{\mathcal{JC}}$ may be represented by a TrOFN $\overleftrightarrow{Tr}(\alpha, \beta, \gamma, \delta)$, where a sequence $(\alpha, \beta, \gamma, \delta)$ is a monotonic permutation of all reported prices. Moreover, it is obvious that the TrOFN $\overleftrightarrow{Tr}(\alpha, \beta, \gamma, \delta)$ should be

oriented from open price P_o to close price P_c. It means that in a general case, any sequence $(\alpha, \beta, \gamma, \delta)$ contains the subsequence (P_o, P_c). From dependences (19)–(22), we get:

$$P_l \leq \min\{P_o, P_c\} \leq \max\{P_o, P_c\} \leq P_h. \tag{35}$$

We see that in a general case, the starting point and ending point are extreme prices. Therefore, any $\widetilde{\mathcal{TC}}$ may be be represented only by a TrOFN $\overleftrightarrow{Tr}(\alpha, P_o, P_c, \delta)$ where the sequence (α, δ) is such a permutation of extreme prices that the sequence $(\alpha, P_o, P_c, \delta)$ is monotonic. Looking for a universal method for determining the permutation (α, δ), in the first step, we will change the identification of extreme prices.

The orientation from open price, P_o, to close price, P_c, may be determined unambiguously only for the case when $\widetilde{\mathcal{TC}}$ has a body. It is equivalent to the condition:

$$P_o \neq P_c. \tag{36}$$

If the extreme price, $P_b \in \{P_l,\ Ph\}$, is closer to the open price, P_o, than to the close price, P_c, then the price, P_b, is called the back price. In other words, the extreme price, $P_b \in \{P_l,\ Ph\}$, is called the back price if it fulfills the condition:

$$|P_b - P_o| \leq |P_b - P_c|. \tag{37}$$

In this way, for the White Candle, the back price, P_b, is equal to the low price, P_l, and for the Black Candle, the back price, P_b, is equal to the high price, P_h. If the extreme price, $P_f \in \{P_l,\ P_h\}$, is closer to the close price, P_c, than to the open price, P_o, then the price, P_f, is called face price. In other words, the extreme price, $Pb \in \{P_l, P_h\}$, is called the face price if it fulfills the condition:

$$\left|P_f - P_c\right| \leq \left|P_f - P_o\right|. \tag{38}$$

In this way, for the White Candle, the face price, P_f, is equal to the high price, P_h, and for the Black Candle, the face price, P_f, is equal to the low price, P_l. Thanks to condition (36), the back price, P_b, and the face price, P_f, are determined explicitly. With this reinterpretation of extreme prices, we get a modernized model of JCs which have a body. The general cases of a White Candle and a Black candle are presented in Figure 7.

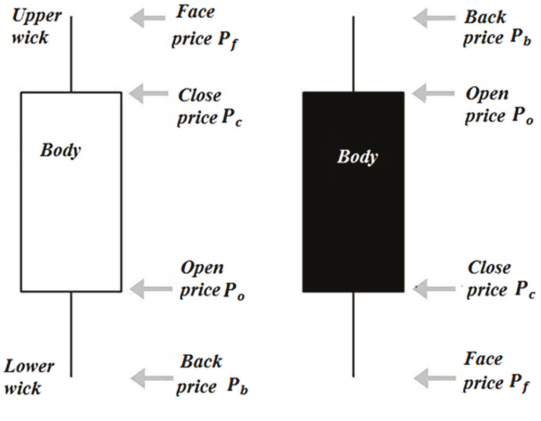
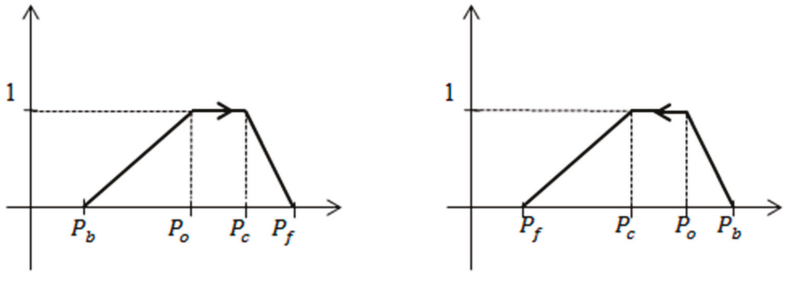


Figure 7. Japanese candles—modernized model. Source: Own elaboration.

The sequence $\left(P_b, P_o, P_c, P_f\right)$ is the unique monotonic permutation of reported prices which contains the subsequence (P_o, P_c). For this reason, we propose to describe the Japanese Candle, $\widetilde{\mathcal{JC}}$,

by the TrOFN, $\overleftrightarrow{Tr}\left(P_b, P_o, P_c, P_f\right)$. Then, any White Candle is represented by positively oriented TrOFN determined by the membership function presented in Figure 8a. In this representation, any Black Candle is described by negatively oriented TrOFN determined by the membership function presented in Figure 8b.



(**a**) (**b**)

Figure 8. Membership function representing modernized JCs: (**a**) White Candle, (**b**) Black Candle. Source: Own elaboration.

Now, we consider the case (25) when the considered $\widetilde{\mathcal{JC}}$ is a Doji. Then, the inequality (35) implies that for any permutation (α, δ) of extreme prices, the sequence $(\alpha, P_o, P_c, \delta) = (\alpha,\ P_c, P_c, \delta)$ is monotonic. Then Doji $\widetilde{\mathcal{JC}}$ may be represented by a TrOFN, $\overleftrightarrow{Tr}(\alpha, P_c, P_c, \delta)$, where the subsequence (α, δ) is any permutation of extreme prices. On the other hand, in line with (5) and (6), the orientation of $\overleftrightarrow{Tr}(\alpha, P_c, P_c, \delta)$ can only be determined as orientation from the starting point to the ending point. Furthermore, for any Doji, the back price, P_b, and the face price, P_f, are not explicitly determined. Thus, the orientation of TrOFN representing any Doji cannot be defined as the direction from a back price, P_b, to a face price, P_f. For these reasons, we propose to present any Doji by TrOFN with orientation from the earlier extreme price, P_{xe}, to the later extreme price, P_{xl}.

The kinds of extreme prices mentioned above are determined in the following way. At the beginning, we assign each quotation $\hat{Q} \in Q([0,T])$ the last moment $\tau\left(\hat{Q}\right)$ of its observation defined by the identity:

$$\tau\left(\hat{Q}\right) = \max\left\{t \in [0,T] : Q(t) = \hat{Q}\right\}. \tag{39}$$

Let us consider now the case when:

$$P_l \neq P_h. \tag{40}$$

Then, we get:

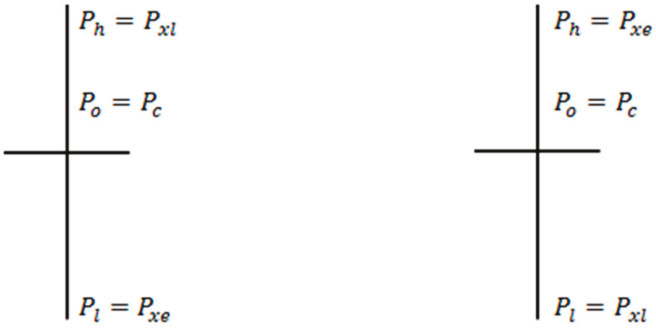
$$\tau(P_l) \neq \tau(P_h) \tag{41}$$

Then, the earlier extreme price, P_{xe}, and later extreme price, P_{xl}, we define as follows:

$$P_{xe} = \begin{cases} P_l, & \tau(P_l) < \tau(P_h), \\ P_h, & \tau(P_l) > \tau(P_h), \end{cases} \tag{42}$$

$$P_{xl} = \begin{cases} P_h, & \tau(P_l) < \tau(P_h), \\ P_l, & \tau(P_l) > \tau(P_h). \end{cases} \tag{43}$$

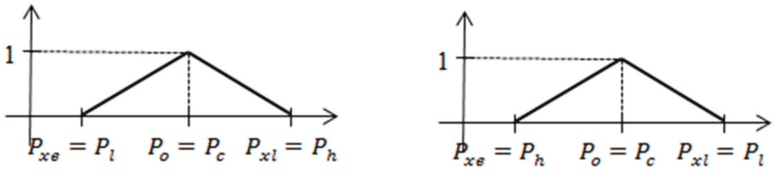
In this way, we interpret extreme prices as an earlier or later one. With this reinterpretation of extreme prices, we get a modernized model of Doji. The general cases Doji Stars are presented in Figure 9.



(**a**) (**b**)

Figure 9. Doji—modernized model: (**a**) positively oriented Doji Star, (**b**) negatively oriented Doji Star. Source: Own elaboration.

We propose to describe any Doji by TrOFN, $\overleftrightarrow{Tr}(P_{xe}, P_c, P_c, P_{xl})$. Then, the Doji Star presented in Figure 9a is described by the positively oriented TrOFN, $\overleftrightarrow{Tr}(P_l, P_c, P_c, P_h)$. The Doji Star presented in Figure 9b is described by the negatively oriented TrOFN, $\overleftrightarrow{Tr}(P_h, P_c, P_c, P_l)$. These representations of Doji are determined by their membership functions presented in Figure 10.



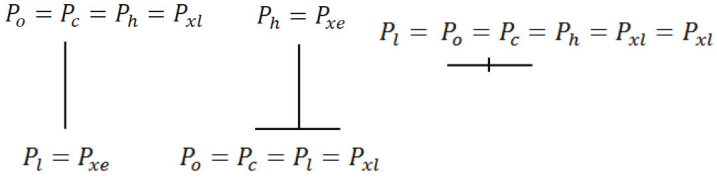
(**a**) (**b**)

Figure 10. Membership function representing the modernized Doji: (**a**) positively oriented Doji Star, (**b**) negatively oriented Doji Star. Source: Own elaboration.

For any Dragonfly Doji we have:

$$\tau(P_h) = \tau(P_c) = T > \tau(P_l). \tag{44}$$

Thus, any Dragonfly Doji is represented by positively oriented TrOFN, $\overleftrightarrow{Tr}(P_l, P_c, P_c, P_c)$. This positive orientation is consistent with the common belief that any Dragonfly Doji could signal a potential bullish reversal of market quotations. The Dragonfly Doji modernized model is shown in Figure 11a.



(**a**) (**b**) (**c**)

Figure 11. Doji—modernized model: (**a**) Dragonfly Doji, (**b**) Gravestone Doji, (**c**) Four Price Doji. Source: Own elaboration.

For any Gravestone Doji we have:

$$\tau(P_l) = \tau(P_c) = T > \tau(P_h). \tag{45}$$

Therefore, any Gravestone Doji is represented by negatively oriented TrOFN, $\overleftrightarrow{Tr}(P_h, P_c, P_c, P_c)$. This negative orientation is consistent with the common belief that any Gravestone Doji could signal a potential bearish reversal of market quotations. The Gravestone Doji modernized model is shown in Figure 11b.

If the condition (40) is not satisfied, then we get:

$$P_l = P_o = P_c = P_h. \tag{46}$$

This case corresponds to the Four Price Doji, the modernized model of which is presented in Figure 11c. Any Four Price Doji is described by the TrOFN, $\overleftrightarrow{Tr}(P_c, P_c, P_c, P_c) = [\![P_c]\!]$, which is not oriented.

It is very easy to check that all types of Japanese candles omitted in the above specification are only special cases of the discussed candles.

Example 3. *We describe each Japanese candle observed in Example 1 by TrOFN determined by means of the method proposed by us. The quotations ACP and OPL are assigned with White Candles represented by TrOFNs in the form* $\overleftrightarrow{Tr}(P_l, P_o, P_c, P_h)$. *The quotations of CPS, JSW, KGH, LTS, and PGE OPL are assigned with Black Candles represented by TrOFNs in the form* $\overleftrightarrow{Tr}(P_h, P_o, P_c, P_l)$. *The quotations ENG and PKO are assigned with Doji Star. Therefore, we should determine an earlier extreme price* P_{xe} *and a later one* P_{xl} *for both companies. We are looking for these prices by studying tick − data. The results of this search are shown in Table 1. For ENG, we get* $P_{xe} = P_l$ *and* $P_{xl} = P_h$. *Therefore, the quotations of ENG are assigned to Doji Star represented by TrOFN in the form* $\overleftrightarrow{Tr}(P_l, P_c, P_c, P_h)$. *For PKO, we obtain* $P_{xe} = P_h$ *and* $P_{xl} = P_l$. *Therefore, the quotations of PKO are assigned to Doji Star represented by TrOFN in the form* $\overleftrightarrow{Tr}(P_h, P_c, P_c, P_l)$. *Obtained representations are presented in Table 2. All positively oriented TrOFNs are indicated in green. All negatively oriented TrOFNs are indicated in blue.*

In summary, any JC can be represented by TrOFN:

$$\overleftrightarrow{JC} = \overleftrightarrow{Tr}(\alpha, P_o, P_c, \delta) \tag{47}$$

where the sequence (α, δ) is such a permutation of extreme prices $\{P_l, P_h\}$ that it is determined by conditions (37) and (38) or by (42) and (43), or by (46).

Any security may be evaluated by a present value (PV) generally defined as a current equivalent value of payments at a fixed point in time (Piasecki 2012). It is commonly accepted that the PV of a future cash flow can be imprecise. The natural consequence of this approach is estimating PV with FNs. A detailed description of the evolution of this particular model can be found, for example in (Łyczkowska-Hanćkowiak and Piasecki 2018a). If an imprecisely estimated PV is additionally equipped with a forecast of its closest changes, then it is called oriented PV (OPV). It is obvious that any OPV should be represented by an OFN $\overleftrightarrow{PV}$.

Let us take into account the fixed security $\mathcal{Y}$. If the volatility of its quotations is characterized by JC (47) then we can determine its OPV as follows:

$$\overleftrightarrow{PV}(\mathcal{Y}) = \overleftrightarrow{JC} = \overleftrightarrow{Tr}(\alpha, P_o, P_c, \delta). \tag{48}$$

Then, the OPV membership function $\mu_{Tr}(\cdot|\alpha, P_o, P_c, \delta)$ is determined by condition (9). Some applications of such an approximated OPV are given in Sections 6 and 7.

6. Oriented Expected Return Determined by Japanese Candlestick

Let us assume that the time horizon $t > 0$ of an investment is fixed. Then, the security $\mathcal{Y}$ considered here is determined by two values:

- Anticipated FV V_t and
- Assessed PV V_0.

The basic characteristic of benefits from owning this security is a return rate r_t given by the identity:

$$r_t = r(V_0, V_t) \tag{49}$$

In the general case, if $(V_0, V_t) \in \mathbb{R}^+ \times \mathbb{R}^+$, then the function: $r : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}$ is a decreasing function of PV and an increasing function of FV. Moreover, in the special case, we have here:

- Simple return rate:

$$r_t = \frac{V_t - V_0}{V_0} = \frac{V_t}{V_0} - 1 \tag{50}$$

- Logarithmic return rate:

$$r_t = \ln \frac{V_t}{V_0}. \tag{51}$$

In this section, we restrict our considerations to the case of any simple return rate given by condition (50). Thanks to this, our considerations will be more clear. However, the results obtained can be easily generalized to the case of generalized return rate given by condition (49).

The security $\mathcal{Y}$ is an authorization to receive future financial revenue, payable to a certain maturity. The value of this revenue is interpreted as the anticipated FV of the security. According to the uncertainty theory introduced by Mises (1962) and Kaplan and Barish (1967), for anyone unknown to us, the future state of affairs is uncertain. The Mises–Kaplan uncertainty is a result of our lack of knowledge about the future state of affairs. Yet, in the researched case, we can point out this particular time in the future, in which the considered state of affairs will already be known to the observer. Behind (Kolmogorov 1933, 1956; Von Mises 1957; Von Lambalgen 1996; Caplan 2001), we will accept that this is a sufficient condition for modeling the uncertainty with probability.

Above is justified in detail that FV is a random variable $\widetilde{V}_t : \Omega \to \mathbb{R}^+$. The set, Ω, is a set of elementary states, ω, of the financial market. In a classical approach to the problem of return rate estimation, a security PV is identified with the observed market price, $\check{C}$. Thus, the return rate is a random variable determined by identity:

$$r_t(\omega) = \frac{\widetilde{V}_t(\omega) - \check{C}}{\check{C}}. \tag{52}$$

In practice of financial markets analysis, the uncertainty risk is usually described by probability distribution of return rate determined by condition (52). Nowadays, we have an extensive knowledge about this subject. Let us assume that the mentioned probability distribution is given by cumulative distribution function $F_r\left(\cdot|\overline{r}\right) : \mathbb{R} \to [0;1]$. We assume that the expected value, $\overline{r}$, of this distribution exists. Moreover, let us note that we have:

$$\widetilde{V}_t(\omega) = \check{C} \cdot (1 + r_t(\omega)). \tag{53}$$

Let us now consider the case when PV is imprecisely estimated by OPV $\overleftrightarrow{PV}(\mathcal{Y})$, determined by conditions (47) and (48). Then, it is represented by its membership function $\mu_{Pv} \in [0,1]^{\mathbb{R}}$ given by identities:

$$\mu_{Pv}(x) = \mu_{Tr}(x|\alpha, P_o, P_c, \delta) \tag{54}$$

and (9). Then, the Zadeh's extension principle, (50) and (52), imply that a return rate is a fuzzy probabilistic set (Hirota 1981) represented by its membership function $\widetilde{\rho} \in [0;1]^{\mathbb{R}\times\Omega}$ as follows:

$$\widetilde{\rho}(r,\omega) = \mu_{\mathcal{PV}}\left(\frac{V_t(\omega)}{1+r}\right) = \mu_{\mathcal{PV}}\left(\frac{\check{C}\cdot(1+r_t(\omega))}{1+r}\right). \tag{55}$$

The membership function, $\rho \in [0;1]^{\mathbb{R}}$, of an expected return rate, $\overset{\leftrightarrow}{\mathcal{R}} \in \mathbb{K}$, is calculated as:

$$\rho(r) = \int_{-\infty}^{+\infty} \mu_{\mathcal{PV}}\left(\frac{\check{C}\cdot(1+y)}{1+r}\right) dF_r(y|\bar{r}) = \mu_{\mathcal{PV}}\left(\frac{\check{C}\cdot(1+\bar{r})}{1+r}\right). \tag{56}$$

According to conditions (9) and (54), the Formula (56) can be transformed into:

$$\rho(r) = \begin{cases} \frac{\check{C}\cdot\frac{1+\bar{r}}{1+r}-\alpha}{P_o-\alpha}, & \text{for} \quad \alpha \le \check{C}\cdot\frac{1+\bar{r}}{1+r} < P_o, \\ 1, & \text{for} \quad P_o \le \check{C}\cdot\frac{1+\bar{r}}{1+r} < P_c, \\ \frac{\check{C}\cdot\frac{1+\bar{r}}{1+r}-\delta}{P_c-\delta}, & \text{for} \quad P_c < \check{C}\cdot\frac{1+\bar{r}}{1+r} \le \delta, \\ 0, & \text{for} \quad \check{C}\cdot\frac{1+\bar{r}}{1+r} > \delta \ \| \ \check{C}\cdot\frac{1+\bar{r}}{1+r} < \alpha. \end{cases} \tag{57}$$

Finally, we get that expected return rate is equal to OFN $\overset{\leftrightarrow}{\mathcal{R}} \in \mathbb{K}$, given as follows:

$$\overset{\leftrightarrow}{\mathcal{R}} = \overset{\leftrightarrow}{L}\left(\frac{\check{C}\cdot(1+\bar{r})}{\alpha} - 1, \frac{\check{C}\cdot(1+\bar{r})}{P_o} - 1, \frac{\check{C}\cdot(1+\bar{r})}{P_c} - 1, \frac{\check{C}\cdot(1+\bar{r})}{\delta} - 1, S_L,\ E_L\right), \tag{58}$$

where,

$$S_L(x) = \frac{\check{C}\cdot\frac{1+\bar{r}}{1+x}-\alpha}{P_o-\alpha}, \tag{59}$$

$$E_L(x) = \frac{\check{C}\cdot\frac{1+\bar{r}}{1+x}-\delta}{P_c-\delta}. \tag{60}$$

The above determined expected return rate, $\overset{\leftrightarrow}{\mathcal{R}}$, is an example of the oriented expected return rate (Piasecki 2017). We see that the expected return rate is not TrOFN. Moreover, the identities (48) and (58) show that JC and the expected return rate determined by it always have opposite orientations. Therefore, we can say:

- Rise in quotations predicted by JC allows us to anticipate a decline in the expected return rate.
- Fall in quotations predicted by JC allows us to anticipate an upturn in the expected return rate.

In theory and practice of finance, both of these facts are well known. This observation proves that the extension of the fuzzy models of imprecise PV and return rate to the case of oriented fuzzy models is the appropriate direction for the development of fuzzy finance theory.

Using the disorientation map (34), we can convert any oriented expected rate to fuzzy return rate. Then, JCs may be applied in fuzzy portfolio analysis (Fang et al. 2008; Gupta et al. 2014) or in (Piasecki 2014).

Example 4. *During the session on the Warsaw Stock Exchange on 23 July, 2019, we observed quotations of ORANGEPL. The volatility of these quotations is characterized* by JC $\overset{\leftrightarrow}{\mathcal{JC}}_{OPL} = \overset{\leftrightarrow}{Tr}(6.17,\ 6.18,\ 6.38,\ 6.47)$. *Then, in agreement with condition (48), the ORANGEPL OPV is equal to positively oriented:*

$$\overset{\leftrightarrow}{PV}(OPL) = \overset{\leftrightarrow}{Tr}(6.17,\ 6.18,\ 6.38,\ 6.47).$$

The expected quarterly return rate from ORANGEPL is determined by the broker's office as follows: $\bar{r} = 0.025$. *We determined the ORANGEPL market price* $\hat{C} = 6.20$ *as open price on 24 July, 2019. From conditions (58)–(60), we get that expected return rate is equal to the negatively oriented OFN* $\overleftrightarrow{\mathcal{R}} \in \mathbb{K}$, *given as follows:*

$$\overleftrightarrow{\mathcal{R}} = \overleftrightarrow{L}(0.0300, 0.0283, -0.0039, -0.0178, S_L,\ E_L),$$

where,

$$S_L(x) = \frac{18.5 - 617.0{\cdot}x}{1+x},$$

$$E_L(x) = \frac{1.278 + 71.68{\cdot}x}{1+x}.$$

Finally, using the disorientation map (34), we calculate fuzzy expected return rate, $\mathcal{R} \in \mathbb{F}$, *in the following way:*

$$\begin{aligned}\mathcal{R} &= L(-0.0178,\ -0.0039, 0.0283, 0.3000, E_L,\ S_L)\\ &= \Psi\Big(\overleftrightarrow{L}(0.0300, 0.0283, -0.0039, -0.0178, S_L,\ E_L)\Big)\end{aligned}$$

7. Case Study

In this section, we present some simple applications of JCs for financial portfolio analysis. The main aim of this section is to show that the proposed JC representation is applicable in quantified analyses of the financial market. Such presentation requires a prior explanation of the theoretical foundations of this case study.

7.1. Theoretical Background

From a financial portfolio, we will understand an arbitrary, finite set of financial assets. In this section, we restrict our considerations to the case of assets given as securities quoted in an assumed time period.

We consider a multi-assets portfolio, π, consisting of assets, $\mathcal{Y}_i$ $(i = 1, 2, \ldots, n)$. Each of the assets $\mathcal{Y}_i \in \pi$ are characterized by OPV:

$$\overleftrightarrow{PV}(\mathcal{Y}_i) = \overleftrightarrow{Tr}\Big(\alpha^{(i)},\ P_o^{(i)},\ P_c^{(i)}, \delta^{(i)}\Big). \tag{61}$$

We distinguish the rising securities' portfolio, $\pi^+ \subset \pi$, and the falling securities' portfolio, $\pi^- \subset \pi$, as follows:

$$\pi^+ = \Big\{\mathcal{Y}_i \in \pi : \overleftrightarrow{PV}(\mathcal{Y}_i) \in \mathbb{K}^+ \cup\ \mathbb{R}\Big\}, \tag{62}$$

$$\pi^- = \Big\{\mathcal{Y}_i \in \pi : \overleftrightarrow{PV}(\mathcal{Y}_i) \in \mathbb{K}^-\Big\}. \tag{63}$$

The portfolio PV is always equal to the sum of its components' PV. For the OPV case, the addition should be modeled by sum $\boxplus$. In (Piasecki 2018), it is shown that a result of multiple additions $\boxplus$ depends on summands permutation. It implies that the portfolio's OPV given as any multiple sum of components' OPV is not explicitly determined. Therefore, in (Łyczkowska-Hanćkowiak and Piasecki 2018a), some reasonable method of calculating a portfolio's OPV is proposed.

At first we calculate OPV of the rising securities' portfolio π^+ :

$$\overleftrightarrow{PV}\big(\pi^+\big) = \boxed{+}_{\mathcal{Y}_i \in \pi^+} \overleftrightarrow{PV}(\mathcal{Y}_i) \tag{64}$$

and OPV of the falling securities' portfolio π^-:

$$\overleftrightarrow{PV}(\pi^-) = \boxed{+}_{\boldsymbol{y}_i \in \pi^-} \overleftrightarrow{PV}(\boldsymbol{y}_i). \tag{65}$$

In (Piasecki 2018), it is proven that both sums (64) and (65) do not depend on permutation of the summand. Therefore, these sums are determined explicitly. Moreover, single addition $\boxplus$ is commutative. Therefore, we can to determine OPV of total portfolio, π, in an explicit manner as the sum:

$$\overleftrightarrow{PV}(\pi) = \overleftrightarrow{PV}\left(\pi^+\right) \boxplus \overleftrightarrow{PV}(\pi^-). \tag{66}$$

The above approach is sufficient to manage portfolio risk, because only rising securities can get BUY or ACCUMULATE recommendations and only falling securities can get SELL or REDUCE recommendations. The complex form of the addition definition (16) allows us to use OPV only for the evaluation of an already constructed portfolio. Such an evaluation may be carried out using the analytical tools described in Section 5 and by (Łyczkowska-Hańćkowiak 2019a, 2019b; Łyczkowska-Hańćkowiak and Piasecki 2018b, 2019a, 2019b).

7.2. Empirical Study

We analyzed the portfolio of shares listed in Example 1. After closing the session on the WSE on July 23, 2019, we evaluated the portfolio, π, of:

- The block B_1 of 10 shares of ACP,
- The block B_2 of 29 shares of CPS,
- The block B_3 of 30 shares of ENG,
- The block B_4 of 5 shares of the JSW,
- The block B_5 of 5 shares of KGH,
- The block B_6 of 10 shares of LTS,
- The block B_7 of 100 shares of OPL,
- The block B_8 of 50 shares of PGE,
- The block B_9 of 10 shares of PKO.

For each security $\boldsymbol{y} \in \pi$, its OPV $\overleftrightarrow{PV}(\boldsymbol{y})$ is estimated by current JC $\overleftrightarrow{\mathcal{JC}}_{\boldsymbol{y}}$, as presented in Example 3 and Table 2. Using the identity (10), for each considered block B_i $(i = 1, 2, \ldots, 9)$ of shares, we calculate their OPV $\overleftrightarrow{PV}(B_i)$ as follows:

$$\overleftrightarrow{PV}(B_1) = 10 \boxdot \overleftrightarrow{\mathcal{JC}}_{ACP} = \overleftrightarrow{Tr}(560.00,\ 560.00,\ 565.50,\ 571.50), \tag{67}$$

$$\overleftrightarrow{PV}(B_2) = 29 \boxdot \overleftrightarrow{\mathcal{JC}}_{CPS} = \overleftrightarrow{Tr}(876.96,\ 874.06,\ 870.00,\ 862.46), \tag{68}$$

$$\overleftrightarrow{PV}(B_3) = 30 \boxdot \overleftrightarrow{\mathcal{JC}}_{ENG} = \overleftrightarrow{Tr}(227.40,\ 228.00,\ 228.00,\ 231.30), \tag{69}$$

$$\overleftrightarrow{PV}(B_4) = 5 \boxdot \overleftrightarrow{\mathcal{JC}}_{JSW} = \overleftrightarrow{Tr}(213.40,\ 213.00,\ 209.60,\ 209.00), \tag{70}$$

$$\overleftrightarrow{PV}(B_5) = 5 \boxdot \overleftrightarrow{\mathcal{JC}}_{KGH} = \overleftrightarrow{Tr}(489.80,\ 487.50,\ 483.10,\ 479.70), \tag{71}$$

$$\overleftrightarrow{PV}(B_6) = 10 \boxdot \overleftrightarrow{\mathcal{JC}}_{LTS} = \overleftrightarrow{Tr}(887.00,\ 887.00,\ 874.80,\ 868.20), \tag{72}$$

$$\overleftrightarrow{PV}(B_7) = 100 \boxdot \overleftrightarrow{\mathcal{JC}}_{OPL} = \overleftrightarrow{Tr}(617.00,\ 618.00,\ 638.00,\ 647.00), \tag{73}$$

$$\overleftrightarrow{PV}(B_8) = 100 \boxdot \overleftrightarrow{\mathcal{JC}}_{PGE} = \overleftrightarrow{Tr}(957.00,\ 948.00,\ 932.00,\ 925.00), \tag{74}$$

$$\overleftrightarrow{PV}(B_9) = 10 \boxdot \overleftrightarrow{\mathcal{JC}}_{PKO} = \overleftrightarrow{Tr}(427.90,\ 426.00,\ 426.00,\ 424.20). \tag{75}$$

In the next step, we distinguish the rising securities' portfolio π^+ and the falling securities' portfolio π^- as follows:

$$\pi^+ = \{ACP,\ ENG,\ OPL,\},\tag{76}$$

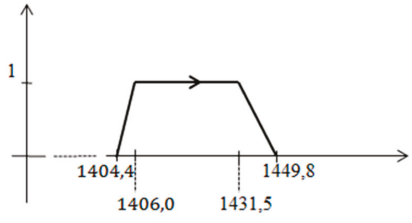
$$\pi^- = \{CPS,\ JSW,\ KGH,\ LTS,\ PGE,\ PKO\}.\tag{77}$$

Then, we calculate OPVs of the rising securities' portfolio $\pi^+ \subset \pi$ and the falling securities' portfolio $\pi^- \subset \pi$. Due to conditions (64) and (65), we obtain:

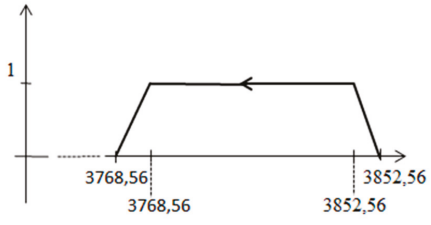
$$\overleftrightarrow{PV}\left(\pi^+\right) = \overleftrightarrow{PV}(B_1) \boxplus \overleftrightarrow{PV}(B_3) \boxplus \overleftrightarrow{PV}(B_7) = \overleftrightarrow{Tr}(1404.40,\ 1406.00,\ 1431.50,\ 1449.80),\tag{78}$$

$$\begin{aligned}\overleftrightarrow{PV}(\pi^-) &= \overleftrightarrow{PV}(B_2) \boxplus \overleftrightarrow{PV}(B_4) \boxplus \overleftrightarrow{PV}(B_5) \boxplus \overleftrightarrow{PV}(B_6) \boxplus \overleftrightarrow{PV}(B_8) \boxplus \overleftrightarrow{PV}(B_9)\\ &= \overleftrightarrow{Tr}(3852.06,\ 3835.56,\ 3795.50,\ 3768.56).\end{aligned}\tag{79}$$

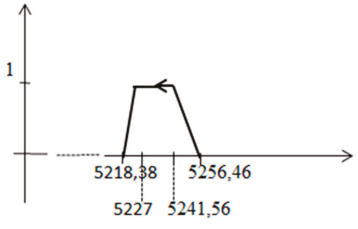
The membership functions of $\overleftrightarrow{PV}(\pi^+)$ and $\overleftrightarrow{PV}(\pi^-)$ are presented in Figure 12.



(**a**)



(**b**)



(**c**)

Figure 12. Membership function of OPV (**a**) for the rising securities' portfolio π^+, (**b**) for the falling securities' portfolio π^-, and (**c**) for total portfolio π. Source: Own elaboration.

In the last step, due to condition (66), we determine OPV of portfolio π as the sum:

$$\overleftrightarrow{PV}(\pi) = \overleftrightarrow{PV}\left(\pi^{+}\right) \boxplus \overleftrightarrow{PV}(\pi^{-}) = \overleftrightarrow{Tr}(5256.46,\ 5241.56,\ 5227.00,\ 5218.38). \tag{80}$$

The membership function of $\overleftrightarrow{PV}(\pi)$ is presented in Figure 12. We notice that the portfolio π is evaluated by a Black Candle, which is a bearish signal. Moreover, the above result may be applied for determining the imprecise expected return from portfolio π. This return rate should be calculated in such a way as that presented in Example 4.

After Klir (1993), we distinguish the ambiguity as a component of imprecision. Ambiguity is understood as a lack of a clear recommendation between one alternative among various others. Ambiguity increases with the amount of the recommended alternatives. With the increase in ambiguity, imprecision is growing too. Comparing the membership function charts, we see that OPV ambiguity of total portfolio, π, is less than OPV ambiguity of the rising securities' portfolio, $\pi^{+} \subset \pi$, and the falling securities' portfolio, $\pi^{-} \subset \pi$. In (Piasecki et al. 2019), the mathematical theorem showing that the analogous result we get for any total portfolio containing simultaneously rising securities and falling securities was proven. In addition, it was proven that this effect is not revealed when we use FNs to determine imprecise PV. Thanks to this, we can conclude that the replacement of FN by OFN reduces the imprecision of the determined PV. This is another benefit of using OFNs.

8. Summary

It is a well-known fact that JCs are a very convenient tool for synthetic recording of high-frequency time series of financial data. We have shown that each JC may be explicitly represented by such a TrOFN that its orientation is always consistent with the closest quotations' changes predicted by the represented JC. The positive orientation of JC representation always means a bullish signal. The negative orientation of JC representation is always a bearish signal. The proposed JC numerical model distinguishes all JC types. In addition, the use of this model does not cause a loss of information about the represented JC. Therefore, the JC representation described here can be recommended as a numerical tool for analyzing high-frequency financial data.

The proposed JC numerical model highlights the imprecision of capital market assessments. In the case of economics and finance, this imprecision is a natural state of affairs. Therefore, we can conclude that the proposed model reflects the essence of information about the observed capital market well. On the other hand, a fuzzy set is a commonly used imprecise information model. The replacement of Kacprzak's model by our model means that we can apply the whole fuzzy sets theory to the analysis of JCs. To our best knowledge, the JC fuzzy numerical model proposed by us is a unique one which is coherent with fuzzy set theory. Some possibilities of application of the usual fuzzy set theory are shown. In a special case, we can use methods dedicated to OFNs.

We showed a simple application of our JC model for financial portfolio analysis. The results obtained have the possibility of further application in any financial analysis. It proves that the JC model proposed by us is applicable for financial analysis.

Each JC chart is a finite sequence of JCs describing the volatility of fixed security quotation in an assumed period of time. Financial practitioners distinguish JC patterns defined as such families of short JC sequences similar to a given reference pattern, which they believe can predict a particular quotations' movement. Using the results obtained in this work, we can recognize any JC chart as a sequence of TrOFNs. Therefore, our model can be used for recognition of JC patterns on a current JC chart. The above remark well justifies the need for further research into determining known reference patterns.

Moreover, our JC model should be used to test high-frequency financial data systems to determine the optimal time horizon described by a single JC. We recommend this research field as a very promising area of scientific considerations and discussion.

The directions of future research described above are very general. We can also suggest more detailed directions of future research. An interesting problem here is the use of the JC numerical model to determine the orientation of behavioral present value (Piasecki 2011; Piasecki and Siwek 2015; Łyczkowska-Hanćkowiak 2017). Another interesting problem is the use of our numerical JC model as a premise in the models described in (Marszałek and Burczyński 2013a, 2013b, 2014). Then, we can compare the effects of the application of our numerical JC model with the effects of the application of Marszałek's fuzzy candles.

Author Contributions: Conceptualization, K.P. and A.Ł.-H.; methodology, K.P.; validation, A.Ł.-H.; formal analysis, A.Ł.-H.; data collection A.Ł.-H.; writing—original draft preparation, K.P.; writing—review and editing, K.P. and A.Ł.-H.; visualization, A.Ł.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors are very grateful to the guest editor and to the anonymous reviewers for their insightful and constructive comments and suggestions. Using these comments allowed us to improve our paper.

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

References

Bankier.pl. n.d. Available online: https://www.bankier.pl/inwestowanie/profile/quote.html?symbol=WIG20 (accessed on 5 August 2019).

Caplan, Bryan. 2001. Probability, common sense, and realism: A reply to Hulsmann and Block. *The Quarterly Journal of Austrian Economics* 4: 69–86.

Delgado, Miguel, Maria Amparo Vila, and Wiliam Voxman. 1998. On a canonical representation of fuzzy numbers. *Fuzzy Sets and Systems* 93: 125–35. [CrossRef]

Detollenaere, Benoit, and Pablo Mazza. 2014. Do Japanese candlesticks help solve the trader's dilemma? *Journal of Banking and Finance* 48: 386–95. [CrossRef]

Dubois, Didier, and Henri Prade. 1978. Operations on fuzzy numbers. *International Journal of Systems Science* 9: 613–29. [CrossRef]

Dubois, Didier, and Henri Prade. 1979. Fuzzy real algebra: Some results. *Fuzzy Sets and Systems* 2: 327–48. [CrossRef]

Fang, Young, King Keung Lai, and Shouyaung Wang. 2008. *Fuzzy Portfolio Optimization. Theory and Methods*. Lecture Notes in Economics and Mathematical Systems 609. Berlin: Springer.

Fock, J. Henning, Christian Klein, and Bernhard Zwergel. 2005. Performance of Candlestick Analysis on Intraday Futures Data. *Journal of Futures Markets* 13: 28–40. [CrossRef]

Goetschel, Roy, and Wiliam Voxman. 1986. Elementary fuzzy calculus. *Fuzzy Sets and Systems* 18: 31–43. [CrossRef]

Gupta, Pankaj, Mukesh Kumar Mehlawat, Masahiro Inuiguchi, and Suresh Chandra. 2014. *Fuzzy Portfolio Optimization. Advances in Hybrid Multi-Criteria Methodologies*. Studies in Fuzziness. Berlin: Springer.

Herrera, Francisco, and Enrique Herrera-Viedma. 2000. Linguistic decision analysis: Steps for solving decision problems under linguistic information. *Fuzzy Sets and Systems* 115: 67–82. [CrossRef]

Hirota, Karou. 1981. Concepts of probabilistic sets. *Fuzzy Sets and Systems* 5: 31–46.

Jasemi, Milad, Ali Mohammad Kimigiari, and Azizollah Memariani. 2011. A modern neural network model to do stock market timing on the basis of the ancient investment technique of Japanese Candlestick. *Expert Systems with Applications* 38: 3884–90. [CrossRef]

Kacprzak, Dariusz, Witold Kosiński, and Witold Konrad Kosiński. 2013. Financial Stock Data and Ordered Fuzzy Numbers. In *Artificial Intelligence and Soft Computing. ICAISC 2013*. Edited by Leszek Rutkowski, Marcin Korytkowski, Rafał Scherer, Ryszard Tadeusiewicz, Lofti A. Zadeh and Jacek M. Zurada. Lecture Notes in Computer Science. Berlin and Heidelberg: Springer, vol. 7894, pp. 1–12. [CrossRef]

Kamo, Takemori, and Cihan Dagli. 2009. Hybrid approach to the Japanese candlestick method for financial forecasting. *Expert Systems with Applications* 36: 5023–30. [CrossRef]

Kaplan, Seymour, and Norman N. Barish. 1967. Decision-Making Allowing Uncertainty of Future Investment Opportunities. *Management Science* 13: 569–77. [CrossRef]

Klir, George J. 1993. Developments in uncertainty-based information. *Advances in Computers* 36: 255–332. [CrossRef]

Kolmogorov, Andrey Nikolaevich. 1933. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Julius Springer.

Kolmogorov, Andrey Nikolaevich. 1956. *Foundations of the Theory of Probability*. New York: Chelsea Publishing Company.

Kosiński, Witold. 2006. On fuzzy number calculus. *International Journal of Applied Mathematics and Computer Science* 16: 51–57.

Kosiński, Witold, and Piotr Słysz. 1993. Fuzzy numbers and their quotient space with algebraic operations. *Bulletin of the Polish Academy of Sciences* 41: 285–95.

Kosiński, Witold, Piotr Prokopowicz, and Dominik Ślęzak. 2002. Fuzzy numbers with algebraic operations: Algorithmic approach. In *Proc.IIS'2002 Sopot*. Edited by Mieczysław Klopotek, Sławomir T. Wierzchoń and Maciej Michalewicz. Heidelberg: Poland Physica Verlag, pp. 311–20.

Kosiński, Witold, Piotr Prokopowicz, and Dominik Ślęzak. 2003. Ordered fuzzy numbers. *Bulletin of the Polish Academy of Sciences* 51: 327–39.

Von Lambalgen, Michiel. 1996. Randomness and foundations of probability: Von Mises' axiomatization of random sequences. In *Probability, Statistics and Game Theory, Papers in Honor of David Blackwell*. Edited by Thomas S. Ferguson, Lloyd Stowell Shapley and Jim B. MacQueen. Harward: Institute for Mathematical Statistics, pp. 347–68.

Lee, Chiung-Hon Leon, Alan Liu, and Wen-Sung Chen. 2006. Pattern discovery of fuzzy time series for financial prediction. *IEEE Transactions on Knowledge and Data Engineering* 18: 613–25.

Łyczkowska-Hanćkowiak, Anna. 2017. Behavioural Present Value determined by ordered fuzzy number. *SSRN Electronic Journal*. [CrossRef]

Łyczkowska-Hanćkowiak, Anna. 2019a. Sharpe's Ratio for Oriented Fuzzy Discount Factor. *Mathematics* 7: 272.

Łyczkowska-Hanćkowiak, Anna. 2019b. Jensen Alpha for Oriented Fuzzy Discount Factor. Paper presented at the 37th International Conference on Mathematical Methods in Economics 2019, České Budějovice, Czech Republic, September 11–13; pp. 79–84.

Łyczkowska-Hanćkowiak, Anna, and Krzysztof Piasecki. 2018a. The present value of a portfolio of assets with present values determined by trapezoidal ordered fuzzy numbers. *Operations Research and Decisions* 28: 41–56. [CrossRef]

Łyczkowska-Hanćkowiak, Anna, and Krzysztof Piasecki. 2018b. The expected discount factor determined for present value given as ordered fuzzy number. In *9th International Scientific Conference "Analysis of International Relations 2018. Methods and Models of Regional Development. Winter Edition" Conference Proceedings*. Edited by Włodzimierz Szkutnik, Anna Sączewska-Piotrowska, Monika Hadaś-Dyduch and Jan Acedański. Katowice: University of Economics in Katowice Publishing, pp. 69–75.

Łyczkowska-Hanćkowiak, Anna, and Krzysztof Piasecki. 2019a. Treynor's ratio for Oriented Fuzzy Discount Factor. In *11th International Scientific Conference "Analysis of International Relations 2019. Methods and Models of Regional Development. Winter Edition" Conference Proceedings*. Edited by Włodzimierz Szkutnik, Anna Sączewska-Piotrowska, Monika Hadaś-Dyduch and Jan Acedański. Katowice: University of Economics in Katowice Publishing, pp. 31–44.

Łyczkowska-Hanćkowiak, Anna, and Krzysztof Piasecki. 2019b. Roy's criterion for Oriented Fuzzy Discount Factor. In *12th International Scientific Conference "Analysis of International Relations 2019. Methods and Models of Regional Development. Summer Edition" Conference Proceedings*. Edited by Włodzimierz Szkutnik, Anna Sączewska-Piotrowska, Monika Hadaś-Dyduch and Jan Acedański. Katowice: University of Economics in Katowice Publishing, pp. 31–43.

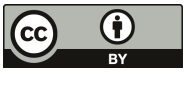
Marshall, Ben R., Martin R. Young, and Lawrence C. Rose. 2006. Candlestick technical trading strategies: Can they create value for investors? *Journal of Banking and Finance* 30: 2303–23. [CrossRef]

Marszałek, Adam, and Tadeusz Burczyński. 2013a. Modelling Financial High Frequency Data Using Ordered Fuzzy Numbers. In *Artificial Intelligence and Soft Computing. ICAISC 2013*. Edited by Leszek Rutkowski, Marcin Korytkowski, Rafał Scherer, Ryszard Tadeusiewicz, Lofti A. Zadeh and Jacek M. Zurada. Lecture Notes in Computer Science. Berlin and Heidelberg: Springer, vol. 7894. [CrossRef]

Marszałek, Adam, and Tadeusz Burczyński. 2013b. Financial Fuzzy Time Series Models Based on Ordered Fuzzy Numbers. In *Time Series Analysis, Modeling and Applications*. Edited by Witold Pedrycz and Shyi-Ming Chen. Intelligent Systems Reference Library. Berlin and Heidelberg: Springer, vol. 47. [CrossRef]

Marszałek, Adam, and Tadeusz Burczyński. 2014. Modelling and forecasting financial time series with ordered fuzzy candlesticks. *Information Sciences* 273: 144–55. [CrossRef]

Von Mises, Richard. 1957. *Probability, Statistics and Truth*. New York: The Macmillan Company.
Mises, Ludwig. 1962. *The Ultimate Foundation of Economic Science an Essay on Method*. Princeton: D. Van Nostrand Company, Inc.
Morris, Gregory L. 2006. *Candlestick Charting Explained: Timeless Techniques for Trading Stocks and Futures*. New York: McGraw-Hill.
Naranjo, Rodrigo, and Matilde Santos. 2019. A fuzzy decision system for money investment in stock markets based on fuzzy candlesticks pattern recognition. *Expert Systems with Applications* 133: 34–48. [CrossRef]
Naranjo, Rodrigo, Javier Arroyo, and Matilde Santos. 2018. Fuzzy modeling of stock trading with fuzzy candlesticks. *Expert Systems with Applications* 93: 15–27. [CrossRef]
Nison, Steve. 1991. *Japanese Candlestick Charting Techniques*. New York: New York Institute of Finance.
Piasecki, Krzysztof. 2011. Behavioural present value. *SSRN Electronic Journal* 1: 1–17. [CrossRef]
Piasecki, Krzysztof. 2012. Basis of financial arithmetic from the viewpoint of the utility theory. *Operations Research and Decisions* 3: 37–53.
Piasecki, Krzysztof. 2014. On Imprecise Investment Recommendations. *Studies in Logic, Grammar and Rhetoric* 37: 179–94. [CrossRef]
Piasecki, Krzysztof. 2017. Expected return rate determined as oriented fuzzy number. Paper presented at the 35th International Conference Mathematical Methods in Economics, Hradec Králové, Czech Republic, September 13–15; pp. 561–65.
Piasecki, Krzysztof. 2018. Revision of the Kosiński's Theory of Ordered Fuzzy Numbers. *Axioms* 7: 16. [CrossRef]
Piasecki, Krzysztof. 2019. Relation "Greater than or Equal to" between Ordered Fuzzy Numbers. *Applied System Innovation* 2: 26. [CrossRef]
Piasecki, Krzysztof, and Joanna Siwek. 2015. Behavioural Present Value Defined as Fuzzy Number—A New Approach. *Folia Oeconomica Stetinensia* 15: 27–41. [CrossRef]
Piasecki, Krzysztof, Ewa Roszkowska, and Anna Łyczkowska-Hanćkowiak. 2019. Impact of the Orientation of the Ordered Fuzzy Assessment on the Simple Additive Weighted Method. *Symmetry* 11. [CrossRef]
Prokopowicz, Piotr. 2016. The Directed Inference for the Kosinski's Fuzzy Number Model. In *Proceedings of the Second International Afro-European Conference for Industrial Advancement AECIA 2015*. Edited by Ajith Abraham, Katarzyna Wegrzyn-Wolska, Aboul Ella Hassanien, Vaclav Snasel and Adel M. Alimi. Advances in Intelligent Systems and Computing. Cham: Springer, vol. 427. [CrossRef]
Prokopowicz, Piotr, Jacek Czerniak, Dariusz Mikołajewski, Łukasz Apiecionok, and Dominik Ślezak. 2017. *Theory and Applications of Ordered Fuzzy Number. Tribute to Professor Witold Kosiński*. Studies in Fuzziness and Soft Computing 356. Berlin: Springer.
Prokopowicz, Piotr, and Witold Pedrycz. 2015. The Directed Compatibility Between Ordered Fuzzy Numbers—A Base Tool for a Direction Sensitive Fuzzy Information Processing. *Artificial Intelligence and Soft Computing* 119: 249–59. [CrossRef]
Tsung-Hsun, Lu, Yung-Ming Shiu, and Tsung-Chi Liu. 2012. Profitable candlestick trading strategies—The evidence from a new perspective. *Review of Financial Economics* 21: 63–68. [CrossRef]
Tsung-Hsun, Lu, Yi-Chi Chen, and Yu-Chin Hsu. 2015. Trend definition or holding strategy: What determines the profitability of candlestick charting? *Journal of Banking and Finance* 61: 172–83. [CrossRef]
Zadeh, Lofti A. 1975a. The concept of a linguistic variable and its application to approximate reasoning. Part I. Information linguistic variable. *Expert Systems with Applications* 36: 3483–88.
Zadeh, Lofti A. 1975b. The concept of a linguistic variable and its application to approximate reasoning. Part II. *Information Sciences* 8: 301–57. [CrossRef]
Zadeh, Lofti A. 1975c. The concept of a linguistic variable and its application to approximate reasoning. Part III. *Information Sciences* 9: 43–80. [CrossRef]



econometrics

MDPI

Article

New Evidence of the Marginal Predictive Content of Small and Large Jumps in the Cross-Section

Bo Yu, Bruce Mizrach and Norman R. Swanson *

Department of Economics, School of Arts and Sciences, Rutgers University, 75 Hamilton Street, New Brunswick, NJ 08901, USA; byu@econ.rutgers.edu (B.Y.); mizrach@econ.rutgers.edu (B.M.)
* Correspondence: nswanson@economics.rutgers.edu

Received: 19 February 2020; Accepted: 2 May 2020; Published: 19 May 2020

Abstract: We investigate the marginal predictive content of small versus large jump variation, when forecasting one-week-ahead cross-sectional equity returns, building on Bollerslev et al. (2020). We find that sorting on signed small jump variation leads to greater value-weighted return differentials between stocks in our highest- and lowest-quintile portfolios (i.e., high–low spreads) than when either signed total jump or signed large jump variation is sorted on. It is shown that the benefit of signed small jump variation investing is driven by stock selection within an industry, rather than industry bets. Investors prefer stocks with a high probability of having positive jumps, but they also tend to overweight safer industries. Also, consistent with the findings in Scaillet et al. (2018), upside (downside) jump variation negatively (positively) predicts future returns. However, signed (large/small/total) jump variation has stronger predictive power than both upside and downside jump variation. One reason large and small (signed) jump variation have differing marginal predictive contents is that the predictive content of signed large jump variation is negligible when controlling for either signed total jump variation or realized skewness. By contrast, signed small jump variation has unique information for predicting future returns, even when controlling for these variables. By analyzing earnings announcement surprises, we find that large jumps are closely associated with "big" news. However, while such news-related information is embedded in large jump variation, the information is generally short-lived, and dissipates too quickly to provide marginal predictive content for subsequent weekly returns. Finally, we find that small jumps are more likely to be diversified away than large jumps and tend to be more closely associated with idiosyncratic risks. This indicates that small jumps are more likely to be driven by liquidity conditions and trading activity.

Keywords: forecasting; integrated volatility; high-frequency data; jumps; realized skewness; cross-sectional stock returns; signed jump variation

1. Introduction

Theoretical models of the risk-return relationship anticipate that volatility should be priced, and that investors should demand higher expected returns for more volatile assets. However, ex-ante risk measures are not directly observable, and must be estimated (see e.g., Rossi and Timmermann (2015)). Given the necessity of estimating volatility, various risk estimators have been used in the empirical literature studying the strength and sign of the risk-return relationship. Unfortunately, the evidence from the literature is mixed, in the sense that researchers have found both negative and positive relationships between return and volatility. One possible reason for these surprisingly contradictory findings is that the risk-return relationship is nonlinear. Examples of papers pursuing this hypothesis include Campbell and Vuolteenaho (2004), who incorporate different factor betas based on good and bad news about cash flows and discount rates; and Woodward and Anderson (2009) who find that bull and bear market betas differ substantially across

most industries. This research has helped to spawn the "smart-beta" approach to factor investing.[1] In related research, Feunou et al. (2013) model the effects of volatility in positive and negative return states separately. They define so-called disappointment aversion preferences, and show that investors should demand a higher return for downside variability. These authors find empirical support for their model in the U.S. and several foreign markets using a bi-normal GARCH process to estimate volatility.

In this paper, we focus on the importance of jumps in volatility for understanding the risk-return relationship. We do this by assessing the marginal predictive content of small versus large jump variation, when forecasting one-week-ahead cross-sectional equity returns. We also examine earnings announcements as well as carry out various Fama–MacBeth type regressions to uncover the linkages between (small and large) jumps and news. Finally, we examine the importance of control variates, including skewness and other firm characteristics, when undertaking to disentangle the relative importance of small, large, positive, and negative jumps for the dynamic evolution of firm specific returns. Much of the empirical research that explores the importance of jumps in this context focuses on the estimation of continuous and jump variation components using nonparametric realized measures constructed with high-frequency financial data. A key paper in this area is Bollerslev et al. (2020), who examine the relationship between signed jumps and future stock returns in the cross-section. They document that signed jump variation, which captures the asymmetric impact of upside and downside jump risk, is a good predictor of future returns, particularly for small and illiquid stocks.[2] We add to this literature by decomposing jump variation into signed small and large components and evaluating the importance of these elements in the cross-section of stock returns. We use the cross-section of individual stocks because aggregate index returns may mask small jump effects on return predictability. Indeed, many studies document that aggregation may diversify away idiosyncratic small jumps in the cross-section (see e.g., Aït-Sahalia and Jacod (2012); Duong and Swanson (2015)).

The motivation for our paper can be traced back to Yan (2011) and Jiang and Yao (2013), who show that large, infrequent jumps are priced in the cross-section of returns. Feunou et al. (2018) take the decomposition used by these authors one step further, and model jumps in the realized semi-variances of market returns. They construct a new measure of the variance risk premium, and find a strong positive premium for downside risk. Fang et al. (2017) find a similar result for Chinese market returns. In a related line of research, various authors study the information content in upside and downside jump variation. For example, Guo et al. (2015) document that at the market level, a negative jump component in realized volatility predicts an increase in future equity premia. Bollerslev et al. (2015) identify both left and right jump tail risks under the risk-neutral measure. They find that the left jump tail risk is an appropriate proxy for market fear. Additionally, they find that including a variance risk premium together with jump tail risk measures as predictors significantly improves market return forecasts. Finally, they show that jump risk helps explain the high–low book-to-market and winners-versus-losers portfolio returns.

Building on the above literature, we decompose jump variation into four distinct components depending on both the direction (upward and downward) and magnitude (small and large) of the jumps.[3] Specifically, we decompose individual stock jump (semi) variances into small and large components, using high-frequency intraday data. We then investigate the relationship

---

1 In 2017, Morningstar reported that this approach to investing has attracted over one trillion dollars in assets (see e.g., Jennifer Thompson, Financial Times, 27 December 2017).

2 In a related paper, Duong and Swanson (2015) construct both small and large jump measures based on some fixed truncation levels. They exploit the risk predictability of different jump measures using both index data and Dow 30 stocks and find that small jump variation has more volatility predictability than large jump variation.

3 The methods that we implement to separate jump variation use recent advances in financial econometrics due to Andersen et al. (2007, 2003), Jacod (2008), Mancini (2009), Barndorff-Nielsen et al. (2010), Todorov and Tauchen (2010), Aït-Sahalia and Jacod (2012), and Patton and Sheppard (2015). Most importantly, the reader is referred to Aït-Sahalia and Jacod (2012), who survey the methods used in this paper.

between these various jump measures and future returns, using single- and double-sorted stock portfolios, and Fama–MacBeth cross-sectional regression analysis. The reason that we decompose jump semi-variances into small and large components is that this decomposition allows us to explore the possibility that they contain different information relevant to investing and return predictability. As Maheu and McCurdy (2004) note, large jumps may reflect important individual stock and market news announcements. Smaller jumps (or continuous variations) may result from liquidity and strategic trading.

Of note is that in our setup, the number of large jumps is always finite, regardless of jump activity. However, there may exist infinitely many small jumps. This framework is surveyed in Aït-Sahalia and Jacod (2012), where cutoff (truncation) methods used for decomposing total jump variation into large and small jump components are outlined. We follow their approach when constructing large and small jump variation measures. Summarizing, we construct jump variation measures by estimating the difference between quadratic variation and power variation, as well as by decomposing quadratic variation into jump and non-jump components using cutoffs. The cutoff (truncation) values that we use are related to those discussed in Li et al. (2017), and in Aït-Sahalia and Jacod (2012). However, we are not interested in jump detection (or jump testing), per-se, in discussed in Bollerslev et al. (2008), but are instead interested in the estimation and analysis of jump variation measures. For this reason, our methodology most closely mirrors that discussed in Aït-Sahalia and Jacod (2012). In contexts where jump testing is of interest, many approaches are available. For example, one may rely on the conservative approach discussed in Bollerslev et al. (2008), or the false discovery control and spurious jump detection approach discussed in Bajgrowicz et al. (2016) and Scaillet et al. (2018). For further discussion of testing, the reader is referred to Lee and Mykland (2008, 2012), Barras et al. (2010), Bajgrowicz and Scaillet (2012), Christensen et al. (2014), and the references cited therein.[4]

Our key findings can be summarized as follows. First, we find that both small and large upside (downside) jump variation negatively (positively) predicts subsequent weekly returns. However, portfolios sorted using signed total jump variation are associated with increased average returns and risk-adjusted alphas for high–low portfolios, relative to the cases where upside or downside jump variation is sorted on. This finding is in accord with the findings of Bollerslev et al. (2020).

Our second finding involves the case where jump variation is further decomposed into "small" and "large" components. In this case, sorting on signed small jump variation leads to value-weighted high–low portfolios with greater average returns and alphas than when either signed total jump or signed large jump variation is sorted on. Indeed, when the truncation parameter used to differentiate small from large jumps is based on a 5 standard-deviation (i.e., $\alpha = 5$) cutoff, we find that average return spreads are 10% higher when signed small jump variation is sorted on rather than signed total jump variation. Moreover, these average return spreads are statistically significant in both cases. However, average return spreads are not significantly different from zero when signed large jump variation is sorted on. Indeed, including large jump variation is actually detrimental to predictive accuracy, as average returns and alphas for high–low portfolios actually decline relative to the case where total variation is instead used in our prediction experiments. These results suggest that there may be a "jump-threshold", beyond which large jump variation contains no marginal predictive ability,

[4] As noted by the editor, an important issue in the context of the estimation of jump variation components is the assessment of the importance of the cutoff methods used in this paper when disentangling small and large jump variations. Our approach to this issue in the sequel is to assess the robustness of our empirical findings to the use of various different cutoffs. However, Monte Carlo simulation may shed further light on the issue. In undertaking Monte Carlo simulations, one must carefully simulate both infinite-activity and finite-activity jump processes, consider various portfolios of simulated assets, and calibrate DGPs using models fitted to panels of high-frequency asset prices and returns. Although this topic is important, it is beyond the scope of the current paper, and is left to future research.

relative to that contained in small jump variation.[5] In summary, we find that large jump variation has little to no marginal predictive content, beyond a certain threshold.[6] Indeed, when the threshold is judiciously selected, one can actually improve predictive performance in our experiments, leading to increased high–low portfolio average returns and alphas, when sorting portfolios based on small jump variation rather than total jump variation.

To understand why small jumps matter in our analysis, note that Scaillet et al. (2018) document that jumps are frequent events and that jump dynamics are not consistent with the compound Poisson processes with constant intensity. In addition, many recent studies find that infinite-activity Levy jump specifications (e.g., Variance Gamma and Normal Inverse Gaussian return jumps) can improve the goodness of fit and option pricing performance of models (e.g., see Yang and Kanniainen (2016)). These results point to the possible importance of small jumps, which may have infinite numbers in a finite time interval. Though more frequent than large jumps, note that we estimate daily jump variation measures and aggregate them into weekly measures for forecasting future weekly returns. Thus, it is not individual small jumps that necessarily matter, but the overall contribution of small jumps to total variation, in the cross-section. Additionally, we show (see below discussion) that small jump variation is more closely associated with idiosyncratic risk than is large jump variation. If idiosyncratic volatility matters (see Ang et al. (2006, 2009)), so do small jumps, at least to some extent. Indeed, our results partially help explain the idiosyncratic volatility puzzle (investors prefer positive skewness/asymmetry, so they will accept lower returns for stocks with higher probabilities of having positive jumps).

Third, industry double-sorts indicate that the benefit of small signed jump variation investing is driven by stock selection within an industry, rather than industry bets. Investors prefer stocks with a high probability of having positive jumps, but they also tend to overweight safer industries.

Fourth, the reason small and large (signed) jump variation measures have differing marginal predictive content for returns is associated with the importance of realized skewness as a control variable in our experiments. Specifically, we find that in double-sorted portfolios, the content of signed large jump variation is negligible when controlling for either signed total jump variation or realized skewness. By contrast, signed small jump variation has unique information for predicting future returns, even when controlling for total jump variation or realized skewness. This finding is consistent with the results from a series of Fama–MacBeth regressions, in which we control for multiple firm characteristics and risk measures.

Finally, small and large jump variation measures are driven by different economic factors and contain different information for predicting future returns. For example, large jumps are closely associated with "big" news. In particular, large earning announcement surprises increase both the magnitude and occurrence of large jumps. While such news-related information is embedded in large jump variation, the information is generally short-lived, and dissipates too quickly to provide marginal predictive content for subsequent weekly returns. This is consistent with our finding that filtering out signed small jump variation (which we know to be useful) from signed total jump variation results

---

[5] When equal-weighted portfolios are instead examined, sorting on total jump variation yields higher average returns and alphas than when sorting on small or large jump variation. However, deeper inspection of our tabulated results in this case reveals that average returns associated with large jump variation sorts are much smaller—around 1/2 of the magnitude of those associated with small and total jump variation sorts. Also, the magnitude of average returns associated with small jump variation sorts is much closer (within 10%) to the average returns associated with total jump variation sorts, when our truncation parameter uses a 5 standard-deviation cutoff instead of a 4 standard-deviation cutoff. This suggests that the jump-threshold differs depending upon portfolio type, and indicates that our findings based on equal-weighted portfolios are largely in accord with the findings elucidated above.

[6] This result is not in contradiction with the extant literature on the importance of large jumps. This is because all our conclusions are for the cross-section (our paper is the first one that investigates the return predictability of large/small jump variation measured in the cross-section). It is true that large jumps have been shown to have return/variance predictability for individual stocks/portfolios. But in the cross-section, we do not observe such return predictability. In this cross-section, this is simply because large jumps are rare, so that they provide little (or at least infrequent) information for future weekly returns, in the cross-section.

in increased predictive ability, relative to the case where only signed total jump variation is used in return forecasting, particularly for big firms. Additionally, this finding is interesting, given that the comparison of aggregated and weighted jump variation measures indicates that small jump variation captures idiosyncratic risk. More specifically, we find that small jumps are more likely to be diversified away than large jumps, thus tend to be more closely associated with idiosyncratic risks, and are therefore more likely to be driven by liquidity conditions and trading activity.[7]

In closing, it is worth stressing that the magnitudes of price movements depend on the magnitude of news (if they are related to news), and on the underlying process describing price evolution, and thus movements may manifest in the form of large/small jumps and/or price drifts. Needless to say, the process associated with how firms and markets digest information is critical to understanding and quantifying what sorts of price movements ensue. For any given news, responses may differ in terms of upside or downside price drifts, or large/small jumps. In this sense, both small and large jumps may matter, as shown in our empirical results, in the cross-section. Also, note that idiosyncratic volatility, one of the most studied asset pricing anomalies, is one of the most successful predictors for stock returns in the cross-section (see Gu et al. (2019)). As we show that small jumps are more likely to be associated with idiosyncratic risk, and since signed small jump variation has a negative relationship with future stock returns, our results help explain the idiosyncratic volatility puzzle. This is because of agent's preference for lottery-like returns (i.e., investors will accept lower returns for stocks with high probabilities of having positive jumps).

The rest of this paper is organized as follows. In Section 2 we discuss the model setup and define the jump risk measures that we use. Section 3 contains a discussion of the data used in our empirical analysis, and highlights key summary statistics taken from our dataset. Section 4 presents our main empirical findings, including discussions of results based on single portfolio sorts, double-sorts, cumulative return and Sharpe ratio analysis, firm-level Fama–MacBeth regressions, and finally, jumps and news announcements. Section 5 concludes.

2. Model Setup and Estimation Methodology

Following Aït-Sahalia and Jacod (2012), assume that the log price, X_t, of a security follows an Itô semimartingale, formally defined as:

$$X_t = X_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + \int_0^t \int_{\{|x| \leq \gamma\}} x(\mu - \nu)(ds, dx) + \int_0^t \int_{\{|x| \geq \gamma\}} x\mu(ds, dx),$$

where b and σ denote the drift and diffusive volatility processes, respectively; W is a standard Brownian motion; μ is a random positive measure with its compensator ν; and γ is the (arbitrary) cutoff level (threshold) used to distinguish between small and large jumps. As pointed out in Aït-Sahalia and Jacod (2012), the continuous part of this model (i.e., the $\int_0^t \sigma_s dW_s$ term) captures the normal hedgeable risk of the asset. The large jumps part of the model (i.e., the $\int_0^t \int_{\{|x| \geq \gamma\}} x\mu(ds, dx)$ term) might capture big news-related events such as default risk, and the small jumps part of the model (i.e., the $\int_0^t \int_{\{|x| \leq \gamma\}} x(\mu - \nu)(ds, dx)$ term) might arise due to other information flows, as discussed below. For cases where jumps are summable (e.g., when jumps have finite activity, so that

[7] This result is consistent with the finding of Amaya et al. (2015) that preference for positive asymmetry (skewness) may partially explain the idiosyncratic volatility puzzle, especially for small firms.

$\sum_{s\leq t}\Delta X_s < \infty$, for all t), then the size of a jump at time s is defined as $\Delta X_s = X_s - X_{s-}$.[8] In this context, the "true" price of risk is often defined by the quadratic variation, QV_t, of the process X_t. Specifically,

$$QV_t = \int_0^t \sigma_s^2 ds + \sum_{s\leq t} \Delta X_s^2,$$

where the variation of the continuous component (i.e., the integrated volatility) is given by $IV_t = \int_0^t \sigma_s^2 ds$, and the variation of the price jump component is given by $QJ_t = \sum_{s\leq t}\Delta X_s^2$.

In the sequel, intraday stock returns are assumed to be observed over equally spaced time intervals in a given day, where the sampling interval is denoted by Δ_n, and the number of intraday observations is n. Thus, the intraday log-return over the ith interval is defined as

$$r_{i,t} = X_{i\Delta_n,t} - X_{(i-1)\Delta_n,t}.$$

It is well known that when the sampling interval goes to zero, the realized volatility, RV_t, which is calculated by summing up all successive intraday squared returns, converges to QV_t, as $n \to \infty$, where

$$RV_t = \sum_{i=1}^{n} r_{i,t}^2 \to_u QV_t = IV_t + QJ_t,$$

where $\to_u$ denotes convergence in probability, uniformly in time.

To separate jump variation from integrated volatility, Andersen et al. (2007) show that the jump and continuous components of realized variance can be constructed as:

$$RVJ_t = \max(RV_t - \widehat{IV}_t, 0)$$

and

$$RVC_t = RV_t - RVJ_t,$$

respectively, where $\widehat{IV}_t$ is an estimator of $\int_0^t \sigma_s^2 ds$. Following Barndorff-Nielsen and Shephard (2004) and Barndorff-Nielsen et al. (2006), we use tripower variation to estimate the integrated volatility. In particular, define

$$\widehat{IV}_t = V_{\frac{2}{3},\frac{2}{3},\frac{2}{3}}\mu_{\frac{2}{3}}^{-3},$$

where $\mu_q = E(|Z|^q)$ is the qth absolute moment of the standard normal distribution, and

$$V_{m_1,m_2,\ldots m_k} = \sum_{i=k}^{n} |r_{i,t}|^{m_1}|r_{i-1,t}|^{m_2}\ldots|r_{i-k+1,t}|^{m_k},$$

where $m_1, m_2 \ldots m_k$ are positive, such that $\sum_1^k m_i = q$.

Now, again following Aït-Sahalia and Jacod (2012), we separate total jump variation into small and large variation measures, using various truncation levels, γ (see Duong and Swanson (2011, 2015) for additional details). In particular, define realized small and large jump variation as follows:

$$RVLJ_{\gamma,t} = \min(RVJ_t, \sum_{i=1}^{n} r_{i,t}^2 I_{\{|r_{i,t}|\geq\gamma\}})$$

and

$$RVSJ_{\gamma,t} = RVJ_t - RVLJ_{\gamma,t},$$

[8] A jump process has finite activity when it makes a finite number of jumps, almost surely, in each finite time interval, otherwise it is said to have infinite activity.

respectively, where $I(\cdot)$ denotes the indicator function, which equals one if the absolute return is larger than the truncation level, and is otherwise equal to zero. We are also interested in upside and downside variation measures associated with positive and negative returns. Following Barndorff-Nielsen et al. (2010), we construct realized semi-variances, defined as: $RS_t^+ = \sum_{i=1}^n r_{i,t}^2 I_{\{r_{i,t}>0\}}$, $RS_t^- = \sum_{i=1}^n r_{i,t}^2 I_{\{r_{i,t}<0\}}$, and $RV_t = RS_t^+ + RS_t^-$. These authors show that the above upside and downside semi-variances (RS_t^+ and RS_t^-, respectively) each converge to the sum of one half of the integrated volatility and the corresponding signed jump variation. Specifically,

$$RS_t^+ \rightarrow_u \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{s\leq t} \Delta X_s^2 I_{\{\Delta X_s>0\}}$$

and

$$RS_t^- \rightarrow_u \frac{1}{2}\int_0^t \sigma_s^2 ds + \sum_{s\leq t} \Delta X_s^2 I_{\{\Delta X_s<0\}}.$$

We construct upside and downside jump variation as follows:

$$RVJP_t = \max(RS_t^+ - \frac{1}{2}\widehat{IV}_t, 0) \tag{1}$$

and

$$RVJN_t = \max(RS_t^- - \frac{1}{2}\widehat{IV}_t, 0). \tag{2}$$

In addition, signed jump variation can be calculated as the difference between these upside and downside jump measures, as follows:

$$SRVJ_t = RVJP_t - RVJN_t. \tag{3}$$

This measure captures asymmetry in upside and downside jump variation.

In our analysis, we decompose the above upside and downside jump variation into small and large components, again using the cutoff (truncation level) discussed in Aït-Sahalia and Jacod (2012). For further related discussion, see Mancini (2009), Duong and Swanson (2015), and Li et al. (2017). In particular, upside and downside large jump variations based on γ are defined as follows:

$$RVLJP_{\gamma,t} = \min(RVJP_t, \sum_{i=1}^n r_{i,t}^2 I_{\{r_{i,t}>\gamma\}}) \tag{4}$$

The corresponding downside measure is defined as follows:

$$RVLJN_{\gamma,t} = \min(RVJN_t, \sum_{i=1}^n r_{i,t}^2 I_{\{r_{i,t}<-\gamma\}}). \tag{5}$$

In all of our analyses, we construct γ by estimating $\alpha\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$, thus accounting for the time-varying diffusive spot volatility of different stocks in the cross-section.[9] In the sequel, we consider

[9] As an example of how our cutoff level compares with others used in the literature, consider Li et al. (2017). These authors use bipower variation as the fixed value for $\widehat{IV}_t^{(i)}$. We instead use bipower variation as the initial value for the integrated volatility $\widehat{IV}_t^{(0)}$, say, and $\widehat{IV}_t^{(i)}$ is estimated using truncated bipower variation with threshold $\gamma^{(i-1)}$, say, where $\gamma^{(i-1)}$ is fixed only when $|\widehat{IV}_t^{(i)} - \widehat{IV}_t^{(i-1)}|$ is smaller than 5% $\times$ $\widehat{IV}_t^{(i-1)}$.

three values for γ, say γ^1 (with $\alpha = 4$), γ^2 (with $\alpha = 5$), and γ^3 (with $\alpha = 6$). Our signed large jump variation (i.e., large jump asymmetry) measures are defined as follows:

$$SRVLJ_t = RVLJP_t - RVLJN_t. \tag{6}$$

Our corresponding small jump variation measures are defined as the difference between total and large jump variation. Specifically,

$$RVSJP_t = RVJP_t - RVLJP_t \tag{7}$$

and

$$RVSJN_t = RVJN_t - RVLJN_t. \tag{8}$$

Additionally, signed small jump variation is defined as:

$$SRVSJ_t = RVSJP_t - RVSJN_t. \tag{9}$$

To analyze the predictability of various jump measures in the cross-section, we normalize each of the jump variation measures discussed above by total realized variation.

An alternative to our approach for calculating the upside and downside jump variation measures in (1) and (2) is to use truncated realized variation (TRV) instead of tripower variation as a consistent estimator of integrated volatility, where $TRV_t = \sum_{i=1}^{n} r_{i,t}^2 I_{\{|r_{i,t}| \leq \alpha_n\}} \rightarrow_u IV_t = \int_0^t \sigma_s^2 ds$. Upside and downside jump variation can then be calculated using:

$$RVJP_t = RS_t^+ - \sum_{i=1}^{n} r_{i,t}^2 I_{\{0 < r_{i,t} \leq \alpha_n\}} \tag{10}$$

and

$$RVJN_t = RS_t^- - \sum_{i=1}^{n} r_{i,t}^2 I_{\{-\alpha_n \leq r_{i,t} < 0\}}, \tag{11}$$

where α_n is the truncation level. Please note that this truncation level is a generic "jump detection" truncation level, whereas γ is a truncation level used to disentangle large and small jump variations from total jump variation.[10] Our empirical findings based on the use of (10) and (11) to define $RVJP_t$ and $RVJN_t$ are qualitatively the same as those reported in Section 4 based on the use of (1) and (2).

To measure skewness and kurtosis, we construct higher order realized return moments. Following Amaya et al. (2015), standardized daily skewness is defined as:

$$RSK_t = \frac{\sqrt{n} \sum_{i=1}^{n} r_{i,t}^3}{RV_t^{\frac{3}{2}}}, \tag{12}$$

and normalized daily realized kurtosis is defined as:

$$RKT_t = \frac{n \sum_{i=1}^{n} r_{i,t}^4}{RV_t^2}. \tag{13}$$

Finally, it should be noted that we follow Amaya et al. (2015) and Bollerslev et al. (2020), and conduct our cross-sectional analysis at the weekly frequency. In particular, on each Tuesday, we compute the following weekly realized measures: $RV_t^W = (\frac{252}{5} \sum_{i=0}^{4} RV_{t-i})^{1/2}$ and $RM_t^W = \frac{1}{5}(\sum_{i=0}^{4} RM_{t-i})$, where RV_t is defined above, and where RM_t denotes any of the realized

[10] Here, the threshold, $\alpha_n = 3\sqrt{\frac{1}{l} \widehat{IV}_t^{(i)}} \Delta_n^{0.49}$, is estimated using the same procedure as in footnote 9.

measures defined above other than RV_t (e.g., $RVJP_t$, $RVJN_t$, $SRVJ_t$, etc.). Hereafter, we shall drop the superscript "W" for the sake of notational brevity. All the descriptors used to denote the various realized measures constructed in our empirical analysis are summarized in Table 1.

As described in detail in Section 4, the realized measures outlined above are used in several different ways in our empirical analysis. First, we carry out single portfolio sorts, in which we sort stock portfolios on the above realized jump measures, and predict average excess returns, one week ahead. In these experiments, we also calculate alphas based on regressions that use the Fama–French and Carhart factors. In this first part of our analysis, we also examine cumulative returns and Sharpe ratios. In addition to the single portfolio sorts, we carry out double portfolio sorts, in which we sort not only on realized jump risk measures, but also on various control variables, including realized skewness and other firm specific characteristics. Using these double-sorts, we also examine the inter-play between individual stock-level jump variations and industry-level jump variations. Needless to say, the purpose of our double-sorts is to examine the robustness of our findings based on single sorts, after controlling for other realized measures. Next, we carry out a series of Fama–MacBeth regressions, in order to check the robustness of our findings to the inclusion of various firm specific characteristics. Finally, we carry out an event study in which the effect of earning surprises on realized jump measures is examined. For complete details, see Section 4.

Table 1. Realized Measures and Firm Characteristics.

Panel A: Realized Measures Used in Portfolio Sorts and Fama–MacBeth Regressions	
RVJP	Positive (upside) jump variation, see (1).
RVJN	Negative (downside) jump variation, see (2).
SRVJ	Signed jump variation, $RVJP - RVJN$, see (3).
RVLJP	Positive (upside) large jump variation, see (4).
RVLJN	Negative (downside) large jump variation, see (5).
SRVLJ	Signed large jump variation, $RVLJP - RVLJN$, see (6).
RVSJP	Positive (upside) small jump variation, see (7).
RVSJN	Negative (downside) small jump variation, see (8).
SRVSJ	Signed small jump variation, $RVSJP - RVSJN$, see (9).
RVOL	Realized volatility
RSK	Realized skewness, see (12).
RKT	Realized kurtosis, see (13).
Panel B: Explanatory Variables and Firm Characteristics Used in Fama–MacBeth Regressions	
BETA	Market beta
log(Size)	Natural logarithm of firm size
BEME	Book-to-market ratio
MOM	Momentum
REV	Short-term reversal
IVOL	Idiosyncratic volatility
CSK	Co-skewness
CKT	Co-kurtosis
MAX	Maximum daily return
MIN	Minimum daily return
ILLIQ	Illiquidity

Notes: The realized measures listed in Panel A of this table are defined and discussed in Section 2. For detailed descriptions of the explanatory variables and firm characteristics listed in Panel B of this table, refer to Bollerslev et al. (2020), and the references cited therein.

3. Data

We use high-frequency trading data obtained from the consolidated Trade and Quote (TAQ) database. In particular, we analyze all stocks in the TAQ database that are listed on the NYSE, Amex, and NASDAQ stock exchanges. There are 15,585 unique stocks during the 1246 weeks analyzed in this

paper.[11] The sample period is from 4 January 1993 to 31 December 2016. Intraday prices are sampled at 5-min intervals from 9:30 a.m. to 4:00 p.m. from Monday to Friday. Overnight returns are not considered in this paper, and days with less than 80 transactions at a 5-min frequency are eliminated. All high-frequency data used in this paper are cleaned to remove trades outside of exchange hours, negative or zero prices or volumes, trade corrections and non-standard sale conditions, using the methodology described in Appendix A.1 in Bollerslev et al. (2020).

We constructed two variants of our dataset. The first is cleaned as discussed above. The second classifies 5-min intraday returns greater than 15% as abnormal and replaces them with zeros. In the sequel, results based on analysis of the second dataset are reported. However, results based on the use of the first dataset are qualitatively the same; and indeed key return results reported in this paper generally change by 1 basis point or less when the former dataset is used in our analysis. Complete results are available upon request from the authors.

Daily and monthly returns, and adjusted numbers of shares for individual securities are collected from the CRSP database. Delisting returns in CRSP are used as returns after the last trading day. Daily Fama–French and Carhart four-factor (FFC) portfolio returns are obtained from Kenneth R. French's website.

Following Amaya et al. (2015) and Bollerslev et al. (2020), we also construct various lower frequency firm-level variables that might be related to future returns, such as the market beta (BETA), the firm size, the book-to-market ratio (BEME), momentum (MOM), short-term reversals (REV), idiosyncratic volatility (IVOL), co-skewness (CSK), co-kurtosis (CKT), maximum (MAX) and minimum (MIN) daily return in the previous week, and the Amihud (2002) illiquidity measure (ILLIQ). For a complete list of these firm specific control variables, refer to Table 1. For a detailed description of these variables, including the methodology used to construct them, see Appendix A.2 in Bollerslev et al. (2020).

Please note that while the majority of our analysis is based on the examination of individual stocks, in our double-sorts, there are some cases (that are reported in Section 4.4) where we examine the inter-play between individual stock-level jump variations and industry-level jump variations, as mentioned above. In this case, we follow the Fama–French industry classification approach, and group stocks into 49 industries based on their SIC codes, which are obtained from CRSP.

3.1. Unconditional Distributions of Realized Measures

Figure 1 displays kernel density estimates of the unconditional distributions of each of our realized measures, across all firms and weeks. The top two panels in the figure show the distributions of signed jump variation and realized skewness. Both distributions are approximately symmetric and peaked around zero. The skewness distribution is more fat-tailed, however.[12] The middle two panels of Figure 1 display the distributions of signed small and large jump variation. Similar to signed jump variation, both signed small and large jump variation measures are approximately symmetric around zero, but signed small jump variation is less fat-tailed.[13] Consistent with the results in Amaya et al. (2015) and Bollerslev et al. (2020), realized volatility and realized kurtosis are both right skewed and very fat-tailed, as shown in the bottom two panels of the figure.[14]

Figure 2 shows the time variation in the cross-sectional distribution of each realized measure using 10-week moving averages. In particular, 10th, 50th, and 90th percentiles for each realized measure in the cross-section are plotted. Thus, dispersion at any given time in these plots reflects

---

[11] In some cases, multiple TAQ symbols are matched with a unique Center for Research in Security Prices (CRSP) PERMNO. Over each quarter, the TAQ symbol which has the most observations is kept and the other overlapping observations are dropped.

[12] The kurtosis of signed jump variation is 4.36. For realized skewness, the analogous statistic is 12.04.

[13] The unconditional kurtosis is 6.43 and 3.51, for signed small and large jump variation, based on truncation level γ^1; and 8.87 and 3.09 based on truncation level γ^2, respectively.

[14] The kurtosis is 15.85 and 27.24, for the unconditional distribution of realized volatility and realized kurtosis, respectively.

information about the cross-sectional distribution of the realized measure. Inspection of Panels A and B in the figure reveals that signed jump variation and realized skewness have stable dispersion, for all three cross-sectional percentiles, over time, while the cross-sectional dispersion in realized volatility and kurtosis are rather time-dependent (see Panels C and D). Additionally, similar to the cross-sectional distribution of signed jump variation, the percentiles for signed small and large jump variation measures are quite steady over time, as indicated in Panels E–H.

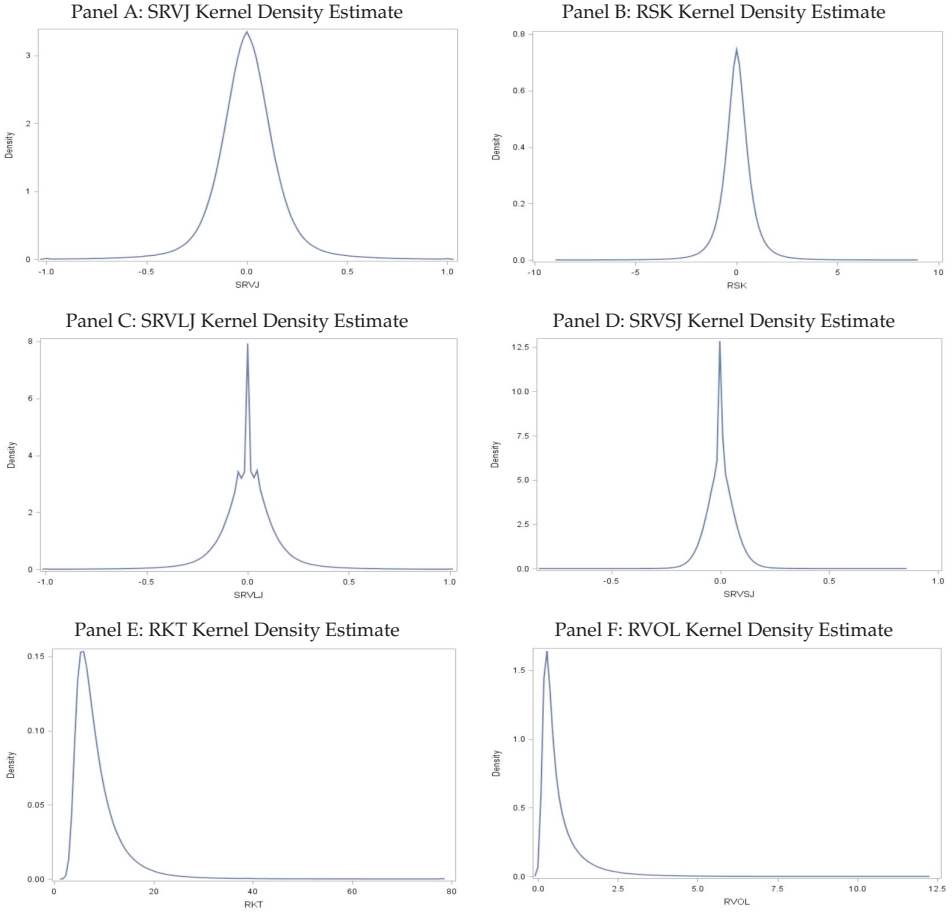


Figure 1. Unconditional Distributions of Realized Measures. Panels A-F display unconditional distribution kernel density estimates of various realized measures, for the cross-section of stock returns for the period January 1993 to December 2016. Signed small and large jump variation measures are constructed using truncation levels $\gamma^1 = 4\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$. Distributions are similar when using $\gamma^2 = 5\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$.

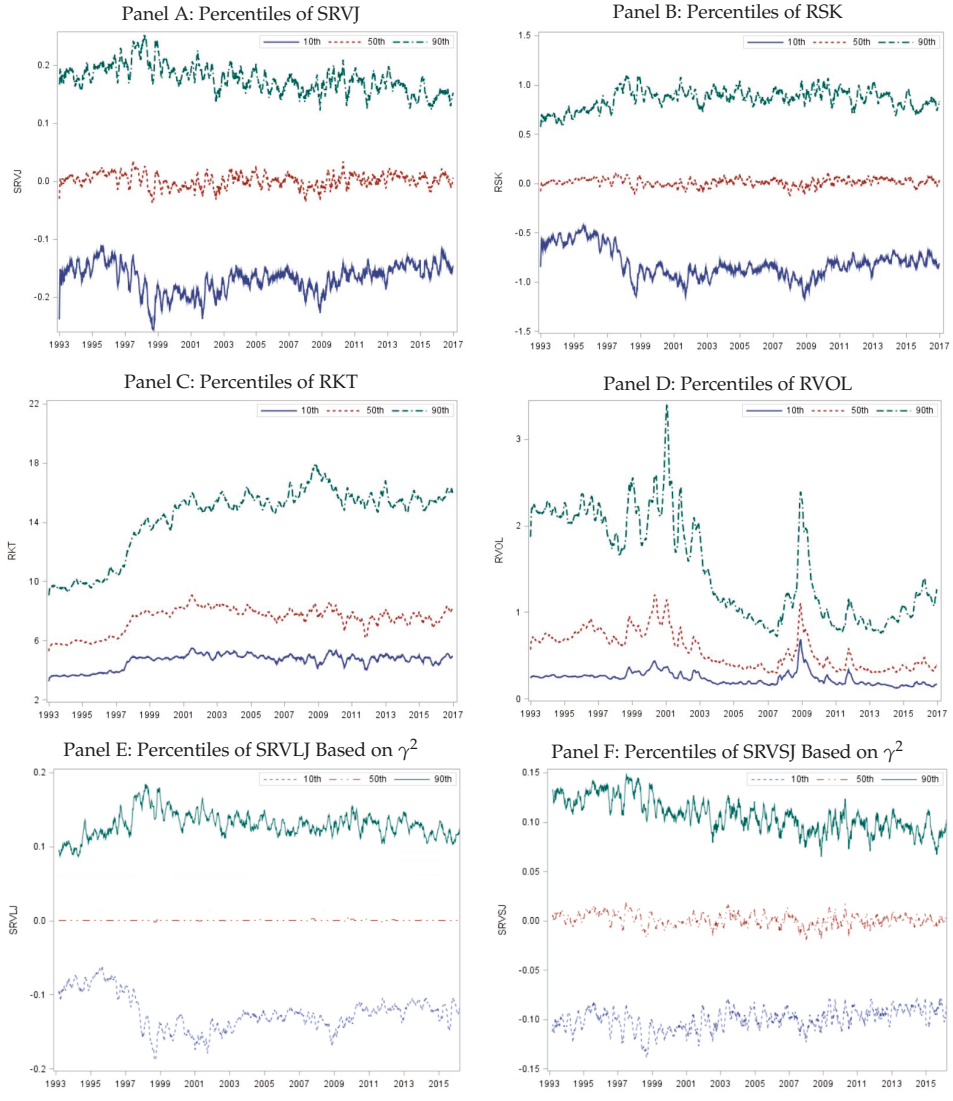


Figure 2. Percentiles of Realized Measures. Panels A-H display 10-week moving averages of percentiles of realized measures, for the cross-section of stocks, for the period January 1993 to December 2016. Signed small and large jump variation measures are contructed based on jump truncation level $\gamma^2 = 5\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$.

3.2. Summary Statistics and Portfolio Characteristics

Table 2 contains various summary statistics for all the realized measures summarized in Table 1. In Panel A, the cross-sectional means and standard errors for each of the realized measures is given. This is done for two different truncation levels, denoted as $\gamma^1 = 4\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$ and $\gamma^2 = 5\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$. As might be expected, jump variation is quite sensitive to the choice of γ.

For example, the (normalized) mean of RVSJP (positive (upside) small jump variation) increases from 0.1180 to 0.1715 when the threshold is increased from γ^1 to γ^2.

Panel B of Table 2 contains cross-sectional correlations for all the realized measures. In accord with the findings reported by Amaya et al. (2015) and Bollerslev et al. (2020), signed jump variation (SRVJ) and realized skewness (RSK) are highly correlated with each other and have significantly positive correlations with the short-term reversal variable (REV); as well as with maximum (MAX) and minimum (MIN) daily returns in the previous week.

Interestingly, we also find that signed large jump variation (SRVLJ) is highly correlated with SRVJ and with RSK. However, signed small jump variation (SRVSJ) has a lower correlation with SRVJ and much smaller positive correlation with RSK. This finding is consistent with our finding discussed below that realized skewness captures information that is primarily contained in large jumps; and serves as an important distinction between the findings in this paper and those reported in the papers discussed above.

Table 3 complements Table 2 by sorting stocks into quintile portfolios based on different realized measures. On each Tuesday, stocks are ranked by the realized variation measures, and we calculate the equal-weighted averages of each firm characteristic in the same week. Panels A, B, C and E report summary statistics for portfolios sorted by SRVJ, SRVLJ, SRVSJ, and RSK, respectively. Consistent with the correlations contained in Table 2, firms with larger signed small and large jump variation measures tend to have higher signed jump variation, realized skewness, REV, MAX and MIN. Firms with high realized volatility and realized kurtosis (see Panels D and F) tend to be illiquid and small.[15]

[15] See the Supplementary Appendix for results based on the examination of additional quintile portfolios that are constructed based on ex-ante risk measures and displayed with ex-post risk measures. It is clear that sorting stocks based on jump risk measures results in portfolios with the desired risk exposures.

Table 2. Summary Statistics for Various Realized Measures and Firm Characteristics Based on Two Jump Truncation Levels.

Panel A: Cross-Sectional Summary Statistics

	SRVJ	RVJP	RVJN	SRVLJ	RVLJP	RVLJN	SRVSJ	RVSJP	RVSJN	RVOL	RSK	RKT	BETA	log(Size)	BEME	MOM	REV	IVOL	CSK	CKT	MAX	MIN	ILLIQ
Part I: Jump Truncation Level = γ^1																							
Mean	0.0061	0.2698	0.2637	0.0045	0.1518	0.1472	0.0015	0.1180	0.1165	0.9489	0.0288	8.2569	1.0794	6.5280	0.5969	2023.8456	70.6077	0.0293	−0.0263	1.1438	412.1094	−346.7608	−5.2826
Std	0.1537	0.1350	0.1347	0.1424	0.1555	0.1523	0.0635	0.0783	0.0783	2.1211	0.8159	4.5706	0.5566	1.8359	0.7224	7464.5273	927.3551	0.0250	0.3283	0.8474	572.1454	359.6789	2.4047
Part II: Jump Truncation Level = γ^2																							
Mean	0.0061	0.2698	0.2637	0.0029	0.0983	0.0954	0.0031	0.1715	0.1684	0.9489	0.0288	8.2569	1.0794	6.5280	0.5969	2023.8456	70.6077	0.0293	−0.0263	1.1438	412.1094	−346.7608	−5.2826
Std	0.1537	0.1350	0.1347	0.1303	0.1401	0.1368	0.0859	0.0911	0.0909	2.1211	0.8159	4.5706	0.5566	1.8359	0.7224	7464.5273	927.3551	0.0250	0.3283	0.8474	572.1454	359.6789	2.4047

Panel B: Cross-Sectional Correlations

	SRVJ	RVJP	RVJN	SRVLJ	RVLJP	RVLJN	SRVSJ	RVSJP	RVSJN	RVOL	RSK	RKT	BETA	log(Size)	BEME	MOM	REV	IVOL	CSK	CKT	MAX	MIN	ILLIQ
Part I: Jump Truncation Level = γ^1																							
SRVJ	1.00	0.57	−0.57	0.91	0.43	−0.40	0.37	0.13	−0.18	−0.02	0.94	0.03	−0.03	0.01	0.01	0.01	0.30	−0.03	0.09	0.00	0.17	0.22	0.00
RVJP		1.00	0.33	0.52	0.85	0.37	0.21	0.04	−0.13	0.22	0.54	0.45	−0.26	−0.49	0.14	−0.10	0.15	0.12	0.04	−0.24	0.15	0.06	0.56
RVJN			1.00	−0.52	0.35	0.84	−0.22	−0.10	0.08	0.24	−0.54	0.41	−0.23	−0.49	0.13	−0.11	−0.19	0.15	−0.06	−0.24	−0.05	−0.20	0.55
SRVLJ				1.00	0.48	−0.44	−0.04	−0.05	−0.01	−0.01	0.92	0.03	−0.02	0.00	0.01	0.00	0.20	−0.02	0.05	0.00	0.12	0.16	0.00
RVLJP					1.00	0.57	−0.02	−0.46	−0.45	0.23	0.44	0.61	−0.25	−0.47	0.12	−0.06	0.09	0.13	0.02	−0.24	0.12	−0.02	0.54
RVLJN						1.00	0.01	−0.44	−0.45	0.24	−0.41	0.59	−0.23	−0.48	0.11	−0.06	−0.10	0.15	−0.03	−0.24	0.01	−0.17	0.54
SRVSJ							1.00	0.42	−0.42	−0.02	0.19	0.00	−0.03	0.01	0.01	0.01	0.26	−0.03	0.08	0.00	0.14	0.18	0.00
RVSJP								1.00	0.64	−0.04	0.06	−0.40	0.06	0.03	0.01	−0.05	0.10	−0.03	0.04	0.04	0.02	0.11	−0.06
RVSJN									1.00	−0.03	−0.10	−0.40	0.08	0.02	0.01	−0.06	−0.12	−0.01	−0.03	0.03	−0.09	−0.04	−0.06
RVOL										1.00	−0.01	0.22	−0.05	−0.55	0.08	−0.12	0.06	0.56	−0.01	−0.27	0.44	−0.47	0.56
RSK											1.00	0.04	−0.02	0.00	0.01	0.00	0.22	−0.02	0.06	0.00	0.13	0.17	0.00
RKT												1.00	−0.20	−0.34	0.09	−0.02	0.00	0.10	−0.01	−0.19	0.08	−0.10	0.40
BETA													1.00	0.10	−0.09	0.00	−0.04	0.06	0.01	0.30	0.03	−0.09	−0.16
ME														1.00	−0.19	0.11	−0.05	−0.52	0.01	0.40	−0.32	0.35	−0.93
BEME															1.00	0.03	0.02	0.05	0.00	−0.06	0.05	−0.03	0.18
MOM																1.00	0.00	−0.08	−0.07	0.06	−0.05	0.05	−0.15
REV																	1.00	0.12	0.16	−0.04	0.49	0.29	0.05
IVOL																		1.00	0.02	−0.35	0.50	−0.47	0.47
CSK																			1.00	0.01	0.07	0.07	0.00
CKT																				1.00	−0.16	0.15	−0.37
MAX																					1.00	−0.28	0.34
MIN																						1.00	−0.35
ILLIQ																							1.00

	SRVJ	RVJP	RVJN	SRVLJ	RVLJP	RVLJN	SRVSJ	RVSJP	RVSJN	RVOL	RSK	RKT	BETA	log(Size)	BEME	MOM	REV	IVOL	CSK	CKT	MAX	MIN	ILLIQ
Part II: Jump Truncation Level = γ^2																							
SRVJ	1.00	0.57	−0.57	0.83	0.40	−0.37	0.52	0.23	−0.27	−0.02	0.94	0.03	−0.03	0.01	0.01	0.01	0.30	−0.03	0.09	0.00	0.17	0.22	0.00
RVJP		1.00	0.33	0.48	0.77	0.33	0.30	0.30	0.01	0.22	0.54	0.45	−0.26	−0.49	0.14	−0.10	0.15	0.12	0.04	−0.24	0.15	0.06	0.56
RVJN			1.00	−0.47	0.31	0.75	−0.31	0.03	0.33	0.24	−0.54	0.41	−0.23	−0.49	0.13	−0.11	−0.19	0.15	−0.06	−0.24	−0.05	−0.20	0.55
SRVLJ				1.00	0.49	−0.44	−0.04	−0.05	−0.01	−0.01	0.89	0.04	−0.01	0.00	0.01	0.00	0.16	−0.02	0.04	0.00	0.09	0.13	0.00
RVLJP					1.00	0.56	−0.02	−0.36	−0.34	0.20	0.43	0.64	−0.24	−0.40	0.11	−0.05	0.06	0.11	0.01	−0.22	0.10	−0.02	0.47
RVLJN						1.00	0.01	−0.34	−0.35	0.21	−0.40	0.62	−0.23	−0.41	0.10	−0.06	−0.09	0.13	−0.03	−0.22	0.01	−0.14	0.47
SRVSJ							1.00	0.47	−0.47	−0.02	0.32	0.00	−0.03	0.01	0.01	0.01	0.30	−0.03	0.09	0.00	0.16	0.21	0.00
RVSJP								1.00	0.55	0.03	0.13	−0.30	−0.02	−0.14	0.04	−0.07	0.13	0.02	0.04	−0.04	0.07	0.10	0.13
RVSJN									1.00	0.05	−0.17	−0.30	0.01	−0.15	0.04	−0.07	−0.15	0.05	−0.04	−0.04	−0.08	−0.09	0.13
RVOL										1.00	−0.01	0.22	−0.05	−0.55	0.08	−0.12	0.06	0.56	−0.01	−0.27	0.44	−0.47	0.56
RSK											1.00	0.04	−0.02	0.00	0.01	0.00	0.22	−0.02	0.06	0.00	0.13	0.17	0.00
RKT												1.00	−0.20	−0.34	0.09	−0.02	0.00	0.10	−0.01	−0.19	0.08	−0.10	0.40

Table 2. *Cont.*

Panel B: Cross-Sectional Correlations																							
	SRVJ	**RVJP**	**RVJN**	**SRVLJ**	**RVLJP**	**RVLJN**	**SRVSJ**	**RVSJP**	**RVSJN**	**RVOL**	**RSK**	**RKT**	**BETA**	**log(Size)**	**BEME**	**MOM**	**REV**	**IVOL**	**CSK**	**CKT**	**MAX**	**MIN**	**ILLIQ**
Part II: Jump Truncation Level = γ^2																							
BETA													1.00	0.10	−0.09	0.00	−0.04	0.06	0.01	0.30	0.03	−0.09	−0.16
ME														1.00	−0.19	0.11	−0.05	−0.52	0.01	0.40	−0.32	0.35	−0.93
BEME															1.00	0.03	0.02	0.05	0.00	−0.06	0.05	−0.03	0.18
MOM																1.00	0.00	−0.08	−0.07	0.06	−0.05	0.05	−0.15
REV																	1.00	0.12	0.16	−0.04	0.49	0.29	0.05
IVOL																		1.00	0.02	−0.35	0.50	−0.47	0.47
CSK																			1.00	0.01	0.07	0.07	0.00
CKT																				1.00	−0.16	0.15	−0.37
MAX																					1.00	−0.28	0.34
MIN																						1.00	−0.35
ILLIQ																							1.00

Notes: See notes to Table 1. This table presents cross-sectional summary statistics and correlations for all realized measures and control variables based on two truncation levels: $\gamma^1 = 4\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$ and $\gamma^2 = 5\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$. The entries in the table for realized measures (see columns 2–13) are constructed using 5-min intraday high-frequency data. Entries for firm characteristics (see columns 14–24) are constructed using daily data, with the exception of BEME, which is constructed using monthly data. For complete details, see Sections 3 and 4.

Table 3. Realized Measures and Firm Characteristics of Portfolios Sorted by Various Realized Measures.

Panel A: Stocks Sorted by SRVJ																							
Quintile	**RVJP**	**RVJN**	**RVLJP**	**RVLJN**	**RVSJP**	**RVSJN**	**SRVLJ**	**SRVSJ**	**SRVJ**	**RVOL**	**RSK**	**RKT**	**BETA**	**log(Size)**	**BEME**	**MOM**	**REV**	**IVOL**	**CSK**	**CKT**	**MAX**	**MIN**	**ILLIQ**
1	0.2021	0.3959	0.1161	0.2723	0.0860	0.1235	−0.1563	−0.0375	−0.1938	0.9394	−0.9324	9.8720	1.0369	6.1326	0.6235	0.2006	−0.0363	0.0317	−0.0708	1.0482	0.0295	−0.0494	−4.6903
2	0.2200	0.2777	0.1015	0.1391	0.1185	0.1385	−0.0376	−0.0200	−0.0576	0.9513	−0.2504	7.1729	1.1351	6.7041	0.5711	0.2051	−0.0111	0.0293	−0.0443	1.1987	0.0346	−0.0383	−5.5835
3	0.2435	0.2399	0.1127	0.1103	0.1308	0.1296	0.0023	0.0012	0.0036	1.0360	0.0194	6.9162	1.1301	6.8171	0.5695	0.2030	0.0077	0.0283	−0.0242	1.2222	0.0397	−0.0328	−5.7301
4	0.2801	0.2138	0.1441	0.1005	0.1360	0.1132	0.0436	0.0227	0.0663	0.9162	0.2954	7.2497	1.1096	6.7855	0.5778	0.2097	0.0266	0.0278	−0.0066	1.2055	0.0454	−0.0282	−5.6668
5	0.4035	0.1914	0.2846	0.1138	0.1189	0.0775	0.1708	0.0413	0.2121	0.9018	1.0138	10.0739	0.9851	6.2007	0.6427	0.1936	0.0485	0.0293	0.0145	1.0443	0.0569	−0.0247	−4.7419
Panel B: Stocks Sorted by SRVLJ																							
Quintile	**RVJP**	**RVJN**	**RVLJP**	**RVLJN**	**RVSJP**	**RVSJN**	**SRVLJ**	**SRVSJ**	**SRVJ**	**RVOL**	**RSK**	**RKT**	**BETA**	**log(Size)**	**BEME**	**MOM**	**REV**	**IVOL**	**CSK**	**CKT**	**MAX**	**MIN**	**ILLIQ**
Part I: Jump Truncation Level = γ^1																							
1	0.2255	0.3959	0.1250	0.3016	0.1005	0.0943	−0.1766	0.0062	−0.1703	0.9429	−0.9075	10.2967	1.0163	6.0403	0.6329	0.1966	−0.0221	0.0316	−0.0548	1.0357	0.0347	−0.0461	−4.5515
2	0.2222	0.2603	0.0874	0.1297	0.1347	0.1307	−0.0422	0.0041	−0.0382	0.9128	−0.2003	6.9378	1.1314	6.7639	0.5678	0.2104	−0.0011	0.0288	−0.0329	1.2050	0.0374	−0.0357	−5.6652
3	0.2296	0.2252	0.0968	0.0936	0.1329	0.1316	0.0032	0.0013	0.0044	1.1510	0.0218	6.5571	1.1308	6.9059	0.5648	0.1997	0.0069	0.0282	−0.0245	1.2380	0.0390	−0.0326	−5.8581
4	0.2726	0.2229	0.1493	0.0981	0.1233	0.1248	0.0512	−0.0016	0.0497	0.8314	0.2602	7.2170	1.1237	6.7993	0.5764	0.2118	0.0169	0.0280	−0.0166	1.2086	0.0421	−0.0304	−5.6962
5	0.4041	0.2153	0.3149	0.1234	0.0891	0.0920	0.1915	−0.0028	0.1887	0.9158	0.9900	10.5179	0.9873	6.0946	0.6471	0.1887	0.0354	0.0299	−0.0021	1.0376	0.0533	−0.0284	−4.6003
Part II: Jump Truncation Level = γ^2																							
1	0.2412	0.3897	0.0841	0.2403	0.1571	0.1494	−0.1561	0.0076	−0.1485	0.9355	−0.8465	10.5599	1.0128	6.0603	0.6360	0.1928	−0.0152	0.0313	−0.0478	1.0403	0.0366	−0.0438	−4.5663
2	0.2189	0.2314	0.0341	0.0519	0.1848	0.1794	−0.0178	0.0054	−0.0124	0.9321	−0.0762	6.4172	1.1366	6.8798	0.5622	0.2132	0.0045	0.0283	−0.0278	1.2310	0.0388	−0.0340	−5.8322
3	0.2438	0.2405	0.0805	0.0770	0.1633	0.1635	0.0035	−0.0001	0.0033	1.6847	0.0202	7.7437	1.0936	6.5831	0.6248	0.1664	0.0063	0.0322	−0.0021	1.2281	0.0426	−0.0360	−5.3661
4	0.2985	0.2611	0.1233	0.0854	0.1752	0.1756	0.0379	−0.0004	0.0375	1.0724	0.2140	8.0742	1.0912	6.4686	0.5995	0.1944	0.0125	0.0294	−0.0213	1.1478	0.0418	−0.0324	−5.2346
5	0.3985	0.2321	0.2526	0.0837	0.1459	0.1484	0.1690	−0.0025	0.1664	0.9201	0.9332	10.8242	0.9935	6.1033	0.6466	0.1896	0.0285	0.0300	−0.0084	1.0418	0.0506	−0.0301	−4.6103

Table 3. *Cont.*

Panel C: Stocks Sorted by SRVSJ																							
Quintile	**RVJP**	**RVJN**	**RVLJP**	**RVLJN**	**RVSJP**	**RVSJN**	**SRVLJ**	**SRVSJ**	**SRVJ**	**RVOL**	**RSK**	**RKT**	**BETA**	**log(Size)**	**BEME**	**MOM**	**REV**	**IVOL**	**CSK**	**CKT**	**MAX**	**MIN**	**ILLIQ**
Part I: Jump Truncation Level = γ^1																							
1	0.2060	0.2784	0.1085	0.0968	0.0975	0.1815	0.0116	−0.0840	−0.0724	0.7622	−0.1832	7.1817	1.1514	6.7378	0.5654	0.2183	−0.0270	0.0292	−0.0637	1.2017	0.0302	−0.0438	−5.6348
2	0.2711	0.2845	0.1758	0.1649	0.0952	0.1197	0.0110	−0.0244	−0.0135	1.0260	−0.0157	8.6959	1.0715	6.4139	0.6062	0.1933	−0.0046	0.0302	−0.0417	1.1242	0.0380	−0.0375	−5.1168
3	0.3077	0.3003	0.2186	0.2126	0.0891	0.0877	0.0060	0.0014	0.0074	1.2309	0.0345	9.7978	1.0053	6.1262	0.6382	0.1754	0.0066	0.0309	−0.0275	1.0469	0.0423	−0.0356	−4.6619
4	0.2743	0.2492	0.1413	0.1438	0.1330	0.1053	−0.0026	0.0277	0.0251	0.9562	0.0709	8.0830	1.0906	6.6614	0.5854	0.2098	0.0190	0.0283	−0.0099	1.1752	0.0436	−0.0305	−5.4837
5	0.2830	0.1981	0.1004	0.1051	0.1826	0.0930	−0.0047	0.0896	0.0849	0.7380	0.2395	7.1974	1.0947	6.8064	0.5783	0.2217	0.0428	0.0272	0.0136	1.1959	0.0517	−0.0251	−5.6801
Part II: Jump Truncation Level = γ^2																							
1	0.2058	0.3081	0.0749	0.0650	0.1309	0.2431	0.0099	−0.1121	−0.1023	0.8104	−0.3171	7.6436	1.1207	6.5435	0.5792	0.2179	−0.0318	0.0301	−0.0682	1.1558	0.0296	−0.0461	−5.3493
2	0.2561	0.2841	0.1103	0.1027	0.1458	0.1814	0.0077	−0.0356	−0.0280	1.0119	−0.0747	8.4222	1.0918	6.5103	0.5957	0.1964	−0.0079	0.0298	−0.0427	1.1482	0.0363	−0.0379	−5.2661
3	0.2906	0.2847	0.1423	0.1381	0.1484	0.1465	0.0041	0.0019	0.0060	1.1729	0.0308	9.3041	1.0385	6.3412	0.6204	0.1788	0.0065	0.0301	−0.0273	1.0955	0.0413	−0.0346	−4.9812
4	0.2821	0.2444	0.0950	0.0979	0.1871	0.1465	−0.0029	0.0405	0.0377	0.9731	0.1235	8.1884	1.0851	6.6358	0.5919	0.2046	0.0220	0.0284	−0.0086	1.1700	0.0447	−0.0299	−5.4350
5	0.3137	0.1959	0.0671	0.0713	0.2466	0.1246	−0.0043	0.1220	0.1178	0.7734	0.3845	7.6692	1.0627	6.6238	0.5954	0.2153	0.0469	0.0278	0.0158	1.1527	0.0543	−0.0245	−5.4040
Panel D: Stocks Sorted by RVOL																							
Quintile	**RVJP**	**RVJN**	**RVLJP**	**RVLJN**	**RVSJP**	**RVSJN**	**SRVLJ**	**SRVSJ**	**SRVJ**	**RVOL**	**RSK**	**RKT**	**BETA**	**log(Size)**	**BEME**	**MOM**	**REV**	**IVOL**	**CSK**	**CKT**	**MAX**	**MIN**	**ILLIQ**
1	0.2255	0.2140	0.1013	0.0943	0.1241	0.1197	0.0070	0.0044	0.0114	0.2290	0.0485	6.8794	0.8390	8.3393	0.5407	0.1686	0.0040	0.0137	−0.0213	1.4325	0.0177	−0.0154	−7.4961
2	0.2375	0.2282	0.1136	0.1071	0.1240	0.1211	0.0065	0.0028	0.0094	0.3596	0.0403	7.2835	1.0471	7.4505	0.5506	0.1853	0.0044	0.0187	−0.0213	1.3708	0.0257	−0.0227	−6.4401
3	0.2567	0.2493	0.1387	0.1327	0.1180	0.1166	0.0060	0.0014	0.0074	0.5331	0.0341	7.9018	1.2246	6.6274	0.5581	0.2598	0.0051	0.0253	−0.0249	1.2373	0.0346	−0.0304	−5.5465
4	0.2864	0.2823	0.1717	0.1675	0.1147	0.1148	0.0042	−0.0001	0.0041	0.8136	0.0216	8.7067	1.2761	5.7649	0.5997	0.3071	0.0058	0.0338	−0.0308	1.0491	0.0464	−0.0405	−4.4310
5	0.3429	0.3449	0.2336	0.2347	0.1093	0.1102	−0.0011	−0.0010	−0.0021	2.8115	−0.0003	10.5156	1.0101	4.4547	0.7443	0.0910	0.0160	0.0548	−0.0331	0.6286	0.0817	−0.0645	−2.4951
Panel E: Stocks Sorted by RSK																							
Quintile	**RVJP**	**RVJN**	**RVLJP**	**RVLJN**	**RVSJP**	**RVSJN**	**SRVLJ**	**SRVSJ**	**SRVJ**	**RVOL**	**RSK**	**RKT**	**BETA**	**log(Size)**	**BEME**	**MOM**	**REV**	**IVOL**	**CSK**	**CKT**	**MAX**	**MIN**	**ILLIQ**
1	0.2077	0.3914	0.1200	0.2822	0.0877	0.1091	−0.1622	−0.0215	−0.1837	0.9182	−0.9829	10.3573	1.0293	6.1615	0.6212	0.2118	−0.0274	0.0312	−0.0611	1.0555	0.0324	−0.0472	−4.7275
2	0.2203	0.2758	0.0967	0.1366	0.1235	0.1393	−0.0398	−0.0157	−0.0556	0.9332	−0.2596	6.8740	1.1303	6.7033	0.5719	0.2052	−0.0077	0.0291	−0.0412	1.1981	0.0353	−0.0374	−5.5757
3	0.2430	0.2394	0.1057	0.1035	0.1373	0.1359	0.0023	0.0014	0.0037	1.1222	0.0189	6.5111	1.1288	6.8070	0.5748	0.1932	0.0073	0.0283	−0.0238	1.2155	0.0391	−0.0325	−5.7121
4	0.2786	0.2142	0.1418	0.0960	0.1368	0.1182	0.0459	0.0186	0.0644	0.8769	0.3040	6.9460	1.1139	6.7671	0.5790	0.2056	0.0231	0.0280	−0.0098	1.2017	0.0444	−0.0290	−5.6422
5	0.3996	0.1978	0.2947	0.1179	0.1049	0.0800	0.1769	0.0249	0.2018	0.8942	1.0656	10.5964	0.9944	6.2012	0.6374	0.1963	0.0400	0.0297	0.0045	1.0482	0.0549	−0.0272	−4.7552
Panel F: Stocks Sorted by RKT																							
Quintile	**RVJP**	**RVJN**	**RVLJP**	**RVLJN**	**RVSJP**	**RVSJN**	**SRVLJ**	**SRVSJ**	**SRVJ**	**RVOL**	**RSK**	**RKT**	**BETA**	**log(Size)**	**BEME**	**MOM**	**REV**	**IVOL**	**CSK**	**CKT**	**MAX**	**MIN**	**ILLIQ**
1	0.1804	0.1785	0.0257	0.0248	0.1548	0.1536	0.0008	0.0011	0.0019	0.6884	0.0110	4.4470	1.1920	7.6130	0.5303	0.1864	0.0054	0.0248	−0.0222	1.3529	0.0339	−0.0290	−6.7592
2	0.2242	0.2206	0.0757	0.0738	0.1484	0.1468	0.0020	0.0017	0.0036	0.7285	0.0167	5.7679	1.1586	6.9411	0.5522	0.2203	0.0077	0.0276	−0.0252	1.2505	0.0387	−0.0324	−5.9541
3	0.2630	0.2582	0.1340	0.1310	0.1291	0.1272	0.0030	0.0019	0.0048	0.8130	0.0215	7.0711	1.1103	6.5028	0.5817	0.2175	0.0083	0.0296	−0.0265	1.1627	0.0417	−0.0347	−5.3464
4	0.3070	0.3004	0.2067	0.2017	0.1004	0.0987	0.0049	0.0017	0.0066	0.9433	0.0292	8.9744	1.0438	6.0841	0.6186	0.2085	0.0076	0.0313	−0.0281	1.0644	0.0440	−0.0370	−4.6922
5	0.3745	0.3612	0.3171	0.3051	0.0574	0.0561	0.0119	0.0013	0.0133	1.5722	0.0660	15.0322	0.8918	5.4974	0.7022	0.1793	0.0064	0.0332	−0.0295	0.8881	0.0478	−0.0403	−3.6585

Notes: See notes to Table 2. Entries in this table are time series averages of equal-weighted realized measures and firm characteristics of stocks sorted by various realized measures. The sample includes all NYSE, NASDAQ and AMEX listed stocks for the period January 1993 to December 2016. At the end of each Tuesday, all the stocks in the sample are sorted into quintile portfolios, based on ascending values of various realized measures. The equal-weighted realized measures and firm characteristics of each quintile portfolio are calculated over the same week. Additionally, $\gamma^1 = 4\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$ and $\gamma^2 = 5\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$ are jump truncation levels. See Sections 2 and 4 for further details.

4. Empirical Results

In this section, we begin by discussing single (univariate) portfolio sort results (see Tables 4–7), and double sort results (see Tables 8–13). We then discuss results based on a firm-level Fama–MacBeth regression analysis and an analysis of jumps and news announcements. We assume a weekly holding period, and return calculations reported in the tables are carried out as follows. At the end of each Tuesday, stocks are sorted into quintile portfolios based on different realized variation measures (see Panel A of Table 1). We then calculate equal-weighted and value-weighted portfolio returns over the subsequent week. We report the time series average of these weekly returns for each portfolio (these returns are called "Mean Return" in the tables). In addition, we regress the excess return of each portfolio on the Fama–French and Carhart (FFC4) factors to control for systematic risks, using regression of the form

$$r_{i,t} - r_{f,t} = \alpha_i + \beta_i^{MKT}(MKT_t - r_{f,t}) + \beta_i^{SMB} SMB_t + \beta_i^{HML} HML_t + \beta_i^{UMD} UMD_t + \epsilon_{i,t} \qquad (14)$$

where $r_{i,t}$ denotes the weekly return for firm i, $r_{f,t}$ is the risk-free rate; and MKT_t, SMB_t, HML_t, and UMD_t denote FFC4 market, size, value and momentum factors, respectively. The intercepts from these regressions (called "Alpha" in our tabulated results), measure risk-adjusted excess returns, and are also reported in Tables 4–13. Needless to say, our objective in these tables is to assess whether predictability exists, after controlling for various systematic risk factors. Finally, in Table 14, we report the results of cross-sectional (firm-level) Fama–MacBeth regressions used to investigate return predictability when simultaneously controlling for multiple realized measures and firm characteristics.

Table 4. Univariate Portfolio Sorts Based on Positive, Negative, and Signed Total Jump Variation.

Panel A: Stocks Sorted by RVJP												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Mean Return	33.65	30.50	33.04	28.08	20.64	−13.01 ***	23.52	19.48	17.93	20.91	18.27	−5.25
	(3.54)	(3.28)	(3.49)	(2.87)	(2.19)	(−2.75)	(3.54)	(3.27)	(2.93)	(3.35)	(2.83)	(−1.35)
Alpha	10.59	7.47	11.24	7.52	2.88	−7.71	2.88	−0.63	−2.30	0.67	−2.75	−5.63 *
	(4.16)	(3.64)	(4.34)	(2.33)	(0.72)	(−1.64)	(2.31)	(−0.44)	(−1.22)	(0.32)	(−1.19)	(−1.87)
Panel B: Stocks Sorted by RVJN												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Mean Return	13.17	23.01	26.77	33.79	49.23	36.06 ***	16.23	25.44	26.41	26.62	31.36	15.13 ***
	(1.52)	(2.59)	(2.84)	(3.36)	(4.62)	(6.47)	(2.55)	(4.11)	(4.08)	(3.93)	(4.29)	(3.75)
Alpha	−9.36	−0.03	4.18	13.24	31.71	41.07 ***	−3.55	4.94	5.27	5.49	10.05	13.60 ***
	(−4.46)	(−0.02)	(1.86)	(3.93)	(6.34)	(7.51)	(−3.08)	(3.02)	(2.64)	(2.37)	(4.13)	(4.52)
Panel C: Stocks Sorted by SRVJ												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Mean Return	51.85	39.02	26.15	17.86	11.02	−40.82 ***	34.67	27.43	19.93	13.64	9.65	−25.02 ***
	(5.14)	(3.85)	(2.70)	(1.98)	(1.33)	(−9.85)	(4.85)	(4.12)	(3.10)	(2.16)	(1.59)	(−5.78)
Alpha	30.54	17.81	4.56	−3.58	−9.64	−40.18 ***	13.44	6.94	−0.52	−6.53	−10.25	−23.69 ***
	(8.40)	(5.78)	(1.74)	(−1.56)	(−4.05)	(−10.10)	(5.01)	(3.95)	(−0.40)	(−4.48)	(−4.47)	(−5.56)

Notes: Entries in this table are average returns and risk-adjusted alphas for single-sorted portfolios based on RVJP, RVJN and SRVJ, which are described in Table 2. The sample includes all NYSE, NASDAQ and AMEX listed stocks for the period January 1993 to December 2016. At the end of each Tuesday, all the stocks in the sample are sorted into quintile portfolios based on ascending values of the various jump variation measures listed in the title of each panel. Each portfolio is held for one week. The row labeled "Mean Return" reports the time series average values of one-week-ahead equal-weighted and value-weighted returns for quintile portfolios. The row labeled "Alpha" reports Fama–French–Carhart four-factor alphas, based on the model (14), for each of the quintile portfolios, as well as for the difference between portfolio 5 and portfolio 1. Newey–West t-statistics are given in parentheses; and *, **, and *** denote means and alphas that are significant at the 10%, 5%, and 1% levels, respectively.

Table 5. Univariate Portfolio Sorts Based on Positive, Negative, and Signed Large Jump Variation.

Panel A: Stocks Sorted by RVLJP												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Part I: Jump Truncation Level = γ^1												
Mean Return	31.40	30.00	29.17	31.04	23.78	−7.61 **	21.78	20.61	20.72	18.56	17.75	−4.03
	(3.31)	(3.19)	(3.10)	(3.28)	(2.48)	(−2.05)	(3.41)	(3.21)	(3.33)	(2.95)	(2.66)	(−1.09)
Alpha	10.07	7.64	6.65	9.04	6.08	−3.98	1.97	0.68	−0.42	−2.32	−2.83	−4.80 *
	(4.20)	(3.44)	(2.81)	(3.35)	(1.53)	(−1.08)	(1.95)	(0.52)	(−0.26)	(−1.27)	(−1.21)	(−1.71)
Part II: Jump Truncation Level = γ^2												
Mean Return	29.42	43.86	30.00	29.12	25.77	−3.65	20.42	27.59	20.12	22.50	20.39	−0.03
	(3.15)	(2.16)	(3.04)	(3.07)	(2.74)	(−1.11)	(3.25)	(1.77)	(2.98)	(3.63)	(3.14)	(−0.01)
Alpha	7.74	32.93	7.28	7.02	7.70	−0.04	0.45	13.80	−0.27	1.60	−0.52	−0.97
	(3.70)	(3.23)	(2.92)	(2.62)	(2.02)	(−0.01)	(0.72)	(1.35)	(−0.13)	(0.83)	(−0.22)	(−0.37)
Panel B: Stocks Sorted by RVLJN												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Part I: Jump Truncation Level = γ^1												
Mean Return	23.80	23.68	27.64	28.85	41.56	17.76 ***	19.65	19.58	25.63	22.38	25.24	5.59
	(2.65)	(2.59)	(2.98)	(2.95)	(4.10)	(4.47)	(3.10)	(3.10)	(4.00)	(3.27)	(3.44)	(1.41)
Alpha	2.31	1.11	5.00	7.10	23.74	21.43 ***	−0.20	−1.32	4.52	0.89	4.34	4.54
	(1.11)	(0.55)	(2.32)	(2.47)	(5.39)	(5.53)	(−0.18)	(−0.92)	(2.47)	(0.40)	(1.82)	(1.61)
Part II: Jump Truncation Level = γ^2												
Mean Return	24.76	6.46	27.89	29.12	38.62	13.86 ***	19.69	13.21	21.46	23.20	22.34	2.66
	(2.71)	(0.33)	(2.75)	(3.03)	(3.88)	(4.01)	(3.13)	(0.93)	(3.05)	(3.61)	(3.15)	(0.75)
Alpha	3.02	6.75	6.58	7.06	20.62	17.60 ***	−0.30	6.52	1.26	2.28	1.58	1.88
	(1.63)	(1.08)	(2.50)	(2.67)	(4.81)	(5.06)	(−0.45)	(0.91)	(0.57)	(1.06)	(0.66)	(0.70)

Table 5. *Cont.*

Panel C: Stocks Sorted by SRVLJ												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Part I: Jump Truncation Level = γ^1												
Mean Return	44.35	32.94	31.08	22.72	16.04	−28.31 ***	26.27	22.99	22.36	17.77	16.27	−10.01 ***
	(4.52)	(3.36)	(3.13)	(2.44)	(1.88)	(−9.00)	(3.92)	(3.58)	(3.29)	(2.79)	(2.71)	(−3.09)
Alpha	23.47	11.38	8.91	0.96	−4.90	−28.36 ***	5.00	2.42	1.83	−2.64	−4.26	−9.25 ***
	(7.13)	(4.20)	(3.04)	(0.40)	(−2.17)	(−9.39)	(2.24)	(1.64)	(1.01)	(−1.82)	(−2.24)	(−2.87)
Part II: Jump Truncation Level = γ^2												
Mean Return	40.55	28.37	33.05	24.16	19.03	−21.52 ***	22.59	20.48	16.45	18.86	20.14	−2.45
	(4.19)	(2.91)	(1.48)	(2.55)	(2.19)	(−8.22)	(3.40)	(3.18)	(1.14)	(3.02)	(3.27)	(−0.80)
Alpha	19.59	8.18	24.26	2.23	−1.97	−21.55 ***	1.82	0.76	6.85	−2.41	−0.26	−2.08
	(6.15)	(3.29)	(2.16)	(0.82)	(−0.84)	(−8.33)	(0.86)	(0.68)	(1.04)	(−1.23)	(−0.13)	(−0.69)

Notes: See notes to Table 4. Entries are average returns and risk-adjusted alphas for single-sorted portfolios based on RVLJP, RVLJN and SRVLJ. Jump truncation levels are $\gamma^1 = 4\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$ and $\gamma^2 = 5\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$.

Table 6. Univariate Portfolio Sorts Based on Positive, Negative, and Signed Small Jump Variation.

Panel A: Stocks Sorted by RVSJP												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Part I: Jump Truncation Level = γ^1												
Mean Return	32.01	32.51	29.03	26.59	25.72	−6.29 **	27.93	26.10	18.44	16.52	15.37	−12.55 **
	(3.44)	(3.43)	(3.09)	(2.85)	(2.65)	(−2.14)	(3.65)	(3.84)	(2.91)	(2.75)	(2.45)	(−2.54)
Alpha	13.11	9.39	6.37	4.77	6.13	−6.98 ***	7.40	5.21	−1.83	−2.89	−4.95	−12.35 ***
	(3.74)	(4.21)	(3.02)	(2.08)	(1.82)	(−2.65)	(2.44)	(3.28)	(−1.39)	(−1.55)	(−2.32)	(−2.93)
Part II: Jump Truncation Level = γ^2												
Mean Return	34.25	31.55	28.43	27.41	24.23	−10.02 ***	29.37	18.76	17.76	14.75	18.98	−10.40 **
	(3.72)	(3.37)	(3.04)	(2.90)	(2.48)	(−3.26)	(4.09)	(2.95)	(2.94)	(2.39)	(2.93)	(−2.25)
Alpha	14.08	8.85	5.90	5.75	5.08	−9.00 ***	8.52	−2.00	−2.01	−5.30	−1.51	−10.02 **
	(4.93)	(4.12)	(2.73)	(2.32)	(1.42)	(−3.13)	(3.87)	(−1.52)	(−1.24)	(−2.67)	(−0.62)	(−2.54)

Table 6. *Cont.*

Panel B: Stocks Sorted by RVSJN												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Part I: Jump Truncation Level = γ^1												
Mean Return	23.66	21.60	27.21	31.73	41.77	18.10 ***	6.94	15.90	21.97	27.77	31.26	24.32 ***
	(2.64)	(2.38)	(2.97)	(3.35)	(3.93)	(5.07)	(1.00)	(2.42)	(3.46)	(4.34)	(4.68)	(5.00)
Alpha	5.26	−1.04	4.35	9.56	21.62	16.36 ***	−13.39	−4.28	1.80	7.41	10.52	23.91 ***
	(1.59)	(−0.51)	(2.22)	(4.02)	(5.49)	(5.46)	(−4.54)	(−2.80)	(1.33)	(3.72)	(4.07)	(5.21)
Part II: Jump Truncation Level = γ^2												
Mean Return	19.42	23.04	26.65	32.48	44.37	24.96 ***	14.22	18.78	25.82	29.05	32.60	18.38 ***
	(2.23)	(2.60)	(2.93)	(3.35)	(4.09)	(6.07)	(2.13)	(3.00)	(4.13)	(4.42)	(4.43)	(3.80)
Alpha	−0.37	0.70	4.18	10.38	24.84	25.22 ***	−5.47	−1.28	5.81	7.80	10.85	16.31 ***
	(−0.14)	(0.37)	(2.18)	(4.02)	(5.85)	(7.21)	(−2.72)	(−1.07)	(3.39)	(3.67)	(4.02)	(4.15)
Panel C: Stocks Sorted by SRVSJ												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Part I: Jump Truncation Level = γ^1												
Mean Return	46.41	40.51	25.12	19.00	13.74	−32.67 ***	34.54	24.08	18.26	17.84	10.42	−24.12 ***
	(4.57)	(4.04)	(2.67)	(2.06)	(1.62)	(−8.60)	(5.00)	(3.58)	(2.81)	(2.85)	(1.72)	(−6.60)
Alpha	23.64	19.68	5.23	−2.09	−8.14	−31.78 ***	13.77	3.25	−2.37	−2.39	−9.27	−23.04 ***
	(7.62)	(6.10)	(1.53)	(−0.85)	(−4.17)	(−9.01)	(6.18)	(1.91)	(−1.07)	(−1.52)	(−5.00)	(−6.54)
Part II: Jump Truncation Level = γ^2												
Mean Return	47.90	41.62	27.23	17.90	11.20	−36.70 ***	36.88	25.13	18.45	14.98	9.41	−27.47 ***
	(4.70)	(4.13)	(2.87)	(1.98)	(1.34)	(−9.06)	(5.31)	(3.79)	(2.86)	(2.39)	(1.52)	(−6.94)
Alpha	25.37	20.76	6.99	−3.03	−10.51	−35.88 ***	16.07	4.52	−1.87	−5.39	−10.34	−26.41 ***
	(7.79)	(6.65)	(2.23)	(−1.25)	(−5.21)	(−9.49)	(6.72)	(2.60)	(−1.23)	(−3.30)	(−5.00)	(−6.72)

Notes: See notes to Table 5.

Table 7. Univariate Portfolio Sorts Based on Realized Volatility, Skewness, Kurtosis and Continuous Variance.

Panel A: Stocks Sorted by RVOL												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Mean Return	23.36	28.00	28.89	31.78	33.91	10.55	20.72	21.42	19.75	26.78	29.19	8.47
	(4.47)	(3.92)	(2.96)	(2.59)	(2.24)	(0.81)	(4.09)	(2.83)	(1.84)	(1.98)	(1.92)	(0.64)
Alpha	4.50	5.01	5.15	8.74	16.33	11.83	1.95	−1.39	−3.94	2.44	5.44	3.49
	(2.07)	(2.94)	(2.57)	(2.54)	(2.11)	(1.37)	(1.35)	(−0.67)	(−1.01)	(0.43)	(0.67)	(0.40)
Panel B: Stocks Sorted by RSK												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Mean Return	47.56	38.06	27.86	19.44	12.98	−34.58 ***	29.45	27.52	19.27	14.68	13.21	−16.24 ***
	(4.85)	(3.82)	(2.86)	(2.12)	(1.54)	(−9.94)	(4.27)	(4.22)	(2.98)	(2.32)	(2.18)	(−4.29)
Alpha	26.22	16.77	6.73	−2.15	−7.90	−34.12 ***	7.87	7.02	−0.82	−5.38	−6.77	−14.64 ***
	(7.93)	(5.66)	(2.41)	(−0.96)	(−3.51)	(−10.08)	(3.30)	(4.44)	(−0.60)	(−3.73)	(−3.23)	(−3.85)
Panel C: Stocks Sorted by RKT												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Mean Return	28.95	28.59	29.91	29.42	29.07	0.12	19.87	21.57	21.12	22.17	19.55	−0.32
	(3.07)	(3.00)	(3.13)	(3.06)	(3.24)	(0.04)	(3.12)	(3.42)	(3.37)	(3.38)	(2.96)	(−0.10)
Alpha	8.55	6.42	7.47	7.92	9.36	0.81	0.21	0.65	−0.10	0.94	−1.92	−2.13
	(3.21)	(2.87)	(3.05)	(2.87)	(3.07)	(0.28)	(0.20)	(0.46)	(−0.06)	(0.49)	(−0.91)	(−0.81)
Panel D: Stocks Sorted by RVC												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**
Mean Return	36.18	31.82	28.23	27.47	22.21	−13.97 **	24.36	24.80	22.94	23.40	20.41	−3.95
	(3.54)	(3.22)	(3.00)	(3.05)	(2.42)	(−2.58)	(3.53)	(3.65)	(3.59)	(3.87)	(3.20)	(−1.00)
Alpha	19.82	10.62	5.54	4.66	−0.94	−20.76 ***	4.27	3.29	2.42	3.27	0.06	−4.22
	(4.04)	(3.36)	(2.48)	(2.47)	(−0.41)	(−3.87)	(1.76)	(1.62)	(1.30)	(2.10)	(0.07)	(−1.46)

Notes: See notes to Table 5.

Table 8. Double-Sorted Portfolios: Portfolios Sorted by Various Jump Variation Measures.

Panel A: Stocks Sorted by SRVLJ, Controlling for SRVJ Based on γ^2												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
	SRVJ Quintile						**SRVJ Quintile**					
SRVLJ Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part I: Mean Return and Alpha												
1 (Low)	47.99	28.73	21.12	12.90	7.82	23.71	41.88	18.83	19.24	13.59	5.04	16.61
2	50.55	37.24	22.38	26.91	12.99	33.21	29.61	25.26	18.26	17.39	14.08	22.70
3	53.07	38.57	23.49	21.73	12.37	29.75	36.31	26.28	8.10	15.50	17.40	24.19
4	56.29	21.42	35.09	14.59	17.86	27.51	36.38	7.94	18.53	13.64	16.59	20.33
5 (High)	47.70	45.92	32.70	26.10	7.59	32.00	32.64	39.20	23.00	20.20	19.84	27.00
High-Low	−0.24	17.19	11.58	13.20	−0.23	8.30	6.38	20.37	3.76	6.61	14.81	10.38
Alpha	−6.48	16.64	10.80	13.88	2.60	7.49	3.69	19.41	2.46	6.06	13.84	9.09
Part II: *t*-Statistics												
1 (Low)	5.15	2.90	2.26	1.48	0.92	2.68	3.45	2.66	2.89	2.08	0.75	2.67
2	4.81	3.62	1.94	0.81	1.43	3.49	3.97	3.40	2.28	0.88	1.99	3.49
3	4.87	3.03	0.75	1.89	1.39	3.03	4.65	3.06	0.37	1.92	2.65	3.73
4	4.64	0.63	3.07	1.53	2.04	2.81	4.26	0.27	2.18	2.05	2.50	3.04
5 (High)	4.39	4.29	3.23	2.71	0.89	3.38	3.77	4.89	3.16	2.78	2.99	4.01
High-Low	−0.04	4.24	3.31	3.87	−0.06	4.18	1.09	3.71	0.79	1.54	3.09	4.15
Alpha	−1.19	4.11	3.06	4.09	0.67	3.80	0.62	3.40	0.49	1.44	2.96	3.53
Panel B: Stocks Sorted by SRVSJ, Controlling for SRVJ Based on γ^2												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
	SRVJ Quintile						**SRVJ Quintile**					
SRVSJ Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part I: Mean Return and Alpha												
1 (Low)	56.90	44.50	34.78	26.13	19.50	36.36	41.88	35.91	26.98	18.28	22.43	29.73
2	55.66	47.02	30.91	18.59	10.34	32.50	40.76	34.11	20.84	13.23	16.52	25.09
3	56.24	41.20	28.46	19.74	10.02	31.13	30.72	23.04	20.39	13.82	14.29	20.45
4	53.58	37.21	16.75	12.62	10.19	26.07	28.23	23.82	12.02	14.94	6.93	17.19
5 (High)	34.30	25.12	19.72	12.12	5.79	19.41	17.03	16.65	18.03	12.69	3.77	13.63
High-Low	−22.60	−19.38	−15.06	−14.01	−13.71	−16.95	−28.00	−19.26	−8.95	−5.59	−18.66	−16.09
Alpha	−19.26	−18.83	−14.19	−14.86	−16.20	−16.67	−25.86	−20.41	−6.87	−4.79	−18.22	−15.23

Table 8. *Cont.*

Panel B: Stocks Sorted by SRVSJ, Controlling for SRVJ Based on γ^2												
	Equal-Weighted Returns and Alphas						Value-Weighted Returns and Alphas					
	SRVJ Quintile						SRVJ Quintile					
SRVSJ Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part II: *t*-Statistics												
1 (Low)	5.44	4.22	3.40	2.67	2.16	3.76	5.47	4.75	3.58	2.54	3.33	4.44
2	5.20	4.21	3.08	1.93	1.17	3.35	5.19	4.65	3.00	1.95	2.33	3.86
3	5.15	3.99	2.77	2.14	1.17	3.26	4.04	3.21	2.89	2.02	2.20	3.20
4	5.33	3.58	1.71	1.42	1.23	2.86	3.39	3.41	1.72	2.22	1.02	2.68
5 (High)	3.46	2.60	2.12	1.39	0.70	2.19	2.22	2.32	2.67	1.89	0.56	2.16
High-Low	−4.99	−5.04	−4.14	−3.79	−3.45	−7.22	−5.37	−3.66	−1.79	−1.19	−3.80	−5.77
Alpha	−4.15	−4.90	−3.91	−4.16	−4.12	−7.42	−4.93	−3.65	−1.30	−1.02	−3.78	−5.32
Panel C: Stocks Sorted by SRVSJ, Controlling for SRVLJ Based on γ^2												
	Equal-Weighted Returns and Alphas						Value-Weighted Returns and Alphas					
	SRVLJ Quintile						SRVLJ Quintile					
SRVSJ Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part I: Mean Return and Alpha												
1 (Low)	60.34	51.38	40.69	44.93	31.91	47.28	41.88	35.33	35.90	34.10	30.94	36.43
2	57.30	38.41	58.52	28.26	28.35	40.27	23.32	25.07	37.49	18.00	25.93	25.12
3	35.25	26.42	4.21	28.76	14.62	25.17	16.19	21.74	9.63	20.36	21.71	19.86
4	28.42	16.02	48.13	10.74	10.78	19.49	17.78	15.39	97.02	14.92	14.12	22.20
5 (High)	18.79	9.45	−1.20	8.01	8.26	10.89	14.98	8.09	6.98	7.78	13.31	11.30
High-Low	−41.55	−41.93	−40.72	−38.27	−23.65	−36.71	−26.63	−27.24	−27.87	−27.67	−17.63	−25.38
Alpha	−40.15	−41.33	−40.17	−37.25	−22.90	−35.45	−25.59	−25.58	−26.01	−26.35	−17.73	−24.07
Part II: *t*-Statistics												
1 (Low)	5.60	4.76	1.56	4.18	3.28	4.58	5.38	4.76	1.72	4.12	4.37	5.13
2	5.40	3.61	2.17	2.64	3.02	3.97	3.03	3.41	1.61	2.34	3.60	3.62
3	3.58	2.60	0.19	2.85	1.62	2.65	2.27	3.21	0.53	2.93	3.11	3.15
4	2.83	1.70	1.20	1.10	1.23	2.01	2.49	2.30	1.11	2.03	2.07	2.35
5 (High)	2.12	1.06	−0.06	0.93	1.01	1.31	2.12	1.22	0.46	1.08	2.02	1.82
High-Low	−8.29	−8.49	−2.04	−6.08	−5.33	−8.59	−5.15	−5.54	−1.50	−4.14	−3.64	−6.61
Alpha	−8.24	−8.75	−2.03	−6.22	−5.50	−8.80	−4.90	−5.18	−1.39	−3.87	−3.66	−6.13

Table 8. *Cont.*

Panel D: Stocks Sorted by SRVLJ, Controlling for SRVSJ Based on γ^2												
	Equal-Weighted Returns and Alphas						Value-Weighted Returns and Alphas					
	SRVSJ Quintile						SRVSJ Quintile					
SRVLJ Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part I: Mean Return and Alpha												
1 (Low)	59.99	60.01	41.84	27.66	16.62	41.23	41.88	30.53	19.67	16.32	12.01	23.80
2	49.75	41.95	27.20	17.65	11.51	29.31	35.69	25.91	18.85	17.68	8.92	20.89
3	76.41	38.05	24.45	6.48	14.41	32.06	45.15	16.78	11.30	10.47	21.45	21.03
4	43.99	30.88	23.29	13.82	11.84	24.86	38.54	18.80	17.59	15.79	10.15	20.55
5 (High)	32.38	28.64	10.59	8.97	6.78	17.47	31.35	25.93	18.89	13.97	10.03	20.03
High-Low	−27.61	−31.38	−31.25	−18.69	−9.84	−23.75	−9.14	−4.60	−0.78	−2.35	−1.98	−3.77
Alpha	−26.51	−31.63	−31.28	−19.37	−9.80	−23.72	−8.62	−3.72	−0.59	−2.79	−1.66	−3.47
Part II: *t*-Statistics												
1 (Low)	5.51	5.59	4.22	2.85	1.85	4.25	5.37	4.03	2.69	2.23	1.71	3.59
2	4.51	3.85	2.62	1.87	1.29	3.05	4.71	3.54	2.63	2.71	1.32	3.28
3	2.31	1.95	1.61	0.41	0.80	2.49	1.53	1.16	1.08	0.78	1.49	2.25
4	4.05	2.90	2.28	1.44	1.31	2.63	4.83	2.51	2.47	2.31	1.39	3.23
5 (High)	3.36	2.99	1.21	1.03	0.83	2.02	4.32	3.63	2.69	2.11	1.55	3.23
High-Low	−6.48	−7.66	−6.92	−4.65	−2.73	−8.80	−1.83	−0.98	−0.16	−0.46	−0.44	−1.49
Alpha	−6.16	−7.50	−6.94	−4.94	−2.83	−8.94	−1.70	−0.79	−0.12	−0.53	−0.37	−1.38

Notes: See notes to Table 5. This table presents average returns (called "Mean Return") and risk-adjusted alphas (called "Alpha") for portfolios sorted by various jump variation measures. The sample includes NYSE, NASDAQ and AMEX listed stocks for the period January 1993 to December 2016. At the end of each Tuesday, all the stocks in the sample are sorted into quintile portfolios based on ascending values of SRVJ (SRVLJ/SRVSJ). Then, within each quintile portfolio, stocks are further sorted based on the values of SRVLJ/SRVSJ (SRVSJ/SRVLJ), resulting in 25 portfolios. Each portfolio is held for one week. The row labeled "High-Low" reports the average values of one-week-ahead returns in Part I (corresponding Newey–West *t*-statistics are given in Part II of the panel). The row labeled "Alpha" reports Fama–French–Carhart four-factor alphas in Part I (corresponding Newey–West *t*-statistics are again given in Part II of the panel) for the double-sorted High-Low portfolios. Please note that entries given in the "Average" column of the table, are average returns across the 5 quintiles. Finally, note that SRVLJ and SRVSJ are constructed based on jump truncation level $\gamma^2 = 5\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$.

Table 9. Double-Sorted Portfolios: Portfolios Sorted by SRVJ and RSK.

Panel A: Stocks Sorted by SRVJ, Controlling for RSK												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
	RSK Quintile						**RSK Quintile**					
SRVJ Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part I: Mean Return and Alpha												
1 (Low)	56.92	55.49	42.81	31.69	20.91	41.57	41.88	39.12	30.49	25.47	17.56	30.90
2	56.77	46.17	33.30	22.89	15.27	34.88	41.63	35.10	21.38	17.20	13.71	25.81
3	49.97	38.92	23.47	17.47	11.61	28.29	37.18	27.01	22.24	14.45	10.52	22.28
4	42.67	29.57	21.68	13.20	12.18	23.86	28.53	20.24	15.10	11.99	11.47	17.47
5 (High)	31.39	20.03	17.94	11.86	4.85	17.21	18.32	21.67	12.23	7.31	12.60	14.42
High-Low	−25.54	−35.46	−24.87	−19.83	−16.06	−24.35	−23.56	−17.46	−18.27	−18.16	−4.95	−16.48
Alpha	−28.79	−36.20	−24.40	−18.40	−12.75	−24.11	−24.22	−18.52	−18.40	−16.82	−4.88	−16.57
Part II: *t*-Statistics												
1 (Low)	6.13	5.04	3.97	3.04	2.22	4.21	5.72	4.87	3.97	3.35	2.51	4.58
2	5.34	4.12	3.14	2.32	1.70	3.52	5.11	4.57	2.96	2.44	2.02	3.89
3	4.67	3.80	2.30	1.90	1.35	2.99	4.90	3.78	3.07	2.14	1.63	3.51
4	4.09	2.98	2.27	1.46	1.44	2.61	3.67	2.84	2.17	1.76	1.72	2.73
5 (High)	3.27	2.20	2.02	1.36	0.59	2.01	2.59	3.24	1.76	1.07	1.90	2.36
High-Low	−5.36	−7.09	−5.12	−4.32	−3.40	−7.70	−4.70	−3.17	−3.32	−3.28	−0.89	−5.24
Alpha	−6.18	−7.57	−5.25	−4.32	−2.80	−8.07	−4.81	−3.50	−3.36	−3.04	−0.90	−5.45
Panel B: Stocks Sorted by RSK, Controlling for SRVJ												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
	SRVJ Quintile						**SRVJ Quintile**					
RSK Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part I: Mean Return and Alpha												
1 (Low)	51.14	29.34	21.71	17.58	12.40	26.43	41.88	18.33	20.91	12.07	4.09	18.48
2	49.50	39.74	23.05	14.15	11.46	27.58	37.29	33.29	14.95	10.36	10.00	21.18
3	49.96	37.41	27.26	18.14	12.93	29.14	28.38	23.82	22.52	14.80	15.23	20.95
4	53.85	43.31	28.11	18.79	12.52	31.32	36.37	28.73	19.57	17.15	10.67	22.50
5 (High)	54.80	45.32	30.65	20.66	5.78	31.44	36.75	32.79	24.63	18.92	14.26	25.47
High-Low	3.66	15.98	8.94	3.08	−6.63	5.01	−0.25	14.46	3.72	6.85	10.17	6.99
Alpha	0.54	16.64	8.57	2.34	−4.54	4.71	−0.71	15.66	4.39	6.30	9.15	6.96

Table 9. *Cont.*

Panel B: Stocks Sorted by RSK, Controlling for SRVJ												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
	SRVJ Quintile						**SRVJ Quintile**					
RSK Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part II: *t*-Statistics												
1 (Low)	5.56	3.07	2.36	1.99	1.42	3.02	4.98	2.61	3.15	1.74	0.60	2.97
2	4.76	3.75	2.33	1.57	1.30	2.94	4.89	4.50	2.06	1.53	1.52	3.29
3	4.76	3.60	2.69	1.94	1.50	3.07	3.59	3.33	3.29	2.24	2.33	3.30
4	4.99	3.98	2.77	1.97	1.49	3.27	4.70	3.89	2.69	2.39	1.63	3.45
5 (High)	5.00	4.19	2.99	2.17	0.68	3.27	4.50	4.20	3.29	2.83	2.09	3.87
High-Low	0.85	3.98	2.31	0.84	−1.61	2.35	−0.05	2.85	0.74	1.48	2.19	2.87
Alpha	0.14	4.40	2.24	0.66	−1.09	2.42	−0.14	3.05	0.89	1.38	2.01	2.92

Notes: See notes to Table 5. This table presents average returns (called "Mean Return") and risk-adjusted alphas (called "Alpha") for portfolios sorted by SRVJ controlling for RSK, and vice versa. The sample includes NYSE, NASDAQ and AMEX listed stocks for the period January 1993 to December 2016. At the end of each Tuesday, all the stocks in the sample are sorted into quintile portfolios based on ascending values of RSK (SRVJ), and then within each quintile portfolio, stocks are further sorted using values of SRVJ (RSK), resulting in 25 portfolios. Each portfolio is held for one week. The row labeled "High-Low" reports the average values of one-week-ahead returns in Part I (corresponding Newey–West *t*-statistics are given in Part II of the panel). The row labeled "Alpha" reports Fama–French–Carhart four-factor alphas in Part I (corresponding Newey–West *t*-statistics are again given in Part II of the panel) for each of the quintile portfolios, as well as for the average across 5 RSK (SRVJ) portfolios.

Table 10. Double-Sorted Portfolios: Portfolios Sorted by SRVLJ/SRVSJ, Controlling for RSK.

Panel A: Stocks Sorted by SRVLJ, Controlling for RSK Based on γ^2												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
	RSK Quintile						**RSK Quintile**					
SRVLJ Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part I: Mean Return and Alpha												
1 (Low)	47.31	31.04	20.83	13.46	9.23	24.38	41.88	20.74	17.63	10.28	10.95	17.04
2	48.24	35.53	18.15	41.82	10.40	30.72	27.52	25.93	13.06	22.40	13.21	21.88
3	46.79	36.39	−2.22	17.53	16.01	27.86	27.77	24.25	23.68	13.64	18.71	21.29
4	47.56	9.61	27.36	19.04	20.92	27.46	27.71	8.29	20.27	17.45	16.06	20.34
5 (High)	43.55	43.48	34.14	26.41	7.84	31.13	35.80	32.07	20.61	23.70	18.89	26.28
High-Low	−3.57	12.44	13.30	12.95	−1.39	6.75	10.42	11.33	2.98	13.42	7.94	9.22
Alpha	−8.76	11.52	13.22	13.09	1.59	6.13	8.26	11.20	1.91	12.69	7.30	8.27

Table 10. *Cont.*

Panel A: Stocks Sorted by SRVLJ, Controlling for RSK Based on γ^2												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
	RSK Quintile						**RSK Quintile**					
SRVLJ Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part II: *t*-Statistics												
1 (Low)	5.10	3.13	2.17	1.51	1.08	2.73	3.34	2.85	2.51	1.59	1.63	2.68
2	4.67	3.47	1.41	1.61	1.16	3.21	3.71	3.71	1.47	1.12	1.98	3.42
3	4.48	2.95	−0.07	1.54	1.79	2.91	3.81	2.87	0.85	1.56	2.89	3.30
4	4.32	0.37	2.27	1.96	2.38	2.89	3.68	0.43	2.12	2.52	2.45	3.17
5 (High)	4.22	4.03	3.42	2.75	0.93	3.34	4.08	3.80	2.81	3.23	2.79	3.90
High-Low	−0.66	2.93	4.28	4.11	−0.37	3.64	1.71	1.89	0.66	3.03	1.62	3.76
Alpha	−1.62	2.63	4.14	4.13	0.42	3.18	1.35	1.76	0.41	2.90	1.51	3.22
Panel B: Stocks Sorted by SRVSJ, Controlling for RSK Based on γ^2												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
	RSK Quintile						**RSK Quintile**					
SRVSJ Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part I: Mean Return and Alpha												
1 (Low)	56.15	51.19	42.56	31.93	24.16	41.20	41.88	40.69	31.16	31.49	21.39	33.55
2	54.34	50.15	35.79	24.76	12.89	35.59	32.45	33.20	21.93	19.73	23.95	26.25
3	55.24	38.53	26.29	18.62	11.09	29.95	27.91	22.64	18.61	12.36	16.37	19.58
4	40.20	31.45	18.77	14.26	12.14	23.35	18.13	22.01	15.30	13.17	10.40	15.79
5 (High)	30.21	18.90	15.72	7.49	4.76	15.42	18.59	19.31	12.59	8.23	6.56	13.06
High-Low	−25.94	−32.29	−26.84	−24.43	−19.40	−25.78	−24.44	−21.37	−18.56	−23.26	−14.84	−20.49
Alpha	−23.86	−31.76	−26.49	−23.94	−20.71	−25.35	−22.27	−22.89	−18.30	−21.83	−14.44	−19.95
Part II: *t*-Statistics												
1 (Low)	5.43	4.75	3.98	3.16	2.62	4.15	5.11	5.25	4.16	4.39	3.15	4.92
2	5.17	4.48	3.42	2.47	1.43	3.61	4.30	4.53	2.96	2.68	3.39	3.95
3	5.25	3.76	2.60	1.96	1.25	3.14	3.75	3.12	2.66	1.72	2.46	3.03
4	4.11	3.13	1.96	1.58	1.45	2.59	2.50	3.20	2.27	1.98	1.57	2.55
5 (High)	3.18	2.06	1.71	0.87	0.58	1.79	2.56	2.86	1.79	1.23	0.97	2.09
High-Low	−5.78	−7.15	−6.04	−5.58	−4.72	−8.41	−4.21	−4.04	−3.67	−4.56	−2.93	−6.31
Alpha	−5.36	−7.20	−6.12	−5.75	−5.27	−8.89	−3.83	−4.29	−3.51	−4.09	−2.83	−6.01

Notes: See notes to Table 8. Portfolios are sorted by SRVLJ/SRVSJ, controlling for RSK, and using truncation level γ^2, as discussed in the notes to Table 2.

Table 11. Double-Sorted Portfolios: Portfolios Sorted by RSK, Controlling for SRVLJ or SRVSJ.

Panel A: Stocks Sorted by RSK, Controlling for SRVLJ Based on γ^2												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
	SRVLJ Quintile						**SRVLJ Quintile**					
RSK Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part I: Mean Return and Alpha												
1 (Low)	51.98	46.26	32.88	42.99	35.40	43.72	41.88	34.45	32.97	31.19	33.46	34.64
2	48.38	38.05	55.49	31.86	20.58	37.03	30.03	23.16	48.38	17.47	16.06	24.56
3	43.16	26.83	8.78	20.69	16.54	25.94	25.91	21.10	10.09	19.92	19.16	21.15
4	37.70	22.51	55.61	11.20	14.84	24.67	18.23	18.69	8.48	13.82	14.50	16.23
5 (High)	21.43	8.10	−0.87	14.29	7.74	12.81	13.59	7.52	−8.20	15.74	15.50	12.31
High-Low	−30.55	−38.16	−30.25	−30.06	−27.66	−31.07	−23.12	−26.93	−37.56	−16.82	−17.96	−22.43
Alpha	−32.89	−37.85	−28.79	−30.01	−24.81	−30.46	−23.34	−24.74	−36.67	−14.80	−18.66	−21.39
Part II: *t*-Statistics												
1 (Low)	5.63	4.46	1.33	4.16	3.61	4.48	4.94	4.65	1.67	4.00	4.80	5.06
2	4.68	3.68	2.16	3.04	2.22	3.72	3.99	3.28	2.04	2.34	2.31	3.63
3	4.14	2.66	0.40	2.04	1.84	2.69	3.32	3.22	0.59	2.66	2.87	3.29
4	3.62	2.29	1.37	1.11	1.73	2.50	2.54	2.82	0.48	1.80	2.22	2.54
5 (High)	2.27	0.89	−0.04	1.63	0.92	1.49	1.91	1.08	−0.52	2.33	2.24	1.98
High-Low	−7.91	−9.22	−1.71	−5.63	−5.85	−9.46	−4.56	−6.02	−2.29	−2.75	−3.47	−6.54
Alpha	−8.43	−9.50	−1.64	−5.86	−5.51	−9.57	−4.59	−5.54	−2.22	−2.37	−3.65	−6.10
Panel B: Stocks Sorted by RSK, Controlling for SRVSJ Based on γ^2												
	Equal-Weighted Returns and Alphas						**Value-Weighted Returns and Alphas**					
	SRVSJ Quintile						**SRVSJ Quintile**					
RSK Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part I: Mean Return and Alpha												
1 (Low)	57.49	56.03	40.07	26.79	17.75	39.62	41.88	24.28	18.84	16.57	12.21	23.20
2	51.21	42.88	31.02	15.27	12.69	30.61	35.14	26.67	14.77	21.10	9.49	21.43
3	54.96	41.34	25.81	20.97	10.12	30.64	37.20	24.66	20.50	10.79	7.58	20.15
4	42.52	38.93	29.50	16.21	9.58	27.35	37.22	23.08	17.57	15.29	11.02	20.84
5 (High)	33.23	28.85	9.62	10.20	5.82	17.54	31.94	30.24	21.43	16.36	8.89	21.77
High-Low	−24.26	−27.17	−30.45	−16.59	−11.93	−22.08	−12.15	5.96	2.59	−0.21	−3.33	−1.43
Alpha	−22.78	−27.83	−30.17	−17.16	−11.85	−21.96	−10.86	7.74	2.94	−0.96	−3.03	−0.83

Table 11. *Cont.*

Panel B: Stocks Sorted by RSK, Controlling for SRVSJ Based on γ^2												
	Equal-Weighted Returns and Alphas						Value-Weighted Returns and Alphas					
	SRVSJ Quintile						SRVSJ Quintile					
RSK Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**Average**
Part II: *t*-Statistics												
1 (Low)	5.53	5.32	4.11	2.82	1.99	4.19	5.79	3.21	2.52	2.21	1.73	3.49
2	4.84	4.00	3.06	1.62	1.44	3.20	4.45	3.56	2.02	3.25	1.42	3.30
3	4.96	3.79	2.58	2.21	1.15	3.15	4.86	3.35	3.01	1.55	1.13	3.12
4	4.01	3.79	2.86	1.73	1.12	2.87	5.03	3.26	2.44	2.17	1.62	3.22
5 (High)	3.41	3.05	1.09	1.17	0.72	2.03	4.19	4.35	2.91	2.49	1.38	3.48
High-Low	−5.79	−6.79	−6.74	−4.40	−3.12	−8.56	−2.27	1.17	0.48	−0.04	−0.69	−0.49
Alpha	−5.33	−6.78	−6.68	−4.61	−3.19	−8.57	−2.06	1.50	0.53	−0.17	−0.61	−0.29

Notes: See notes to Table 8. Portfolios are sorted by RSK, controlling for SRVLJ/SRVSJ, and using truncation level γ^2, as discussed in the notes to Table 2.

Table 12. Double-Sorted Portfolios: Portfolios Independently Sorted by Stock- and Industry-Level SRVJ.

	Equal-Weighted Returns and Alphas							Value-Weighted Returns and Alphas						
	Industry-Level Quintile							Industry-Level Quintile						
Stock-Level Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**
Part I: Mean Return and Alpha														
1 (Low)	39.23	47.42	53.00	61.43	72.18	32.94	34.74	36.50	33.40	32.78	36.58	25.03	10.19	9.28
2	30.55	33.28	38.69	48.44	55.77	25.22	25.86	28.37	27.31	20.24	37.10	14.64	7.20	6.68
3	14.07	21.47	24.27	30.98	43.15	29.07	29.60	18.16	24.85	17.56	20.35	8.69	10.64	11.98
4	6.90	12.45	15.65	19.51	34.51	27.61	28.30	12.97	11.96	10.38	14.58	2.67	9.54	9.99
5 (High)	−7.35	1.99	9.90	16.40	25.94	33.29	32.92	−4.30	9.86	4.86	13.25	−3.95	19.82	21.72
High-Low	−46.58	−45.43	−43.09	−45.03	−46.24			−40.80	−23.54	−27.92	−23.33	−31.16		
Alpha	−44.64	−45.23	−41.18	−45.96	−46.46			−41.42	−22.73	−26.52	−23.09	−28.98		
Industry-Level Effect (average of High-Low column; Alpha column)						29.63	30.29						11.48	11.93
Stock-Level Effect (average of High-Low row; Alpha row)						−45.28	−44.70						−29.35	−28.55

Table 12. *Cont.*

	Equal-Weighted Returns and Alphas							Value-Weighted Returns and Alphas						
	Industry-Level Quintile							Industry-Level Quintile						
Stock-Level Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**
Part II: t—Statistics														
1 (Low)	3.82	4.37	4.98	5.87	6.83	5.20	5.35	4.61	3.98	3.93	4.55	4.71	1.40	1.27
2	2.90	3.13	3.49	4.48	5.22	3.50	3.35	3.82	3.56	2.35	4.58	3.16	1.14	1.00
3	1.39	2.09	2.28	3.11	4.43	4.60	4.53	2.34	3.23	2.25	2.76	1.87	1.54	1.73
4	0.70	1.28	1.60	2.12	3.93	4.40	4.47	1.69	1.62	1.36	1.87	0.68	1.50	1.54
5 (High)	−0.79	0.22	1.07	1.88	3.16	5.88	5.93	−0.51	1.27	0.62	1.93	−1.00	2.82	3.06
High-Low	−9.33	−8.12	−8.27	−7.80	−8.01			−7.19	−3.92	−5.22	−4.31	−5.33		
Alpha, FFC4	−9.00	−8.16	−7.97	−7.93	−8.34			−7.37	−3.73	−5.06	−4.31	−5.47		
Industry-Level Effect (average of High-Low column; Alpha column)						5.66	5.57						2.23	2.20
Stock-Level Effect (average of High-Low row; Alpha row)						−11.39	−11.50						−8.88	−8.91

Notes: See notes to Table 8. This table presents average returns and risk-adjusted alphas for portfolios sorted by stock-level and industry-level SRVJ. The sample includes all NYSE, NASDAQ and AMEX listed stocks for the period January 1993 to December 2016. A stock's industry signed jump variation (SRVJ) is the capitalization-weighted average of the SRVJ of all stocks within the industry. At the end of each Tuesday, all stocks in the sample are sorted into quintile portfolios based on stock-level and industry-level SRVJ, independently, resulting in 25 portfolios. Each portfolio is held for one week. The row labeled "Industry-Level Effect" reports average values of one-week ahead returns (and Fama–French–Carhart four-factor alphas in the High-Low (Alpha) column) in Part I (corresponding Newey–West *t*-statistics are given in Part II). The row labeled "Stock-Level Effect" reports the average values of one-week-ahead returns (and alphas) in Part I (corresponding Newey–West *t*-statistics are again given in Part II).

Table 13. Double-Sorted Portfolios: Portfolios Sorted by Stock- and Industry-Level SRVLJ/SRVSJ Independently.

Panel A: Portfolios Sorted Based on SRVLJ														
	Equal-Weighted Returns and Alphas							Value-Weighted Returns and Alphas						
	Industry-Level Quintile							Industry-Level Quintile						
Stock-Level Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**
Part I: Mean Return and Alpha														
1 (Low)	34.64	35.34	35.04	51.12	51.00	16.36	17.03	22.26	24.61	20.46	26.39	25.19	2.92	0.73
2	25.37	25.36	22.38	34.21	36.00	10.64	9.43	18.78	22.97	19.49	23.51	27.29	8.51	7.46
3	17.55	0.89	26.31	34.26	31.56	14.40	11.15	18.90	15.95	11.59	4.81	17.06	−6.76	−10.28
4	13.59	20.91	21.20	27.71	34.08	19.69	19.75	11.26	21.71	10.01	12.25	25.88	13.89	15.35
5 (High)	12.22	11.90	14.43	25.57	25.57	13.36	12.93	14.16	21.60	16.13	20.63	22.85	8.69	11.30
High-Low	−22.42	−23.44	−20.61	−25.55	−25.42			−8.10	−3.00	−4.33	−5.75	−2.33		
Alpha	−22.02	−23.37	−20.45	−25.74	−26.12			−9.58	−3.52	−4.34	−5.40	1.00		
Industry-Level Effect (average of High-Low column; Alpha column)						14.83	14.31						7.24	6.90
Stock-Level Effect (average of High-Low row; Alpha row)						−23.49	−23.54						−4.70	−4.37

Table 13. *Cont.*

Panel A: Portfolios Sorted Based on SRVLJ														
	Equal-Weighted Returns and Alphas							**Value-Weighted Returns and Alphas**						
	Industry-Level Quintile							**Industry-Level Quintile**						
Stock-Level Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**
Part II: *t*-Statistics														
1 (Low)	3.63	3.39	3.33	4.94	5.23	3.40	3.41	3.25	3.06	2.48	3.25	3.16	0.49	0.12
2	2.55	2.44	2.14	3.46	3.73	2.17	1.90	2.70	3.17	2.57	3.20	3.73	1.60	1.32
3	0.78	0.04	0.93	1.41	1.46	0.85	0.65	1.08	0.73	0.51	0.25	1.05	−0.37	−0.54
4	1.32	1.95	1.94	2.74	3.73	3.67	3.73	1.49	2.62	1.20	1.45	3.69	2.43	2.65
5 (High)	1.31	1.24	1.50	2.84	2.99	2.97	2.91	1.83	2.83	2.12	2.83	3.56	1.50	2.09
High-Low	−6.14	−5.13	−4.66	−5.85	−6.19			−1.57	−0.59	−0.90	−1.19	−0.45		
Alpha, FFC4	−6.09	−5.03	−4.58	−5.91	−6.38			−1.93	−0.68	−0.87	−1.09	0.21		
Industry-Level Effect (average of High-Low column; Alpha column)						3.77	3.55						1.77	1.57
Stock-Level Effect (average of High-Low row; Alpha row)						−9.05	−9.14						−2.01	−1.91
Panel B: Portfolios Sorted Based on SRVSJ														
	Equal-Weighted Returns and Alphas							**Value-Weighted Returns and Alphas**						
	Industry-Level Quintile							**Industry-Level Quintile**						
Stock-Level Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**
Part I: Mean Return and Alpha														
1 (Low)	40.39	42.50	43.55	55.02	68.45	28.07	29.88	44.09	37.87	32.07	36.14	45.53	1.44	0.72
2	32.52	40.60	41.29	45.73	59.67	27.15	28.20	24.11	24.09	23.82	27.45	42.09	17.99	18.36
3	15.39	22.96	24.90	30.69	45.29	29.90	30.43	20.22	15.93	21.91	17.63	25.37	5.16	4.88
4	9.07	9.08	17.26	17.20	32.70	23.63	23.94	11.38	12.88	22.53	11.01	25.43	14.04	15.23
5 (High)	−2.44	3.16	7.13	12.21	22.30	24.73	26.19	8.91	7.91	11.60	9.76	11.23	2.32	5.33
High-Low	−42.83	−39.34	−36.43	−42.82	−46.16			−35.18	−29.95	−20.47	−26.38	−34.30		
Alpha	−42.27	−39.03	−34.75	−42.27	−45.96			−36.51	−30.26	−18.68	−25.89	−31.90		
Industry-Level Effect (average of High-Low column; Alpha column)						26.69	27.73						8.19	8.90
Stock-Level Effect (average of High-Low row; Alpha row)						−41.51	−40.86						−29.26	−28.65

Table 13. *Cont.*

Panel B: Portfolios Sorted Based on SRVSJ														
	Equal-Weighted Returns and Alphas							**Value-Weighted Returns and Alphas**						
	Industry-Level Quintile							**Industry-Level Quintile**						
Stock-Level Quintile	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**	**1 (Low)**	**2**	**3**	**4**	**5 (High)**	**High-Low**	**Alpha**
Part II: *t*-Statistics														
1 (Low)	3.91	3.90	3.98	5.06	6.20	4.04	3.99	5.58	4.66	3.66	4.40	5.62	0.21	0.10
2	3.00	3.83	3.88	4.42	5.68	3.89	3.88	3.19	3.04	2.90	3.50	5.39	2.84	2.77
3	1.55	2.26	2.46	3.11	4.68	4.91	4.94	2.52	1.97	2.73	2.33	3.43	0.70	0.68
4	0.92	0.93	1.74	1.84	3.64	3.72	3.74	1.46	1.78	2.85	1.47	3.72	2.18	2.43
5 (High)	−0.25	0.33	0.77	1.41	2.70	3.94	4.15	1.06	0.96	1.49	1.36	1.74	0.34	0.76
High-Low	−8.38	−7.85	−7.04	−7.99	−7.60			−6.21	−5.45	−3.70	−5.15	−6.15		
Alpha, FFC4	−8.38	−7.95	−6.85	−8.19	−8.02			−6.55	−5.68	−3.41	−5.15	−5.69		
Industry-Level Effect (average of High-Low column; Alpha column)						5.02	5.01						1.60	1.68
Stock-Level Effect (average of High-Low row; Alpha row)						−10.65	−11.14						−8.78	−8.75

Notes: See notes to Table 12. Jumps are decomposed using truncation level γ^2, as discussed in the footnote to Table 2.

Table 14. Fama–MacBeth Cross-Sectional Regressions.

Panel A: Regressions Without Control Variables																
	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**	**XIII**	**XIV**	**XV**	**XVI**
Intercept	18.54	27.95	23.99	15.77	30.31	31.03	31.46	31.20	19.74	32.04	20.01	20.04	28.88	30.01	30.48	30.17
	(1.94)	(3.07)	(2.82)	(1.69)	(3.32)	(3.37)	(3.41)	(3.39)	(2.03)	(3.73)	(2.14)	(2.08)	(3.29)	(3.40)	(3.44)	(3.41)
RVJP	−63.86								−128.25							
	(−6.00)								(−6.24)							
RVJN	107.11								196.57							
	(8.29)								(8.98)							
RVLJP		−53.42		−44.85						76.84		−79.63				
		(−6.46)		(−4.46)						(6.58)		(−3.94)				
RVLJN		71.27		83.09						−30.83		149.92				
		(8.12)		(7.45)						(−2.40)		(7.18)				
RVSJP			−130.77	−99.16							−88.56	−129.39				
			(−8.97)	(−6.24)							(−6.64)	(−6.33)				
RVSJN			165.05	161.39							129.24	195.31				
			(8.22)	(7.19)							(7.94)	(8.26)				
SRVLJ					−50.07		−53.94						72.19		−82.69	
					(−7.98)		(−8.37)						(6.60)		(−4.48)	
SRVSJ						−141.69	−144.75							−103.72	−149.56	
						(−9.25)	(−9.32)							(−8.25)	(−7.66)	
SRVJ								−81.15								−150.59
								(−10.15)								(−7.80)
RVOL									−8.94	−7.46	−6.74	−9.08	−5.90	−6.31	−6.40	−6.38
									(−1.60)	(−1.32)	(−1.21)	(−1.62)	(−1.05)	(−1.12)	(−1.14)	(−1.13)
RSK									16.12	−22.16	−9.87	9.12	−24.75	−10.15	4.08	14.02
									(5.59)	(−9.55)	(−9.49)	(3.07)	(−10.41)	(−9.72)	(1.39)	(4.91)
RKT									−0.68	−0.68	0.46	−0.68	0.12	0.09	0.08	0.09
									(−2.25)	(−2.27)	(1.45)	(−2.24)	(0.42)	(0.30)	(0.28)	(0.32)
Adjusted R^2	0.0063	0.0033	0.0035	0.0082	0.0005	0.0019	0.0024	0.0016	0.0204	0.0175	0.0185	0.0214	0.0160	0.0168	0.0172	0.0168

Table 14. *Cont.*

Panel B: Regressions with Control Variables																
	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**	**XIII**	**XIV**	**XV**	**XVI**
Intercept	100.76	100.26	92.97	98.02	97.60	97.72	98.57	98.26	89.27	92.87	93.62	89.60	94.59	94.82	95.19	94.94
	(4.24)	(5.45)	(5.25)	(4.15)	(5.67)	(5.65)	(5.69)	(5.67)	(3.30)	(4.26)	(3.93)	(3.27)	(4.37)	(4.33)	(4.30)	(4.29)
RVJP	−30.35								−33.59							
	(−3.04)								(−1.87)							
RVJN	28.77								50.58							
	(3.26)								(3.48)							
RVLJP		−27.56		−27.07						11.55		−28.36				
		(−4.46)		(−2.81)						(1.20)		(−1.53)				
RVLJN		16.42		23.05						−0.32		48.59				
		(2.67)		(2.61)						(−0.03)		(2.89)				
RVSJP			−26.94	−34.15							−24.07	−38.20				
			(−2.78)	(−2.51)							(−2.20)	(−2.04)				
RVSJN			44.62	45.83							26.48	52.34				
			(4.25)	(3.81)							(2.67)	(3.42)				
SRVLJ					−22.63		−25.76						9.90		−31.02	
					(−5.18)		(−5.67)						(1.10)		(−1.93)	
SRVSJ						−33.16	−38.71							−23.75	−41.74	
						(−3.92)	(−4.45)							(−2.74)	(−2.83)	
SRVJ								−28.64								−39.38
								(−6.26)								(−2.69)
RVOL									4.79	5.07	4.59	4.68	4.94	4.87	4.86	4.84
									(0.79)	(0.84)	(0.76)	(0.77)	(0.82)	(0.80)	(0.80)	(0.79)
RSK									3.02	−4.90	−3.58	2.67	−5.53	−3.67	1.42	2.46
									(1.23)	(−3.01)	(−4.69)	(0.98)	(−3.45)	(−4.79)	(0.50)	(0.95)
RKT									−0.53	−0.61	−0.38	−0.56	−0.49	−0.46	−0.44	−0.44
									(−2.00)	(−2.27)	(−1.08)	(−1.87)	(−1.78)	(−1.63)	(−1.49)	(−1.46)
Beta	−8.28	−8.11	−8.09	−8.27	−8.07	−8.26	−8.29	−8.16	−7.71	−7.75	−8.18	−7.72	−8.10	−8.18	−8.22	−8.14
	(−1.46)	(−1.40)	(−1.38)	(−1.46)	(−1.37)	(−1.41)	(−1.41)	(−1.39)	(−1.36)	(−1.35)	(−1.42)	(−1.37)	(−1.40)	(−1.42)	(−1.43)	(−1.41)
log(Size)	−14.93	−14.77	−14.73	−14.94	−14.76	−14.76	−14.67	−14.67	−14.36	−14.25	−14.29	−14.48	−14.13	−14.11	−14.04	−14.05
	(−5.24)	(−5.25)	(−5.21)	(−5.32)	(−5.15)	(−5.14)	(−5.13)	(−5.13)	(−5.19)	(−5.15)	(−5.16)	(−5.26)	(−5.08)	(−5.07)	(−5.06)	(−5.07)
BE/ME	−0.76	−0.75	−0.65	−0.76	−0.67	−0.61	−0.59	−0.62	−0.72	−0.57	−0.57	−0.69	−0.56	−0.54	−0.56	−0.58
	(−0.37)	(−0.36)	(−0.32)	(−0.37)	(−0.33)	(−0.30)	(−0.29)	(−0.30)	(−0.34)	(−0.27)	(−0.27)	(−0.33)	(−0.27)	(−0.26)	(−0.27)	(−0.28)
MOM	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	(0.96)	(0.97)	(0.94)	(0.98)	(0.93)	(0.93)	(0.93)	(0.93)	(1.26)	(1.19)	(1.18)	(1.22)	(1.19)	(1.21)	(1.20)	(1.21)
REV	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
	(−5.63)	(−5.83)	(−5.63)	(−5.53)	(−5.84)	(−5.66)	(−5.55)	(−5.64)	(−5.74)	(−5.97)	(−5.82)	(−5.70)	(−5.99)	(−5.83)	(−5.77)	(−5.80)
IVOL	−301.46	−293.89	−297.02	−304.39	−291.92	−292.47	−298.32	−298.14	−425.97	−416.77	−422.83	−424.54	−419.05	−421.52	−424.24	−425.20
	(−2.18)	(−2.15)	(−2.17)	(−2.21)	(−2.13)	(−2.12)	(−2.16)	(−2.17)	(−4.69)	(−4.57)	(−4.63)	(−4.67)	(−4.59)	(−4.62)	(−4.65)	(−4.67)
CSK	−7.52	−8.67	−8.13	−7.29	−8.69	−8.23	−7.62	−7.87	−7.00	−7.84	−7.44	−6.89	−7.91	−7.45	−7.35	−7.52
	(−1.75)	(−2.01)	(−1.89)	(−1.70)	(−2.01)	(−1.91)	(−1.77)	(−1.82)	(−1.66)	(−1.86)	(−1.77)	(−1.64)	(−1.88)	(−1.77)	(−1.74)	(−1.78)
CKT	2.34	2.29	2.24	2.36	2.32	2.32	2.43	2.38	1.74	1.72	1.69	1.74	1.71	1.77	1.83	1.81

Table 14. *Cont.*

Panel B: Regressions with Control Variables																
	I	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**	**IX**	**X**	**XI**	**XII**	**XIII**	**XIV**	**XV**	**XVI**
	(1.19)	(1.15)	(1.13)	(1.20)	(1.15)	(1.16)	(1.21)	(1.18)	(0.91)	(0.89)	(0.87)	(0.91)	(0.88)	(0.91)	(0.94)	(0.93)
MAX	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03	−0.03
	(−5.33)	(−5.66)	(−5.57)	(−5.31)	(−5.64)	(−5.71)	(−5.55)	(−5.54)	(−7.37)	(−7.55)	(−7.45)	(−7.35)	(−7.59)	(−7.55)	(−7.57)	(−7.55)
MIN	−0.02	−0.02	−0.02	−0.02	−0.02	−0.02	−0.02	−0.02	−0.01	−0.02	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
	(−2.79)	(−3.11)	(−3.14)	(−2.79)	(−3.11)	(−3.12)	(−2.83)	(−2.85)	(−2.81)	(−2.94)	(−2.80)	(−2.81)	(−2.88)	(−2.80)	(−2.73)	(−2.75)
ILLIQ	−7.84	−7.68	−8.08	−7.86	−7.99	−8.03	−7.87	−7.88	−8.94	−8.87	−8.68	−9.10	−8.54	−8.51	−8.41	−8.43
	(−5.24)	(−5.12)	(−5.22)	(−5.26)	(−5.15)	(−5.16)	(−5.08)	(−5.10)	(−4.79)	(−5.12)	(−4.96)	(−4.87)	(−4.97)	(−4.95)	(−4.86)	(−4.88)
Adjusted R^2	0.0602	0.0597	0.0597	0.0609	0.0590	0.0592	0.0594	0.0592	0.0647	0.0641	0.0642	0.0652	0.0636	0.0637	0.0639	0.0638

Notes: See notes to Tables 1 and 5. This table reports results for cross-sectional Fama–MacBeth regressions, based on the regression model depicted as Equation (15) in Section 4.5. In these regression models, future weekly returns are regressed on various realized measures and control variates. The two panels use jump truncation level γ^2, as discussed in the footnote to Table 2. The regressions that are reported on are of the form: $r_{i,t+1} = \gamma_{0,t} + \sum_{j=1}^{K_1} \gamma_{j,t} X_{i,j,t} + \sum_{s=1}^{K_2} \phi_{s,t} Z_{i,s,t} + \epsilon_{i,t+1}$, $t = 1, \ldots, T$, where $r_{i,t+1}$ denotes the stock return for firm i in week $t+1$, K_1 is the number of potential variation measures, and $X_{i,j,t}$ denotes a relevant realized measure at the end of week t. In addition, there are K_2 variables measuring firm characteristics, which are denoted by $Z_{i,j,t}$ (see Section 3 for details). In the table, time series averages of the coefficient estimates ($\frac{1}{T}\sum_{t=1}^{T} \widehat{\gamma}_{j,t}$ and $\frac{1}{T}\sum_{t=1}^{T} \widehat{\phi}_{j,t}$) are reported, along with Newey–West t-statistics (in parentheses). For complete details, see Section 4.

4.1. Single (Univariate) Portfolio Sorts Based on Realized Measures

In this section, we first discuss the results contained in Table 4. Recall that the Mean Return in this table is an average taken over our entire time series of equal-weighted and value-weighted portfolio returns, for single-sorted portfolios based on positive jump variation (RVJP), negative jump variation (RVJN) and signed jump variation (SRVJ). Values in parentheses are Newey–West t-statistics (see Bollerslev et al. (2015) and Petersen (2009) for further discussion). Panel A provides results for portfolios sorted by RVJP. Inspection of the entries in this panel indicates that mean returns and alphas of high–low portfolios (i.e., the difference in returns (alphas) between the fifth and first quintiles) are all negative, indicating a negative association between RVJP and subsequent stock returns. Interestingly, the alpha of -7.71 basis points (bps) is insignificant for the high–low spread for the equal-weighted portfolio, while the alpha of -5.63 bps is only significant at a 10% level for the high–low spread for the value weighted portfolio.

The lack of statistical significance for some of the mean return values reported in Panel A does not characterize our findings when negative and signed jump variation measures are used for sorting. Moreover, the magnitudes of the mean returns and alphas are usually three or more times larger when sorting on negative and signed jump variation (to see this, turn to Panels B and C of Table 4). In Panel B, the high–low spread of mean returns equals 36.06 bps, with a t-statistic of 6.47 for the equal-weighted portfolio, and 15.13 bps with a t-statistic of 3.75 for the value-weighted portfolio. Moreover, both equal-weighted and value-weighted portfolios generate significant positive abnormal future returns measured by the alphas. These results clearly point to a statistically significant positive association between negative jump variation and the following week's returns.

Panel C in Table 4 contains results for portfolios sorted by signed jump variation. The negative high–low spreads indicate a statistically significant negative association between signed jump variation and future returns. In particular, a strategy buying stocks in the lowest signed jump variation quintile and selling stocks in the highest signed jump variation quintile earns a mean return of 40.82 bps with a t-statistic of 9.85 each week for the equal-weighted portfolio and 25.02 bps with a t-statistic of 5.78 for the value-weighted portfolio. These results are consistent with the results reported in Bollerslev et al. (2020). Interestingly, almost all of the mean returns listed in Table 4 are "alpha" (see tabulated average alphas in the table), and cannot be explained by standard portfolio risk factors using regressions of the type given above as Equation (14).

A key question that we provide evidence on in this paper is whether the results summarized in Table 4 carry over to the case where small and large jump variation is separately sorted on. First, consider large jumps. Table 5 reports the results for portfolios sorted by positive, negative and signed large jump variation, respectively. Similar to positive jump variation, positive large jumps negatively predict subsequent returns, but the predictability is not significant, regardless of the truncation level (γ) used to separate small and large jumps, and regardless of portfolio weighting used. This is demonstrated by the fact that the t-statistics for mean returns and alphas of high–low portfolios all indicate insignificance, at a 5% testing level, regardless of truncation level. Thus, there is no ambiguity, as in Panel A of Table 4. Positive jump variation is not a significant predictor, under our large jump scenario. On the other hand, we shall see that sorting on small and large negative variation measures yields significant excess returns, as does sorting on small positive jump variation, under both equal and value weighting schemes.

As just noted, equal-weighted high–low portfolios sorted on large negative jump variation generate significant positive returns and alphas (see Panel B of Table 5). However, analogous returns and alphas under value weighting are not significant. Signed large jump variation is sorted on in Panel C of Table 5. Signed large jump variation is useful for undertaking a long-short trading strategy based on the difference between large upside and downside jump variation measures. Inspection of the results in this panel of the table reveals that the high–low spread for the equal-weighted portfolio generates an average risk-adjusted weekly return of -28.36 bps (with a t-statistics of -9.39) and -9.25 bps (with a t-statistics of -2.87) for the value-weighted portfolio, for truncation level equal to γ^1.

Results based on γ^2 (i.e., our larger truncation level) are also significant, although magnitudes are lesser and only for our equal-weighted portfolio.[16] In particular, observe that when large jump variation is constructed using γ^2, the high–low spreads for value-weighted portfolios sorted by downside or signed large jump variation measures are insignificant, suggesting that small firms have stronger relationships (than larger firms) between signed (or negative) large jump variation and subsequent returns. This may be because smaller firms are in some ways more susceptible to changing market conditions than larger firms.

Table 6 summarizes results analogous to those reported in Table 5, but for positive, negative and signed small jump variation measures. Similar to large jump measures, positive and signed small jump variation measures negatively predict future returns, and negative small jump variation measures positively predict returns in the following week. By contrast, the differences in average (risk-adjusted) returns between equal-weighted and value-weighted long-short portfolios based on RVSJP and RVSJN are smaller than those for portfolios based on large jumps (compare the entries for the high–low quintiles under the two weighting schemes in Panels A and B of Table 6 with like entries in Panels A and B of Table 5). These results indicate that big firms have a stronger relationship between small jump variation and future returns than that between large jumps and subsequent weekly returns. Since stocks for big firms are more liquid and price discovery is more rapid, predictability associated with large jumps is much weaker or insignificant for big firms. This finding is in line with Bollerslev et al. (2020), who document that the predictability of signed jump variation is stronger for small and illiquid firms and is driven by investor overreaction. In addition, when using our larger truncation level, γ^2, value-weighted high–low spreads based on signed small jump variation are larger than those based on signed total jump variation and signed large jump variation. This result implies that a long-short strategy associated with signed small jump variation generates the highest value-weighted risk-adjusted returns, given the use of an appropriate truncation level to separate small and large jumps.

As discussed above, in order to understand why small jumps matter in our analysis, note that Scaillet et al. (2018) document that jumps are frequent events and that jump dynamics are not consistent with the compound Poisson processes with constant intensity. In addition, many recent studies find that infinite-activity Levy jump specifications (e.g., Variance Gamma and Normal Inverse Gaussian return jumps) can improve the goodness of fit and option pricing performance of models (for example, see Yang and Kanniainen (2016)). These results point to the possible importance of small jumps, which may have infinite numbers in a finite time interval. Though more frequent than large jumps, note that we estimate daily jump variation measures and aggregate them into weekly measures for forecasting future weekly returns. Thus, it is not individual small jumps that necessarily matter, but the overall contribution of small jumps to total variation, in the cross-section. Additionally, we show (see below discussion) that small jump variation is more closely associated with idiosyncratic risk than are large jump variation. If idiosyncratic volatility matters (see Ang et al. (2006, 2009)), so do small jumps, at least to some extent. Indeed, our results partially help explain the idiosyncratic volatility puzzle (investors prefer positive skewness/asymmetry, so they will accept lower returns for stocks with a higher probability of having positive jumps).

Table 7 reports results for portfolios sorted by realized volatility, realized skewness, realized kurtosis, and continuous variance. Consistent with the results in Amaya et al. (2015) and Bollerslev et al. (2020), there is a significant negative relationship between realized skewness and future returns, while the association is not significant between either realized volatility or realized kurtosis and returns in the following week, regardless of portfolio weighting scheme. In addition,

[16] Empirical findings based on γ^3 are similar to those discussed above, and hence are not reported. This robustness of our findings to the choice of γ also characterizes the other empirical findings discussed in the sequel.

continuous variance significantly and negatively predicts one-week-ahead returns for equal-weighted portfolios, but this negative association is not significant for value-weighted portfolios.

4.2. Cumulative Returns and Sharpe Ratios

Not surprisingly, our findings based on univariate portfolio sorts suggest that strategies that use different realized measures deliver different risk-adjusted average returns. In order to investigate this result further, we calculate cumulative returns and Sharpe ratios for short-long portfolios, sorted on various risk measures that are described in Table 1, including SRVJ, RSJ, SRVLJ, SRVSJ, and RSK. In addition, for comparison purposes, we also carry out our analysis using the relative signed jump variation measure (called RSJ) that is examined by Bollerslev et al. (2020). Our experiments are carried out as follows. Beginning in January 1993, various short-long portfolios are constructed, with an initial investment of \$1. These portfolios are re-balanced and accumulated at a weekly frequency, until the end of 2016.[17] Figure 3 plots portfolio values over time. Consistent with our results based on single portfolio sorts, inspection of the plots in this figure indicates that for equal-weighted portfolios sorting on signed jump variation (SRVJ) yields the largest portfolio accumulations; and for value-weighted portfolios, sorting on signed small jump variation (SRVSJ) yields the largest portfolio accumulations.[18]

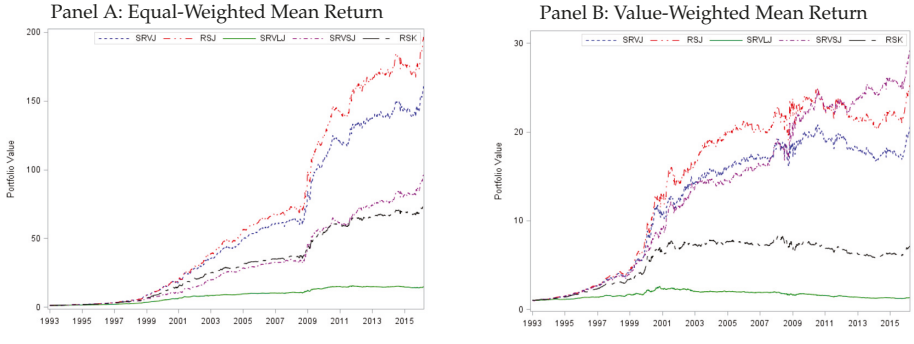


Figure 3. Cumulative Gains of Short-Long Portfolios. Panels A–B display cumulative gains of equal-weighted and value-weighted short-long portfolios constructed using SRVJ, SRVLJ, SRVSJ, and RSK (see Table 1 and Section 2 for a discussion of these measures). RSJ is the relative signed jump variation measure defined and analyzed in Bollerslev et al. (2020), who include the risk-free rate in all of their calculations, while we do not (refer to Bollerslev et al. (2020) for complete details). In all experiments, the initial investment, made on January 1993, is \$1. Each portfolio is re-balanced and accumulated on a weekly basis, through 2016. Signed small and large jump variation measures used in the experiment reported on in this figure are constructed based on truncation level $\gamma^2 = 5\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\,\Delta_n^{0.49}$. See Section 4.2 for further discussion.

Now, consider the Sharpe ratios reported below, which are reported for various jump measures, and are constructed based on truncation level $\gamma^2 = 5\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\,\Delta_n^{0.49}$.

---

[17] Cumulative returns calculations do not include the risk-free rate. For a definition of cumulative returns both with and without the weekly risk-free rate, see Bollerslev et al. (2020).

[18] Please note that RSJ, which measures the same signed jump variation as SRVJ, although using different estimation methodology, generates the highest cumulative return for equal-weighted portfolios, but is dominated by SRVSJ for value-weighted portfolios.

Table 15. Sharpe Ratios.

	SRVJ	RSJ	SRVLJ	SRVSJ	RSK
Equal-Weighted	2.1342	2.1363	1.8556	1.8161	2.2234
Value-Weighted	1.1322	1.1310	0.1611	1.2755	0.8665

The entries in Table 15 are Sharpe ratios for equal- and value-weighted short-long portfolios constructed using SRVJ, RSJ, SRVLJ, SRVSJ, and RSK. Recall that RSK is realized skewness (see Table 1 for definitions of these measures). The sample of stocks used for Sharpe ratio calculations includes all NYSE, NASDAQ and AMEX listed stocks for the period January 1993 to December 2016. At the end of each Tuesday, all the stocks in the sample are sorted into quintile portfolios based on ascending values of various realized risk measures. A high–low spread portfolio is then formed as the difference between portfolio 1 and portfolio 5, and held for one week, where 1 and 5 refer to quintiles, as in Tables 3–7. The Sharpe ratio is calculated with the one-week-ahead returns.

Interestingly, for equal-weighted portfolios, the RSK-based short-long strategy yields the highest Sharpe ratio (i.e., 2.2234), although the ratio of 2.1342 for SRVJ is approximately the same. Still, the success of the RSK measure is likely due to its relatively stable performance, compared with other jump-based strategies. This finding is similar to the findings discussed in Xiong et al. (2016), who show that tail risks can be substantially reduced by forecasting skewness. Note also that the signed small jump variation (SRVSJ)-based portfolio has the highest Sharpe ratio, among all value-weighted portfolios. However, it is clear that all equal-weighted portfolios outperform their corresponding value-weighted counterparts. This result is consistent with the finding discussed above that small and illiquid firms tend to react more strongly to realized risk measures.

4.3. Double Portfolio Sorts Based on Realized Measures

To further investigate whether small and large jumps are priced differently, we use double portfolio sorts. In particular, we carry out double-sorts in order to examine the robustness of our findings based on single sorts, after controlling for other realized measures. Table 8 reports returns and alphas from various of these sorts in which we alternate the sorting order among SRVJ, SRVLJ and SRVSJ. When we first sort by total jump variation, and then sort stocks based on SRVLJ or SRVSJ, a negative relation only exists between SRVSJ and subsequent weekly returns (see Panels A and B of the table). This result indicates that there is no marginal predictive content associated with large jumps, when conditioning on the predictive content associated with total jump variation, while small jumps have unique information for predicting future returns, even compared to total jumps.

Panel C reports returns and alphas based on sorting on SRVSJ after controlling for SRVLJ. Both the equal- and value-weighted high–low spreads and alphas are statistically significant in this case, while this is not the case if stocks are first sorted by SRVSJ and then by SRVLJ, as shown in Panel D. More specifically, the high–low return is -25.38 bps (with a t-statistic of -6.61), for the value-weighted portfolio in Panel C, and is -3.77 bps with a t-statistics of -1.49 in Panel D, for the value-weighted portfolio. This indicates that the predictable content in large jumps becomes negligible after controlling for small jumps.

Bollerslev et al. (2020) document that the negative association between realized skewness and one-week-ahead returns is reversed when controlling for the signed jump variation. To further investigate the relationship between skewness and different jump variation measures, we use double portfolio sorts to control for different effects that are associated with cross-sectional variation in future returns.

Panel A of Table 9 reports average returns and corresponding t-statistics for 25 portfolios sorted by SRVJ (signed jump variation), controlling for realized skewness (RSK). Inspection of the results in this table indicates that the negative association between SRVJ and future returns still exists, after controlling for RSK, indicating that there is unique predictive information contained in signed

jump variation. Panel B in this table reports results for portfolios sorted first by SRVJ and then by RSK. The high–low spreads of the averaged portfolios are positive after controlling for SRVJ, confirming the results reported in Bollerslev et al. (2020).

Panel A of Table 10 contains results for portfolios sorted by SRVLJ (signed large jump variation) after controlling for RSK. As noted above, the negative association between SRVLJ and future returns is reversed after controlling for skewness. By contrast, this issue does not exist for portfolios sorted by SRVSJ (signed small jump variation) when controlling for skewness, as shown in Panel B of Table 10, indicating that signed small jump variation has unique information about future return premia. However, first accounting for skewness negates the usefulness that signed large jump variation has for predicting future returns. This finding serves as an important distinction between the predictive content of small and large jumps.

Finally, Table 11 contains results for portfolios sorted on RSK, after controlling for SRVLJ and SRVSJ, respectively. Inspection of the entries in this table indicates that the high–low spreads are negative, except in select value-weighted portfolio cases, when controlling for SRVSJ. This is not surprising since skewness captures information from both SRVLJ and SRVSJ, while the negative association between realized skewness and subsequent returns remains, when controlling for either SRVLJ or SRVSJ, in most cases. Of note is that this negative association disappears for some value-weighted portfolios, when controlling for SRVSJ, suggesting that signed small jump variation (especially for big firms) is the main driver of the signed total jump variation. These findings are consistent with the findings documented by Bollerslev and Todorov (2011) that S&P 500 market portfolios tend to have symmetric jump tails (large jumps).

4.4. Using Double Portfolio Sorts to Examine Stock-Level versus Industry-Level Predictability

In this section, we carry out an additional set of double portfolio sort experiments, in which industry-based investing is compared with individual stock-based investing. Our earlier findings indicate that low signed jump variation investing (buying stocks with low signed jump variation and shorting stocks with high signed jump variation) can deliver significant risk-adjusted returns (this is a result also found by Bollerslev et al. (2020), for example). In order to examine whether this investment strategy relies on industry betting or stock selection within industries (or both), we form double-sorted portfolios based on industry-level and stock-level signed jump risk variation. In particular, each Tuesday we group stocks into 49 industries based on SIC codes. Industry-level signed jump risk is calculated as the value-weighted average of signed (large/small) jump variation measures for stocks within each industry. Thus, stocks in the same industry have the same industry signed (large/small) jump variation during a given week. Stock-level signed jump risk is calculated as outlined in the above. Double sorts are then used to investigate the selection effects at industry- and stock-level. Specifically, stocks are sorted into 25 portfolios based on industry- and stock-level signed (large/small) jump variation quintiles. With this particular variety of sorting, results are independent of the order in which stocks are sorted.

Figure 4 depicts the percentage of stocks in each portfolio (see Panel A), and the market capitalization in these portfolios (see Panel B). If industry-level selection and stock-level selection lead to different quintile portfolios (i.e., off-diagonal portfolios in the figures have non-zero membership), it is possible to separate these two effects using double-sorts. Specifically, there are different industry- and stock-level effects. Both panels indicate this to be the case.

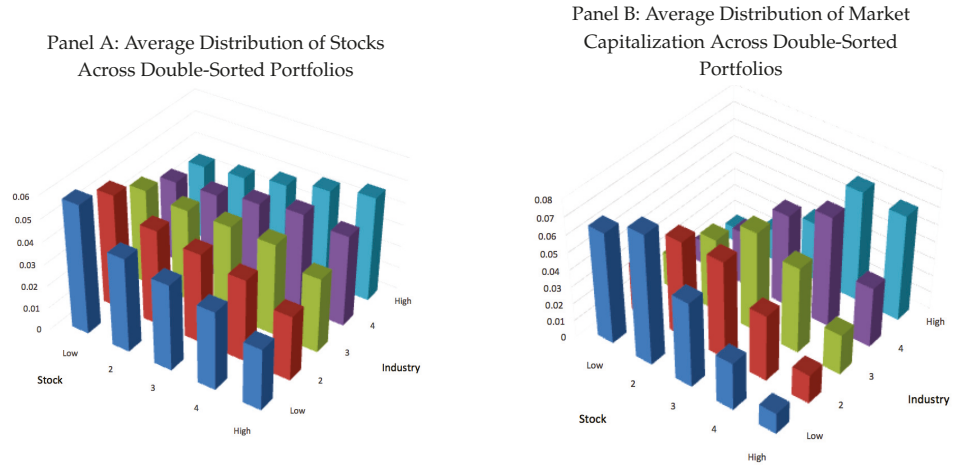


Figure 4. Distribution of Stocks in Portfolios Formed Based on Stocks' Signed Jump Variation (SRVJ) and Industry Signed Jump Variation. The vertical axis in Panels A and B measures time series average proportions of stocks and market capitalizations, across double-sorted portfolios.

Tables 12 and 13 report our empirical findings based on our double portfolio sort experiments. In particular, Table 12 reports results for sorting done on signed jump variation (SRVJ), while Table 13 reports results for sorting done on signed large jump variation (SRVLJ) and signed small jump variation (SRVSJ), respectively. Entries in the tables are mean returns and alphas, as in previous tables. However, in these tables we also report industry-level effects and stock-level effects. These are reported in the last two rows of entries in each panel of the tables. The first of these two rows, called "Industry-Level Effect" reports average high–low returns and alphas by averaging across quintiles in the high–low and alpha columns of the table (these are industry-level results). The second of these two rows, called "Stock-Level Effect" reports average high–low returns and alphas by averaging across quintiles in the high–low and alpha rows of the table (these are stock-level results). Summarizing, rows in these tables display portfolios formed by stocks in the same stock-level SRVJ, SRVLJ, or SRVSJ quintiles, while columns report results for portfolios formed by stocks in the same industry-level SRVJ, SRVLJ, or SRVSJ quintiles.

Turning to Table 12, notice, for example, that a strategy of buying stocks in the highest industry SRVJ quintile and selling stocks in the lowest industry SRVJ quintile generates an equal-weighted average return of 29.63 bps with a t-statistic of 5.66, and the corresponding value-weighted average return is 11.48 bps with a t-statistics of 2.23 (see Table 12). This finding is interesting, as it suggests that the negative association between SRVJ and future returns is reversed at the industry level. The equal-weighted average of the high–low row (i.e., the average stock-level effect) is -45.28 bps with a t-statistic of -11.39 and the alpha is -44.70 bps with a t-statistic of -11.50, indicating that the stock-level effect is economically significant. At the stock-level, investors prefer stocks with high SRVJ, requiring lower returns under higher SRVJ, given that there is a large probability of extremely large positive jumps. By contrast, when sorting at the industry-level, investors are more interested in industry exposure with lower SRVJ, or in return distributions concentrated to the right. Lottery-like payoff exposure comes from individual stocks, not from industry bets. These results are mirrored in Table 13, where SRVLJ and SRVSJ are the sorting measures. However, average stock- and industry-level returns and alphas are much higher under SRVSJ sorting than under SRVLJ sorting. For example, buying stocks in the highest industry SRVLJ quintile and selling stocks in the lowest industry SRVLJ quintile generates an equal-weighted average return of 14.83 bps with a t-statistic of 3.77 under SRVLJ

sorting (see Panel A of Table 13), versus an equal-weighted average return of 26.69 bps with a t-statistic of 5.02 under SRVSJ sorting (see Panel B of Table 13).

4.5. Firm-Level Fama–MacBeth Regressions

Table 14 gathers results based on firm-level Fama–MacBeth regressions, which we run in order to investigate the return predictability associated with variation measures, when controlling for multiple firm specific characteristics. Regressions are carried out as follows. At the end of each Tuesday, we run the cross-sectional regression,

$$r_{i,t+1} = \gamma_{0,t} + \sum_{j=1}^{K_1} \gamma_{j,t} X_{i,j,t} + \sum_{s=1}^{K_2} \phi_{s,t} Z_{i,s,t} + \epsilon_{i,t+1}, \quad t = 1, \ldots, T, \tag{15}$$

where $r_{i,t+1}$ denotes the stock return for firm i in week $t+1$, K_1 is the number of potential variation measures, and $X_{i,j,t}$ denotes a relevant realized measure at the end of week t. In addition, there are K_2 variables measuring firm characteristics, which are denoted by $Z_{i,j,t}$ (see Section 3 for details). After estimating the cross-sectional regression coefficients on a weekly basis, we form the time series average of the resulting T weekly $\widehat{\gamma}_{j,t}$ and $\widehat{\phi}_{s,t}$ values, in order to estimate the average risk premium associated with each risk measure. Specifically, we construct

$$\widehat{\gamma}_j = \frac{1}{T} \sum_{t=1}^{T} \widehat{\gamma}_{j,t}, \quad \text{and} \quad \widehat{\phi}_s = \frac{1}{T} \sum_{t=1}^{T} \widehat{\phi}_{s,t}, \quad \text{for} \quad j = 1, \ldots, K_1, \; s = 1, \ldots, K_2.$$

Panel A of Table 14 reports results for regressions on various realized variation measures, without controlling for firm specific characteristics. Consistent with our results based on univariate sorting, signed jump variation (SRVJ) significantly negatively predicts cross-sectional variation, in these weekly returns regressions. Additionally, both signed small and large jump variation measures negatively predict future weekly returns. Finally, both small and large upside (downside) jump variation measures negatively (positively) predict subsequent weekly returns. However, when including measures that contain information from both small and large jump variation measures, as well as realized skewness, the negative association between skewness and future returns is reversed (see the results for the regressions labeled IX, XII, XV, XVI). In particular, skewness drives out signed large jump variation in regression XIII by reverting the negative association between the latter and future returns. If only small jumps are considered to be control variables, skewness still negatively predicts future returns. This again indicates that signed small jump variation has unique and significant information about future returns.

Panel B of Table 14 reports regression results for the same set of regressions in Panel A, but controlling for various firm specific characteristics, ranging from BETA to ILLIQ (see Table 1 for details). In these regressions, signed (small) jump variation is always significant. Additionally, skewness significantly negatively predicts future returns in regressions that only include small jump variation. This provides yet further evidence that signed small jump variation has unique and significant information about future returns, while large jumps have information in common with realized skewness.

4.6. Pricing Distinctions between Small and Large Jumps

The results in previous sections show that small and large jump variation measures contain different information, and thus have different predictive content. To further investigate whether the differences are driven by distinct economic factors, we provide empirical evidence on the inter-relationship between jumps and news.

4.6.1. Jumps and News Announcement

We begin by examining the relationship between jumps and firm-level news announcements. In order to do this, we construct event windows using the approach of Bernard and Thomas (1989). We then plot the dynamics of SRVJ, SRVLJ, and SRVSJ around earnings announcements. In particular, following Livnat and Mendenhall (2006), the earning surprise (SUE) for each stock is defined as

$$SUE_{j,t} = \frac{(X_{j,t} - E_{j,t})}{P_{j,t}}, \tag{16}$$

where $E_{j,t}$ and $X_{j,t}$ denote the analysts' expectations and reported actual earnings per share, respectively. Here, $P_{j,t}$ is the price per share for stock j at the end of quarter t. In a $[-12, 12]$ week event window, where week zero denotes the earning announcement week, stocks are sorted into tertile portfolios by the value of SUE at the end of week zero. We then calculate the equal-weighted and value-weighted average of jump measures for each tertile portfolio at each week. Figure 5 displays various jump variation measures of portfolios with the most negative, median, and positive earning surprises. It turns out that large (both positive and negative) jump variation measures are higher during announcement weeks, regardless of news sentiment (i.e., regardless of whether SUE is positive or negative). However, positive large jump variation (RVLJP) is higher on days with the most positive earning surprises, and negative large jump variation (RVLJN) reaches its peak on days with the most negative earning surprises. In contrast, both small positive and negative jump variation measures (RVSJP and RVSJN) have lower magnitudes during announcement weeks. The size of the reduction associated with small positive jump variation (RVSJP) is larger on days with the most negative earning surprises, while small negative jump variation (RVSJN) decreases the most on days with the most positive surprises. For signed jump variation, jump magnitudes increase (relative to non-earnings-surprise weeks) on positive surprise days and decrease on negative surprising days. These results indicate that big news, regardless of sentiment, simultaneously leads to increases in the magnitude of large jump variation, and reductions in the level of small jump variation.

The other direction in which we investigate the linkage between news announcements and jump variation is based on an exploration of whether news announcements affect the frequency of occurrence of either small or large jumps. Table 16 reports the average percentage of firms exhibiting particular types of jumps on days with and without earning surprises. Specifically, on each announcement date, all stocks exhibiting earnings are sorted into tertile portfolios based on the absolute value of the earning surprise (SUE). The categories sorted on are denoted as "small", "medium", and "large", with tertiles calculated by appropriate sorting of the firms based on the absolute values of the firms' earnings surprise magnitudes. Then, within each tertile, the percentage of firms exhibiting a particular type of jump (averaged across all earnings surprise days) is calculated and reported. For these calculations, only days in which at least 3 firms report earning surprises and included in our sample.[19] Turning to the results in the table, note, for example, that the entry 0.3042 in the sixth column of Panel A indicates that 30.42% of firms in the "small surprise" tertile portfolio recorded a large jump (measured by SRVLJ) on small surprise days, on average, across the entire daily sample. By contrast, 89.83% of firms exhibit small jumps (measured by SRVSJ) on days with small surprises.

Two clear conclusions emerge upon examination of the results in this table. First, when the magnitude of earning surprises increases, the average percentage of firms with large jumps (SRVLJ) increases from 30.42% to 37.37%. In particular, in Panel A, note that for the small tertile, the percentage of firms exhibiting large jumps (SRVLJ) is 30.42%, while for the large tertile, the percentage is 37.37%. By contrast, the percentage of firms with small jumps decreases as the relative magnitude of earnings surprises increases (i.e., the percentage of firms associated with SRVSJ decreases from 89.83% to

[19] Results are virtually identical if we only include days in which at least 12 or 24 firms report earnings surprises.

88.29%). This result indicates that "big news" is associated with an increase in the prevalence of large jumps. Second, the prevalence of jumps differs depending upon whether one tabulates results on earnings surprise days (Panel A) or on non-earnings surprise days (Panel B). For example, large news surprises are associated with large jumps for 31.07% of firms on non-announcement days (see Panel B) and 37.37% of firms on announcement days (Panel A). This result is consistent with event study finding that jump magnitudes are larger on announcement days than non-announcement days.

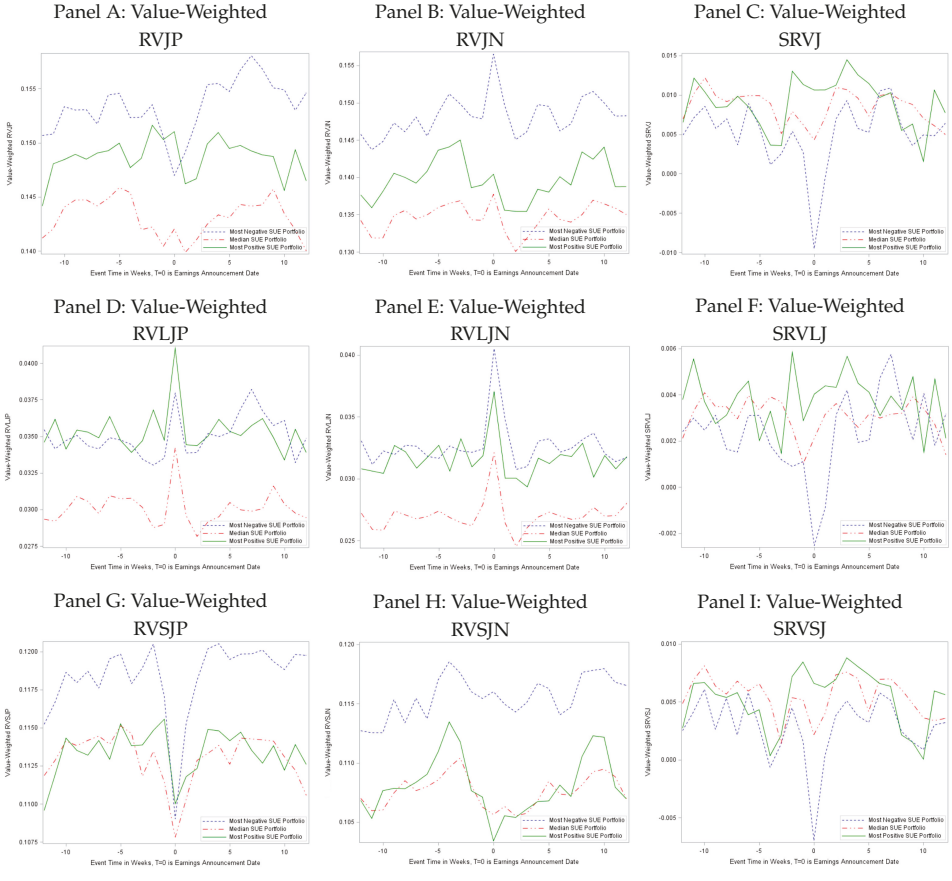


Figure 5. Jump Variation Measures Around Earnings Announcement. Panels A–I display value-weighted averages of various weekly jump variation measures in a [−12, 12] week window around earnings announcement.

Table 16. Jumps Associated with (Absolute) Magnitude of Earning Surprises.

Panel A: Daily Average Percentage of Firms Exhibiting Various Types of Jumps, on Days Characterized by Earnings Surprises									
A-SUE	**RVJP**	**RVJN**	**SRVJ**	**RVLJP**	**RVLJN**	**SRVLJ**	**RVSJP**	**RVSJN**	**SRVSJ**
Small	0.8099	0.8180	0.9849	0.1951	0.2004	0.3042	0.7233	0.7258	0.8983
Medium	0.8310	0.8232	0.9841	0.2216	0.2173	0.3289	0.7319	0.7232	0.8928
Large	0.8605	0.8621	0.9909	0.2618	0.2572	0.3737	0.7455	0.7488	0.8829
Panel B: Daily Average Percentage of Firms Exhibiting Various Types of Jumps, on Days Characterized by No Earnings Surprises									
	RVJP	**RVJN**	**SRVJ**	**RVLJP**	**RVLJN**	**SRVLJ**	**RVSJP**	**RVSJN**	**SRVSJ**
	0.8836	0.8786	0.9884	0.2252	0.2220	0.3107	0.7941	0.7900	0.9095
Panel C: *t*-Statistics Associated with the Difference in Jump Size Percentages Between Portfolios									
			Difference	**SRVJ**	**SRVLJ**	**SRVSJ**			
			Medium-Small	−0.55	4.98	−1.68			
			Large-Medium	5.76	9.02	−3.06			
			Large-None	3.54	16.85	−10.85			

Notes: See notes to Table 1. Panels A and B of this table report daily average percentages of firms exhibiting various types of jumps, on days with (Panel A) and without (Panel B) earnings surprises. On earning announcement dates for which at least 3 stocks report earning, the "reporting" stocks are sorted into tertile portfolios (called "Small", "Medium", and "Large"), based on the absolute value of earning surprise (A-SUE), where SUE is defined in Equation (16). Thus, small, medium and large portfolios are only constructed on days for which at least 3 firms are characterized by an earnings surprise. Then, the percentage of firms exhibiting jumps in each of the three earnings surprise size categories is calculated, for various jump types (i.e., RVJP, RVJN, etc.) Finally, percentages are averages over all reporting days in the sample. Finally, various Newey–West *t*-statistics measuring the significance of the differences in jump size percentages for SRVJ, SRVLJ, and SRVSJ type jumps are reported in Panel C of the table.

It is also worth noting that Panel C of Table 16 reports *t*-statistics that test whether the differences in percentages of jumps in different portfolios are significant. In this table, "None" refers to the case where percentages are calculated on non-earnings-announcement days. Thus, the fact that the "Large-None" *t*-statistic associated with SRVLJ is 16.85, indicates that the percentage of large jumps on "large-surprise" earnings announcement days is significantly greater than the percentage of large jumps on non-earnings-announcement days. This in turn implies that large jumps tend to occur on large-surprise earnings announcement days. On the other hand, the reverse is true in the case of small jumps. In particular, the "Large-None" *t*-statistic associated with SRVSJ is -10.85, indicating that small jumps tend to occur on non-earnings-announcement days.

As discussed above, it is worth closing this section by stressing that the magnitudes of price movements depend on the magnitude of news (if they are related to news), and on the underlying process describing price evolution, and thus movements may manifest in the form of large/small jumps and/or price drifts. Clearly, the process associated with how firms and markets digest information is critical to understanding and quantifying what sorts of price movements ensue. For any given news, responses may differ in terms of upside or downside price drifts, or large/small jumps). In this sense, both small and large jumps may matter, as shown in our empirical results, in the cross-section.

4.6.2. Systematic versus Idiosyncratic Risks

To further explore the unique information embedded in either large or small jump variation measures, and examine their association with systematic and idiosyncratic risks, we identify the effect of diversification on both small and large jumps. In order to do this, we construct two alternative measures of SRVLJ and SRVSL. The ratio of these is plotted in Figure 6.

Method 1: For jump measures using this method, we simply construct SRVLJ and SRVSJ as done earlier in the paper. Specifically, we sort stocks into quintiles based on either weekly SRVLJ or SRVSJ. Then, we construct daily ratios of SRVLJ to SRVSJ for each individual stock in a given quintile. Finally, these ratios are aggregated, forming weekly measures of SRVLJ/SRVSJ. These measures are then used to form equal- or value-weighted ratios of SRVLJ to SRVSJ. These values are depicted in red (solid line) in Figure 6.

Method 2: For jump measures using this method, we start by constructing the same quintiles (based on weekly SRVLJ and SRVSJ) as done above. Then, we use the 5-min returns for each stock in a given quintile to construct 5-min aggregate portfolio returns for that quintile. We then construct daily jump measures using these portfolio returns (called SRVLJ and SRVSJ, and SRVLJ/SRVSJ), which are portfolio versions of the similar measures constructed using Method 1. Finally, daily measures are aggregated into weekly measures. These values are depicted in blue (dotted line) in Figure 6.

Comparing jump variation ratios constructed in these two different ways allows us to explore the importance of diversification when measuring jump variation. Turning to our findings, Figure 6 shows the time series of aggregated (Method 2) and weighted average (Method 1) jump variation measures for the first quintile portfolios. The fact that Method 1 (red line) is much smoother than Method 2 (blue line) means that the small jump component in the ratio of SRVLJ/SRVSJ remains much larger than in the other case. Thus, the obvious difference between aggregated and weighted averages of SRVLJ/SRVSJ indicates that small jump variation is more likely to be diversified away than large jump variation. This can be immediately seen upon examination of the plots in any of the four panels in the figure. Small jump variation is therefore more closely related to firm specific or idiosyncratic risks, while large jump variation is more likely to be systematic risks. [20]

[20] See the Supplementary Appendix for plots of jump variation measures for the other quintile portfolios.

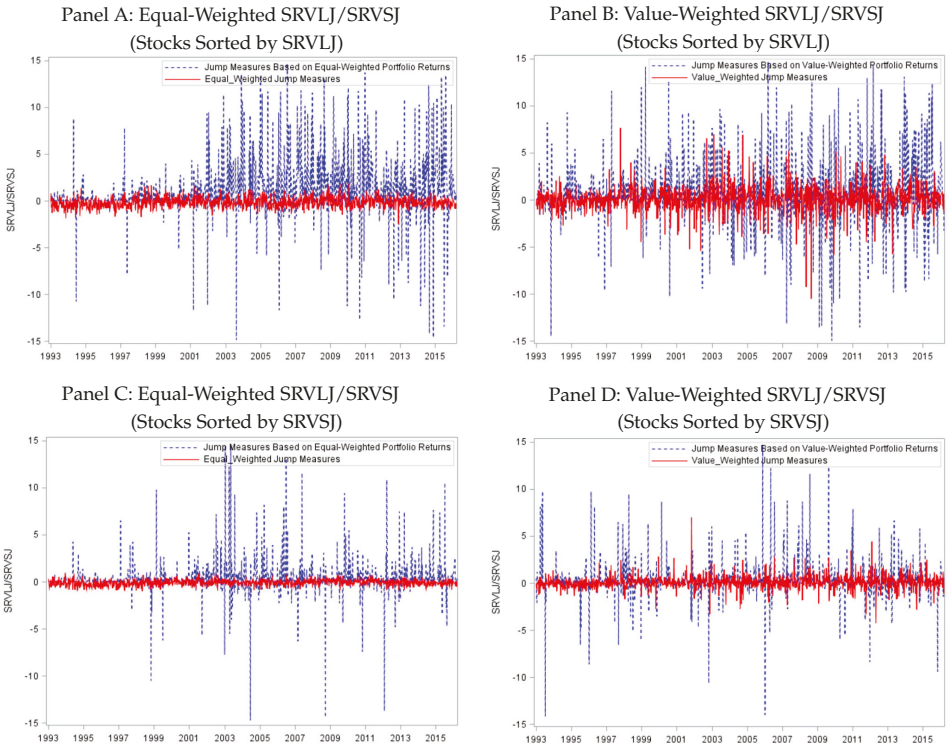


Figure 6. Aggregated and Weighted Average of Jump Variation Measures. Panels A-D display weekly aggregated and weighted averages of the ratio of SRVLJ to SRVSJ for 1st quintile stocks, sorted on SRVLJ and SRVSJ. Aggregated jump measures are depicted in blue (dotted line), and are constructed using 5-min portfolio returns. Weighted average jump measures are depicted in red (solid line) and are constructed using individual daily jump measures, and then aggregating to weekly. All calculation use jump truncation level $\gamma^1 = 4\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\Delta_n^{0.49}$. For complete details, refer to Section 4.6.2.

Another way to explore the relationship between systematic and idiosyncratic risks is to carry out Fama–MacBeth type regressions where the dependent variable is one of our jump variation measures and the independent variables are firm characteristics.[21] The results from several these sorts of regressions are reported in Table 17. Evidently, the firm characteristics always explain more of the dynamics associated with small jumps than with large jumps. This finding is supported by the fact that the adjusted R^2 is higher when the dependent variable is a small jump variation measure (compare the results of regressions I and II with III and IV). This again suggests that small jump variation is more likely to be associated with idiosyncratic risks. [22]

[21] Specifically, our objective in this section is to discuss regressions of the form given in Equation (15), with the dependent variable replaced by various realized variables.

[22] See the Supplementary Appendix for results from double-sorted portfolios that condition on various control variables. In these tables, it is noteworthy that when stocks are first sorted by a control variable (e.g., illiquidity, volatility, firm size and reversal), the SRVJ (SRVLJ and SRVSJ) effect is much higher within quintile portfolios with high illiquidity, high volatility, small firm size, and low reversal. This result suggests that all these control variables significantly contribute to the predictability of jump variation measures. This result provides additional confirmation to earlier findings reported in Bollerslev et al. (2020).

Table 17. Fama–MacBeth Type Regressions Using Various Jump Variation Measures as Dependent Variable.

	SRVLJ	**SRVLJ**	**SRVSJ**	**SRVSJ**	**SRVJ**	**SRVJ**
	I	**II**	**III**	**IV**	**V**	**VI**
Intercept	0.0080	0.0176	0.0034	0.0159	0.0115	0.0335
	(5.81)	(10.78)	(2.59)	(11.64)	(4.63)	(12.31)
RVOL	−0.0050	0.0025	−0.0065	0.0006	−0.0115	0.0032
	(−11.02)	(5.36)	(−15.37)	(1.88)	(−14.81)	(4.71)
Beta		0.0014		−0.0011		0.0003
		(4.40)		(−3.76)		(0.79)
log(Size)	0.0013	0.0003	0.0030	0.0017	0.0043	0.0020
	(5.79)	(1.62)	(15.19)	(9.21)	(12.71)	(6.67)
BE/ME		0.0007		0.0003		0.0010
		(4.11)		(2.72)		(4.58)
MOM		0.0007		0.0011		0.0018
		(3.54)		(5.94)		(6.13)
REV	0.25166	0.1176	0.3172	0.1882	0.5688	0.3058
	(61.05)	(31.43)	(65.55)	(41.52)	(69.66)	(39.62)
IVOL		−0.1424		−0.1880		−0.3303
		(−17.05)		(−23.23)		(−24.43)
CSK		0.0142		0.0206		0.0349
		(19.14)		(28.33)		(26.38)
CKT		−0.0008		−0.0008		−0.0015
		(−2.28)		(−1.96)		(−2.31)
MAX		0.2456		0.2248		0.4704
		(43.77)		(25.28)		(35.84)
MIN		0.5115		0.4365		0.9480
		(54.63)		(48.47)		(58.64)
ILLIQ	0.0020	0.0016	0.0032	0.0023	0.0052	0.0039
	(6.33)	(5.32)	(18.60)	(16.36)	(12.20)	(10.28)
Adjusted R^2	0.0322	0.0492	0.1070	0.1473	0.1049	0.1517

Notes: See notes to Tables 1, 5 and 14. This table reports results for cross-sectional Fama–MacBeth type regressions using various jump variation measures (listed across the first row of entries in the table) as dependent variables, and for various control variables (listed in the first column of the table). Thus, the regressions in this table mirror those reported in Table 14, with one difference. Specifically, the dependent variable in the regressions is either SRVLJ, SRVSJ, or SRVJ. Here, SRVLJ and SRVSJ are contructed using jump truncation level $\gamma^2 = 5\sqrt{\frac{1}{t}\widehat{IV}_t^{(i)}}\,\Delta_n^{0.49}$.

5. Concluding Remarks

In this paper, we add to the literature that explores the relationship between equity returns and volatility. In particular, we focus on the strand of this literature that explores the data for evidence of asymmetry (non-linearity) in the return volatility trade-off. Following Bollerslev et al. (2020), we decompose realized variation into upside and downside semi-variances (good and bad volatilities). We then take the additional step of partitioning the semi-variances into small and large components. Within this context, we examine the marginal predictive content of small and large jump variation measures. We also examine the importance of earnings announcements for examining the linkages between small and large jumps and news.

We find that sorting on signed small jump variation leads to value-weighted high–low portfolios with greater average returns and alphas than when either signed total jump or signed large jump variation is sorted on. We also find that there is a threshold, beyond which large jump variation contains no marginal predictive ability, relative to that contained in small jump variation. Indeed, including large jump variation can actually be detrimental to predictive accuracy, as average returns and alphas for high–low portfolios actually decline when total variation is instead used in some of our prediction experiments. Analysis of returns and alphas based on industry double-sorts indicates that the benefit of small signed jump variation investing is driven by stock selection within an industry,

rather than industry bets. Investors prefer stocks with a high probability of having positive jumps, but they also tend to overweight safer industries. Additionally, we find that the content of signed large jump variation is negligible when controlling for either signed total jump variation or realized skewness. By contrast, signed small jump variation has unique information for predicting future returns, even when controlling for total jump variation or realized skewness. Finally, we find that large jumps are closely associated with "big" news, or large earnings surprises, as might be expected. We also find that small jumps are more likely to be diversified away than large jumps, thus tend to be more closely associated with idiosyncratic risks, and are therefore more likely to be driven by liquidity conditions and trading activity. One reason for this is that while news-related information is embedded in large jump variation, the information is generally short-lived, and dissipates too quickly to provide marginal predictive content for subsequent weekly returns, in the particular cross-sectional setup that we use in our analysis.

Supplementary Materials: The Supplementary Materials available at http://www.mdpi.com/2225-1146/8/2/19/s1.

Author Contributions: The authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors are grateful to Mingmian Cheng, Yuan Liao, Xiye Yang and seminar participants at the 2019 SoFiE Conference, the 2019 Asia Meeting of the Econometric Society, and the 2019 Midwest Economics Association Conference for numerous useful comments. The authors are also grateful to the editor, Marc Paolella, and 3 referees for the many useful comments and suggestions that they made on earlier versions of this paper.

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

References

Aït-Sahalia, Yacine, and Jean Jacod. 2012. Analyzing the spectrum of asset returns: Jump and volatility components in high frequency data. *Journal of Economic Literature* 50: 1007–50. [CrossRef]

Amaya, Diego, Peter Christoffersen, Kris Jacobs, and Aurelio Vasquez. 2015. Does realized skewness predict the cross-section of equity returns? *Journal of Financial Economics* 118: 135–67. [CrossRef]

Amihud, Yakov. 2002. Illiquidity and stock returns: Cross-section and time-series effects. *Journal of Financial Markets* 5: 31–56. [CrossRef]

Andersen, Torben G., Tim Bollerslev, and Francis X. Diebold. 2007. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics 89*: 701–20. [CrossRef]

Andersen, Torben G., Tim Bollerslev, Francis X. Diebold, and Paul Labys. 2003. Modeling and forecasting realized volatility. *Econometrica* 71: 579–625. [CrossRef]

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2006. The cross-section of volatility and expected returns. *The Journal of Finance* 61: 259–99. [CrossRef]

Ang, Andrew, Robert J. Hodrick, Yuhang Xing, and Xiaoyan Zhang. 2009. High idiosyncratic volatility and low returns: International and further u.s. evidence. *Journal of Financial Economics* 91: 1–23. [CrossRef]

Bajgrowicz, Pierre, and Olivier Scaillet. 2012. Technical trading revisited: False discoveries, persistence tests, and transaction costs. *Journal of Financial Economics* 106: 473–91. [CrossRef]

Bajgrowicz, Pierre, Olivier Scaillet, and Adrien Treccani. 2016. Jumps in high-frequency data: Spurious detections, dynamics, and news. *Management Science* 62: 2198–217. [CrossRef]

Barndorff-Nielsen, Ole E., Svend Erik Graversen, Jean Jacod, and Neil Shephard. 2006. Limit theorems for bipower variation in financial econometrics. *Econometric Theory* 22: 677–719. [CrossRef]

Barndorff-Nielsen, Ole E., Silvia Kinnebrouk, and Neil Shephard. 2010. Measuring downside risk: Realised semivariance. In *Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle*. Edited by Tim Bollerslev, Jeffrey Russell and Mark Watson. Oxford: Oxford University Press, pp. 117–36.

Barndorff-Nielsen, Ole E., and Neil Shephard. 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2: 1–37. [CrossRef]

Barras, Laurent, Olivier Scaillet, and Russ Wermers. 2010. False discoveries in mutual fund performance: Measuring luck in estimated alphas. *The Journal of Finance* 65: 179–216. [CrossRef]

Bernard, Victor L., and Jacob K. Thomas. 1989. Post-earnings-announcement drift: Delayed price response or risk premium? *Journal of Accounting Research* 27: 1–36. [CrossRef]

Bollerslev, Tim, Tzuo Hann Law, and George Tauchen. 2008. Risk, jumps, and diversification. *Journal of Econometrics* 144: 234–56. [CrossRef]

Bollerslev, Tim, Sophia Zhengzi Li, and Bingzhi Zhao. 2020. Good volatility, bad volatility, and the cross-section of stock returns. *Journal of Financial and Quantitative Analysis* 55: 751–81. [CrossRef]

Bollerslev, Tim, and Viktor Todorov. 2011. Estimation of jump tails. *Econometrica* 79: 1727–83.

Bollerslev, Tim, Viktor Todorov, and Lai Xu. 2015. Tail risk premia and return predictability. *Journal of Financial Economics* 118: 113–34. [CrossRef]

Campbell, John Y., and Tuomo Vuolteenaho. 2004. Bad beta, good beta. *American Economic Review* 94: 1249–75. [CrossRef]

Christensen, Kim, Roel C. A. Oomen, and Mark Podolskij. 2014. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics* 114: 576–99. [CrossRef]

Duong, Diep, and Norman R. Swanson. 2011. Volatility in discrete and continuous-time models: A survey with new evidence on large and small jumps. In *Missing Data Methods: Time-Series Methods and Applications*. Edited by David M. Drukker. Advances in Econometrics, vol. 27B. Bingley: Emerald Press, pp. 179–233.

Duong, Diep, and Norman R. Swanson. 2015. Empirical evidence on the importance of aggregation, asymmetry, and jumps for volatility prediction. *Journal of Econometrics* 187: 606–21. [CrossRef]

Fang, Nengsheng, Wen Jiang, and Ronghua Luo. 2017. Realized semivariances and the variation of signed jumps in china's stock market. *Emerging Markets Finance and Trade* 53: 563–86. [CrossRef]

Feunou, Bruno, Mohammad R. Jahan-Parvar, and Romeo Tedongap. 2013. Modeling market downside volatility. *Review of Finance* 17: 443–81. [CrossRef]

Feunou, Bruno, Mohammad R. Jahan-Pravar, and Cédric Okou. 2018. Downside variance risk premium. *Journal of Financial Econometrics* 16: 341–83. [CrossRef]

Gu, Shihao, Bryan Kelly, and Dacheng Xiu. 2019. *Empirical Asset Pricing via Machine Learning*. Working Paper. Chicago: University of Chicago.

Guo, Hui, Kent Wang, and Hao Zhou. 2015. *Good Jumps, Bad Jumps, and Conditional Equity Premium*. Working Paper. Cincinnati: University of Cincinnati.

Jacod, Jean. 2008. Asymptotic properties of realized power variations and related functionals of semimartingales. *Stochastic Processes and their Applications* 118: 517–59. [CrossRef]

Jiang, George J., and Tong Yao. 2013. Stock price jumps and cross-sectional return predictability. *Journal of Financial and Quantitative Analysis* 48: 1519–44. [CrossRef]

Lee, Suzanne S., and Per A. Mykland. 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies* 21: 2535–63. [CrossRef]

Lee, Suzanne S., and Per A. Mykland. 2012. Jumps in equilibrium prices and market microstructure noise. *Journal of Econometrics* 168: 396–406. [CrossRef]

Li, Jia, Viktor Todorov, George Tauchen, and Rui Chen. 2017. Mixed-scale jump regressions with bootstrap inference. *Journal of Econometrics* 201: 417–32. [CrossRef]

Livnat, Joshua, and Richard R. Mendenhall. 2006. Comparing the post-earnings announcement drift for surprises calculated from analyst and time series forecasts. *Journal of Accounting Research* 44: 177–205. [CrossRef]

Maheu, John M., and Thomas H. McCurdy. 2004. News arrival, jump dynamics, and volatility components for individual stock returns. *Journal of Finance* 59: 755–93. [CrossRef]

Mancini, Cecilia. 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* 36: 270–96. [CrossRef]

Patton, Andrew J., and Kevin Sheppard. 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics* 97: 683–97. [CrossRef]

Petersen, Mitchell A. 2009. Estimating standard errors in finance panel data sets: Comparing approaches. *Review of Financial Studies* 22: 435–80. [CrossRef]

Rossi, Alberto G., and Allan Timmermann. 2015. Modeling covariance risk in merton's icapm. *Review of Financial Studies* 28: 1428–61. [CrossRef]

Scaillet, Olivier, Adrien Treccani, and Christopher Trevisan. 2018. High-frequency jump analysis of the bitcoin market. *Journal of Financial Econometrics*. [CrossRef]

Todorov, Viktor, and George Tauchen. 2010. Activity signature functions for high-frequency data analysis. *Journal of Econometrics* 154: 125–38. [CrossRef]

Woodward, George, and Heather M. Anderson. 2009. Does beta react to market conditions? Estimates of 'bull' and 'bear' betas using a nonlinear market model with an endogenous threshold parameter. *Quantitative Finance* 9: 913–24. [CrossRef]

Xiong, James X., Thomas M. Idzorek, and Roger G. Ibbotson. 2016. The economic value of forecasting left-tail risk. *Journal of Portfolio Management* 42: 114–23. [CrossRef]

Yan, Shu. 2011. Jump risk, stock returns, and slope of implied volatility smile. *Journal of Financial Economics* 99: 216–33. [CrossRef]

Yang, Hanxue and Juho Kanniainen. 2016. Jump and volatility dynamics for the s&p 500: Evidence for infinite-activity jumps with non-affine volatility dynamics from stock and option markets. *Review of Finance* 21: 811–44.

econometrics

MDPI

Article

Reducing the Bias of the Smoothed Log Periodogram Regression for Financial High-Frequency Data

Erhard Reschenhofer * and Manveer K. Mangat

Department of Statistics and Operations Research, University of Vienna, Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria; manveer.mangat@univie.ac.at

* Correspondence: erhard.reschenhofer@univie.ac.at; Tel.: +43-1-4277-38646

Received: 8 March 2020; Accepted: 27 September 2020; Published: 10 October 2020

Abstract: For typical sample sizes occurring in economic and financial applications, the squared bias of estimators for the memory parameter is small relative to the variance. Smoothing is therefore a suitable way to improve the performance in terms of the mean squared error. However, in an analysis of financial high-frequency data, where the estimates are obtained separately for each day and then combined by averaging, the variance decreases with the sample size but the bias remains fixed. This paper proposes a method of smoothing that does not entail an increase in the bias. This method is based on the simultaneous examination of different partitions of the data. An extensive simulation study is carried out to compare it with conventional estimation methods. In this study, the new method outperforms its unsmoothed competitors with respect to the variance and its smoothed competitors with respect to the bias. Using the results of the simulation study for the proper interpretation of the empirical results obtained from a financial high-frequency dataset, we conclude that significant long-range dependencies are present only in the intraday volatility but not in the intraday returns. Finally, the robustness of these findings against daily and weekly periodic patterns is established.

Keywords: long-range dependence; log periodogram regression; smoothed periodogram; subsampling; intraday returns

JEL Classification: C13; C14; C22; C58

1. Introduction

After Mandelbrot (1971) had discussed the possibility that the strength of the statistical dependence of stock prices decreases very slowly, several researchers investigated this issue empirically. For example, Greene and Fielitz (1977) found indications of long-range dependence when they applied a technique called range over standard deviation (R/S) analysis (Hurst 1951; Mandelbrot and Wallis 1969; Mandelbrot 1972, 1975) to daily stock return series. This technique is based on the R/S statistic Q_n, which is defined as the range of all partial sums of a time series of length n from its mean divided by its standard deviation. For a large class of short-range dependent processes, Q_n/n^H converges to a non-degenerate random variable if $H = 0.5$. An analogous result with $H \neq 0.5$ holds for long-range dependent processes. The parameter H is called the Hurst coefficient and is used as a measure of long-range dependence. However, Lo (1991) pointed out that the results obtained with this technique may be misleading because of the sensitivity of Q_n to short-range dependence (see also Davis and Harte 1987; Hauser and Reschenhofer 1995) and proposed, therefore, a Newey and West (1987) type modification for the denominator of the R/S statistic, which is appropriate for general forms of short-range dependence. Contrary to the findings of Greene and Fielitz (1977) and others, he found no evidence of long-range dependence in daily and monthly index returns once the possible short-range dependence was properly taken care of. A disadvantage of Lo's (1991) modified R/S analysis is its dependence on an important tuning parameter, namely the truncation lag q,

which determines the number of included autocovariances. The general conditions that ensure the consistency of the Newey and West estimator provide little guidance in selecting q in finite samples. Additionally, Andrews's (1991) data-dependent rule for choosing q is based on asymptotic arguments.

Long-range dependence can not only be characterized by a Hurst coefficient $H \neq 0.5$ but also by a slowly decaying autocorrelation function ρ or a spectral density f that is steep in a small neighborhood of frequency zero, i.e.,

$$\rho(k)k^{1-2d} \;\rightarrow\; c \; as \; k \rightarrow \infty, \; c > 0, \; d \neq 0, \tag{1}$$

and:

$$f(\omega) \;\sim\; c\omega^{-2d}, \; \omega \in (0, \varepsilon), c > 0, \; d \neq 0, \tag{2}$$

respectively. The parameter d is called a memory parameter (or fractional differencing parameter) and is related to H by $d = H - 0.5$. It can be estimated by replacing the unknown spectral density f in (2) by the periodogram (Geweke and Porter-Hudak 1983) or a more sophisticated estimate of f (Hassler 1993; Peiris and Court 1993; Reisen 1994), taking the log of both sides, and regressing the log estimate on a deterministic regressor. Robustness against short-range dependence can be achieved by using only the $K = n^{\alpha}$ lowest Fourier frequencies in the regression. A popular choice for the tuning parameter α is 0.5. For the purpose of testing, the asymptotic error variance is used. Applying the log periodogram regression method of Geweke and Porter-Hudak (1983) to the daily returns of the 30 components of the Dow Jones Industrials index and several indices, Barkoulas and Baum (1996) found no convincing evidence in favor of long-range dependence, which is not surprising in light of the finding of Mangat and Reschenhofer (2019) that the test based on the asymptotic error variance has very low power. Unfortunately, using the standard variance formula of the least squares estimator of the slope in a simple linear regression instead of the asymptotic error variance is also problematic because it leads to overrejecting the true null hypothesis (see Mangat and Reschenhofer 2019).

The negative results of Lo (1991) and Barkoulas and Baum (1996) are in line with the results obtained by Cheung and Lai (1995) with both modified R/S analysis and log periodogram regression for stock return data from eighteen countries and by Crato (1994) with fractionally integrated ARMA (ARFIMA) models (Granger and Joyeux 1980; Hosking 1981) for stock indices of the G-7 countries. Using not only the log periodogram regression with the asymptotic error variance but additionally also nonparametric techniques such as R/S analysis and modified R/S analysis as well as parametric techniques, Grau-Carles (2000) also found little evidence of long-range dependence in index returns but strong evidence of persistence in volatility measured as squared returns and absolute returns, respectively, which corroborates earlier findings of Crato and de Lima (1994) and Lobato and Savin (1998). In general, results obtained with ARFIMA models must be treated with caution. Firstly, the true model dimension is unknown in practice and reliable inference after automatic model selection is illusory. Secondly, Pötscher (2002) has shown that the problem of estimating the memory parameter d falls into the category of ill-posed estimation problems when the class of data generating processes is too rich. For example, Grau-Carles (2000) considered all ARFIMA(p,q) processes with $p \leq 3$ and $q \leq 3$, which is possibly an unnecessarily large class for return series.

While the bulk of empirical research focused on major capital markets, Barkoulas et al. (2000) examined an emerging capital market, namely the Greek stock market, with the log periodogram regression and obtained significant estimates of d in the range between 0.20 and 0.30 for values of the tuning parameter α between 0.5 and 0.6. However, their sample period is relatively short and the sampling frequency is weekly rather than daily. Even less confidence-inspiring are the positive results obtained by Henry (2002) with monthly data from several international stock markets. Clearly, methods that have been designed for large samples should not be applied to small and medium samples. Recently, small-sample tests for testing hypotheses about the memory parameter d have been proposed (Mangat and Reschenhofer 2019; Reschenhofer and Mangat 2020). When applied to asset returns, these tests produced negative results throughout. Cajueiro and Tabak (2004), Carbone et al. (2004), Batten and Szilagyi (2007), Batten et al. (2008), Souza et al. (2008), Batten et al. (2013), and

Auer (2016a, 2016b) observed time-variability of the Hurst exponent in stock returns, currency prices, and the prices of precious metals, respectively. These apparent changes were occasionally interpreted as indications of changing market efficiency or even used for the construction of trading strategies. Although it cannot be ruled out that some erratic estimator for the memory parameter d catches signals that are useful for trading purposes even when in fact there is no long-range dependence, there still seems to be a need for a more efficient estimator that actually allows to get some information about the true nature of the data generating process.

In general, there is always a trade-off between bias and variance. Estimators for the memory parameter d that are based on a smooth estimate of the spectral density have typically a smaller variance and a larger bias than those based on the periodogram (Chen et al. 1994; Reschenhofer et al. 2020), which is advantageous in situations where the squared bias is small relative to the variance. However, in the case of high-frequency financial data, there are usually gaps between the individual trading sessions, which make it necessary to estimate d separately for each trading session and compute the final estimate by averaging the individual estimates. Here, the variance decreases with the number of trading sessions but the bias remains fixed; hence, conventional smoothing methods, which achieve a reduction in the variance at the expense of an increase in the bias, are of no use. The goal of this paper is therefore to introduce a new method of smoothing that does not systematically have a negative impact on the bias. This method will be described in detail in the next section. Section 3 presents the results of an extensive simulation study, which compares the performance of various estimators for the memory parameter in terms of bias, variance, and root-mean-square error (RMSE). Using limit order book data obtained from Lobster, Section 4 searches for indications of long-range dependence both in the intraday volatility and in the intraday returns. Section 5 provides a conclusion.

2. Methods

2.1. Log Periodogram Regression

Fractionally integrated white noise satisfies the difference equation:

$$y_t = (1-L)^{-d} u_t, \tag{3}$$

where L is the lag operator and u_t is white noise with mean zero and variance σ^2 (Adenstedt 1974). Its spectral density is given by:

$$f(\omega) = \frac{\sigma^2}{2\pi}\left|1-e^{-i\omega}\right|^{-2d} = \frac{\sigma^2}{2^{1+2d}\,\pi}\left(\sin^2\left(\frac{\omega}{2}\right)\right)^{-d}. \tag{4}$$

The memory parameter d, which represents the degree of long memory if $d \neq 0$, can be estimated by regressing the log periodogram of the time series $y_1, \ldots, y_n$ on a deterministic regressor (Geweke and Porter-Hudak 1983). Indeed, we have:

$$L_j = \log I\left(\omega_j\right) = c + d x_j + v_j, \tag{5}$$

where:

$$I(\omega) = \frac{1}{2\pi n}\left|\sum_{t=1}^{n} y_t e^{-i\omega t}\right|^2. \tag{6}$$

is the periodogram,

$$\omega_j = 2\pi j/n,\ j = 1, \ldots, K \leq m = [(n-1)/2], \tag{7}$$

are the first K Fourier frequencies between 0 and π,

$$x_j = -2\log\left(\sin\left(\frac{\omega_j}{2}\right)\right) \tag{8}$$

is a deterministic regressor,

$$c = \log\left(\sigma^2/\left(2^{1+2d}\,\pi\right)\right) \tag{9}$$

is a constant, and

$$v_j = \log\left(I\left(\omega_j\right)/f\left(\omega_j\right)\right). \tag{10}$$

are random perturbations. Choosing $K \ll m$ rather than $K = m$ is advisable when it is suspected that not only long-term dependencies are present but also short-term dependencies, e.g., when the data come from an ARFIMA process:

$$y_t = \left(1 - \phi_1 L - \ldots - \phi_p L^p\right)^{-1}(1-L)^{-d}\left(1 + \theta_1 L + \ldots + \theta_q L^q\right)u_t \tag{11}$$

(Granger and Joyeux 1980; Hosking 1981), where the parameter d takes care of the former dependencies and the parameters $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ take care of the latter. It is assumed that $d < 0.5$ (stationarity condition), $d > -0.5$ (invertibility condition), and all roots of the lag operator polynomials $\Phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ and $\Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$ lie outside the unit circle (causality condition and invertibility condition, respectively).

In the special case of $d = p = q = 0$ and Gaussianity, the ratios $I\left(\omega_j\right)/f\left(\omega_j\right)$ are independent and identically distributed (i.i.d.) standard exponential and $v_1, \ldots, v_m$ are, therefore, i.i.d. Gumbel with mean $-\gamma$ and variance $\pi^2/6$, where $\gamma = 0.57721\ldots$ is Euler's constant. The variance of the Geweke Porter-Hudak (GPH) estimator $\hat{d}_{GPH}$ is then identical to the variance of the ordinary least squares (OLS) estimator for the slope in a simple regression model, i.e.,

$$var\left(\hat{d}_{GPH}\right) = \frac{\sigma_v^2}{S_{xx}} = \frac{\pi^2}{6S_{xx}}, \tag{12}$$

where:

$$S_{xx} = \sum_{t=1}^{K}(x_t - \overline{x})^2. \tag{13}$$

In a neighborhood of frequency zero:

$$\sin(\omega) \approx \omega, \tag{14}$$

Hence:

$$S_{xx} \approx 4\sum_{t=1}^{K}\left(\log(t) - \overline{\log(t)}\right)^2. \tag{15}$$

Furthermore:

$$\begin{array}{c}\int_1^K \log^2(t) - \frac{1}{K}\left(\int_1^K \log(t)\right)^2 = K\log^2(K) - 2K\log(K) + 2(K-1) \\ -\frac{1}{K}(K\log(K) - (K-1))^2 \;=\; K + o(K).\end{array} \tag{16}$$

Indeed, we have:

$$s_{xx} = 4(K + o(K)) \tag{17}$$

If:

$$K\log(K)/n \to 0 \tag{18}$$

(see Hurvich and Beltrao 1993), hence, the variance formula (10) becomes:

$$var\left(\hat{d}_{GPH}\right) \approx \frac{\pi^2}{24K} \tag{19}$$

in line with the asymptotic result:

$$\sqrt{K}\left(\hat{d}_{GPH} - d\right) \xrightarrow{d} N\left(0, \frac{\pi^2}{24}\right) \tag{20}$$

derived by Hurvich et al. (1998) under the assumption that $K = o\left(n^{4/5}\right)$ and $\log^2(n) = o(K)$ for a class of stationary Gaussian long-memory processes with spectral densities of the form:

$$f(\omega) = \left|1 - e^{-i\omega}\right|^{-2d} f^*(\omega), \tag{21}$$

which includes all stationary ARFIMA processes.

If $d \neq 0$, the ratios $I\left(\omega_j\right)/f\left(\omega_j\right)$ are neither independent nor identically distributed, not even asymptotically (Robinson 1995). The problem is the irregular behavior of the spectral density in the neighborhood of frequency zero, i.e., $f(\omega) \to \infty$ as $\omega \to 0$ if $d > 0$ and $f(\omega) \to 0$ as $\omega \to 0$ if $d < 0$. Robinson (1995), therefore, proposed to remove the lowest Fourier frequencies from the log periodogram regression. Künsch (1986) showed that in the case of ARFIMA processes, the ratios $I\left(\omega_j\right)/f\left(\omega_j\right)$, $j = H+1, \ldots, H+K$ are indeed asymptotically i.i.d. standard exponential provided that $(H+1)/\sqrt{n} \to \infty$ and $(H+K)/n \to 0$. However, Reisen et al. (2001) and Mangat and Reschenhofer (2019) found that even the removal of only the first Fourier frequency already has a negative effect on the performance of the estimator $\hat{d}_{GPH}$.

2.2. Smoothing the Periodogram

An obvious possibility to further develop the estimator $\hat{d}_{GPH}$ is to smooth the periodogram before it is used in the regression (5) (Hassler 1993; Peiris and Court 1993; Reisen 1994). In order to illustrate the effect of smoothing, we consider the simple case of $K/3$ non-overlapping averages:

$$\left(I\left(\omega_{j-1}\right) + I\left(\omega_j\right) + I\left(\omega_{j+1}\right)\right)/3, \; j = 2, 5, 8, \ldots, K-1. \tag{22}$$

In this case, the sample size is divided by three but at the same time the variance of the error term decreases approximately from:

$$\operatorname{var}\left(\log\left(\frac{I\left(\omega_j\right)}{f\left(\omega_j\right)}\right)\right) \approx \frac{\pi^2}{6} \tag{23}$$

to the variance of the log chi-square distribution with 6 degrees of freedom because:

$$\operatorname{var}\left(\log\left(\frac{I\left(\omega_{j-1}\right) + I\left(\omega_j\right) + I\left(\omega_{j+1}\right)}{3\, f\left(\omega_j\right)}\right)\right) \approx \operatorname{var}\left(\log\left(\frac{2I\left(\omega_{j-1}\right)}{f\left(\omega_{j-1}\right)} + \frac{2I\left(\omega_j\right)}{f\left(\omega_j\right)} + \frac{2I\left(\omega_{j+1}\right)}{f\left(\omega_{j+1}\right)}\right)\right). \tag{24}$$

Noting that the mean (first cumulant) and the variance (second cumulant) of the log chi-square distribution with k degrees of freedom are given by:

$$\kappa_1 = \log(2) + \psi(k/2) \tag{25}$$

and:

$$\kappa_2 = \psi'(k/2), \tag{26}$$

respectively. we obtain for $k = 6$, $\kappa_1 = 1.615932$ and $\kappa_2 = 0.3949341$. Here, ψ is the digamma function and ψ' is its first derivative. Overall, the (approximate) variance of the least squares estimator of the memory parameter d decreases from

$$\frac{\pi^2}{6}\frac{1}{4K} = 1.644934\frac{1}{4K} \tag{27}$$

to

$$\psi'(3)\frac{1}{4K/3} = 1.184802\frac{1}{4K}, \tag{28}$$

where we have assumed that

$$\frac{1}{K/3}\sum\nolimits_{t=2,5,\ldots}(x_t-\overline{x})^2 \approx \frac{1}{K}\sum\nolimits_{t=1}^{K}(x_t-\overline{x})^2 \approx 4. \tag{29}$$

The little practical relevance of asymptotic results such as (20) can be seen when the asymptotic values are confronted with the actual values obtained by simulations. In the simplest case of Gaussian white noise, we do not have to safeguard against short-range dependence and can therefore choose a value of α slightly below 4/5. Choosing $\alpha = 0.7$ and $K \approx n^{\alpha}$, we obtain 0.00857 (27) and 0.00617 (28) vs. 0.01148 and 0.00885 (simulated) for $n = 250$ and $K = 48$, 0.00326 and 0.00235 vs. 0.00381 and 0.00282 for $n = 1000$ and $K = 126$, 0.00065 and 0.00047 vs. 0.00068 and 0.00050 for $n = 10,000$ and $K = 630$, and 0.00021 and 0.00015 vs. 0.00021 and 0.00015 for $n = 50,000$ and $K = 1947$. Obviously, huge sample sizes are required for good agreement. In the case of a nontrivial ARFIMA process, this problem will become even more serious because a smaller value of α must be chosen.

More sophisticated further developments of the estimator $\hat{d}_{GPH}$ are obtained by using more than three periodogram ordinates, allowing for overlaps, and introducing weights, or, equivalently, by using a lag-window estimator of the form:

$$\hat{f}\left(\omega_j\right) = \frac{1}{2\pi}\sum\nolimits_{s=-m}^{m} w(s/m)\hat{\gamma}(s)e^{-i\omega_j s},\ j = 1,\ldots,K, \tag{30}$$

where $\hat{\gamma}(s)$ denotes the sample autocovariance at lag s and the lag window w satisfies $w(0) = 1$, $\left|w(s)\right| \leq 1$, and $w(-s) = w(s)$ (see Hassler 1993; Peiris and Court 1993; Reisen 1994). A disadvantage of these estimation procedures is that they require the specification of a second tuning parameter, namely the length of the weighted averages in the former case and $m \leq n-1$ in the latter case, in addition to K. Of course, suitable weights and a suitable lag window, respectively, must be chosen too. Carrying out an extensive simulation study to compare various frequency-domain estimators for d, Reschenhofer et al. (2020) found that too strong smoothing, e.g., caused by choosing a too small value for m, entails an extremely large bias. Hunt et al. (2003) derived an approximation for the bias and observed generally a good agreement between their approximation and the corresponding value obtained by simulations when $d > 0$. However, the practical relevance of this approximation is limited because of its dependence on characteristics of the data generating process, which are unknown in practice.

2.3. Using Subsamples

A simple method of smoothing without introducing a bias is to average estimates obtained from different subsamples. Assume, for example, that the final estimate $\hat{d}$ is obtained by averaging over N preliminary estimates $\hat{d}_1,\ldots,\hat{d}_N$ obtained from independent subsamples $y_{11},\ldots,y_{n1},\ldots,y_{1N},\ldots,y_{nN}$; then, the variance of $\hat{d}$ vanishes as N increases while the bias remains unchanged. Of course, artificially splitting a long, homogeneous time series into non-overlapping subseries does not necessarily have a positive effect. For illustration, consider the simplest case where the time series $y_1,\ldots,y_n$ is split into two disjoint subseries $y_1,\ldots,y_{n/2}$ and $y_{n/2+1},\ldots,y_n$ of equal length. To allow a fair comparison, the frequency range $(0,\omega_K]$, is kept constant, which implies that in the case of the two subseries the number of used Fourier frequencies is $K/2$. Under the simplistic and mostly unrealistic assumption that the two subseries are independent, the (approximate) variance of the mean of the two GPH estimators based on the two subseries is given by:

$$\frac{1}{4}\left(\frac{\pi^2}{6}\frac{1}{4K/2}+\frac{\pi^2}{6}\frac{1}{4K/2}\right)=\frac{\pi^2}{6}\frac{1}{4K} \tag{31}$$

which is, therefore, of the same size as that of the original estimator, which is based on the whole time series. However, there is still room for improvement. A reduction in the variance may be achieved by

allowing for overlaps between the subseries, e.g., with a rolling estimation window or a combination of different partitions.

At first glance, the idea of improving an OLS estimator by averaging the OLS estimators obtained from the whole sample and the first and second halves, respectively, seems to be at odds with the Gauß-Markov theorem because the combined estimator is still linear. However, the crucial point here is that only the observations are partitioned and not the log periodogram, which is used as dependent variable in the regression and is obtained from the observations through nonlinear transformations. For illustration, consider an estimator of the form:

$$\widetilde{d_2} = (1-2\lambda)\hat{d}_1 + \lambda\hat{d}_{21} + \lambda\hat{d}_{22}, \tag{32}$$

where $\hat{d}_1, \hat{d}_{21}, \hat{d}_{22}$ are the OLS estimators for d based on the log periodograms L^1, L^{21}, L^{22} of the whole sample and the first and second halves, respectively. In the special case of Gaussian white noise with variance 2π, the constant c in the regression (3) vanishes, and we may, therefore, use the simpler estimators:

$$\check{d}_1 = \frac{\sum_{j=1}^{K} \underline{x}_j L_j^1}{\sum_{j=1}^{K} \underline{x}_j^2} \approx \frac{1}{4K}\sum_{j=1}^{K} \underline{x}_j L_j^1, \tag{33}$$

and:

$$\check{d}_{2s} = \frac{\sum_{j=1}^{K/2} \underline{x}_{2j} L_j^{2s}}{\sum_{j=1}^{K/2} \underline{x}_{2j}^2} \approx \frac{1}{2K}\sum_{j=1}^{K/2} \underline{x}_{2j} L_j^{2s},\ s = 1, 2, \tag{34}$$

where $\underline{x}_j = x_j - \bar{x}$. For the variances of the simplistic estimators $\check{d}_1$ and:

$$\check{d}_2 = (1-2\lambda)\check{d}_1 + \lambda\check{d}_{21} + \lambda\check{d}_{22}, \tag{35}$$

we obtain approximately:

$$var(\check{d}_1) \approx \left(\frac{1}{4K}\right)^2 \sum_{j=1}^{K} \underline{x}_j^2 \frac{\pi^2}{6} \approx \frac{\pi^2}{24K} \tag{36}$$

and:

$$\begin{aligned} var(\check{d}_2) &\approx \tfrac{\pi^2}{24K}((1-2\lambda)^2 + 4\lambda^2) + 4\lambda(1-2\lambda)cov(\check{d}_1, \check{d}_{21}) \\ &\approx \tfrac{\pi^2}{24K}((1-2\lambda)^2 + 4\lambda^2 + 4\lambda(1-2\lambda)(\rho_0 + \rho_1)) \\ &\approx 0.69\ \tfrac{\pi^2}{24K}, \text{ if } \lambda = \tfrac{1}{4}, \end{aligned} \tag{37}$$

respectively, where we have used that $cov(\check{d}_1, \check{d}_{21}) = cov(\check{d}_1, \check{d}_{22})$ and $cov(\check{d}_{21}, \check{d}_{22}) = 0$ as well as the rough approximations:

$$\sum_{j=1}^{\frac{K}{2}} \underline{x}_{2j}^2 \approx \sum_{j=1}^{\frac{K}{2}} \underline{x}_{2j}\underline{x}_{2j-1} \approx \sum_{j=1}^{\frac{K}{2}-1} \underline{x}_{2j}\underline{x}_{2j+1} \approx 2K, \tag{38}$$

$$cor(L_j^1, L_k^{2s}) \approx \begin{cases} \rho_0 = 0.35,\ if\ 2k = j, \\ \rho_1 = 0.13,\ if\ \left|2k - j\right| = 1, \\ 0,\ else \end{cases} \tag{39}$$

(see Table 1), and:

$$\begin{aligned} cov(\check{d}_1, \check{d}_{21}) &\approx \tfrac{1}{8K^2} cov(\textstyle\sum_{j=1}^{\frac{K}{2}} \underline{x}_{2j} L^1_{2j} + \sum_{j=1}^{\frac{K}{2}} \underline{x}_{2j-1} L^1_{2j-1},\ \sum_{k=1}^{\frac{K}{2}} \underline{x}_{2k} L^{21}_k) \\ &\approx \tfrac{1}{8K^2}\Big(\textstyle\sum_{j=1}^{\frac{K}{2}} \sum_{k=1}^{\frac{K}{2}} \underline{x}_{2j}\underline{x}_{2k}\, cov(L^1_{2j}, L^{21}_{2k}) + \sum_{j=1}^{\frac{K}{2}} \sum_{k=1}^{\frac{K}{2}} \underline{x}_{2j-1}\underline{x}_{2k}\, cov(L^1_{2j-1}, L^{21}_{2k})\Big) \\ &\approx \tfrac{1}{8K^2}\Big(\rho_0 \tfrac{\pi^2}{6} \textstyle\sum_{j=1}^{\frac{K}{2}} \underline{x}^2_{2j} + \rho_1 \tfrac{\pi^2}{6} \sum_{j=1}^{\frac{K}{2}} \underline{x}_{2j}\underline{x}_{2j-1}\Big) \\ &\approx \tfrac{\pi^2}{24K}(\rho_0 + \rho_1) \end{aligned} \tag{40}$$

For a further reduction of the variance, we may consider more general estimators of the form:

$$\widetilde{d}_k = \frac{1}{k}\Big(\hat{d}_1 + \sum\nolimits_{j=2}^{k} \frac{1}{j}(\hat{d}_{j1} + \ldots + \hat{d}_{jj})\Big), \tag{41}$$

which are based on k partitions. The next section examines whether this possible reduction actually materializes and whether it is accompanied by an increase in the bias. All computations are carried out with the free statistical software R (R Core Team 2018).

Table 1. Sample correlations between L_j, $j = 1, \ldots, 20$, and L^1_k, $k = 1, \ldots, 10$, obtained from 10,000,000 realizations of Gaussian white noise ($n = 400$).

	1	2	3	4	5	6	7	8	9	10
1	0.1475	0.0186	0.0072	0.0044	0.0027	0.0014	0.0013	0.001	0.0008	0.0005
2	0.3541	0.0002	−0.0001	−0.0004	0	0.0003	−0.0003	0.0002	0.0002	0.0005
3	0.1364	0.133	0.0154	0.006	0.0032	0.0025	0.0009	0.001	0.0007	0.0003
4	−0.0001	0.3541	−0.0001	−0.0002	0.0002	0.0008	−0.0005	−0.0002	−0.0004	−0.0003
5	0.0164	0.1316	0.1307	0.0144	0.005	0.0027	0.0019	0.0016	0.0008	0.0008
6	−0.0001	−0.0003	0.354	0.0002	0.0002	−0.0004	0.0004	0.0001	−0.0005	0.0005
7	0.007	0.0147	0.1311	0.1308	0.014	0.0043	0.0025	0.0021	0.0013	0.0011
8	0	0.0001	0.0004	0.3541	0.0004	0.0001	−0.0001	−0.0002	−0.0001	−0.0002
9	0.0035	0.0054	0.0143	0.1302	0.1302	0.0139	0.0051	0.003	0.0016	0.0009
10	−0.0003	0	−0.0001	0.0004	0.3539	−0.0003	0.0003	0.0001	−0.0005	0.0003
11	0.0023	0.0033	0.0047	0.0138	0.1301	0.13	0.0133	0.0054	0.0025	0.0014
12	−0.0004	−0.0001	−0.0004	−0.0001	0.0003	0.3542	0.0001	−0.0001	0.0002	0
13	0.0013	0.002	0.0032	0.0053	0.0137	0.1305	0.1309	0.0147	0.004	0.003
14	−0.0004	0.0001	0.0003	0.0004	0.0008	0.0002	0.3544	−0.0002	0.0005	−0.0002
15	0.0011	0.0016	0.002	0.0025	0.0059	0.014	0.1304	0.1297	0.0141	0.0055
16	−0.0006	0.0001	−0.0004	0	0.0002	−0.0001	−0.0001	0.354	0.0002	0.0002
17	0.0011	0.0009	0.0009	0.0021	0.0025	0.0049	0.0138	0.1305	0.1304	0.0137
18	0.0003	−0.0002	0	−0.0001	−0.0006	−0.0004	−0.0002	−0.0004	0.3541	−0.0001
19	0.0008	0.0005	0.0011	0.0015	0.0019	0.0026	0.0046	0.0138	0.1306	0.1302
20	−0.0001	0.0005	0.0001	0.0002	0.0008	0.0001	0.0007	−0.0003	−0.0005	0.3541

3. Simulations

In this section, we compare the new estimator $\widetilde{d}_k$ (41) for $k = 2,\ 3,\ 5,\ 10$ with Geweke and Porter-Hudak's (1983) estimator $\hat{d}_{GPH}$, which is based on the log periodogram regression (5), and the estimators $\hat{d}_{sm}$ and $\hat{d}^{\beta}_{smP}$, which are obtained by replacing the periodogram ordinates in (5) by simple moving averages of neighboring periodogram ordinates and lag-window estimates of the form (30) with truncation lags $m = \lfloor n^{\beta} \rfloor$, $\beta = 0.5,\ 0.7,\ 0.9, 1$, respectively. In the latter case, the Parzen window is used, which is given by:

$$w(z) = \begin{cases} 1 - 6z^2 + 6|z|^3, & |z| < \frac{1}{2}, \\ 2(1 - |z|)^3, & \frac{1}{2} \le |z| \le 1. \end{cases} \tag{42}$$

With a view to the later application of the estimators to 1-min intraday returns in Section 4, the sample size $n = 390$ is chosen for our simulation study because there are 390 min in a regular

trading session for U.S. stocks, which starts at 9:30 a.m. and ends at 4:00 p.m. The number K of Fourier frequencies included in the log periodogram regression is defined by setting $K = 20 \approx [n^{\alpha}]$, $\alpha = 0.5$. For $k = 2$, the first $K/k = 10$ Fourier frequencies of the two disjoint subseries of length $n/k = 195$ are given by $\omega_2, \omega_4, \ldots, \omega_K$, and for $k = 10$, the first $K/k = 2$ Fourier frequencies of the 10 disjoint subseries of length $n/k = 39$ are given by ω_{10}, ω_K. Clearly, we cannot go beyond $k = 10$ because at least two frequencies are required to carry out the log periodogram regression. Additionally, using frequencies outside the interval $(0, \omega_K]$ is not an option because this would amount to an unfair advantage, particularly when there are no short-term dependencies which have to be taken into account.

With the help of the R-package 'fracdiff', 10,000 realizations of length $n = 390$ of ARFIMA(1,d,0) processes with standard normal innovations and parameter values $d = -0.25,\ -0.1,\ 0,\ 0.1,\ 0.25$ and $\phi_1 = -0.25,\ -0.1,\ 0,\ 0.1,\ 0.25$, respectively, are generated using a burn-in period of 10,000. For each realization, the estimators $\hat{d}_{GPH}$, $\hat{d}_{sm}$, $\hat{d}^{\beta}_{smP}$, $\beta = 0.5,\ 0.7,\ 0.9, 1$, $\widetilde{d}_k$, $k = 2,\ 3,\ 5,\ 10$, are employed for the estimation of the memory parameter d. The competing estimators are compared with respect to bias (Table 2), variance (Table 3), and RMSE (Table 4). Table 3 shows that $\widetilde{d}_2$ has indeed a smaller variance than $\hat{d}_{GPH} = \widetilde{d}_1$. The variance keeps decreasing as the number of partitions increases from two to 10. Table 2 shows that this improvement does in general not come at the cost of a greater bias. In contrast, the reduction in the variance achieved in the case of the estimator $\hat{d}^{\beta}_{smP}$ by increasing the degree of smoothing from $\beta = 0.9$ to $\beta = 0.5$ is for $d \neq 0$ accompanied by a dramatic increase in the bias. Overall, in terms of the RSME, the best results are obtained with $\hat{d}^{0.5}_{smP}$ for small values of d and with $\hat{d}^{0.7}_{smP}$ for larger value of d. However, this is only relevant in the standard case where only a single time series is available. When a large number of time series are examined simultaneously (as in the empirical study of Section 4), the bias is the decisive factor and the new estimators $\widetilde{d}_k$ are therefore more appropriate than the conventional estimators $\hat{d}^{\beta}_{smP}$.

Since values of β such as 0.5, 0.7, or 0.9 are usually chosen to minimize the MSE for a single sample, we may suspect that the estimator $\hat{d}^{\beta}_{smP}$ becomes more competitive in the case of multiple samples when the averaging is taken into account. This can be done by further reducing the degree of smoothing. Unfortunately, there is a limit to what can be achieved by increasing the value of β. Table 2 shows that large biases are still obtained with the maximum possible value of β, i.e., $\beta = 1$. This is due to the fact that global smoothing inevitably causes local distortions and cutting off higher-order sample autocovariances is not the only source of smoothing. Downweighting the sample autocovariances with the Parzen window also has a strong smoothing effect, even when all sample autocovariances are used.

Table 2. Bias of the estimators $\hat{d}_{GPH}$ (log periodogram regression), $\hat{d}_{sm}$ (simple smoothing), $\hat{d}^{\beta}_{smP}$, $\beta = 1,\ 0.9,\ 0.7,\ 0.5$ (smoothing with Parzen window and truncation lag $m = [n^{\beta}]$), and $\widetilde{d}_k$, $k = 2,\ 3,\ 5,\ 10$ (k partitions) obtained from 10,000 realizations (length: $n = 390$, number of used Fourier frequencies: $K = 20$) of Gaussian ARFIMA(1,d,0) processes with $d = -0.25,\ -0.1,\ 0,\ 0.1,\ 0.25$ and $\phi_1 = -0.25,\ -0.1,\ 0,\ 0.1,\ 0.25$.

d	ϕ_1	$\hat{d}_{GPH}$	$\hat{d}_{sm}$	$\hat{d}^{1}_{smP}$	$\hat{d}^{0.9}_{smP}$	$\hat{d}^{0.7}_{smP}$	$\hat{d}^{0.5}_{smP}$	$\widetilde{d}_2$	$\widetilde{d}_3$	$\widetilde{d}_5$	$\widetilde{d}_{10}$
−0.25	−0.25	0.0074	−0.0001	−0.0073	−0.0099	0.0345	0.1609	0.0087	0.0084	0.0098	0.0107
	−0.1	0.0050	0.0002	−0.0083	−0.0107	0.0345	0.1625	0.0080	0.0084	0.0087	0.0092
	0	0.0042	−0.0031	−0.0098	−0.0124	0.0337	0.1641	0.0065	0.0065	0.0076	0.0086
	0.1	0.0097	0.0036	−0.0049	−0.0073	0.0380	0.1664	0.0126	0.0120	0.0128	0.0140
	0.25	0.0151	0.0110	0.0006	−0.002	0.0436	0.1717	0.0165	0.0179	0.0201	0.0216
−0.1	−0.25	0.0002	−0.0029	−0.0211	−0.0280	−0.008	0.0570	0.0008	0.0016	0.0006	0.0002
	−0.1	0.0015	−0.0028	−0.0212	−0.0286	−0.0085	0.0578	−0.0001	0.0005	0.0001	−0.0001
	0	0.0039	0.0017	−0.0184	−0.0251	−0.0053	0.0601	0.0038	0.0052	0.0060	0.0057
	0.1	0.0014	0.0007	−0.0197	−0.0263	−0.0056	0.0612	0.0024	0.0028	0.0039	0.0037
	0.25	0.0055	0.0059	−0.0148	−0.0215	−0.0003	0.0666	0.0086	0.0099	0.0093	0.0101

Table 2. *Cont.*

d	ϕ_1	$\hat{d}_{GPH}$	$\hat{d}_{sm}$	$\hat{d}^{1}_{smP}$	$\hat{d}^{0.9}_{smP}$	$\hat{d}^{0.7}_{smP}$	$\hat{d}^{0.5}_{smP}$	$\tilde{d}_2$	$\tilde{d}_3$	$\tilde{d}_5$	$\tilde{d}_{10}$
0	−0.25	−0.0043	−0.0035	−0.0282	−0.0376	−0.0321	−0.0107	−0.0038	−0.0039	−0.0048	−0.0049
	−0.1	−0.0011	0.0006	−0.0258	−0.0353	−0.0299	−0.0096	−0.0004	−0.0007	−0.0004	−0.0010
	0	−0.0011	−0.0001	−0.0265	−0.0361	−0.0305	−0.0087	−0.0016	−0.0004	−0.0006	−0.0006
	0.1	−0.0001	0.0009	−0.0235	−0.0333	−0.0278	−0.0063	0.0016	0.0025	0.0019	0.0025
	0.25	0.0040	0.0064	−0.0214	−0.0309	−0.0250	−0.0022	0.0033	0.0060	0.0053	0.0073
0.1	−0.25	0.0009	0.0057	−0.0274	−0.039	−0.0475	−0.0762	0.0009	−0.0003	0.0008	−0.0001
	−0.1	0.0016	0.0056	−0.0277	−0.0396	−0.0478	−0.0754	−0.0003	0.0002	−0.0007	−0.0006
	0	−0.0005	0.0043	−0.0277	−0.0396	−0.0479	−0.0745	−0.0012	−0.0012	−0.0012	−0.0010
	0.1	0.0029	0.0059	−0.0250	−0.0374	−0.0458	−0.0727	0.0020	0.0028	0.0038	0.0034
	0.25	0.0097	0.0149	−0.0186	−0.0305	−0.0392	−0.0685	0.0088	0.0096	0.0114	0.0115
0.25	−0.25	0.0006	0.0102	−0.0314	−0.0451	−0.0690	−0.1748	0.0021	0.0018	0.0009	0.0006
	−0.1	0.0016	0.0112	−0.0314	−0.0453	−0.0689	−0.1744	0.0006	0.0011	0.0014	0.0010
	0	0.0044	0.0140	−0.0281	−0.0420	−0.0656	−0.1730	0.0032	0.0037	0.0040	0.0039
	0.1	0.0049	0.0162	−0.0269	−0.0408	−0.0649	−0.1718	0.0049	0.0065	0.0061	0.0060
	0.25	0.0079	0.0229	−0.0228	−0.0364	−0.0600	−0.1682	0.0105	0.0120	0.0130	0.0137

Table 3. Variance of the estimators $\hat{d}_{GPH}$ (log periodogram regression), $\hat{d}_{sm}$ (simple smoothing), $\hat{d}^{\beta}_{smP}$, $\beta = 1,\ 0.9,\ 0.7,\ 0.5$ (smoothing with Parzen window and truncation lag $m = [n^{\beta}]$), and $\tilde{d}_k$, $k = 2,\ 3,\ 5,\ 10$ (k partitions) obtained from 10,000 realizations (length: $n = 390$, number of used Fourier frequencies: $K = 20$) of Gaussian ARFIMA(1,d,0) processes with $d = -0.25,\ -0.1,\ 0,\ 0.1,\ 0.25$ and $\phi_1 = -0.25,\ -0.1,\ 0,\ 0.1,\ 0.25$.

d	ϕ_1	$\hat{d}_{GPH}$	$\hat{d}_{sm}$	$\hat{d}^{1}_{smP}$	$\hat{d}^{0.9}_{smP}$	$\hat{d}^{0.7}_{smP}$	$\hat{d}^{0.5}_{smP}$	$\tilde{d}_2$	$\tilde{d}_3$	$\tilde{d}_5$	$\tilde{d}_{10}$
−0.25	−0.25	0.0330	0.0328	0.0201	0.018	0.0106	0.0011	0.0287	0.0259	0.0254	0.0238
	−0.1	0.0334	0.0339	0.0207	0.0185	0.0110	0.0012	0.0297	0.0266	0.0261	0.0245
	0	0.0342	0.0337	0.0209	0.0185	0.0108	0.0011	0.0296	0.0267	0.0262	0.0248
	0.1	0.0327	0.0330	0.0202	0.0180	0.0107	0.0011	0.0287	0.0262	0.0257	0.0240
	0.25	0.0323	0.0325	0.0199	0.0178	0.0106	0.0011	0.0287	0.0260	0.0258	0.0242
−0.1	−0.25	0.0333	0.0327	0.0211	0.0187	0.0114	0.0011	0.0295	0.0268	0.0264	0.0250
	−0.1	0.0332	0.0317	0.0209	0.0186	0.0114	0.0011	0.0291	0.0264	0.0260	0.0250
	0	0.0334	0.0330	0.0212	0.0189	0.0115	0.0012	0.0298	0.0271	0.0267	0.0251
	0.1	0.0330	0.0315	0.0208	0.0185	0.0112	0.0011	0.0289	0.0262	0.0258	0.0246
	0.25	0.0328	0.0320	0.0209	0.0185	0.0112	0.0011	0.0291	0.0266	0.0263	0.0248
0	−0.25	0.0333	0.0322	0.0212	0.0191	0.0120	0.0012	0.0296	0.0268	0.0263	0.0250
	−0.1	0.0328	0.0320	0.0212	0.0191	0.0120	0.0012	0.0293	0.0268	0.0261	0.0252
	0	0.0335	0.0319	0.0214	0.0192	0.0119	0.0012	0.0297	0.0271	0.0266	0.0254
	0.1	0.0338	0.0323	0.0217	0.0195	0.0122	0.0012	0.0299	0.0271	0.0270	0.0260
	0.25	0.0332	0.0324	0.0213	0.0192	0.0120	0.0012	0.0300	0.0273	0.0269	0.0255
0.1	−0.25	0.0332	0.0327	0.0218	0.0198	0.0130	0.0012	0.0299	0.0274	0.0271	0.0260
	−0.1	0.0327	0.0321	0.0218	0.0199	0.0130	0.0012	0.0294	0.0269	0.0262	0.0252
	0	0.0328	0.0317	0.0214	0.0194	0.0127	0.0012	0.0293	0.0264	0.0263	0.0250
	0.1	0.0331	0.0321	0.0215	0.0195	0.0129	0.0012	0.0295	0.0269	0.0267	0.0256
	0.25	0.0326	0.0321	0.0217	0.0197	0.0130	0.0012	0.0293	0.0268	0.0263	0.0254
0.25	−0.25	0.0333	0.0315	0.0220	0.0202	0.0145	0.0013	0.0300	0.0271	0.0271	0.0260
	−0.1	0.0327	0.0323	0.0222	0.0205	0.0148	0.0013	0.0302	0.0278	0.0275	0.0265
	0	0.0328	0.0312	0.0219	0.0202	0.0146	0.0012	0.0297	0.0268	0.0264	0.0255
	0.1	0.0333	0.0325	0.0226	0.0207	0.0147	0.0013	0.0301	0.0274	0.0274	0.0262
	0.25	0.0339	0.0319	0.0226	0.0208	0.0150	0.0012	0.0302	0.0275	0.0272	0.0261

Table 4. RMSE of the estimators $\hat{d}_{GPH}$ (log periodogram regression), $\hat{d}_{sm}$ (simple smoothing), $\hat{d}^{\beta}_{smP}$, $\beta = 1,\ 0.9,\ 0.7,\ 0.5$ (smoothing with Parzen window and truncation lag $m = [n^{\beta}]$), and $\widetilde{d}_k$, $k = 2,\ 3,\ 5,\ 10$ (k partitions) obtained from 10,000 realizations (length: $n = 390$, number of used Fourier frequencies: $K = 20$) of Gaussian ARFIMA(1,d,0) processes with $d = -0.25,\ -0.1,\ 0,\ 0.1,\ 0.25$ and $\phi_1 = -0.25,\ -0.1,\ 0,\ 0.1,\ 0.25$.

d	ϕ_1	$\hat{d}_{GPH}$	$\hat{d}_{sm}$	$\hat{d}^{1}_{smP}$	$\hat{d}^{0.9}_{smP}$	$\hat{d}^{0.7}_{smP}$	$\hat{d}^{0.5}_{smP}$	$\widetilde{d}_2$	$\widetilde{d}_3$	$\widetilde{d}_5$	$\widetilde{d}_{10}$
−0.25	−0.25	0.1818	0.1811	0.1421	0.1344	0.1084	0.1643	0.1697	0.1612	0.1595	0.1545
	−0.1	0.1827	0.1840	0.1442	0.1365	0.1104	0.1661	0.1724	0.1634	0.1618	0.1567
	0	0.1851	0.1837	0.1449	0.1368	0.1092	0.1674	0.1721	0.1635	0.1621	0.1578
	0.1	0.1812	0.1816	0.1423	0.1343	0.1103	0.1698	0.1700	0.1623	0.1607	0.1555
	0.25	0.1803	0.1807	0.1412	0.1335	0.1119	0.1751	0.1701	0.1621	0.1619	0.1571
−0.1	−0.25	0.1825	0.1808	0.1466	0.1396	0.1070	0.0663	0.1717	0.1636	0.1624	0.1581
	−0.1	0.1823	0.1782	0.1460	0.1394	0.1072	0.0669	0.1705	0.1625	0.1611	0.1580
	0	0.1829	0.1816	0.1467	0.1398	0.1072	0.0691	0.1727	0.1647	0.1636	0.1585
	0.1	0.1817	0.1775	0.1454	0.1386	0.1061	0.0698	0.1699	0.1618	0.1607	0.1569
	0.25	0.1811	0.1789	0.1451	0.1378	0.1059	0.0745	0.1707	0.1634	0.1625	0.1578
0	−0.25	0.1826	0.1796	0.1481	0.1431	0.1142	0.0360	0.1721	0.1639	0.1624	0.1583
	−0.1	0.1812	0.1790	0.1479	0.1426	0.1137	0.0359	0.1713	0.1638	0.1615	0.1588
	0	0.1831	0.1785	0.1486	0.1433	0.1132	0.0351	0.1723	0.1646	0.1630	0.1593
	0.1	0.1837	0.1796	0.1491	0.1435	0.1139	0.0351	0.1729	0.1647	0.1645	0.1611
	0.25	0.1824	0.1801	0.1475	0.1418	0.1123	0.0345	0.1731	0.1653	0.1640	0.1599
0.1	−0.25	0.1822	0.1810	0.1502	0.146	0.1237	0.0837	0.1728	0.1657	0.1646	0.1612
	−0.1	0.181	0.1793	0.1502	0.1464	0.1237	0.0831	0.1715	0.1639	0.1617	0.1588
	0	0.181	0.1781	0.1490	0.1448	0.1226	0.0820	0.1711	0.1624	0.1622	0.1582
	0.1	0.1819	0.1792	0.1489	0.1446	0.1223	0.0805	0.1717	0.1641	0.1633	0.1599
	0.25	0.1808	0.1799	0.1485	0.1437	0.1206	0.0768	0.1713	0.1640	0.1626	0.1596
0.25	−0.25	0.1824	0.1778	0.1517	0.1493	0.1390	0.1784	0.1733	0.1648	0.1647	0.1612
	−0.1	0.1809	0.1800	0.1522	0.1502	0.1398	0.1780	0.1738	0.1666	0.1657	0.1629
	0	0.1810	0.1772	0.1505	0.1483	0.1375	0.1765	0.1723	0.1636	0.1626	0.1598
	0.1	0.1824	0.1809	0.1526	0.1495	0.1377	0.1754	0.1737	0.1657	0.1657	0.1621
	0.25	0.1842	0.1799	0.1522	0.1487	0.1363	0.1718	0.1740	0.1663	0.1654	0.1623

4. Empirical Results

In this section, we employ the estimators discussed in the previous sections for the search of possible long-range dependencies in intraday returns and absolute intraday returns. For this purpose, the limit order book data from 27 June 2007 to 30 April 2019 (2980 trading days) of the iShares Core S&P 500 ETF (IVV) are downloaded from Lobster (https://lobsterdata.com). In the process of data cleaning, 27 early-closure days (the day before Independence Day, the day after Thanksgiving, and Christmas Eve) are removed as well as 9 January 2019 because of a large number of missing values. For each of the remaining days, the first mid-quotes (midpoints of the best bid and ask quotes) in each minute and the last mid-quote in the last minute are computed and subsequently used to obtain 1-min log returns. Finally, another three days are omitted because of extreme returns, namely 19 September 2008, 6 May 2010, and 24 August 2015, which leaves 2949 days for our analysis. Estimates are computed for each day, divided by the number of days, and plotted cumulatively; hence, the last values correspond to the averages of the estimates. The validity of these values is reinforced by the striking linearity of the curves. This linearity also implies that the possible long-range dependence is not changing over time; hence, there appears to be no such thing as fractal dynamics. Figure 1a suggests that d is close to zero in case of the 1-min log returns. The large negative values obtained with $\hat{d}^{0.9}_{smP}$ and $\hat{d}^{0.7}_{smP}$ as well as the comparatively inconspicuous values obtained with $\hat{d}^{0.5}_{smP}$ can be explained with the help of the results of our simulation study. According to Table 2, they are indicative for $d = 0$. In contrast, there is strong evidence of long-range dependence in the volatility (see Figure 1b). Most estimators suggest

that the memory parameter d is approximately in the range between 0.3 and 0.4. Only the estimator $\hat{d}^{0.5}_{smP}$, which is severely downward biased in case of positive d (see Table 2), favors a smaller value.

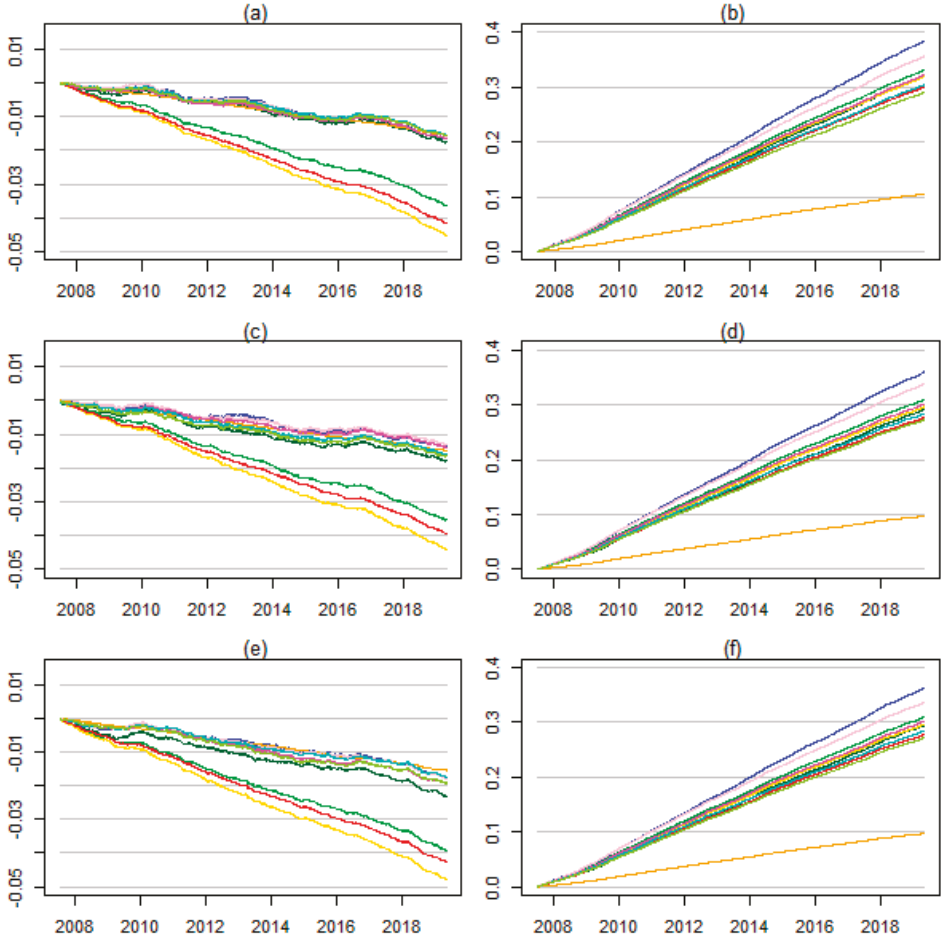


Figure 1. Cumulative plots of the estimates obtained by applying $\hat{d}_{GPH}$ (blue), $\hat{d}_{sm}$ (darkgreen), $\hat{d}^{1}_{smP}$ (green), $\hat{d}^{0.9}_{smP}$ (gold), $\hat{d}^{0.7}_{smP}$ (red), $\hat{d}^{0.5}_{smP}$ (orange), $\hat{d}_2$ (pink), $\hat{d}_3$ (magenta), $\hat{d}_5$ (turquoise), $\hat{d}_{10}$ (yellowgreen) to the (**a**) 1-min intraday log returns $r_t(s)$, $s = 1, \ldots, 390$, (**b**) absolute 1-min intraday log returns $\left|r_t(s)\right|$, (**c**) $r_t(s) - r_{t-1}(s)$, (**d**) $\left|r_t(s) - r_{t-1}(s)\right|$, (**e**) $r_t(s) - r_{t-5}(s)$, (**f**) $\left|r_t(s) - r_{t-5}(s)\right|$.

Visual significance of the differences between certain estimates can be ascertained just by observing the large differences between the slopes of the corresponding lines in Figure 1 and noting the striking stability of these lines over time. However, we still might want to augment our visual analysis with a formal statistical test. A simple way to accomplish that is to calculate the difference between two estimates separately for each trading day and compare the number of positive differences to the number of negative differences (sign test). Not surprisingly, the resulting p-values are infinitesimal. For example, even in the case of the two neighboring lines corresponding to $\hat{d}_{GPH}$ and $\widetilde{d}_2$ in Figure 1b, the p-value is less than 2.2×10^{-16}. It is still less than 9.7×10^{-8} when we omit most of the trading days and use only Wednesdays in order to ensure approximate independence of the subsamples. Note that there are $4 \times 390 = 1560$ 1-min returns between the last 1-min return of some Wednesday

and the first 1-min return of the next Wednesday plus five overnight breaks and a whole weekend. Even for a relatively large value of the memory parameter such as $d = 0.3$, the autocorrelation of an ARFIMA(0,*d*,0) process at lag $j = 1561$ is quite small, i.e.,

$$\rho(j) = \frac{\Gamma(j+d)\Gamma(1-d)}{\Gamma(j-d+1)\Gamma(d)} = \prod_{h=1}^{j} \frac{h-1+d}{h-d} \approx 0.023. \tag{43}$$

Finally, in order to check the robustness of our findings against daily and weekly periodic patterns, we repeat the graphical analyses with suitably transformed data. Replacing the 1-min log returns $r_t(s)$, $s = 1, \ldots, 390$, by the daily differences $r_t(s) - r_{t-1}(s)$ and the weekly differences $r_t(s) - r_{t-5}(s)$, respectively, ensures that any daily or weekly periodic patterns are erased while long-range dependencies remain unaffected. Figure 1c,e are very similar to Figure 1a, which shows that the insights gained from Figure 1a are genuine. Analogously, comparing Figure 1d,f with Figure 1b, we see that the same is true for the absolute returns

5. Discussion

In this paper, we have introduced a new estimator for the memory parameter d, which is based on running a log periodogram regression repeatedly for different partitions of the data. In contrast to conventional smoothing methods, which manage to achieve a reduction in the variance at the expense of an increase in the bias, our approach does not systematically have a negative impact on the bias, which makes it particularly useful for applications where the bias is the decisive factor. For example, intraday returns are usually only available during trading hours and estimation must therefore be carried out separately for each trading day. When the individual estimates are eventually combined by averaging, the variance decreases as the sample size increases, but the bias does not change. The results of an extensive simulation study confirm the good performance of the new estimator. It outperforms all of its competitors when both bias and variance are taken into account, but the bias is weighted more heavily.

The importance of results obtained with the help of simulations is due to the fact that reliable inference on the memory parameter d is not possible under general conditions. Some asymptotic results can be obtained under very restrictive conditions though. Unfortunately, convergence is typically very slow (recall the discussion in Section 2.2). Indeed, Pötscher (2002) showed that many common estimation problems in statistics and econometrics, which include the estimation of d, are ill-posed in the sense that the minimax risk is bounded from below by a positive constant independent of n and does, therefore, not converge to zero as $n \to \infty$. In particular, he found that for any estimator $\hat{d}_n$ for d based on a sample of size n from a Gaussian process with spectral density f:

$$\sup_{f \in \mathcal{F}} E\left|\hat{d}_n - d\right|^r \geq \frac{1}{2^r} > 0, \tag{44}$$

where $1 \leq r < \infty$ and $\mathcal{F}$ is the set of all ARFIMA spectral densities ($p \geq 0$, $q \geq 0$), ARFI spectral densities ($p \geq 0$, $q = 0$), or FIMA spectral densities ($p = 0$, $q \geq 0$). Furthermore, he showed that for every $f_0 \in \mathcal{F}$, (44) holds also "locally," when the supremum is taken over an arbitrarily small L_1-neighborhood of f_0. Finally, he established that confidence intervals for d coincide with the entire parameter space for d with high probability and are therefore uninformative. Nevertheless, it may be possible to formally derive the statistical properties of our new estimator for a rather narrow class of processes such as low order ARFI processes. However, this is left for future research. The current paper provides just a proof of concept.

In our empirical investigation of high-frequency data of an index ETF, we have applied the competing estimators to 1-min log returns and absolute 1-min log returns separately for each day. The results are quite stable over time and across estimation methods. The few deviations are due to conventional smoothing methods and can easily be explained by the size of their bias as shown

in Table 2. We may, therefore, safely conclude that significant long-range dependencies are present only in the intraday volatility but not in the intraday returns. These findings are genuine and not just due to daily or weekly periodic patterns because similar results are obtained when daily and weekly differences are investigated instead of the original intraday returns.

Author Contributions: Both authors contributed equally to the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors thank the academic editor and three anonymous reviewers for helpful comments and suggestions. Open Access Funding by the University of Vienna.

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

References

Adenstedt, Rolf K. 1974. On Large-Sample Estimation for the Mean of Stationary Random Sequence. *The Annals of Statistics* 2: 1095–107. [CrossRef]

Andrews, Donald W. K. 1991. Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59: 817–58. [CrossRef]

Auer, Benjamin R. 2016a. On the performance of simple trading rules derived from the fractal dynamics of gold and silver price fluctuations. *Finance Research Letters* 16: 255–67. [CrossRef]

Auer, Benjamin R. 2016b. On time-varying predictability of emerging stock market returns. *Emerging Markets Review* 27: 1–13. [CrossRef]

Barkoulas, John T., and Christopher F. Baum. 1996. Long-term dependence in stock returns. *Economic Letters* 53: 253–59. [CrossRef]

Barkoulas, John T., Christopher F. Baum, and Nickolas Travlos. 2000. Long memory in the Greek stock market. *Applied Financial Economics* 10: 177–84. [CrossRef]

Batten, Jonathan A., and Peter G. Szilagyi. 2007. Covered interest parity arbitrage and long-term dependence between the US dollar and the Yen. *Physica A* 376: 409–21. [CrossRef]

Batten, Jonathan A., Craig A. Ellis, and Thomas A. Fethertson. 2008. Sample period selection and long-term dependence: New evidence from the Dow Jones Index. *Chaos, Solitons Fractals* 36: 1126–40. [CrossRef]

Batten, Jonathan, Cetin Ciner, Brian M. Lucey, and Peter G. Szilagyi. 2013. The structure of gold and silver spread returns. *Quantitative Finance* 13: 561–70. [CrossRef]

Cajueiro, Daniel O., and Benjamin M. Tabak. 2004. The Hurst exponent over time: Testing the assertion that emerging markets are becoming more efficient. *Physica A* 336: 521–37. [CrossRef]

Carbone, Anna, Giuliano Castelli, and H. Eugene Stanley. 2004. Time-dependent Hurst exponent in financial time series. *Physica A* 344: 267–71. [CrossRef]

Chen, Gemai, Bovas Abraham, and Shelton Peiris. 1994. Lag window estimation of the degree of differencing in fractionally integrated time series models. *Journal of Time Series Analysis* 15: 473–87. [CrossRef]

Cheung, Yin-Wong, and Kon S. Lai. 1995. A search for long memory in international stock market returns. *Journal of International Money and Finance* 14: 597–615. [CrossRef]

Crato, Nuno. 1994. Some international evidence regarding the stochastic behaviour of stock returns. *Applied Financial Economics* 4: 33–39. [CrossRef]

Crato, Nuno, and Pedro J. F. de Lima. 1994. Long-range dependence in the conditional variance of stock returns. *Economics Letters* 45: 281–85. [CrossRef]

Davis, Robert, and David Harte. 1987. Tests of the Hurst effect. *Biometrika* 74: 95–101. [CrossRef]

Geweke, John, and Susan Porter-Hudak. 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4: 221–38. [CrossRef]

Granger, Clive W. J., and Roselyne Joyeux. 1980. An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis* 1: 15–29. [CrossRef]

Grau-Carles, Pilar. 2000. Empirical evidence of long-range correlations in stock returns. *Physica A* 287: 396–404. [CrossRef]

Greene, Myron T., and Bruce D. Fielitz. 1977. Long term dependence in common stock eturns. *Journal of Financial Economics* 4: 339–49. [CrossRef]

Hassler, Uwe. 1993. Regression of spectral estimators with fractionally integrated time series. *Journal of Time Series Analysis* 14: 369–80. [CrossRef]

Hauser, Michael A., and Erhard Reschenhofer. 1995. Estimation of the fractionally differencing parameter with the R/S method. *Computational Statistics & Data Analysis* 20: 569–79.

Henry, Ólan T. 2002. Long memory in stock returns: Some international evidence. *Applied Financial Economics* 12: 725–29. [CrossRef]

Hosking, Jonathan R. M. 1981. Fractional differencing. *Biometrika* 68: 165–76. [CrossRef]

Hunt, R. L., M. Shelton Peiris, and N. C. Weber. 2003. The bias of lag window estimators of the fractional difference parameter. *Journal of Applied Mathematics and Computing* 12: 67–79. [CrossRef]

Hurst, Harold E. 1951. Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 116: 770–99.

Hurvich, Clifford M., and Kaizo I. Beltrao. 1993. Asymptotics for the low-freqeuncy ordinates of the periodogram of a long-memory time series. *Journal of Time Series Analysis* 14: 455–72. [CrossRef]

Hurvich, Clifford M., Rohit Deo, and Julia Brodsky. 1998. 'The mean square error of Geweke and Porter-Hudak's estimator of the memory parameter of a long-memory time series. *Journal of Time Series Analysis* 19: 19–46. [CrossRef]

Künsch, Hans-Rudolf. 1986. Discrimination between monotonic trends and long-range dependence. *Journal of Applied Probability* 23: 1025–30. [CrossRef]

Lo, Andrew. 1991. Long-term memory in stock market prices. *Econometrica* 59: 1279–313. [CrossRef]

Lobato, Ignacio N., and N. E. Savin. 1998. Real and spurious long-memory properties of stock-market data. *Journal of Business & Economic Statistics* 16: 261–68.

Mandelbrot, Benoît. 1971. When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models. *The Review of Economics and Statistics* 53: 225–36. [CrossRef]

Mandelbrot, Benoît. 1972. Statistical methodology for non-periodic cycles: From the covariance to R/S analysis. *Annals of Economic and Social Measurement* 1: 259–90.

Mandelbrot, Benoît. 1975. Limit theorems on the delf.-normalized range for weakly and strongly dependent processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 31: 271–85. [CrossRef]

Mandelbrot, Benoît, and James Wallis. 1969. Computer experiments with fractional Gaussian noises. Parts 1, 2, 3. *Water Resources Research* 4: 909–18. [CrossRef]

Mangat, Manveer K., and Erhard Reschenhofer. 2019. Testing for long-range dependence in financial time series. *Central European Journal of Economic Modelling and Econometrics* 11: 93–106.

Newey, Whitney K., and Kenneth D. West. 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55: 703–8. [CrossRef]

Peiris, M. Shelton, and J. R. Court. 1993. A note on the estimation of degree of differencing in long memory time series analysis. *Probability and Mathematical Statistics* 14: 223–29.

Pötscher, Benedikt M. 2002. Lower risk bounds and properties of confidence sets for ill-posed estimation problems with applications to spectral density and persistence estimation, unit roots, and estimation of long memory parameters. *Econometrica* 70: 1035–65. [CrossRef]

R Core Team. 2018. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing.

Reisen, Valderio A. 1994. Estimation of the fractional difference parameter in the ARIMA(p,d,q) model using the smoothed periodogram. *Journal of Time Series Analysis* 15: 335–50. [CrossRef]

Reisen, Valderio A., Bovas Abraham, and Silvia Lopes. 2001. Estimation of parameters in ARFIMA processes: A simulation study. *Communications in Statistics: Simulation and Computation* 30: 787–803. [CrossRef]

Reschenhofer, Erhard, and Manveer K. Mangat. 2020. Detecting long-range dependence with truncated ratios of periodogram ordinates. *Communications in Statistics—Theory and Methods*. [CrossRef]

Reschenhofer, Erhard, Manveer K. Mangat, and Thomas Stark. 2020. Improved estimation of the memory parameter. *Theoretical Economics Letters* 10: 47–68. [CrossRef]

Robinson, Peter M. 1995. Log-periodogram regression of time series with long range dependence. *Annals of Statistics* 23: 1048–72. [CrossRef]

Souza, Sergio, Benjamin M. Tabak, and Daniel O. Cajueiro. 2008. Long-range dependence in exchange rates: The case of the European monetary system. *International Journal of Theoretical and Applied Finance* 11: 199–223. [CrossRef]

econometrics

MDPI

Article

Regularized Maximum Diversification Investment Strategy †

N'Golo Koné

Department of Economics, Queen's University, 94 University Avenue Kingston, Kingston, ON K7L 3N6, Canada; ngk1@queensu.ca; Tel.: +1-514-652-1005

† I am greatly indebted to Marine Carrasco for her invaluable guidance. I thank Georges Dionne, Tony S. Wirjanto, Jade Wei, and two anonymous referees for their helpful comments.

Abstract: The maximum diversification has been shown in the literature to depend on the vector of asset volatilities and the inverse of the covariance matrix of the asset return covariance matrix. In practice, these two quantities need to be replaced by their sample statistics. The estimation error associated with the use of these sample statistics may be amplified due to (near) singularity of the covariance matrix, in financial markets with many assets. This, in turn, may lead to the selection of portfolios that are far from the optimal regarding standard portfolio performance measures of the financial market. To address this problem, we investigate three regularization techniques, including the ridge, the spectral cut-off, and the Landweber–Fridman approaches in order to stabilize the inverse of the covariance matrix. These regularization schemes involve a tuning parameter that needs to be chosen. In light of this fact, we propose a data-driven method for selecting the tuning parameter. We show that the selected portfolio by regularization is asymptotically efficient with respect to the diversification ratio. In empirical and Monte Carlo experiments, the resulting regularized rules are compared to several strategies, such as the most diversified portfolio, the target portfolio, the global minimum variance portfolio, and the naive 1/N strategy in terms of in-sample and out-of-sample Sharpe ratio performance, and it is shown that our method yields significant Sharpe ratio improvements.

Keywords: portfolio selection; maximum diversification; regularization

JEL Classification: G11; C16; C52

Citation: Koné, N'Golo. 2021. Regularized Maximum Diversification Investment Strategy. *Econometrics* 9: 1. http://doi.org/10.3390/econometrics9010001

Received: 17 August 2020
Accepted: 22 December 2020
Published: 29 December 2020

1. Introduction

Since the seminal work of Markowitz (1952), which offers essential basis to portfolio selection, diversification issues have been in the center of many problems in the financial market. According to Markowitz's portfolio theory, a portfolio is diversified if its variance could not be reduced any further at the same level of the expected return.The fundamental objective of this diversification is to construct a portfolio with various assets that earns the highest return for the least volatility that may be a good alternative to the market cap-weighted portfolios. In fact, there is evidence that market portfolios are not as efficient as assumed by Sharpe (1964) in his Capital Asset Price Model (CAPM). The CAPM model as introduced by Sharpe (1964) implies that the tangency portfolio is the only efficient one and should produce the greatest returns relative to risk. Nonetheless, several empirical studies have shown that investing in the minimum variance portfolio yields better out-of-sample results than does an investment in the tangency portfolio (see Haugen and Baker 1991; Choueifaty et al. 2013; Lohre et al. 2014).

Even if these surprising results seem to be due to the high estimation risk associated with the expected returns (according to Kempf and Memmel (2006)), the efficiency of the market capitalization-weighted index has been questioned motivating numerous investment alternatives (see Arnott et al. (2005); Clarke et al. (2006); Maillard et al. (2010)). Subsequently, Choueifaty (2011) introduced the concept of maximum diversification, via

a formal definition of portfolio diversification: the diversification ratio (DR) and claimed that portfolios with maximal DRs were maximally diversified and provided an efficient alternative to market cap-weighted portfolios.

This optimal maximum diversification portfolio is shown to be a function of the inverse of the covariance matrix of asset returns (see Theron and Van Vuuren 2018), which is unknown and needs to be estimated. Solving for the maximum diversification portfolio leads to estimate the covariance matrix of returns and take its inverse. This results in estimation error, amplified due to (near) singularity of the covariance matrix, in financial markets with many assets. This, in turn, may lead to the selection of portfolios that are far from the optimal regarding standard portfolio performance measures of the financial market. Therefore, Choueifaty et al. (2013) propose the most diversified portfolio (MDP) by imposing a non-negative constraint on the maximum diversification problem[1]. However, this ad hoc constraint suggests that the MDP is unlikely to represent the final word of diversification. Without the ability to short securities it may be impossible to unlock the full range of uncorrelated risk sources present in the market (see Maguire et al. 2014). Therefore, this paper proposes a more general method to control for estimation error in the covariance matrix of asset returns without restricting the ability to short sell in the financial market. This method is fundamentally based on different ways to stabilize the inverse of the covariance matrix particularly useful when the number of assets in the financial market increases considerably compared with the estimation window. More precisely, as in Carrasco (2012) and Carrasco and Tchuente (2015) we investigate three regularization techniques including the spectral cut-off, the Tikhonov, and Landweber–Fridman approaches in order to stabilize the inverse of the covariance matrix. This procedure has been used by Carrasco et al. (2019) to stabilize the inverse of the covariance matrix in the mean-variance portfolio.

These regularization schemes involve a tuning parameter that needs to be chosen. Therefore, we propose a data-driven method for selecting the tuning parameter that minimizes the distance between the inverse of the estimated covariance matrix and the inverse of the population covariance matrix.

We show, under appropriate regularity conditions, that the selected strategy by regularization is asymptotically efficient with respect to the diversification ratio for a wide choice of the tuning parameter. Meaning that, even if the optimal diversified portfolio is unknown, there exists a feasible portfolio obtained by regularization capable of reaching similar level of performance in terms of the diversification ratio.

To evaluate the performance of our procedures, we implement a simulation exercise based on a three-factor model calibrated on real data from the US financial market. We obtain by simulation that our procedure significantly improve the performance of the proposed strategy with respect to the Sharpe ratio. Moreover, the regularized rules are compared to several strategies such as the most diversified portfolio, the target portfolio, the global minimum variance portfolio, and the naive 1/N strategy in terms of in-sample and out-of-sample Sharpe ratio, and it is shown that our method yields significant Sharpe ratio improvements. To confirm our simulations, we do an empirical analysis using Kenneth R. French's 30-industry portfolios, 100 portfolios formed on size and book-to-market, and a subset of the S&P500 index constituents. The empirical results show that by stabilizing the inverse of the covariance matrix in the maximum diversification portfolio, we considerably improve the performance of the selected strategy in terms of maximizing the Sharpe ratio.

The main finding of this paper is that by stabilizing the inverse of the covariance matrix in the maximum diversification portfolio, we considerably improve the performance of the selected portfolio with respect to several statistics in the financial market including the diversification ratio, the Sharpe ratio, and the rebalancing costs (turnover) as shown by extensive simulations and empirical study. Therefore, our methods are highly

[1] The objective is to reduce the effect of estimation error on the performance of selected maximum diversification portfolio.

recommended for investors in the sense that these procedures help them to select very effective strategies with lower rebalancing cost.

The rest of the paper is organized as follows. Section 2 presents the economy. The regularized portfolio is presented in Section 3. Section 4 gives some asymptotic properties of the selected strategy and proposes a data-driven method to select the tuning parameter. Section 5 presents some simulation results and an empirical study. Section 6 concludes the paper.

2. The Model

We consider a simple economy with N risky assets with random returns vector R_{t+1} and a risk-free asset. Let us denote R_f the gross return on this risk-free asset. Empirically with monthly data, R_f is calibrated to be the mean of the one-month Treasury-Bill (T-B) rate observed in the data. The number of risky assets in our economy N is assumed to be large for diversification issue.

We assume that the excess returns $r_{t+1} = R_{t+1} - R_f 1_N$ are independent and identically distributed with the mean and the covariance matrix given by μ and $\Sigma = \left\{\sigma_{i,j}\right\}_{i,j\in N}$, respectively. Let us denote by $\omega = (\omega_1, ..., \omega_N)^{'}$ the vector of portfolio weights that represents the amount of the capital to be invested in the risky assets and the remain $1 - \omega^{'} 1_N$ is allocated to the risk-free asset. Short-selling is allowed in the financial market, i.e., some of the weights ω_i could be negative. Let us denote $\sigma = (\sigma_{1,1}, ..., \sigma_{N,N})^{'}$ the vector of asset volatilities.

According to Choueifaty (2011), the diversification ratio (DR) of any portfolio ω is given by

$$DR(\omega) = \frac{\omega^{'}\sigma}{\sqrt{\omega^{'}\Sigma\omega}} \tag{1}$$

which is the ratio of weighted average of volatilities divided by the portfolio volatility.

Using the relation in Equation (1), the maximum diversification portfolio is obtained by solving the following optimization problem

$$\max_{\omega} DR(\omega). \tag{2}$$

As the DR is invariant by scalar multiplication (for instance see Choueifaty et al. (2013)), solving the problem in Equation (2) is equivalent of solving the following new problem according to Theron and Van Vuuren (2018)

$$\min_{\omega^{'}\sigma=1} \frac{1}{2}\omega^{'}\Sigma\omega. \tag{3}$$

This new optimization problem is very close to the global minimum variance portfolio. The only difference is that the constraint $\omega^{'}1 = 1$ in the global minimum variance problem is replaced by $\omega^{'}\sigma = 1$. The solution of this new optimization problem is given by

$$\omega = \frac{\Sigma^{-1}\sigma}{\sigma^{'}\Sigma^{-1}\sigma} = \left(\sigma^{'}\Sigma^{-1}\sigma\right)^{-1}\left(\Sigma^{-1}\sigma\right). \tag{4}$$

The solution in (4) is unknown because it depends on the covariance matrix of asset returns and the vector of volatilities that are unknown and need to be estimated from available data set. We need, in particular, to estimate the covariance matrix and take its inverse. The sample covariance may not be appropriate because it may be nearly singular, and sometimes not invertible. The issue of ill-conditioned covariance matrix must be

addressed because inverting such matrix increases dramatically the estimation error and then makes the maximum diversification portfolio unreliable. Many techniques have been proposed in the literature to stabilize the inverse of the covariance matrix in the solution in (4). According to Carrasco et al. (2007), an interesting way to stabilize the inverse of the covariance matrix consists of dampening the explosive effect of the inversion of the singular values of $\hat{\Sigma}$. It consists in replacing the sequence $\{1/\lambda_j\}$ of explosive inverse singular values by a sequence $\{q(\alpha, \lambda_j)/\lambda_j\}$, where the damping function $q(\alpha, \lambda)$ is chosen such that

1. $q(\alpha, \lambda)/\lambda$ remains bounded when $\lambda \to 0$
2. for any λ, $\lim_{\alpha \to 0} q(\alpha, \lambda) = 1$

where α is the regularization parameter. The damping function is specific to each regularization.

In this paper, we propose a consistent way to estimate the solution in (4) using three regularization schemes based on three different ways of inverting the covariance matrix of asset returns. These regularization techniques are the spectral cut-off, the Tikhonov, and the Landweber–Fridman. The spectral cut-off regularization scheme is based on principal components whereas the Tikhonov's one is based on Ridge regression (also called Bayesian shrinkage) and the last one is an iterative method.

3. The Regularized Portfolio

The regularization methods used in this paper are drawn from the literature on inverse problems (see Kress (1999)). They are designed to stabilize the inverse of Hilbert–Schmidt operators (operators for which the eigenvalues are square summable). These regularization techniques will be applied to the sample covariance matrix of asset returns to stabilize its inverse in the selected portfolio.

Let $\hat{\lambda}_1^2 \geq \hat{\lambda}_2^2 \geq ... \geq \hat{\lambda}_N^2 \geq 0$ be the eigenvalues of the sample covariance matrix $\hat{\Sigma}$. By spectral decomposition, we have that $\hat{\Sigma} = PDP'$ with $PP' = I_N$, where P is the matrix of eigenvectors and D the diagonal matrix with eigenvalues $\hat{\lambda}_j^2$ on the diagonal. Furthermore, let $\hat{\Sigma}^\alpha$ be the regularized inverse of $\hat{\Sigma}$.

$$\hat{\Sigma}^\alpha = PD^\alpha P'$$

where D^α is the diagonal matrix with elements $q(\alpha, \hat{\lambda}_j^2)/\hat{\lambda}_j^2$. The positive parameter α is the regularization parameter, a kind of smoothing parameter which is unknown and needs to be selected. $q(\alpha, \hat{\lambda}_j^2)$ is the damping function that depends on the regularization scheme used.

3.1. Tikhonov Regularization (TH)

This regularization scheme is close to the well known ridge regression used in presence of multicollinearity to improve properties of OLS estimators. In Tikhonov's regularization scheme, the real function $q(\alpha, \hat{\lambda}_j^2)$ is given by

$$q(\alpha, \hat{\lambda}_j^2) = \frac{\hat{\lambda}_j^2}{\hat{\lambda}_j^2 + \alpha}$$

3.2. The Spectral Cut-Off (SC)

It consists in selecting the eigenvectors associated with the eigenvalues greater than some threshold.

$$q(\alpha, \hat{\lambda}_j^2) = I\left\{\hat{\lambda}_j^2 \geq \alpha\right\}$$

The explosive influence of the factor $1/\hat{\lambda}_j^2$ is filtered out by imposing $q(\alpha, \hat{\lambda}_j^2) = 0$ for small $\hat{\lambda}_j^2$, that is, $\hat{\lambda}_j^2 < \alpha$. α is a positive regularization parameter such that no bias is introduced when $\hat{\lambda}_j^2$ exceeds the threshold α. Another version of this regularization scheme is the Principal Components (PC) which consists in using a certain number of eigenvectors to compute the inverse of the operator. The PC and the SC are perfectly equivalent, only the definition of the regularization term α differs. In the PC, α is the number of principal components. In practice, both methods will give the same estimator.

3.3. Landweber–Fridman Regularization (LF)

In this regularization scheme, $\hat{\Sigma}^{\alpha}$ is computed by an iterative procedure with the formula

$$\begin{cases} \hat{\Sigma}_l^{\alpha} = \left(I_N - c\hat{\Sigma}^{\alpha}\right)\hat{\Sigma}_{l-1} + c\hat{\Sigma} & \text{for } l = 1, 2, ...1/\alpha - 1 \\ \hat{\Sigma}_0^{\alpha} = c\hat{\Sigma} \end{cases}$$

The constant c must satisfy $0 < c < 1/\hat{\lambda}_1^2$. Alternatively, we can compute this regularized inverse with

$$q(\alpha, \hat{\lambda}_j^2) = 1 - \left(1 - c\hat{\lambda}_j^2\right)^{\frac{1}{\alpha}}$$

The basic idea behind this procedure is similar to the spectral cut-off method but with a smooth bias function.

See Carrasco et al. (2007) for more details about these regularization techniques. The regularized diversified portfolio for a given regularization scheme is

$$\hat{\omega}_{\alpha} = \frac{\hat{\Sigma}^{\alpha}\hat{\sigma}}{\hat{\sigma}'\hat{\Sigma}^{\alpha}\hat{\sigma}} = \left(\hat{\sigma}'\hat{\Sigma}^{\alpha}\hat{\sigma}\right)^{-1}\hat{\Sigma}^{\alpha}\hat{\sigma}. \tag{5}$$

This regularized portfolio depends on an unknown tuning parameter that needs to be selected through a data-driven method.

4. Asymptotic Properties of the Selected Portfolio

In this section, we will look at the efficiency of the regularized portfolio with respect to the diversification ratio. We will also propose a data-driven method to select the tuning parameter.

4.1. Efficiency of the Regularized Diversified Portfolio

To obtain the efficiency of the selected portfolio, we need to impose some regularity conditions, in particular we will need the following assumption.

Assumption A: $\frac{\Sigma}{N}$ is a trace class operator.

A a trace class operator K is a compact operator with a finite trace, i.e., $Tr(K) = O(1)$. This assumption is more realistic than assuming that Σ is a Hilbert–Schmidt operator. Moreover, Carrasco et al. (2019) show that Assumption A holds when the returns are generated from a standard factor model.

Under Assumption A, the following proposition presents information about the asymptotic property of the diversification ratio associated with the selected portfolio.

Proposition 1. *Under Assumption A we have that*

$$DR(\hat{\omega}_{\alpha}) \to_p DR(\omega_t), \tag{6}$$

if $\frac{N}{\alpha\sqrt{T}} \to 0$ *as* T *goes to infinity.*

Proof. In Appendix A. □

Comment on Proposition 1. The regularity condition behind proposition 1 implies several things: First, $\alpha\sqrt{T} \to +\infty$ implies that the estimation window should go to infinity faster than the optimal tuning parameter goes to zero. Second, $\frac{N}{\alpha\sqrt{T}} \to 0$ implies that $\alpha\sqrt{T}$ should go to infinity faster than the number of assets in the financial market. Therefore, the number of assets should be limited asymptotically compared with the estimation window. As the regularization parameter α is in $(0,1)$, $\frac{N}{\alpha\sqrt{T}} \to 0$ is implied by the following condition $\frac{N}{\sqrt{T}} \to 0$. However, the regularity condition $\frac{N}{\sqrt{T}} \to 0$ seems to be more restrictive than assuming that $\frac{N}{T} \to Constant$. One way to avoid this regularity condition will be to assume that the covariance matrix of assets returns is a trace class operator or to assume that this covariance matrix is a Hilbert–Schmidt operator. These assumptions seem to be more restrictive than assuming that $\frac{N}{\sqrt{T}} \to 0$, which seems to be close to the reality asymptotically. Moreover, $\frac{N}{\sqrt{T}} \to 0$ is only an asymptotic assumption and we do not need to have $\frac{N}{\sqrt{T}}$ close to zero in practice to obtain good performance with the regularized portfolio. Particularly, in finite sample, N could be larger than T or too close to T. Proposition 1 shows that the regularized diversified portfolio is asymptotically efficient in terms of the diversification ratio for a wide choice of the tuning parameter. Meaning that, even if the optimal diversified portfolio in Equation (4) is unknown, there exists a feasible portfolio obtained by regularization capable of reaching similar level of performance in terms of the diversification ratio.

4.2. Data-Driven Method for Selecting the Tuning Parameter

We show in the previous sections that the selected portfolio depends on a certain smoothing parameter $\alpha \in (0,1)$. We have derived the efficiency of the selected portfolio assuming that this tuning parameter is given. However, in practice, the regularization parameter is unknown and needs to be selected. Therefore, we propose a data-driven selection procedure to obtain an approximation of this parameter.

Our objective here is to select the tuning parameter which minimizes the distance between the inverse of the estimated covariance matrix and the inverse of the true covariance matrix. According to Ledoit and Wolf (2003), most of the existing shrinkage estimators from finite-sample statistical decision theory as well as in Frost and Savarino (1986) break down when $N \geq T$ because their loss functions involve the inverse of the sample covariance matrix which is a singular matrix in this situation. Therefore, to avoid this problem, they propose a loss function that does not depend on this inverse. This loss function is a quadratic measure of distance between the true and the estimated covariance matrices based on the Frobenius norm. Unlike in Ledoit and Wolf (2003), we will use a loss function that depends on the inverse of the covariance matrix under the assumption that the true covariance matrix is invertible. One important thing to notice here is that the regularized covariance matrix is always invertible even if $N \geq T$ meaning that our loss function exists for $N \geq T$. In fact, we know that the optimal diversified portfolio as given by Equation (4) depends on the inverse of the covariance matrix of assets returns. Moreover, because our objective is to stabilize the inverse of this covariance matrix in the estimated portfolio by regularization, we propose to use a loss function that minimizes a quadratic distance between the regularized inverse and the theoretical covariance matrix.

The loss function we consider here is given by

$$\mu' \left[\left(\hat{\Sigma}^{\alpha} - \Sigma^{-1} \right)' \Sigma \left(\hat{\Sigma}^{\alpha} - \Sigma^{-1} \right) \right] \mu \tag{7}$$

where μ is the expected excess return. The choice of this specific quadratic distance is useful to obtain a criterion that can easily be approximated by generalized cross validation approach.

Therefore, the objective is to select the tuning parameter that minimizes

$$E\left\{\mu'\left[\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)'\Sigma\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)\right]\mu\right\}. \tag{8}$$

It implies that

$$\hat{\alpha}=arg\min_{\alpha\in H_T}E\left\{\mu'\left[\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)'\Sigma\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)\right]\mu\right\} \tag{9}$$

To obtain a better approximation of the tuning parameter based on a generalized cross-validation criterion, we need additional assumptions. Therefore, let us start with some useful notations.

We denote by $\Omega=E\left(r_t r_t'\right)=E\left(X'X\right)/T$ and $\beta=\Omega^{-1}\mu=E(X'X)^{-1}E(X'1_T)$ where $r_t, t=1,\cdots,T$ are the observations of the excess returns and X the $T\times N$ matrix with tth row given by r_t'.

Assumption B

For some $\nu>0$, we have that

$$\sum_{j=1}^{N}\frac{<\beta,\phi_j>^2}{\eta_j^{2\nu}}<\infty$$

where ϕ_j and η_j^2 denote the eigenvectors and eigenvalues of $\frac{\Omega}{N}$.

The regularity condition in Assumption B can be found in Carrasco et al. (2007) and Carrasco (2012). Moreover, Carrasco et al. (2019) show that Assumption B hold if the returns are generated by a factor model. Assumption B is used combined with Assumption A to derive the rate of convergence of the mean squared error in the OLS estimator of β. These two assumptions imply in particular that $\|\beta\|^2<+\infty$ such that we have the following relations,

$$\|\beta-\beta_\alpha\|^2=\begin{cases}O(\alpha^{\nu}) & \text{for } SC, LF\\ O\left(\alpha^{min(\nu,2)}\right) & \text{for } T\end{cases}$$

β_α is the regularized version of β.

The following result gives us a very nice equivalent of the objective function. We can easily apply a cross-validation approximation procedure on this expression of the objective function.

Proposition 2. *Under Assumptions A and B we have that*

$$E\left\{\mu'\left[\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)'\Sigma\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)\right]\mu\right\}\sim E\left\{\left(\hat{\Sigma}^{\alpha}\hat{\mu}-\Sigma^{-1}\mu\right)'\Sigma\left(\hat{\Sigma}^{\alpha}\hat{\mu}-\Sigma^{-1}\mu\right)\right\}$$

if $\frac{1}{\alpha^2 T}\to 0$ *and* $\sqrt{N}\alpha^{\min(\frac{\nu}{2},1)}\to 0$ *as* T *goes to infinity.*

Proof. In Appendix B. □

We obtain the following corollary from this proposition.

Corollary 1. *Under Assumptions A and B we have that*

$$E\left\{\mu'\left[\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)'\Sigma\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)\right]\mu\right\}\sim\frac{1}{T}E\|X(\hat{\beta}_\alpha-\beta)\|^2+\frac{(\mu'(\beta_\alpha-\beta))^2}{(1-\mu'\beta)}$$

if $\frac{1}{\alpha^2 T}\to 0$ *and* $\sqrt{N}\alpha^{\min(\frac{\nu}{2},1)}\to 0$ *as* T *goes to infinity.*

The result in Corollary 1 is obtained by using Proposition 2 combined with Proposition 1 in Carrasco et al. (2019).

From Corollary 1, it follows that minimizing $E\left\{\mu'\left[\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)'\Sigma\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)\right]\mu\right\}$ is equivalent to minimizing

$$\frac{1}{T}E\|X(\hat{\beta}_{\alpha}-\beta)\|^2 \tag{10}$$

$$+\frac{(\mu'(\beta_{\alpha}-\beta))^2}{(1-\mu'\beta)}. \tag{11}$$

Terms (10) and (11) depend on the unknown β, and therefore need to be approximated. The approximation of these two quantities is borrowed from Carrasco et al. (2019). More precisely, the rescaled MSE

$$\frac{1}{T}E\left[\left\|X\left(\widehat{\beta}_{\alpha}-\beta\right)\right\|^2\right]$$

can be approximated by generalized cross-validation criterion:

$$GCV(\alpha)=\frac{1}{T}\frac{\|(I_T-M_T(\alpha))1_T\|^2}{(1-tr(M_T(\alpha))/T)^2}.$$

Using the fact that

$$\hat{\mu}'(\beta_{\alpha}-\beta)=\frac{1_T'}{T}(M_T(\alpha)-I_T)X\beta,$$

(11) can be estimated by plug-in:

$$\frac{\left(1_T'(M_T(\alpha)-I_T)X\hat{\beta}_{\tilde{\alpha}}\right)^2}{T^2\left(1-\hat{\mu}'\hat{\beta}_{\tilde{\alpha}}\right)} \tag{12}$$

where $\hat{\beta}_{\tilde{\alpha}}$ is an estimator of β obtained for some consistent $\tilde{\alpha}$ ($\tilde{\alpha}$ can be obtained by minimizing $GCV(\alpha)$).

The optimal value of τ is defined as

$$\hat{\alpha}=\arg\min_{\tau\in H_T}\left\{GCV(\alpha)+\frac{\left(1_T'(M_T(\alpha)-I_T)X\hat{\beta}_{\tilde{\alpha}}\right)^2}{T^2\left(1-\hat{\mu}'\hat{\beta}_{\tilde{\alpha}}\right)}\right\}$$

where $H_T=\{1,2,...,T\}$ for spectral cut-off and Landweber–Fridman and $H_T=(0,1)$ for Ridge.

5. Simulations and Empirical Study

We start this section by a simulation exercise to set up the performance of our procedure and compare our result to the existing methods. In particular, we compare our method to the most diversified portfolio proposed by Choueifaty and Coignard (2008). More precisely, we focus on how our procedure performs in terms of the Sharpe ratio and the diversification ratio. To end this section, we analyze the out-of-sample performance of the selected portfolio.

5.1. Data

In our simulations and empirical analysis, various forms of monthly data will be used from July 1980 to June 2016. The one-month Treasury-Bill (T-Bill) rate is used as a proxy for the risk-free rate, and R_f is calibrated to be the mean of the one-month Treasury-Bill rate observed in the data. We use monthly returns of Fama–French three factors and of 30 industry portfolios from the Kenneth R. French data library in order to calibrate unknown

parameters of the simulation model. In the empirical study, we also use monthly data for the 100 portfolios formed on size and book-to-market from the Kenneth R. French data Library and the CRSP monthly data for the S&P500 index constituents.

5.2. Simulation

We implement a simple simulation exercise to assess the performance of our procedure and compare it with the existing procedures. Let us consider for this purpose a simple economy with $N \in \{10, 20, 40, 60, 80, 90, 100\}$ risky assets. We use several values of N to see how the size of the financial market (defined by the number of assets in the economy) could affect the performance of the selected strategy. Let T be the sample size used to estimate the unknown parameters in the investment process. Following Chen and Yuan (2016) and Carrasco et al. (2019), we simulate the excess returns at each simulation step from the following three-factor model for $i = 1, ..., N$ and $t = 1, ..., T$

$$r_{it} = b_{i1} f_{1t} + b_{i2} f_{2t} + b_{i3} f_{3t} + \epsilon_{it} \tag{13}$$

$f_t = (f_{1t}, f_{2t}, f_{3t})'$ is the vector of common factors, $b_i = (b_{i1}, b_{i2}, b_{i3})'$ is the vector of factor loadings associated with the ith asset, and ϵ_{it} is the idiosyncratic component of r_{it} satisfying $E(\epsilon_{it}|f_t) = 0$. We assume that $f_t \sim \mathcal{N}\left(\mu_f, \Sigma_f\right)$, where μ_f and Σ_f are calibrated on the monthly data of the market portfolio, the Fama–French size, and the book-to-market portfolio from July 1980 to June 2016. Moreover, we assume that $b_i \sim \mathcal{N}(\mu_b, \Sigma_b)$ with μ_b and Σ_b calibrated using data of 30 industry portfolios from July 1980 to June 2016. Idiosyncratic terms ϵ_{it} are supposed to be normally distributed. The covariance matrix of the residual vector is assumed to be diagonal and given by Σ_ϵ=diag$\left(\sigma_1^2, ..., \sigma_N^2\right)$ with the diagonal elements drawn from a uniform distribution between 0.10 and 0.30 to yield an average cross-sectional volatility of 20%.

In the compact form (13) can be written as follows,

$$R = BF + \epsilon \tag{14}$$

where B is a $N \times 3$ matrix whose ith row is b_i'. The covariance matrix of the vector of excess return r_t is given by

$$\Sigma = B\Sigma_f B' + \Sigma_\epsilon.$$

The mean of the excess return is given by $\mu = B\mu_f$. The return on the risk-free asset R_f is calibrated to be the mean of the one-month T-B observed in the data from July 1980 to June 2016.

The calibrated parameters used in our simulation process are given in Table 1. The gross return on the risk-free asset calibrated on the data is given by $R_f = 1.0036$. Once generated, the factor loadings are kept fixed over replications, while the factors differ from simulations and are drawn from a trivariate normal distribution.

Table 1. Calibrated parameters.

Parameters for Factors Loadings				**Parameters for Factors Returns**			
μ_b		Σ_b		μ_f		Σ_f	
1.0267	0.0422	0.0388	0.0115	0.0063	0.0020	0.0003	−0.0004
0.0778	0.0388	0.0641	0.0162	0.0011	0.0003	0.0009	−0.0003
0.2257	0.0115	0.0162	0.0862	0.0028	−0.0004	−0.0003	0.0009

Let SR(ω_t) be the Sharpe ratio associated with the optimal portfolio ω_t, then SR(ω_t) is given as follows,

$$SR(\omega_t) = \left[\mu' \Sigma \mu\right]^{1/2}$$

To evaluate the performance of our procedure in terms of the Sharpe ratio, we focus on the actual Sharpe ratio associated with the selected portfolio. The actual Sharpe ratio at time point t is given by

$$SR(\hat{\omega}_t) = \frac{\hat{\omega}_t' \mu}{\left[\hat{\omega}_t' \Sigma \hat{\omega}_t'\right]^{1/2}}$$

We consider the following portfolio selection procedures.

- The sample-based diversified portfolio (SbDP). This strategy is obtained using sample moments to estimate the unknown parameters in the maximum diversification portfolio.

$$SbDP = \frac{\hat{\Sigma}^{-1}\hat{\sigma}}{\hat{\sigma}'\hat{\Sigma}^{-1}\hat{\sigma}}$$

- The most diversified portfolio (MDP) proposed by Choueifaty et al. (2013). This strategy is obtained by solving the optimization problem in Equation (2) under the following constraint,
  $\omega_i \geq 0 \; for \; i = 1, ..., N.$
  The closed form associated with this new optimization problem is given as follows,

$$MDP = diag(\Sigma) C^{-1} 1$$

  where $diag(\Sigma)$ is a diagonal matrix of assets volatilities, C the correlation matrix, and 1 a $N \times 1$ vector of ones. The MDP is then estimated by replacing the unknown parameters by their empirical counterparts.
- The global minimum variance portfolio (GMVP) obtained by minimizing the variance of the return on the optimal selected portfolio. By solving this optimization problem, the following closed form is obtained,

$$GMVP = \frac{\Sigma^{-1}1}{1'\Sigma^{-1}1}$$

  This solution is then estimated by replacing the covariance matrix by the sample covariance matrix.
- The regularized strategies such as: the ridge regularized diversified portfolio (RdgDP), the spectral cut regularized diversified portfolio (SCDP), and the Landweber–Fridman regularized diversified portfolio (LFDP).
- The equal-weighted portfolio which is also called the naive portfolio (XoNP) which allocates a constant amount 1/N+1 in each asset.
- The target (or the maximum Sharpe ratio) portfolio (TgP). The closed form of the target portfolio is

$$TgP = \frac{\Sigma^{-1}\mu}{\mu'\Sigma^{-1}1}$$

  This portfolio is also estimated using sample moments such as the sample mean and the sample covariance matrix to estimate the unknown parameters.

- The linear factor-based shrinkage estimators proposed by Ledoit and Wolf (2003) (LWP). It consists of replacing the sample covariance matrix in the selected portfolio by an optimally weighted average of two existing estimators: the sample covariance matrix and single-index covariance matrix. This method involves also a tuning parameter that is unknown and has been selected by the authors. The tuning parameter selection procedure proposed in Ledoit and Wolf (2003) is based on minimizing the distance between the population covariance matrix and the regularized one. This implies that the way they select the turning parameter is different from our data-driven method. Therefore, the LWP will be considered here as a very good benchmark (and it will be the only benchmark that we consider) to evaluate the ability of our data-driven method to deliver additional performance compare to other data-driven methods.

We perform 1000 simulations and estimate our statistics over replications. We obtain the following results about the actual Sharpe ratio.

Table 2 contains the results about the average monthly Sharpe ratio obtained by simulations. The results show that the sample based diversified portfolio performs very poorly in terms of maximizing the Sharpe ratio in the financial market with large number of assets. This result is essentially due to the fact that the estimation error from estimating the vector of assets volatilities is amplified by using the sample covariance matrix of assets returns closed to a singular matrix when N becomes too large compared with the sample size. Therefore, even if this strategy is supposed to be the maximum diversification's one with the highest Sharpe ratio, the SbDP is dominated by several other strategies such as the GMVP, the XoNP, and the TgP. Therefore, this strategy cannot be consider as the maximum diversification strategy in practice. To solve this problem, Choueifaty et al. (2013) proposes the most diversified portfolio (MDP) obtained by maximizing the diversification ratio under a non-negative constraint on the portfolio weights. This additional constraint in the investment process may help to reduce the effect of estimation error on the performance of the selected portfolio. The results of this analysis are in Table 2. By imposing the non-negative constraint, investors considerably improve the performance of the selected portfolio in terms of the Sharpe ratio. This new strategy even outperforms the global minimum variance portfolio. However, this procedure is still dominated by the target portfolio and the equal weighted portfolio meaning that much remains to be done about finding the maximum diversification strategy in practice. One explanation to this result is that imposing the non-negative constraint on the portfolio weight may limit the ability of the selected portfolio to be fully diversified. Therefore, one needs to find a more general estimation procedure for the maximum diversified portfolio that allows for short selling.

Table 2. The average monthly Actual Sharpe ratio from optimal strategies using a three-factor model as a function of the number of assets in the economy with the sample size $n = 120$, over 1000 replications. True SR is the true actual Sharpe ratio.

Strategies	**Number of Risky Assets**						
	10	**20**	**40**	**60**	**80**	**90**	**100**
SbDP	0.1549	0.0906	0.0889	0.0779	0.0652	0.0719	0.0704
XoNP	0.2604	0.2604	0.2415	0.2525	0.2406	0.2461	0.2467
GMVP	0.2227	0.2338	0.2098	0.2298	0.1710	0.1640	0.1449
MDP	0.2514	0.2545	0.2410	0.2544	0.1778	0.1821	0.1935
TgP	0.2608	0.2818	0.2662	0.2687	0.2026	0.1925	0.1699
LWP	0.2589	0.2702	0.2688	0.2704	0.2628	0.2521	0.2507
RdgDP	0.2587	0.2785	0.2817	0.2907	0.2947	0.2830	0.2991
SCDP	0.2592	0.2872	0.2993	0.2898	0.2746	0.2887	0.2853
LFDP	0.2605	0.2765	0.2840	0.2870	0.2850	0.2912	0.2980
True SR	0.2626	0.2922	0.3393	0.3379	0.3592	0.3477	0.3657

For this purpose, we propose a new way to estimate the optimal diversified portfolio by stabilizing the inverse of the sample covariance matrix without imposing a non-negative constraint on the portfolio weights in the investment process. The results of these methods are also in Table 2. The first thing to point out about these results is that the regularized diversified portfolio outperforms the most diversified portfolio in terms of maximizing the Sharpe ratio. For instance, we obtain an average Sharpe ratio of 0.2514, 0.2587, 0.2592, and 0.2605 for the MDP, the RdgDP, the SCDP, and the LFDP, respectively, when only 10 assets are considered in the economy. The difference in terms of the actual Sharpe ratio performance between our procedure and the most diversified portfolio significantly increases with the number of assets in the financial market. For example, for 100 assets, the average Sharpe ratio is about 0.1935, 0.2991, 0.2853, and 0.2980 for the MDP, the RdgDP, the SCDP, and the LFDP, respectively. This results may be due to the fact that when the number of assets in the economy increases, the degree of diversification of the selected strategy may deteriorate with non-negative constraints on the investment process that may reduce the ability to find a strategy that performs the Sharpe ratio. Moreover, the regularized diversified portfolio outperforms the target strategy and the equal-weighted portfolio when the number of assets in the financial market exceeds 40. Nonetheless, for 10 assets in the economy, the target portfolio outperforms the RdgDP and the SCDP but is dominated by the LFDP. With 20 assets the target portfolio dominates the RdgDP and the LFDP and is dominated by the SCDP. The equal-weighted portfolio outperforms some regularized strategies such as the RdgDP and the SCDP only for 10 assets in the financial market. The fact that the regularized strategies give very interesting results in terms of maximizing the Sharpe ratio (compared with the existing strategies) for large N is because these methods are essentially used to address estimation issues in large dimensional problems. The performance of these procedures seems to be independent of the size of the financial market. In fact, with a reasonable choice of the tuning parameter, each of these methods can achieve satisfactory performance in terms of the Sharpe ratio even if the number of assets in the economy is large.

Our regularized portfolio also outperforms the selected strategy obtained using the linear shrinkage estimator of Ledoit and Wolf (2003) to estimate the covariance matrix of asset returns. The difference in terms of performance between these two portfolios tends to become large when the number of assets we consider in the economy increases. This result can be due to the fact that the estimation error associated with estimating the single-index covariance matrix may be important for very large assets. One other thing that could explain this result comes from the fact that our tuning parameter is selected to minimize the distance between the regularized inverse of the covariance matrix and the inverse of the population's one. Moreover, because the optimal portfolio depends on the inverse of the covariance matrix, selecting a tuning parameter that minimizes the estimation error in the inverse of the covariance matrix seems to be more appropriate than choosing this parameter to minimize the estimation error in the covariance matrix. One important thing to point out is that the Ridge regularized portfolio is a special case of the selected portfolio with the linear shrinkage estimation of the covariance matrix. In this case, the structural covariance matrix is replaced by the identity to avoid the potential estimation error which may be associated with this covariance matrix.

Similar results are obtained when choosing the estimation window to be 1000 and by increasing the number of assets in the economy from 150 to 999 ($N \in \{150, 250, 400, 550, 700, 850, 950, 999\}$) as given in Table 3.

Table 3. The average monthly Actual Sharpe ratio from optimal strategies using a three-factor model as a function of the number of assets in the economy with the sample size $n = 1000$, over 1000 replications. True SR is the true actual Sharpe ratio.

Strategies	Number of Risky Assets							
	150	**250**	**400**	**550**	**700**	**850**	**950**	**999**
SbDP	0.1230	0.1104	0.103	0.0998	0.060	0.03	0.012	0.008
XoNP	0.2630	0.2640	2507	0.240	0.238	0.2207	0.2180	0.220
GMVP	0.3080	02908	0.2890	0.2780	0.250	0.1980	0.1017	0.095
MDP	0.3280	0.3305	0.3198	0.309	0.2679	0.2892	0.1985	0.120
TgP	0.3290	0.3105	0.307	0.3100	0.2608	0.210	0.180	0.098
LWP	0.3302	0.3408	0.3318	0.3070	0.415	0.4504	0.4601	0.4807
RdgDP	0.3702	0.3850	0.3980	0.458	0.524	0.540	0.558	0.601
SCDP	0.3689	0.3860	0.3980	0.460	0.5230	0.535	0.590	0.608
LFDP	0.3704	0.3840	0.3984	0.4560	0.5250	0.538	0.585	0.595
True SR	0.3758	0.3904	0.407	0.489	0.5480	0.588	0.608	0.618

To analyze the statistical significance of the regularized portfolio over the other strategies, we implement the following test procedure about the Sharpe ratio,

$$H_0 : RSR \leq SR_0 vs H_1 : RSR > SR_0$$

where RSR is the regularized Sharpe ratio and SR_0 the Sharpe ratio of the portfolio under comparison. This test is conducted using the same procedure as in Ao et al. (2019). For more information about this test procedure see Jobson and Korkie (1981) and Memmel (2003). The fundamental objective of this test procedure is to confirm the domination of our method over the existing strategies with a statistic test. The p-values associated with this test procedure for each of the regularized portfolios are given in Tables 4–6. According to these results, our regularized portfolio dominates the other strategies in terms of maximizing the Sharpe ratio at the significant level 5%. In particular, our method outperforms the LW portfolio in the large financial market setting.

Table 4. The p-value associated with performance hypothesis testing with the Sharpe ratio from Ridge regularized strategy using a three-factor model as a function of the number of assets in the economy with the sample size $n = 1000$, over 1000 replications.

Strategies	Number of Risky Assets							
	150	**250**	**400**	**550**	**700**	**850**	**950**	**999**
SbDP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
XoNP	0.004	0.002	0.007	0.005	0.000	0.000	0.000	0.000
GMVP	0.008	0.004	0.006	0.007	0.000	0.000	0.000	0.000
MDP	0.003	0.001	0.002	0.000	0.000	0.000	0.000	0.000
TgP	0.009	0.003	0.008	0.004	0.001	0.000	0.008	0.000
LWP	0.089	0.013	0.001	0.012	0.035	0.003	0.043	0.008

Table 5. The *p*-value associated with Performance hypothesis testing with the Sharpe ratio from Landweber–Fridman regularized strategy using a three-factor model as a function of the number of assets in the economy with the sample size $n = 1000$, over 1000 replications.

Strategies	Number of Risky Assets							
	150	**250**	**400**	**550**	**700**	**850**	**950**	**999**
SbDP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
XoNP	0.003	0.001	0.008	0.007	0.001	0.000	0.000	0.000
GMVP	0.010	0.003	0.007	0.002	0.001	0.000	0.000	0.000
MDP	0.005	0.001	0.004	0.000	0.000	0.000	0.000	0.000
TgP	0.008	0.004	0.005	0.004	0.002	0.000	0.008	0.000
LWP	0.090	0.014	0.003	0.009	0.040	0.007	0.001	0.007

Table 6. The *p*-value associated with performance hypothesis testing with the Sharpe ratio from spectral cut-off regularized strategy using a three-factor model as a function of the number of assets in the economy with the sample size $n = 1000$, over 1000 replications.

Strategies	Number of Risky Assets							
	150	**250**	**400**	**550**	**700**	**850**	**950**	**999**
SbDP	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
XoNP	0.004	0.003	0.006	0.005	0.000	0.000	0.000	0.000
GMVP	0.020	0.003	0.005	0.001	0.000	0.000	0.000	0.000
MDP	0.003	0.002	0.003	0.002	0.001	0.000	0.000	0.000
TgP	0.003	0.002	0.004	0.002	0.001	0.000	0.001	0.000
LWP	0.104	0.043	0.002	0.008	0.032	0.004	0.002	0.006

We also compute in Table 7 the average monthly diversification ratio associated with the selected portfolio. We obtain similar results to what has been obtained in Table 2. The regularized portfolio performs well in terms of maximizing the diversification ratio and dominated most of the existing methods in the large financial market. The diversification ratio that we obtain with our method is very close to the true one. This implies that in addition to the asymptotic results obtained in the Section 4, the regularized portfolio has very good finite sample properties. This result shows that we do not need $N/\sqrt{T}$ to be close to zero to improve the finite sample performance of the selected portfolio.

Table 7. The average monthly Actual monthly diversification ratio from optimal strategies using a three-factor model as a function of the number of assets in the economy with the sample size $n = 120$, over 1000 replications. True DR is the true diversification ratio.

Strategies	Number of Risky Assets						
	10	**20**	**40**	**60**	**80**	**90**	**100**
SbDP	2.315	2.307	2.304	2.08	1.308	1.128	1.098
XoNP	3.103	3.140	3.180	3.184	3.325	3.288	3.154
GMVP	3.242	3.241	3.150	3.185	3.147	3.155	3.093
MDP	3.252	3.320	3.240	3.290	3.320	3.265	3.254
TgP	3.240	3.170	3.105	3.050	3.132	3.149	3.080
LWP	3.345	3.360	3.320	3.380	3.398	3.403	3.420
RdgDP	3.325	3.428	3.480	3.590	3.598	3.602	3.640
SCDP	3.347	3.435	3.446	3.570	3.589	3.615	3.625
LFDP	3.289	3.405	3.470	3.548	3.604	3.509	3.638
True DR	3.45	3.56	3.57	3.68	3.8	3.7	3.9

5.3. Empirical Study

In this empirical section, our objective is to use the real data (unlike in the simulation part) to estimate the unknown parameters of the optimal portfolio and then to evaluate the

performance of each estimation procedure based on the same statistics as in the simulation section. Note that our purpose in this paper lies not in forecasting but proposing a consistent way that allows us to correctly estimate the portfolio in Equation (4) in large dimensional setting.

We apply our method to several sets of portfolios from Kenneth R. French's website. In particular, we apply our procedure to the following portfolios: the 30-industry portfolios and the 100 portfolios formed on size and book-to-market. We allow investors to rebalance their portfolios every month. This implies that the optimal portfolio is constructed at the end of each month for a given estimation window M by maximizing the diversification ratio. The investor holds this portfolio for one month, realizes gains and losses, updates information, and then recomputes optimal portfolio weights for the next period using the same estimation window. This procedure is repeated each month, generating a time series of out-of-sample returns. This time series can then be used to analyze the out-of-sample performance of each strategy based on several statistics such as the out-of-sample Sharpe ratio. For this purpose, we use data from July 1980 to June 2018.

Table 8 contains some results of the out-of-sample analysis in terms of the Sharpe ratio for two different data sets: the FF30 and the FF100. The empirical results in this table confirm what we have obtained in the simulation part. According to this result, by stabilizing the inverse of the covariance matrix in the maximum diversification portfolio, we considerably improve the performance of the selected strategy in terms of maximizing the Sharpe ratio. Moreover, our regularized strategies outperform the most diversified strategy, the target portfolio, The LW portfolio, and the global minimum variance portfolio for each data set. The most diversified strategy outperforms the global minimum variance portfolio but is dominated by the Equal-Weight portfolio for each data set. These results of the most diversified portfolio can essentially be explained by the fact that by imposing a non-negative constraint in the investment process, one cannot fully diversify the optimal portfolio. The LWP outperforms the other strategies, in particular, this method dominates the most diversified strategy of Choueifaty et al. (2013). The return of the regularized portfolio is less volatile than what we obtain with the most diversified portfolio, the target one, and the LW strategy.

Table 8. Out-of-sample performance in terms of the Sharpe ratio applied on the 30 industry portfolios (FF30) and the 100 portfolios formed on size and book-to-market (FF100) with a rolling window of 120.

Strategies		**XoNP**	**GMVP**	**MDP**	**TGP**	**RdgP**	**LFP**	**SCP**	**LWP**
FF30	ER	0.0110	0.01134	0.0121	0.017	0.0149	0.014	0.014	0.014
	V	0.0540	0.0630	0.058	0.076	0.063	0.057	0.061	0.067
	SR	0.204	0.180	0.209	0.224	0.237	0.246	0.2295	0.209
FF100	ER	0.0103	0.0127	0.015	0.0173	0.0200	0.0201	0.0203	0.019
	V	0.0485	0.075	0.088	0.091	0.0772	0.0770	0.078	0.082
	SR	0.212	0.1693	0.1705	0.1901	0.2590	0.2610	0.2602	0.2317

We are also interested in how our procedure can perform in terms of minimizing the rebalancing cost at a given period. The rebalancing cost at the time t can be naturally measured by

$$Cost_t = \sum_{j=1}^{N} \left| \omega_{t,j} - \omega_{t-1,j} \right|.$$

This measure of the trading cost is, in fact, the turnover. The transaction cost can be measured using the turnover in the sense that these costs are positively related to the

turnover. Therefore, in the rest of the paper the turnover will be called transaction costs. The average trading cost over the investment horizon is given by

$$TradingCost = \frac{1}{Q}\sum_{t=1}^{Q} Cost_t$$

where Q is the number of rebalancing periods. This quantity can be interpreted as the average percentage of wealth traded at each period. The average monthly rebalancing costs are given in Table 9. These results show that by stabilizing the inverse of the covariance matrix by regularization, we help investors to select strategies that significantly reduce the rebalancing cost. The regularized portfolio outperforms the other strategies in terms of minimizing the trading costs faced by investors in their investment process.

Table 9. Out-of-sample performance in terms of rebalancing cost (turnover) applied on the 30 industry portfolios (FF30) and the 100 portfolios formed on size and book-to-market (FF100) for two different rolling windows.

P	EW	Strategies							
		SbDP	**GMVP**	**MDP**	**TgP**	**LWP**	**RdgDP**	**SCDP**	**LFDP**
FF30	60	6.890	4.329	2.809	4.209	1.0328	0.9952	0.989	0.9872
	120	5.605	3.901	2.087	3.290	0.9892	0.7140	0.7203	0.6450
FF100	120	9.789	6.2390	5.978	6.309	1.7808	1.3267	1.3890	1.2078
	240	7.089	4.297	3.879	4.2870	1.3065	1.0349	1.0398	1.096

The evolution of the share of the selected assets in the optimal portfolio in Figure 1 shows that by regularizing the covariance matrix, we considerably reduce extreme positions in the selected strategy. Therefore, we significantly reduce the transaction costs faced by investors when they decide to take positions in the financial market. Moreover, the return on the selected portfolio becomes less volatile in such a situation.

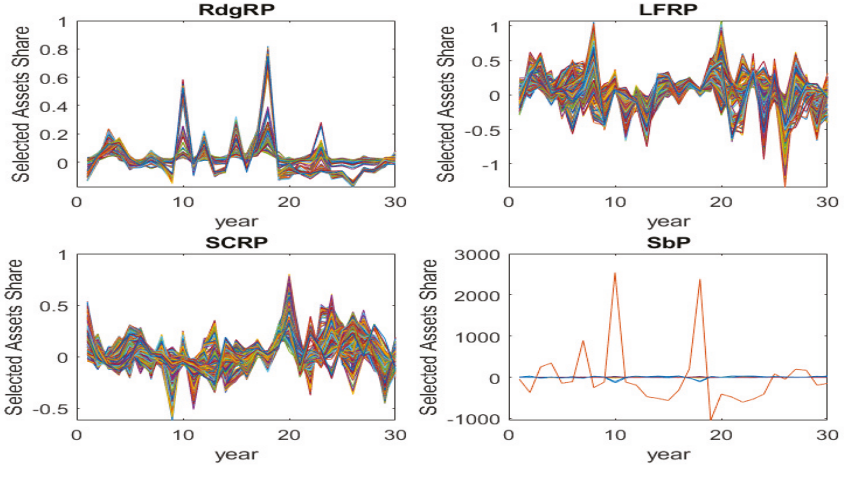


Figure 1. The evolution of the selected assets in the optimal portfolio. We obtain this figure using the 30 industry portfolios with an estimation window of $n = 120$.

Tables 10 and 11 contain the Fama–French monthly regression coefficients for the 100 portfolios formed on size and book-to-market and the 30-industry portfolios, respectively. Monthly data are used from July 1990 to June 2018. According to the result in

Table 10, only the return on the Equal-Weight portfolio can be explained by the Fama–French three-factor model for the 100 portfolios formed on size and book-to-market. The return obtained with the regularized portfolios and the most diversified portfolio can be explained only with the return on the market portfolio (a one-factor model) through a positive relation. However, the return of the most diversified portfolio and the global minimum variance portfolio can be explained with a two factors model when using the 30-industry portfolios. The return of the other strategies such as the regularized portfolios, the Equal-Weight portfolio, and the target portfolio can be explained by the Fama–French three-factor model.

Table 10. Fama–French Monthly Regression Coefficients for the 100 portfolios formed on size and book-to-market from July 1990 to June 2018.

Strategies	Market	HML	SMB	Intercept
Rdg-regularized Portfolio	0.9168	0.079	−0.139	0.0075
	(0.000)	(0.531)	(0.302)	(0.057)
LF- regularized Portfolio	0.823	0.174	−0.1651	0.0125
	(0.000)	(0.153)	(0.204)	(0.001)
SC-regularized Portfolio	1.02	−0.127	−0.133	0.0077
	(0.000)	(0.177)	(0.189)	(0.010)
Most-Diversified Portfolio	0.72	0.13	0.098	0.007
	(0.000)	(0.344)	(0.506)	(0.002)
Equal-Weight-Portfolio	1.002	0.5104	0.33	0.0001
	(0.000)	(0.000)	(0.000)	(0.815)
Global-Minimum-Variance Portfolio	0.416	−0.125	0.155	0.0094
	(0.000)	(0.319)	(0.247)	(0.000)
Target-Portfolio	0.43	0.144	0,207	0.010
	(0.000)	(0.367)	(0.226)	(0.000)
LW-Portfolio	0.802	0.074	0.207	0.0082
	(0.000)	(0.247)	(0.226)	(0.067)

Table 11. Fama–French Monthly Regression Coefficients for the 30-industry portfolios from July 1990 to June 2018.

Strategies	Market	HML	SMB	Intercept
Rdg-regularized Portfolio	1.03	0.24	0.36	0.0007
	(0.000)	(0.003)	(0.000)	(0.767)
LF- regularized Portfolio	0.93	0.22	0.25	0.0046
	(0.000)	(0.003)	(0.001)	(0.042)
SC-regularized Portfolio	0.86	0.27	0.21	0.0054
	(0.000)	(0.000)	(0.031)	(0.053)
Most-Diversified Portfolio	0.46	−0.285	0.070	0.002
	(0.000)	(0.000)	(0.391)	(0.001)
Equal-Weight-Portfolio	0.983	0.061	0.265	0.0013
	(0.000)	(0.006)	(0.000)	(0.050)
Global-Minimum-Variance Portfolio	0.46	−0.146	0.077	0.0021
	(0.000)	(0.008)	(0.188)	(0.017)
Target-Portfolio	0.54	−0.44	−0.21	0.013
	(0.000)	(0.000)	(0.019)	(0.000)
LW-Portfolio	0.982	0.272	0.4112	0.0006
	(0.000)	(0.0098)	(0.0301)	(0.429)

As the portfolio optimization is generally based on individual stocks instead of aggregate portfolios as the Fama–French portfolio, we apply also our method to a subset of the S&P500 index constituents to see how our method performs in such universe. We use for this purpose monthly data from March 1986 to December 2019. At the beginning of this empirical analysis, we randomly form pools of 100 or 150 stocks from the S&P500 index constituents for which there are complete return data for the prior 120 or 240 months. The

optimal portfolio will then be constructed using the same procedure as before. We then compute the out-of-sample performance in terms of the Sharpe ratio and the turnover. The results of this empirical analysis are given in Tables 12 and 13. We obtain similar results as in the case of the Fama–French portfolios proving that our method also performs well when the optimal portfolio is formed with individual stocks from S&P500.

Table 12. Out-of-sample performance in terms of Sharpe ratio applied on two subsets of S&P500 constituents for two different rolling windows.

P	EW	Strategies							
		SbDP	**GMVP**	**MDP**	**TgP**	**LWP**	**RdgDP**	**SCDP**	**LFDP**
100 A	120	0.0850	0.1506	0.2458	0.1983	0.3702	0.4382	0.4380	0.4397
	240	0.0982	0.1604	0.260	0.2028	0.3809	0.4565	0.4567	0.4578
150 A	180	0.0750	0.1204	0.309	0.1407	0.4108	0.5353	0.5320	0.5462
	240	0.0895	0.1750	0.320	0.1890	0.4208	0.5603	0.5609	0.5579

Table 13. Out-of-sample performance in terms of rebalancing cost (turnover) applied on two subsets of S&P500 constituents for two different rolling windows.

Assets	EW	Strategies							
		SbDP	**GMVP**	**MDP**	**TgP**	**LWP**	**RdgDP**	**SCDP**	**LFDP**
100 Assets	120	9.450	6.786	4.675	6.679	3.348	2.1067	2.0801	2.0682
	240	6.978	5.308	3.892	5.234	3.078	1.491	1.608	1.569
150 Assets	180	10.489	7.345	6.782	7.328	3.897	2.678	2.780	2.8960
	240	8.0789	5.542	4.032	5.438	3.057	2.104	2.0978	2.0956

6. Conclusions

This paper addresses the estimation issue that exists in the maximum diversification portfolio framework in the large financial market. We propose to stabilize the inverse of the covariance matrix in the diversified portfolio using regularization techniques from inverse problem literature. These regularization techniques, namely, the ridge, the spectral cut-off, and Landweber–Fridman, involve a regularization parameter or penalty term whose optimal value is selected to minimize the expected distance between the inverse of the estimated covariance matrix and the inverse of the true covariance matrix. We show, under appropriate regularity conditions, that the selected strategy by regularization is asymptotically efficient with respect to the diversification ratio for a wise choice of the tuning parameter. Meaning that, even if the diversified portfolio is unknown, there exists a feasible portfolio obtained by regularization capable of reaching a similar level of performance in terms of the diversification ratio.

To evaluate the performance of our procedures, we implement a simulation exercise based on a three-factor model calibrated on real data from the US financial market. We obtain by simulation that our procedure significantly improves the performance of the selected strategy with respect to the Sharpe ratio. Moreover, the regularized rules are compared to several strategies such as the most diversified portfolio, the target portfolio, the global minimum variance portfolio, and the naive 1/N strategy in terms of in-sample and out-of-sample Sharpe ratio, and it is shown that our method yields significant Sharpe ratio improvements. To confirm our simulations, we do an empirical analysis using Kenneth R. French's 30-industry portfolios and 100 portfolios formed on size and book-to-market. According to this empirical result, by stabilizing the inverse of the covariance matrix in the maximum diversification portfolio, we considerably improve the performance of the selected strategy in terms of maximizing the Sharpe ratio.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

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

Appendix A. Proof of Proposition 1

By definition we have that

$$DR(\hat{\omega}_\alpha) = \frac{\hat{\omega}_\alpha' \sigma}{\sqrt{\hat{\omega}_\alpha' \Sigma \hat{\omega}_\alpha}}.$$

Let us first look at $\hat{\omega}_\alpha' \Sigma \hat{\omega}_\alpha$

$$\begin{aligned} \hat{\omega}_\alpha' \Sigma \hat{\omega}_\alpha &= [(\hat{\omega}_\alpha - \omega) + \omega]' \Sigma [(\hat{\omega}_\alpha - \omega) + \omega] \\ &= \omega' \Sigma \omega + \underbrace{(\hat{\omega}_\alpha - \omega)' \Sigma (\hat{\omega}_\alpha - \omega)}_{(a)} + 2 \underbrace{(\hat{\omega}_\alpha - \omega)' \Sigma \omega}_{(b)}. \end{aligned}$$

Now we are going to look at the properties of (a) and (b). We know that

$$\hat{\omega}_\alpha = \left(\underbrace{\hat{\sigma}' \hat{\Sigma}^\alpha \hat{\sigma}}_{(c)} \right)^{-1} \underbrace{\hat{\Sigma}^\alpha \hat{\sigma}}_{(d)}.$$

$$\begin{aligned} (c) &= \sigma' \hat{\Sigma}^\alpha \sigma + (\hat{\sigma} - \sigma)' \hat{\Sigma}^\alpha (\hat{\sigma} - \sigma) + 2(\hat{\sigma} - \sigma)' \hat{\Sigma}^\alpha \sigma \\ \hat{\Sigma}^\alpha &= \left(\hat{\Sigma}^\alpha - \Sigma^\alpha + \Sigma^\alpha \right). \end{aligned}$$

$$\begin{aligned} \left\| (\hat{\sigma} - \sigma)' \hat{\Sigma}^\alpha (\hat{\sigma} - \sigma) \right\| &= \left\| \frac{(\hat{\sigma} - \sigma)'}{\sqrt{N}} \left(\frac{\hat{\Sigma}}{N} \right)^\alpha \frac{(\hat{\sigma} - \sigma)}{\sqrt{N}} \right\| \\ &= O_p \left(\frac{\|\hat{\sigma} - \sigma\|^2}{N\alpha} \right) \\ &= O_p \left(\frac{\left\| \frac{\hat{\sigma} - \sigma}{\sqrt{N}} \right\|^2}{\alpha} \right). \end{aligned}$$

By Assumption A $\left\| \frac{\sigma}{\sqrt{N}} \right\| = O(1)$. Therefore, we obtain that

$$\begin{aligned} \left\| (\hat{\sigma} - \sigma)' \hat{\Sigma}^\alpha \sigma \right\| &= \left\| \frac{(\hat{\sigma} - \sigma)'}{\sqrt{N}} \left(\frac{\hat{\Sigma}}{N} \right)^\alpha \frac{\sigma}{\sqrt{N}} \right\| \\ &= O_p \left(\frac{\|\hat{\sigma} - \sigma\|}{\sqrt{N}\alpha} \right) \\ &= O_p \left(\frac{\left\| \frac{\hat{\sigma} - \sigma}{\sqrt{N}} \right\|}{\alpha} \right). \end{aligned}$$

Using those information combine with the fact that $\hat{\Sigma}^{\alpha} = \hat{\Sigma}^{\alpha} - \Sigma^{\alpha} + \Sigma^{\alpha}$, we have that

$$(c) = \sigma^{'}\Sigma^{\alpha}\sigma + \sigma^{'}\left(\hat{\Sigma}^{\alpha} - \Sigma^{\alpha}\right)\sigma + O_p\left(\frac{\left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\| + \left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\|^2}{\alpha}\right).$$

$$\begin{aligned}\left\|\sigma^{'}\left(\hat{\Sigma}^{\alpha} - \Sigma^{\alpha}\right)\sigma\right\| &= \left\|\frac{\sigma^{'}}{\sqrt{N}}\left[\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right]\frac{\sigma}{\sqrt{N}}\right\| \\ &\leq \left\|\frac{\sigma}{\sqrt{N}}\right\|^2\left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\| \\ &= O_p\left(\left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\|\right).\end{aligned}$$

$$\left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\| \leq \left\|\left(\frac{\Sigma}{N}\right)^{\alpha}\right\|\left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\right\|\left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\|.$$

Hence,

$$\left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\| = O_p\left(\frac{\left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\|}{\alpha}\right)$$

which implies that

$$(c) = \sigma^{'}\Sigma^{\alpha}\sigma + O_p\left(\frac{\left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\| + \left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\| + \left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\|^2}{\alpha}\right).$$

As $T \to \infty$ we have that $\alpha \to 0 \Rightarrow$

$$(c) = \sigma^{'}\Sigma^{-1}\sigma + O_p\left(\frac{\left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\| + \left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\| + \left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\|^2}{\alpha}\right).$$

Using Assumption A combined with Theorem 4 of Carrasco and Florens (2000), we have that

$$\left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\| = O_p\left(\frac{1}{\sqrt{T}}\right).$$

Moreover, as $\left\|\frac{\hat{\sigma}-\sigma}{\sqrt{N}}\right\|^2 = O_p\left(\frac{1}{T}\right)$ by Assumption A, we have that

$$(c) = \sigma^{'}\Sigma^{-1}\sigma + O_p\left(\frac{1}{\alpha\sqrt{T}}\right).$$

$$\begin{aligned}(d) &= \hat{\Sigma}^{\alpha}\hat{\sigma} \\ &= \hat{\Sigma}^{\alpha}\sigma + \hat{\Sigma}^{\alpha}(\hat{\sigma} - \sigma) \\ &= \Sigma^{\alpha}\sigma + \left(\hat{\Sigma}^{\alpha} - \Sigma^{\alpha}\right)\sigma + \hat{\Sigma}^{\alpha}(\hat{\sigma} - \sigma).\end{aligned}$$

As $\alpha \to 0$ as $T \to \infty$, we have that

$$(d) = \Sigma^{-1}\sigma + \left(\hat{\Sigma}^{\alpha} - \Sigma\right)\sigma + \hat{\Sigma}^{\alpha}(\hat{\sigma} - \sigma).$$

We know that

$$\begin{aligned} \left\|\hat{\Sigma}^{\alpha}(\hat{\sigma} - \sigma)\right\| &= \left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\frac{(\hat{\sigma} - \sigma)}{N}\right\| \\ &\leq \left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\right\|\left\|\frac{(\hat{\sigma} - \sigma)}{N}\right\| \\ &= O_p\left(\frac{1}{\alpha\sqrt{TN}}\right). \end{aligned}$$

Using the fact that

$$\begin{aligned} \left\|(\hat{\Sigma}^{\alpha} - \Sigma)\sigma\right\| &= \left\|\left\{\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\}\frac{\sigma}{N}\right\| \\ &\leq \left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha} - \left(\frac{\Sigma}{N}\right)^{\alpha}\right\|\left\|\frac{\sigma}{N}\right\| \\ &= O_p\left(\frac{\left\|\frac{\hat{\Sigma}}{N} - \frac{\Sigma}{N}\right\|}{\alpha\sqrt{N}}\right) \\ &= O_p\left(\frac{1}{\alpha\sqrt{TN}}\right) \end{aligned}$$

we obtain that

$$(d) = \Sigma^{-1}\sigma + O_p\left(\frac{1}{\alpha\sqrt{TN}}\right).$$

Under the assumption that $\frac{1}{\alpha\sqrt{T}} \to 0$, we have that

$$\hat{\omega}_{\alpha} = \omega + o_p(1). \tag{A1}$$

By Assumption A we have that $\|\Sigma\| = O(N)$. Therefore, using (A1), we obtain that

$$\hat{\omega}_{\alpha}^{'}\Sigma\hat{\omega}_{\alpha} = \omega^{'}\Sigma\omega + o_p(1) \tag{A2}$$

if $\frac{N}{\alpha\sqrt{T}} \to 0$. Therefore,

$$DR(\hat{\omega}_{\alpha}) \to_p DR(\omega_t).$$

Appendix B. Proof of Proposition 2

$$(A) = \mu^{'}\left[\left(\hat{\Sigma}^{\alpha} - \Sigma^{-1}\right)^{'}\Sigma\left(\hat{\Sigma}^{\alpha} - \Sigma^{-1}\right)\right]\mu$$

We also know that $\mu = \hat{\mu} + (\mu - \hat{\mu})$, so

$$\begin{aligned}(A) &= \mu'\left[\left(\hat{\Sigma}^{\alpha} - \Sigma^{-1}\right)'\Sigma\left(\hat{\Sigma}^{\alpha} - \Sigma^{-1}\right)\right]\mu \\ &= \left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu}) + \left(\hat{\Sigma}^{\alpha}\hat{\mu} - \Sigma^{-1}\mu\right)\right]'\Sigma\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu}) + \left(\hat{\Sigma}^{\alpha}\hat{\mu} - \Sigma^{-1}\mu\right)\right] \\ &= \left(\hat{\Sigma}^{\alpha}\hat{\mu} - \Sigma^{-1}\mu\right)'\Sigma\left(\hat{\Sigma}^{\alpha}\hat{\mu} - \Sigma^{-1}\mu\right) + \left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\Sigma\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right] \\ &+ 2\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\Sigma\left(\hat{\Sigma}^{\alpha}\hat{\mu} - \Sigma^{-1}\mu\right)\end{aligned}$$

Let denote by $x = \Sigma^{-1}\mu$ and $\hat{x} = \hat{\Sigma}^{\alpha}\hat{\mu}$; therefore,

$$(A) = (\hat{x} - x)'\Sigma(\hat{x} - x) + \left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\Sigma\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right] + 2\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\Sigma(\hat{x} - x)$$

As $\|\mu - \hat{\mu}\|^2 = O_p\left(\frac{N}{T}\right)$, $\left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\right\|^2 = O_p\left(\frac{1}{\alpha^2}\right)$, we have that

$$\begin{aligned}\left\|\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\right\| &= \left\|\left[\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\frac{(\mu - \hat{\mu})}{N}\right]'\right\| \\ &\leq \left\|\left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\right\|\left\|\frac{(\mu - \hat{\mu})}{N}\right\| \\ &= O_p\left(\frac{1}{\alpha\sqrt{TN}}\right)\end{aligned}$$

$$\begin{aligned}\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\Sigma\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right] &= O_p\left(\left\|\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\right\|^2\|\Sigma\|\right) \\ &= O_p\left(\frac{\|\Sigma\|}{\alpha^2 TN}\right)\end{aligned}$$

Using the fact that $\|\Sigma\| = O(N)$ by Assumption A, we obtain that

$$\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\Sigma\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right] = O_p\left(\frac{N}{\alpha^2 TN}\right) = O_p\left(\frac{1}{\alpha^2 T}\right)$$

$$\begin{aligned}\hat{x} - x &= \hat{\Sigma}^{\alpha}\hat{\mu} - \Sigma^{-1}\mu \\ \hat{\mu} &= (\hat{\mu} - \mu) + \mu \Rightarrow \\ \hat{x} - x &= \hat{\Sigma}^{\alpha}(\hat{\mu} - \mu) + \left(\hat{\Sigma}^{\alpha} - \Sigma^{-1}\right)\mu \Rightarrow\end{aligned}$$

$$\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\Sigma(\hat{x} - x) = \left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\Sigma\left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right] + \left[\hat{\Sigma}^{\alpha}(\mu - \hat{\mu})\right]'\Sigma\left(\hat{\Sigma}^{\alpha} - \Sigma^{-1}\right)\mu$$

$$\begin{aligned}\left(\hat{\Sigma}^{\alpha} - \Sigma^{-1}\right)\mu &= \left(\frac{\hat{\Sigma}}{N}\right)^{\alpha}\left[\frac{\Sigma}{N} - \frac{\hat{\Sigma}}{N}\right]\left(\frac{\Sigma}{N}\right)^{-1}\frac{\mu}{N} \\ &= O_p\left(\frac{1}{\alpha\sqrt{TN}}\right)\end{aligned}$$

which implies that

$$\begin{aligned}(A) &= (\hat{x}-x)'\Sigma(\hat{x}-x)+O_p\left(\frac{2}{\alpha^2 T}\right)\\ &= (\hat{x}-x)'\Sigma(\hat{x}-x)+O_p\left(\frac{1}{\alpha^2 T}\right)\end{aligned}$$

Therefore, we obtain that

$$\begin{aligned}& E\left\{\mu'\left[\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)'\Sigma\left(\hat{\Sigma}^{\alpha}-\Sigma^{-1}\right)\right]\mu\right\}\\ \sim\ & E\left\{(\hat{x}-x)'\Sigma(\hat{x}-x)\right\}\end{aligned}$$

if $\frac{1}{\alpha^2 T}\to 0$.

References

Ao, Mengmeng, Li Yingying, and Xinghua Zheng. 2019. Approaching mean-variance efficiency for large portfolios. *The Review of Financial Studies* 32: 2890–919. [CrossRef]

Arnott, Robert D., Jason Hsu, and Philip Moore. 2005. Fundamental indexation. *Financial Analysts Journal* 61: 83–99. [CrossRef]

Carrasco, Marine. 2012. A regularization approach to the many instruments problem. *Journal of Econometrics* 170: 383–98. [CrossRef]

Carrasco, Marine, and Jean-Pierre Florens. 2000. Generalization of gmm to a continuum of moment conditions. *Econometric Theory* 16: 797–834. [CrossRef]

Carrasco, Marine, Jean-Pierre Florens, and Eric Renault. 2007. Linear inverse problems in structural econometrics estimation based on spectral decomposition and regularization. *Handbook of Econometrics* 6: 5633–751.

Carrasco, Marine, N'golo Koné, and Nérée Noumon. 2019. Optimal portfolio selection using regularization. Available online: https://www.ngolokone.com/research (accessed on 15 January 2020).

Carrasco, Marine, and Guy Tchuente. 2015. Regularized liml for many instruments. *Journal of Econometrics* 186: 427–42. [CrossRef]

Chen, Jiaqin, and Ming Yuan. 2016. Efficient portfolio selection in a large market. *Journal of Financial Econometrics* 14: 496–524. [CrossRef]

Choueifaty, Yves. 2011. Methods and Systems for Providing an Anti-Benchmark Portfolio. U.S. Patent 7,958,038, June 7.

Choueifaty, Yves, and Yves Coignard. 2008. Toward maximum diversification. *The Journal of Portfolio Management* 35: 40–51. [CrossRef]

Choueifaty, Yves, Tristan Froidure, and Julien Reynier. 2013. Properties of the most diversified portfolio. *Journal of Investment Strategies* 2: 49–70. [CrossRef]

Clarke, Roger G., Harindra De Silva, and Steven Thorley. 2006. Minimum-variance portfolios in the us equity market. *The Journal of Portfolio Management* 33: 10–24. [CrossRef]

Frost, Peter A., and James E. Savarino. 1986. An empirical bayes approach to efficient portfolio selection. *Journal of Financial and Quantitative Analysis* 21: 293–305. [CrossRef]

Haugen, Robert A., and Nardin L. Baker. 1991. The efficient market inefficiency of capitalization–weighted stock portfolios. *The Journal of Portfolio Management* 17: 35–40. [CrossRef]

Jobson, J. Dave, and Bob M. Korkie. 1981. Performance hypothesis testing with the sharpe and treynor measures. *Journal of Finance*, 889–908. [CrossRef]

Kempf, Alexander, and Christoph Memmel. 2006. Estimating the global minimum variance portfolio. *Schmalenbach Business Review* 58: 332–48. [CrossRef]

Kress, Rainer. 1999. *Linear Integral Equations*. Berlin/Heidelberg: Springer.

Ledoit, Olivier, and Michael Wolf. 2003. Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of Empirical Finance* 10: 603–21. [CrossRef]

Lohre, Harald, Heiko Opfer, and Gabor Orszag. 2014. Diversifying risk parity. *Journal of Risk* 16: 53–79. [CrossRef]

Maguire, Phil, Philippe Moser, Kieran O'Reilly, Conor McMenamin, Robert Kelly, and Rebecca Maguire. 2014. Maximizing positive porfolio diversification. Paper presented at the 2014 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr), London, UK, March 27–28; pp. 174–81.

Maillard, Sébastien, Thierry Roncalli, and Jérôme Teïletche. 2010. The properties of equally weighted risk contribution portfolios. *The Journal of Portfolio Management* 36: 60–70. [CrossRef]

Markowitz, Harry. 1952. Portfolio selection. *The Journal of Finance* 7: 77–91.

Memmel, Christoph. 2003. Performance Hypothesis Testing with the Sharpe Ratio. *Finance Letters* 1. Available online: https://ssrn.com/abstract=412588 (accessed on 15 August 2020).

Sharpe, William F. 1964. Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance* 19: 425–42.

Theron, Ludan, and Gary Van Vuuren. 2018. The maximum diversification investment strategy: A portfolio performance comparison. *Cogent Economics & Finance* 6: 1427533.

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
Tel. +41 61 683 77 34
Fax +41 61 302 89 18
www.mdpi.com

Econometrics Editorial Office
E-mail: econometrics@mdpi.com
www.mdpi.com/journal/econometrics

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland

Tel: +41 61 683 77 34
Fax: +41 61 302 89 18

www.mdpi.com

ISBN 978-3-0365-0853-5