risks

Special Issue Reprint

Data Analysis for Risk Management - Economics, Finance and Business

Edited by
Krzysztof Jajuga and Józef Dziechciarz

mdpi.com/journal/risks

MDPI

Data Analysis for Risk Management—Economics, Finance and Business

Data Analysis for Risk Management—Economics, Finance and Business

Editors

Krzysztof Jajuga
Józef Dziechciarz

MDPI

Basel • Beijing • Wuhan • Barcelona • Belgrade • Novi Sad • Cluj • Manchester

Editors

Krzysztof Jajuga
Department of Financial Investments and Risk Management
Wroclaw University of Economics and Business
Wroclaw
Poland

Józef Dziechciarz
Department of Econometrics and Operations Research
Wroclaw University of Economics and Business
Wroclaw
Poland

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 *Risks* (ISSN 2227-9091) (available at: https://www.mdpi.com/journal/risks/special_issues/Data_Analysis_for_Risk_Management).

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

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

ISBN 978-3-7258-1416-9 (Hbk)
ISBN 978-3-7258-1415-2 (PDF)
doi.org/10.3390/books978-3-7258-1415-2

Contents

About the Editors . . . . **vii**

Preface . . . . **ix**

Krzysztof Jajuga
Data Analysis for Risk Management—Economics, Finance and Business: New Developments and Challenges
Reprinted from: *Risks* **2023**, *11*, 70, doi:10.3390/risks11040070 . . . . **1**

Joanna Górka and Katarzyna Kuziak
Volatility Modeling and Dependence Structure of ESG and Conventional Investments
Reprinted from: *Risks* **2022**, *10*, 20, doi:10.3390/risks10010020 . . . . **6**

Victor Shevchuk and Roman Kopych
Exchange Rate Volatility, Currency Misalignment, and Risk of Recession in the Central and Eastern European Countries
Reprinted from: *Risks* **2021**, *9*, 82, doi:10.3390/risks9050082 . . . . **31**

Ewa Dziwok and Marta A. Karaś
Systemic Illiquidity Noise-Based Measure—A Solution for Systemic Liquidity Monitoring in Frontier and Emerging Markets
Reprinted from: *Risks* **2021**, *9*, 124, doi:10.3390/risks9070124 . . . . **50**

Orlando Rivera-Escobar, John Willmer Escobar and Diego Fernando Manotas
Measurement of Systemic Risk in the Colombian Banking Sector
Reprinted from: *Risks* **2022**, *10*, 22, doi:10.3390/risks10010022 . . . . **79**

Aneta Ptak-Chmielewska and Paweł Kopciuszewski
New Definition of Default—Recalibration of Credit Risk Models Using Bayesian Approach
Reprinted from: *Risks* **2022**, *10*, 16, doi:10.3390/risks10010016 . . . . **106**

Tomasz Berent and Radosław Rejman
Bankruptcy Prediction with a Doubly Stochastic Poisson Forward Intensity Model and Low-Quality Data
Reprinted from: *Risks* **2021**, *9*, 217, doi:10.3390/risks9120217 . . . . **122**

Aldona Fraczkiewicz-Wronka, Tomasz Ingram, Karolina Szymaniec-Mlicka and Piotr Tworek
Risk Management and Financial Stability in the Polish Public Hospitals: The Moderating Effect of the Stakeholders' Engagement in the Decision-Making
Reprinted from: *Risks* **2021**, *9*, 87, doi:10.3390/risks9050087 . . . . **146**

Yaser Ahmad Arabyat, Ahmad Ali AlZubi, Dyala M. Aldebei and Samerra'a Ziad Al-oqaily
An Efficient Method for Pricing Analysis Based on Neural Networks
Reprinted from: *Risks* **2022**, *10*, 151, doi:10.3390/risks10080151 . . . . **169**

Gero Szepannek
An Overview on the Landscape of R Packages for Open Source Scorecard Modelling
Reprinted from: *Risks* **2022**, *10*, 67, doi:10.3390/risks10030067 . . . . **183**

Adam Śliwiński, Joanna Dropia and Norbert Duczkowski
Risk Factors Affecting Bancassurance Development in Poland
Reprinted from: *Risks* **2021**, *9*, 130, doi:10.3390/risks9070130 . . . . **216**

Błażej Kochański
Which Curve Fits Best: Fitting ROC Curve Models to Empirical Credit-Scoring Data
Reprinted from: *Risks* **2022**, *10*, 184, doi:10.3390/risks10100184 . . . . **235**

About the Editors

Krzysztof Jajuga

Krzysztof Jajuga is president of CFA Society Poland and a full professor of finance at Wroclaw University of Economics and Business, Poland. He holds master's, doctoral, and habilitation degree from Wroclaw University of Economics and Business, was awarded the title of titular professor by President of Poland, has an honorary doctorate from Cracow University of Economics and University WSB, and has an honorary professorship from Warsaw University of Technology, University of Warmia and Mazury in Olsztyn. He is president-elect of International Federation of Classification Societies. He carries out research and has published numerous papers and monographs in the area of financial markets, risk management, household finance, multivariate statistics, and quantitative methods in economic sciences. He is editor in chief of *Argumenta Oeconomica* (JCR journal). He collaborates with many financial institutions and enterprises.

Józef Dziechciarz

Józef Dziechciarz is a professor at Wroclaw University of Economics and Business and is a member of the Committee of Statistics and Econometrics of the Polish Academy of Sciences. His research is in the area of econometric models.

Preface

This volume of “*Risks*” focusses on the methods of data analysis for risk management in economics, finance, and business. The main motivation for presenting this research is to advance the dynamic development of data analysis methods, including classical statistical methods, and machine learning methods that have emerged from statistics and are being applied using high-speed computers, considering the availability of big data.

This volume contains 12 papers, written by authors from several countries. The first paper discussed the main driving forces of the evolution of risk management in economics, finance, and business, providing a foundation for the subsequent research papers. It addresses the following topics:

- The development of theoretical tools of risk analysis;
- The development of instruments of risk management;
- The development of regulations in the area of financial risk.

The other papers can be classified into four groups.

The first group contains empirical papers in which volatility measures are applied to risk analysis:

- Exchange rate volatility and its impact on the business cycle in several countries;
- Volatility of returns of selected ESG indices and conventional indices;
- Identification of risk factors affecting bancassurance development in Poland.

The second group contains papers discussing the important topic of systemic risk:

- Measure of systemic illiquidity for frontier and emerging markets;
- Assessment of systemic risk in the Colombian banking system.

The third group contains papers concentrated on credit risk:

- Modelling default probability using a doubly stochastic Poisson process;
- Bayesian approach for PD recalibration based on both simulated and empirical data;
- Theoretical models for an ROC curve;
- Methods for analysing the financial stability of hospitals.

The fourth group contains papers concerned with using advanced methods, strongly depending on computer technology:

- Using Artificial Neural Networks for stock price forecasting;
- Overview of R packages used in risk management.

The target groups of this volume are researchers, risk managers, other financial practitioners, doctoral and master students, and policy makers.

Krzysztof Jajuga and Józef Dziechciarz
Editors

risks

Editorial

Data Analysis for Risk Management—Economics, Finance and Business: New Developments and Challenges

Krzysztof Jajuga

Department of Financial Investments and Risk Management, Wroclaw University of Economics and Business, Komandorska 118/120, 53-345 Wroclaw, Poland; krzysztof.jajuga@ue.wroc.pl

1. Introduction

The development of the theory and practice of risk management is closely related to the emergence of different risks. Some risks have existed since very ancient times. This refers in particular to natural catastrophe risk resulting from, for example, floods, earthquakes or hurricanes. Eventually, risks related to business activities started to play a significant role in risk management.

The evolution of risk management in economics, finance and business can be presented with regard to several driving forces:

- The development of theoretical tools of risk analysis;
- The development of the instruments of risk management;
- The development of regulations in the area of financial risk.

Theoretical tools of financial risk analysis have been used since the beginning of the twentieth century, such as:

- Macaulay duration, used to measure interest rate risk;
- Markowitz portfolio theory and extension, proposed by James Tobin;
- The Capital Asset Pricing Model;
- The option-pricing model developed by Fischer Black, Myron Scholes and Robert Merton;
- The advanced model to estimate volatility developed by Robert Engle.

The instruments of financial risk management are mainly derivatives. Despite the fact that they have existed from at least the nineteenth century, the dynamic development of these instruments has primarily taken place in the last 50 years in the exchanges, such as Chicago Board of Options Exchange or Chicago Mercantile Exchange, but they have also significantly developed in the OTC markets. This refers in particular to equity derivatives, currency derivatives, interest rate derivatives and commodity derivatives in the form of options, futures, forwards and swaps.

Finally, in the last decade of the twentieth century, the other classes of derivatives were introduced: credit derivatives, catastrophe derivatives, weather derivatives and property derivatives. Another direction of the development of derivatives has been the introduction of exotic derivatives characterized by the enhanced flexibility of their parameters, such as: payoff profile, number of underlying instruments, contingency of the execution, etc.

Regulations in the area of financial risk have been introduced mainly for financial institutions, particularly banks. This has been developed by the Basel Committee for Banking Supervision under the well-known Basel Capital Acord, New Capital Accord and follow up documents. They can be considered as good practices, but in many countries, financial supervisory authorities regard these documents as "soft" regulations. The rules proposed by the Basel Committee initially covered credit risk, but then grew to include market risk, operational risk and liquidity risk. These extensions were introduced after risky events in the financial markets, e.g., the collapse of Barings Bank, the financial crisis

Citation: Jajuga, Krzysztof. 2023. Data Analysis for Risk Management—Economics, Finance and Business: New Developments and Challenges. *Risks* 11: 70. https://doi.org/10.3390/risks11040070

Received: 25 March 2023
Accepted: 28 March 2023
Published: 30 March 2023

of 2007–2008, etc. Very similar regulations, under the name of solvency, were introduced for insurance companies.

In the twenty-first century, much attention has been paid to macroprudential regulations being a response to the threat of systemic risk, as the collapse of large banks (or other financial institutions) could lead to a domino effect in the whole financial sector with an impact on the real economy.

Sometimes, regulations have emerged because they followed the good practices of financial risk management established in both financial institutions and non-financial corporations. These good-practice standards were also introduced in the public sector. Finally, efforts have been made to increase households' awareness of the need for financial risk management.

The risk management process contains several stages: identification, assessment, steering, monitoring and control. Successful risk steering (mitigation) depends on the assessment (measurement), which depends on the application of data analysis methods. The development of computer technology has enabled the effective use of advanced data analysis methods. This has been possible because of several driving forces:

- The increased speed of computers;
- The growth of the available data;
- The increased speed of data transfers across the world;
- The growth of social network connections.

The data analysis methods used today are often classified as so-called artificial intelligence methods, particularly machine learning methods. Most of these methods were developed many years ago (e.g., in the 1960s) in the area of statistical data analysis. Their effective use can be observed in recent years thanks to the increase of computer speed and the growth of available data.

This has provided great opportunities to analyze large sets of data (Big Data) but has also created some challenges, namely:

- The aggregation of classical data (numerical data and text data) with alternative data through video files, audio files, images, sensor data, etc.;
- The detection of fake data, mostly from social media;
- The suitable methodology for extreme risk, which is defined as resulting from an event that has a very low probability of occurrence and leads to very large losses;
- The transparency of machine learning algorithms, which should be understood by the end user;
- The customizability of machine learning algorithms, which should be fitted to the needs of the end user; in risk management, "one size fits all" does not apply;
- The clear assessment of model risk, which is defined as "a risk resulting from erratic model used in real world".

All of the mentioned challenges are related to data analysis in economics, finance and business, and are the main topic of the present Special Issue.

2. A Short Review of the Contributions in This Issue

The present Issue contains 11 papers written by authors affiliated with universities in several countries. All papers contain valuable contributions in the area of applications of data analysis methods for risk management in finance, business and economics.

These papers can be classified into several areas, namely:

- Credit risk;
- Systemic risk;
- Risk analysis in financial investments;
- Risk management at the macro level;
- Risk management at the micro level.

Credit risk is a basic type of risk which has a long history of involvement with scientific methods. This volume contains four papers discussing the different problems of credit risk analysis.

Two of these papers present methods of modelling credit risk.

In their paper, Berent and Rejman (2021) apply a doubly stochastic process (based on Duffie–Duan proposal) to determine the probability of default. This model can be classified as a so-called intensity model, an alternative to scoring models in predicting the probability of default. The authors use the data available for more than 15,000 non-financial companies from emerging markets for the period of 2007–2017, resulting in high out-of-sample accuracy ratios.

The paper by Ptak-Chmielewska and Kopciuszewski (2022) also contains insights related to the estimation of probability of default. The authors apply a Bayesian approach for the recalibration of probability of default through Long-Term Average, considering the number and timing of defaults using simulated and empirical data.

The other two papers in this group concern implementation issues.

The paper by Kochanski (2022) reviews the ROC curve models (binormal, bigamma, bibeta, bilogistic, power and bifractal curves) proposed in the literature, and fits them to actual credit-scoring ROC data (based on publicly available data) in order to determine which models can be used in credit risk management practice.

In his paper, Szepannek (2022) gives a thorough overview of a number of open-source software packages designed for credit scorecard modelling, which can be considered as a link between academic research and practitioner implementation.

Systemic risk has shown increasing popularity in academic research after the financial crisis at the end of the first decade of the twenty-first century. This volume contains two papers discussing the issues of systemic risk.

The paper by Dziwok and Karaś (2021) presents an alternative approach to measuring systemic illiquidity applicable to countries with frontier and emerging financial markets where other existing methods are not applicable. The proposed measure, called Systemic Illiquidity Noise, uses Nelson–Siegel–Svensson methodology. Then, it is applied to a set of 10 Central and Eastern European countries in the period of 2006–2020. The results show three periods of increased risk in the sample period: the global financial crisis, the European public debt crisis, and the COVID-19 pandemic, as well as three sets of countries with different systemic liquidity risk characteristics.

Rivera-Escobar et al. (2022), in their paper, use three well-known systemic risk measures to analyze systemic risk in the banking sector in Columbia during 2008–2017. The main finding is that the Colombian banking sector does not show a significant systemic risk.

Risk in financial investments has been studied for a very long time, and thus many methods to analyze this risk have been developed. This volume contains two papers in this area.

The paper by Górka and Kuziak (2022) concerns ESG (Environmental, Social, Governance) investments. This type of investment has gained a lot of attention in research over the past several years. The paper presents the comparison of the volatility of rates of return of selected ESG indices and the classical types of indices, as well as the dependence between them. The conditional volatility models from the GARCH family and tail-dependence coefficients from the copula-based approach are applied. The analysis period covered 2007 until 2019. The results of the research confirm the higher dependence of extreme values in the crisis period.

Arabyat et al. (2022), in their paper, consider the standard problem of stock price forecasting. They give a proposal based on an artificial neutral network. The proposal is used to transform high-dimensional input into low-dimensional space to then be used for stock price forecasting.

The paper by Shevchuk and Kopych (2021) considers risk management at the macro level. The authors estimate the exchange rate volatility using the EGARCH (1,1) model and its impact on the business cycle fluctuations in four Central and Eastern European countries. The main findings of the paper are the impact of the components of the Index of Economic Freedom, inflation and crisis on exchange rate volatility.

This volume also contains two papers where the risk management is studied at the micro level.

The paper by Frączkiewicz-Wronka et al. (2021) considers the relationship between risk management and the financial stability of public hospitals in Poland. This was empirically studied using data from more than 100 hospitals. The results show that risk management practices are positively related to financial stability and that the hospitals with well-developed risk management practices are better prepared to keep their financial stability at the required level.

The other paper in this group, by Śliwiński et al. (2022), aims to identify the risk factors affecting bancassurance development in Poland. The group of risk factors contained the factors directly related to the insurance product and those resulting from the specificity of the bancassurance channel. The study was conducted on the basis of data on the gross premiums written in Poland in the years 2004–2019.

3. Conclusions

The importance of risk management methodology has grown significantly in the past thirty years. In the first section, the main challenges to be faced by risk managers are indicated. To conclude, the most important risks that should be managed on all levels are:

- Climate and environmental risk;
- Cyber risk;
- Media (and social media) risk—many media go viral by providing fake news, and unfortunately the recipients of this news use it to make decisions;
- Technology risk (non-transparent and non-professional computer algorithms).

All the mentioned risks have an impact on the economy and financial sector, and therefore education and data analysis methodology play a crucial role.

Acknowledgments: The author and his co-editor are grateful to the MDPI Publisher for the invitation to act as guest editors of this Special Issue and are indebted to the editorial staff of "*Risks*" for their kind co-operation, patience and committed engagement.

Conflicts of Interest: The author declares no conflict of interest.

References

Arabyat, Yaser Ahmad, Ahmad Ali AlZubi, Dyala M. Aldebei, and Samerra'a Ziad Al-oqaily. 2022. An Efficient Method for Pricing Analysis Based on Neural Networks. *Risks* 10: 151. [CrossRef]

Berent, Tomasz, and Radosław Rejman. 2021. Bankruptcy Prediction with a Doubly Stochastic Poisson Forward Intensity Model and Low-Quality Data. *Risks* 9: 217. [CrossRef]

Dziwok, Ewa, and Marta A. Karaś. 2021. Systemic Illiquidity Noise-Based Measure—A Solution for Systemic Liquidity Monitoring in Frontier and Emerging Markets. *Risks* 9: 124. [CrossRef]

Frączkiewicz-Wronka, Aldona Kinga, Tomasz Ingram, Karolina Barbara Szymaniec-Mlicka, and Piotr Tworek. 2021. Risk Management and Financial Stability in the Polish Public Hospitals: The Moderating Effect of the Stakeholders' Engagement in the Decision-Making. *Risks* 9: 87. [CrossRef]

Górka, Joanna, and Katarzyna Kuziak. 2022. Volatility Modeling and Dependence Structure of ESG and Conventional Investments. *Risks* 10: 20. [CrossRef]

Kochanski, Błażej. 2022. Which Curve Fits Best: Fitting ROC Curve Models to Empirical Credit-Scoring Data. *Risks* 10: 184. [CrossRef]

Ptak-Chmielewska, Aneta, and Paweł Kopciuszewski. 2022. New Definition of Default—Recalibration of Credit Risk Models Using Bayesian Approach. *Risks* 10: 16. [CrossRef]

Rivera-Escobar, Orlando, John Willmer Escobar, and Diego Fernando Manotas. 2022. Measurement of Systemic Risk in the Colombian Banking Sector. *Risks* 10: 22. [CrossRef]

Shevchuk, Victor, and Roman Kopych. 2021. Exchange Rate Volatility, Currency Misalignment, and Risk of Recession in the Central and Eastern European Countries. *Risks* 9: 82. [CrossRef]

Śliwiński, Adam, Joanna Dropia, and Norbert Duczkowski. 2022. Risk Factors Affecting Bancassurance Development in Poland. *Risks* 9: 130. [CrossRef]

Szepannek, Gero. 2022. An Overview on the Landscape of R Packages for Open Source Scorecard Modelling. *Risks* 10: 67. [CrossRef]

risks

Article

Volatility Modeling and Dependence Structure of ESG and Conventional Investments

Joanna Górka [1] and Katarzyna Kuziak [2,*]

[1] Department of Econometrics and Statistics, Nicolaus Copernicus University in Torun, 87-100 Toruń, Poland; joanna.gorka@umk.pl
[2] Department of Financial Investments and Risk Management, Wroclaw University of Economics and Business, 53-345 Wrocław, Poland
* Correspondence: katarzyna.kuziak@ue.wroc.pl

Abstract: The question of whether environmental, social, and governance investments outperform or underperform other conventional financial investments has been debated in the literature. In this study, we compare the volatility of rates of return of selected ESG indices and conventional ones and investigate dependence between them. Analysis of tail dependence is important to evaluate the diversification benefits between conventional investments and ESG investments, which is necessary in constructing optimal portfolios. It allows investors to diversify the risk of the portfolio and positively impact the environment by investing in environmentally friendly companies. Examples of institutions that are paying attention to ESG issues are banks, which are increasingly including products that support sustainability goals in their offers. This analysis could be also important for policymakers. The European Banking Authority (EBA) has admitted that ESG factors can contribute to risk. Therefore, it is important to model and quantify it. The conditional volatility models from the GARCH family and tail-dependence coefficients from the copula-based approach are applied. The analysis period covered 2007 until 2019. The period of the COVID-19 pandemic has not been analyzed due to the relatively short time series regarding data requirements from models' perspective. Results of the research confirm the higher dependence of extreme values in the crisis period (e.g., tail-dependence values in 2009–2014 range from 0.4820/0.4933 to 0.7039/0.6083, and from 0.5002/0.5369 to 0.7296/0.6623), and low dependence of extreme values in stabilization periods (e.g., tail-dependence values in 2017–2019 range from 0.1650 until 0.6283/0.4832, and from 0.1357 until 0.6586/0.5002). Diversification benefits vary in time, and there is a need to separately analyze crisis and stabilization periods.

Keywords: ESG; risk management; volatility; GARCH; copula; tail dependence

Citation: Górka, Joanna, and Katarzyna Kuziak. 2022. Volatility Modeling and Dependence Structure of ESG and Conventional Investments. *Risks* 10: 20. https://doi.org/10.3390/risks10010020

Academic Editor: Wing-Keung Wong

Received: 14 December 2021
Accepted: 6 January 2022
Published: 12 January 2022

1. Introduction

The pandemic has highlighted social and global inequality and spiked interests in environmental, social, and governance (ESG) investing. ESG assets reached $35.3 trillion in 2020 from around $30.7 trillion in 2018, reaching a third of current total global assets under management, according to the Global Sustainable Investment Association (http://www.gsi-alliance.org/, accessed on 18 November 2021).

According to the 2020 Trends Report, investors are considering ESG factors across USD 17 trillion of professionally managed assets, a 42% increase since 2018. Such continued growth is expected over the long term, too. Since 1995, the value of US sustainable investment assets (USD 639 billion) has increased more than 25-fold (USD 16.6 trillion in 2020), at a compound annual growth rate of 14% (US SIF 2020).

A few terms are used interchangeably to describe environmental, social, and governance investments, e.g., socially responsible investing (SRI), responsible investing, sustainable investing, and impact investing. In this paper, we understand ESG investing in terms of the ESG factors, which enhance traditional financial analysis, making it more complete.

There are some ESG factors that are helpful in the evaluation of investment performance. Investments with high ESG scores can increase rates of return, while those with poor ESG scores may inhibit these rates. One may find energy consumption, pollution, climate change, waste production, natural resource preservation (deforestation, carbon emission reduction), and animal welfare among the environmental factors. The social factors include human rights, child and forced labor, community engagement, health and safety, stakeholder relations and employee relations, customer success, data hygiene, and security. The governance factors contain the following: quality of management, board independence, conflict of interest, executive compensation, hiring and onboarding best practices, transparency, and disclosure. In this paper, we use the term ESG to refer to these indices, which include the companies that disclose these factors.

Morgan Stanley's research on nearly 11,000 mutual funds between 2004 and 2018 indicates there is no financial trade-off in returns of sustainable funds compared to traditional ones. Moreover, sustainable funds showed lower downside risk. The number of ESG focused funds has been growing, since 2004 by 144% (Morgan Stanley, Institute for Sustainable Investing, Sustainable Reality. Analyzing Risk and Returns of Sustainable Funds, 2019. www.morganstanley.com, accessed on 18 November 2021).

A growing number of investors not only focus on the profitability of investment strategies but also look for their social value. ESG investing fulfills this goal. ESG looks at the company's environmental, social, and governance practices, as well as traditional ones. ESG investors believe that investments in companies employing ESG practices may have a material impact on their investments' profitability and risk. Moreover, they accept a lower return in the short term and even slightly higher risk because of the additional future value of their investment.

Considering ESG benefits in investing for economic value is not a new concept. The report "Who Cares, Wins—Connecting Financial Markets to a Changing World", which provided guidelines for companies to incorporate ESG into their operations, was published in 2004. The first ESG index, the Domini 400 Social Index (now known as the MSCI KLD 400 Social Index), was launched in May 1990. Since then, the ESG indices have evolved to meet investors' unique needs in investing.

ESG equity indices are used as benchmarks for ESG investment, as the underlying assets of passive ESG investment tools (such as exchange-traded funds), and as related risk-management tools (ESG index futures). ESG equity indices are usually constructed based on parent equity indices incorporating ESG investment styles. The construction process may comprise screening out the companies with negative ESG impacts and including the companies with positive ESG impacts by adjusting their relative weights. Therefore, the risk-return performances of these indices may be different from those of their parent indices with traditional investment strategies (HKEX 2020).

Making investment decisions, including portfolio construction, requires knowledge of volatility and dependence between financial time series. Determination of the dependence structure is essential for portfolio management, and the misinterpretation of the strength of this dependence can lead to wrong investment decisions. Pearson's linear correlation coefficient is the most common and widely used correlation measure. Because of its linearity, it is a tool appropriate only for measuring the dependence between variables of elliptic distributions. In situations where empirical data are characterized by, for example, asymmetric and non-elliptic distributions, with high kurtosis or skewness, the use of linear correlation coefficient is not advisable. In such instances, copula functions are better tools for investigating the dependence structure between the time series (Messaoud and Aloui 2015).

The objective of the paper is to evaluate the attractiveness of ESG investments as a potential diversifier for conventional investments. In this study, we examine the following hypothesis:

Hypothesis 1 (H1). *ESG investments outperform conventional investments in terms of risk.*

Hypothesis 2 (H2). *Asymptotic dependence increases during the crisis on the market (declines of stock market indices), stabilizing during the non-crisis periods.*

There are many ESG indices; in this paper, we select five ESG indices (see Table A1 for details). Conventional stock indices are represented by the Dow Jones Industry Average (DJI) and the S&P 500 (GSPC).

Because little is known about the dependence structure between ESG and conventional investments, the contribution of this paper is to fill this research gap. The pioneering character of this research—application of tail dependence to ESG investments, is highly contributing both to the knowledge of the field as well as to the practice of assets selection in portfolio construction. The paper also contributes to improving the understanding of volatility and the (tail) dependence structure between ESG and conventional investments. By quantifying the overall and lower tail risk between these two assets, we indicate that these dependencies exist, can be quantified, and are not negligible, especially in times of crisis.

The research results confirm the higher dependence of extreme values in the crisis period (declines of indices); however, this is also due to increased volatility during this period.

The paper is structured as follows: the introduction, literature review, research methodology, data description, empirical results, discussion, conclusions, and references.

Literature Review

Managi et al. (2012) report no statistical difference in means and volatilities generated from the SRI indices and conventional indices in neither of the studied regions (the US, the UK, and Japan). Furthermore, they found strong comovements between the two indices in both regimes (bear and bull). In contrast, Ortas et al. (2014) found that in the period of the global financial crisis of 2008, social and responsible investment strategies were less risky in comparison to conventional investments.

There is no consensus in the literature as to whether ESG investments are characterized by very high returns and very low risks compared to conventional ones.

Weber and Ang (2016) analyzed the performance of an emerging market SRI index concerning its financial performance compared to conventional indices. Their results indicate that the SRI index outperformed in terms of mean return the majority of the conventional emerging market portfolios. Similarly, Verheyden et al. (2016) found that both global and developed-markets portfolios (a 10% best-in-class ESG screening approach) show higher returns, lower (tail) risk, and no significant reduction of diversification potential.

On the other hand, Giese and Lee (2019) reported no clear consensus on whether ESG criteria have enhanced risk-adjusted returns. The empirical findings in the report on ESG indices performance (HKEX 2020), indicate that many ESG indices tended to have similar, if not better, risk-return performances then their parent indices. This implies that investment in ESG indices may provide equally good or even better returns while pursuing an ethics-focused investment strategy.

Ouchen (2021) empirically verified whether the series of returns of an ESG index was less volatile than that of a conventional stock index. He concluded that the ESG index was relatively less turbulent than the stock index. Jain et al. (2019) reported there is no significant difference in the performance between sustainable indices and conventional indices. Plastun et al. (2022) investigated returns on ESG and conventional indices. They showed no significant differences between ESG and conventional indices. The types of price effects detected by them were the same for the cases of ESG and conventional indices (but their power was different in some cases).

Charles et al. (2016) compared the risk-adjusted performance between ESG and conventional indices, as well as within the ESG indices, examining it based on standard and tail risk measures. They showed that the ESG screens for equities lead neither to a significant outperformance nor an underperformance compared to the benchmarks. They also indicated that the weights used to construct these indices (sustainability-score weights vs. market cap-weights) seemed to impact their risk and performance.

Apergis et al. (2015) employed a standard cointegration methodology and a novel time-varying quantile cointegration approach to investigate whether the US Dow Jones Sustainability Index and its conventional parent index are integrated. The results confirmed the presence of an asymmetric long-run relationship between these indices that is not detected by the standard methodology of cointegration.

Classical Markowitz portfolio theory does not consider the role currently played by the ESG investments on the market. It applies risk and return as single criteria, assuming that investors are rational and seek the highest return at the lowest level of risk, and their utility functions are convex (Markowitz 1952). Incorporating the ESG-based factor into the portfolio selection problem, Pedersen et al. (2021) proposed a hypothesis that explained how the increasingly widespread adoption of ESG affected portfolio choice and equilibrium asset prices.

In the case of dependency modeling for classical portfolio theory, linear correlation coefficients were used (assuming elliptic distributions of returns). The problem appears because the returns are not correlated strongly when they are around zero; however, the correlation increases in the tails. Then an appropriate tool for dependency modeling is the copula function (Sklar 1959).

Empirical research indicates that stock returns also display an asymmetric dependence in growing and declining markets, i.e., this dependence may be stronger in bearish markets than in bullish markets and tends to increase in the periods of violent fluctuations of prices (Ang and Bekaert 2002; Jondeau 2016; Longin and Solnik 2001). Because of the detected asymptotic dependence of random variables in tails (Patton 2006), the authors apply copula functions. This approach allows investigating the dependence in variance and in tails, which is not possible using standard dependence measures.

While more than 2000 empirical studies have been conducted analyzing the ESG factors and financial performance, still little is known about the dependence structure and the associated risks (Friede 2019). This is especially important as ESG scores are often related to investment risk (Bax et al. 2021).

Due to the imperfections of financial time series (they are not normally distributed) and correlation coefficients (which measure linear relationship and are constant in time), we applied GARCH family models and copula functions accompanied with some heavy-tailed marginal distributions.

2. Data and Methods

2.1. Research Methodology

The ability to forecast volatility of assets is vital for portfolio selection and asset management, as well as for the pricing of primary and derivative assets (Engle and Ng 1993).

Early studies point to volatility clustering, leptokurtosis, and the leverage effect in stock-returns time series (Mandelbrot 1963) and (Fama and Fama 1965). The additional features of financial time series observed across different financial assets (stocks, stock indices, exchange rates) are as follows: stationarity, fat tails, asymmetry, aggregational Gaussianity, quasi-long-range dependence, and seasonality (e.g., Rydberg 2000; Taylor 1986). In the GARCH model, the variance is influenced by the square of the lagged innovation. However in the equity returns the leverage effect (higher impact of negative shocks on volatility) is observed, the simple GARCH model fails to describe it. GARCH (1,1) with a generalized residuals distribution can capture more volatility assessment than other models. On the other hand, the impact of asymmetry on stock market volatility and return analysis is beyond the descriptive power of the asymmetric GARCH models, which could capture more details.

There are several limitations to GARCH models. The most important one is the inability to capture the asymmetric performance. For that reason, EGARCH, GJR-GARCH, and APGARCH models were proposed. Furthermore, the asymmetric GARCH models can measure the effect of positive or negative shocks on stock market returns and volatility incompletely, and the GARCH (1,1) comparatively fails to accomplish this. The GJR-GARCH model performs better in the face of asymmetry, producing a predictable conditional variance during the period of high volatility. In addition, among the asymmetric GARCH models, the performance of the EGARCH model appeared to be superior.

Based on the properties of the studied time series: volatility clustering, leptokurtosis, asymmetry, leverage effects, mean-reversion, and stationarity—we apply the following models from the GARCH family: GARCH, EGARCH, GJR-GARCH, APGARCH, and AVGARCH in the study. We selected the GARCH model using the Akaike (AIC) and Bayesian (BIC) information criteria. Tables A2–A5 present only the results of estimation for the best-fitted models according to these criteria (with the minimum criteria values).

Base ARMA(p, q) model is as follows (Box and Jenkins 1983; Brockwell and Davis 1991):

$$r_t = \phi_1 r_{t-1} + \ldots + \phi_p r_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \ldots - \theta_q \varepsilon_{t-q}$$

where $\varepsilon_t \sim i.i.d.(0, h_t)$.

The autoregressive conditional heteroskedasticity (ARCH) models were introduced by Engle (1982) and their generalization, the GARCH models, by Bollerslev (1986). The standard GARCH(q, p) model (Bollerslev 1986) may be written as:

$$h_t = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i}$$

where h_t is the conditional variance, ω the intercept, and ε_t^2 the residuals from the ARMA model.

Some researchers pointed out limitations of the GARCH model. The most important one is that GARCH cannot capture asymmetric performance. Later, for improving this problem, EGARCH, GJR-GARCH, and APGARCH were proposed.

The exponential GARCH (EGARCH) is designed to model the logarithm of the variance rather than the level, and this model accounts for an asymmetric response to a shock. The exponential GARCH model of Nelson (1991) is defined as:

$$\ln h_t = \omega + \sum_{i=1}^{q} (\alpha_i z_{t-i} + \gamma_i (|z_{t-i}| - E|z_{t-i}|)) + \sum_{i=1}^{p} \beta_i \ln h_{t-i}$$

where the coefficient α_i captures the sign effect and γ_i the size effect. $E|z_{t-i}|$ is the expected value of the absolute standardized innovation z_t.

The Glosten–Jagannathan–Runkle GARCH (GJR-GARCH) as GARCH model captures features of financial time series like leptokurtic returns and volatility clustering.However, the GJR-GARCH model of Glosten et al. (1993) models positive and negative shocks on the conditional variance asymmetrically by the use of the indicator function I:

$$h_t = \omega + \sum_{i=1}^{q} \left(\alpha_i \varepsilon_{t-i}^2 + \gamma_i I_{t-i} \varepsilon_{t-i}^2 \right) + \sum_{i=1}^{p} \beta_i h_{t-i}$$

where γ_i now represents the 'leverage' term. The indicator function I takes on value of 1 for $\varepsilon \leq 0$ and zero otherwise.

The asymmetric power ARCH (APARCH) model of Ding et al. (1993) allows for both leverage and the Taylor effect, named after Taylor (1986) who observed that the sample autocorrelation of absolute returns was usually larger than that of squared returns:

$$\left(\sqrt{h_t} \right)^{\delta} = \omega + \sum_{i=1}^{q} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^{\delta} + \sum_{i=1}^{p} \beta_i \left(\sqrt{h_{t-i}} \right)^{\delta},$$

where $\delta \in \mathbb{R}^+$, being a Box–Cox transformation of $\sqrt{h_t}$, and γ_i the coefficient in the leverage term.

The absolute value GARCH (AVGARCH) model of Taylor (1986) and Schwert (1990):

$$\sqrt{h_t} = \omega + \sum_{i=1}^{q} \alpha_i \sqrt{h_{t-i}}(|z_{t-i} - \eta_{2i}| - \eta_{1i}(z_{t-i} - \eta_{2i})) + \sum_{i=1}^{p} \beta_i \sqrt{h_{t-i}},$$

where η_{1i} and η_{2i} are rotations and shifts parameters respectively.

This paper examines the structure of interdependence between ESG and conventional indices. In order to achieve this goal, we fit different theoretical distributions to the series of returns. Then, we assumed the best-fitted distribution describing the process of returns as a marginal distribution applied for our copula estimation. We used the two-stage maximum likelihood method to estimate the parameters of the considered two-dimensional copulas. In addition to testing the goodness of fit of alternative copulas, we also verified a set of hypotheses relating to the correlation matrix.

The concept of copula, was first introduced by Sklar (1959). The theoretical background for copulas is provided by Sklar's theorem. There exists a copula C such that:

$$\forall_{x_1 \in X_1 |\mathrm{i}| x_2 \in X_2} \quad F(x_1, x_2) = C(F_1(x_1), F_2(x_2))$$

where F is two-dimensional joint distribution with the marginal distributions F_1, F_2 of random variables (X_1, X_2). If F_1, F_2 are continuous, the copula C is unique:

$$C(u_1, u_2) = F\left(F_1^{-1}(u_1), F_2^{-1}(u_2)\right)$$

where $(u_1, u_2) \in [0, 1]$, $F_i^{-1}(u) = \inf\{x; F_i(x) \geq u\}$ for $i = 1, 2$.

The proof is provided by, e.g., Nelsen (2006).

One may use a wide range of parametric copula families to capture the different structures of dependence (e.g., Gaussian, Archimedean). In the paper, we applied the following copulas: Student's t, Joe–Clayton (a combination of the Joe copula and the Clayton copula), Gumbel–Clayton (a combination of the Clayton copula and the Gumbel copula), and survival. The Gaussian copula does not capture tail dependence, Student's t-copula has symmetric tail dependence in both lower and upper tails, and the Clayton and Gumbel copulas have only lower and upper tail dependence, respectively. Survival copulas correspond to rotation by 180 degrees.

For two dimensions following copulas are defined as (Patton 2006):

(1) Gaussian/normal (N) copula

$$C_N(u_1, u_2;) = N\left(\Phi^{-1}(u_1), \Phi^{-1}(u_2)\right)$$

where N is the normal joint distribution and Φ^{-1} is the quantile of the univariate normal distribution;

(2) Student's t/t (t) copula

$$C_t(u_1, u_2; \nu, \rho) = t_{\nu,\rho}\left(t_\nu^{-1}(u_1), t_\nu^{-1}(u_2)\right)$$

where $\rho \in [-1, 1]$, $t_{\nu,\rho}$ is the joint Student's t distribution and t_ν^{-1} is the univariate Student's t distribution with ν degrees of freedom;

(3) Clayton–Gumbel (*BB1*) copula

$$C_{\{BB1\}}(u_1, u_2; \theta, \delta) = \left(1 + \left[\left(u_1^{-\theta} - 1\right)^{\delta} + \left(u_2^{-\theta} - 1\right)^{\delta}\right]^{\frac{1}{\delta}}\right)^{-\frac{1}{\theta}}$$

where $\theta \geq 0$, $\delta \geq 1$;

(4) Joe–Clayton (*BB7*) copula

$$C_{\{BB7\}}(u_1, u_2; \theta, \delta) = 1 - \left(1 - \left[\left(1 - \overline{u}_1^{\theta}\right)^{-\delta} + \left(1 - \overline{u}_2^{\theta}\right)^{-\delta} - 1\right]^{-\frac{1}{\delta}}\right)^{\frac{1}{\theta}}$$

where $\theta \geq 0$, $\delta \geq 1$, $\overline{u}_1 = 1 - u_1$, $\overline{u}_2 = 1 - u_2$.

Survival copula is the copula of $(1 - u_1)$ and $(1 - u_2)$ instead of u_1 and u_2, respectively. Its function can measure the asymmetric dependence on the opposite side of the distribution as compared to the original function.

Our focus was on the extreme downside market risk, so we investigated the lower tail dependence in detail. The tail-dependence coefficients (Patton 2006) are:

- Lower tail-dependence coefficient:

$$\lambda_L = \lim_{u \to 0^+} P(X_2 \leq F_2^{-1}(u) | X_1 \leq F_1^{-1}(u)) = \lim_{u \to 0^+} \frac{C(u,u)}{u}$$

- Upper tail-dependence coefficient:

$$\lambda_U = \lim_{u \to 1^-} P(X_2 > F_2^{-1}(u) \, | X_1 \rangle F_1^{-1}(u)) = \lim_{u \to 1^-} \frac{1 - 2u + C(u,u)}{1 - u}$$

in the case that the limit exists, $\lambda_L, |\lambda_U \in [0,1]$ and $(\lambda_L \neq 0 \vee \lambda_U \neq 0)$, dependence is present.

The bivariate normal distribution is tail independent, the bivariate Student's t-distribution exhibits the same upper and lower tail dependence, the bivariate Joe–Clayton and Gumbel–Clayton distributions have both lower and upper tail dependence. The concept of tail dependence is embedded within the copula theory.

Instead of Pearson's correlation coefficient in the copula theory, we use the Kendall's τ coefficient (Nelsen 2006; Patton 2006). For each pair, the Kendall's τ is estimated and the *p*-values of the independence test based on the Kendall's τ were determined in the study.

In this paper, we used functions from the VineCopula library in R (Stoeber et al. 2018). We made the following assumptions during the estimation process: we took 39 copulas into consideration, we applied the Maximum Likelihood Estimation (MLE) method, and we used the AIC selection criterion to select the best-fitted copula. To measure the discrepancy between a hypothesized model and the empirical model, we used the goodness-of-fit (GoF) statistics based on the Kendall's process as proposed by Wang and Wells (2000). For computation of *p*-values, the parametric bootstrap described by Genest et al. (2006) was used.

2.2. Data Description

In the study, we used five selected ESG indices (presented in Table A1) and two stock indices—Dow Jones Industry and the S&P 500. Daily logarithmic rates of return for them were calculated, as a difference between logarithms of two consecutive closing prices multiplied by 100 (they can be interpreted as percentage changes). The data set was retrieved from the Thomson Reuters database. In Table 1, Table 4, Table 7 and Table 10 descriptive statistics for rates of return of selected indices were given. The period of analysis covered 3 July 2007 until 31 December 2019. We considered, based on important market events (e.g., global financial crisis, debt crisis in the EU, fall in oil prices) and data requirements from the models' perspectives, the following four subperiods:

1. 3 July 2007–30 January 2009 (388 observations)—the global financial crisis period;
2. 2 December 2009–30 July 2014 (1140 observations)—the period of debt crisis in the EU;
3. 3 June 2014–31 January 2017 (633 observations)—the Russian financial crises, fall in oil prices;
4. 3 January 2017–31 December 2019 (746 observations)—stabilization period.

3. Results

3.1. General Remarks

In order to identify the processes of the series of daily rates of return of the ESG indices and conventional indices (S&P 500, DJI) during the period between 3 July 2007 and 31 December 2019 (excluding the COVID-19 pandemic), it is essential to conduct graphical, statistical, and econometric examinations of these time series to check for their stationarity and the presence of the ARCH effect. We conduct such analysis for four subperiods.

The graphic examination of values of all indices shows that they are not stationary, but their daily rates of return are stationary (see Figures 1, 3, 5 and 7). We applied unit root tests, i.e., the Augmented Dickey–Fuller, Phillips–Perron, and KPSS to confirm stationarity.

At the same time, the distribution of the returns has tails, which are heavier than the tails of the normal distribution. To confirm this, we used the Jarque–Bera test and Q-Q plots.

Generally, the daily returns for both types of indices exhibit no significant autocorrelation, supporting the hypothesis that the returns are uncorrelated across time. To confirm this, we applied the Ljung–Box test. Finally, we checked whether the rates of return are characterized by the ARCH effect using the McLeoda and Li test. The null hypothesis is that the rates of return do not have the ARCH effect, while the alternative hypothesis is that they have the ARCH effect. In all subperiods, the ARCH effect was detected in the returns.

In the case of the standardized innovations from the ARMA-GARCH models, at the assumed 5% level of significance, the *p*-values for the Engle test are greater than 0.05 (see Tables A3–A5), which means that the null hypothesis was not rejected, i.e., the ARCH effect is not present in the innovations. Hence, the models are free from conditional heteroskedasticity in almost all cases (only in the first period for three indices the ARCH effect is present—Table A2). To verify the autocorrelation in the innovations, we used the Ljung–Box test. The null hypothesis that the innovations are independently distributed was not rejected in all the cases (see Tables A2–A5). It means that the models have been well-chosen and fitted.

Finally, the persistence parameters are close to one, which is high (see Tables A2–A5), meaning that the variance moves slowly through time.

We employ not only normally distributed innovations, but also the Student's t-distributions, the generalized error distribution (GED) and a skewed version of both. The reason for considering distributions other than normal is that a GARCH model with conditional normal errors has fatter tails than the normal distribution, and for many financial time series the standardized innovations still appear to be leptokurtic. Therefore, assuming a leptokurtic unconditional distribution for the innovations seems more appropriate.

Because the market returns are not normally distributed, the Gaussian copula would not capture tail dependence. Therefore, we fitted the Student's t-copula and the combination of Clayton and Gumbel copulas. At a 1% significance level for the chosen copulas, we cannot reject the null hypothesis (GoF test), i.e., our copulas are the true copulas. Due to the observed nonnormality in the returns distribution, we measured the dependence by using Kendall's τ coefficient. All results for the Kendall's τ coefficient are statistically significant. In the first two subperiods, dependencies were higher than in the next two subperiods (see Table 2, Table 3, Table 5, Table 6, Table 8, Table 9, Table 11 and Table 12). To quantify the degree of tail dependence in each pair, the Kendall's τ is estimated and the *p*-values of the independence test based on the Kendall's τ are determined (see Figure 2 and Tables 2 and 3). The results indicate the existence of positive, significant dependence.

We analyzed the dependence for 10 pairs of indices, namely S&P and ESG—5 pairs and DJI and ESG—5 pairs. Dependence between S&P and DJI exhibited lower tail dependence in three subperiods; only in the 2008 global financial crisis was the tail dependence symmetric. In the first and last subperiod, ESG and conventional investments showed lower and symmetric tail dependence—in the first subperiod for 5 pairs in the lower tail, for 5 pairs symmetric; in the third subperiod for 3 pairs in the lower tail and for 7 pairs symmetric. In the second period, ESG and conventional investments exhibited lower and upper tail dependence—for 6 pairs in the lower tail, for 3 pairs symmetric (see Tables 2, 3, 5, 6, 8, 9, 11 and 12).

3.2. The Global Financial Crisis Period (3 July 2007–30 January 2009)

In the global financial crisis period, all indices are not normally distributed (high kurtosis, mostly positive asymmetry, as confirmed by the Jarque–Bera test) with negative means and similar standard deviations (lower only for SGESGSEP and TRESQ1—Table 1). The year 2008 was characterized by high volatility. Volatility clustering was observed for all indices (Figure 1). There is no consensus in modeling the volatility of conventional indices. There were APGARCH (innovations with normal distribution—norm) for S&P 500 and EGARCH (innovations with skewed Student's t-distribution—sstd) for the fitted DJI. For ESG indices, GJR-GARCH (innovations with normal distribution) for A1SGI and for other AVGARCH (innovations with normal distribution) fitted best. From Table A2, we can see that all the parameters have very small p-values, which shows their significance. Persistence ranges from 0.9654 to 0.9818 for S&P 500 and DJI, and from 0.9649 to 0.9730 for ESG indices, indicating the slow movement of variance through time.

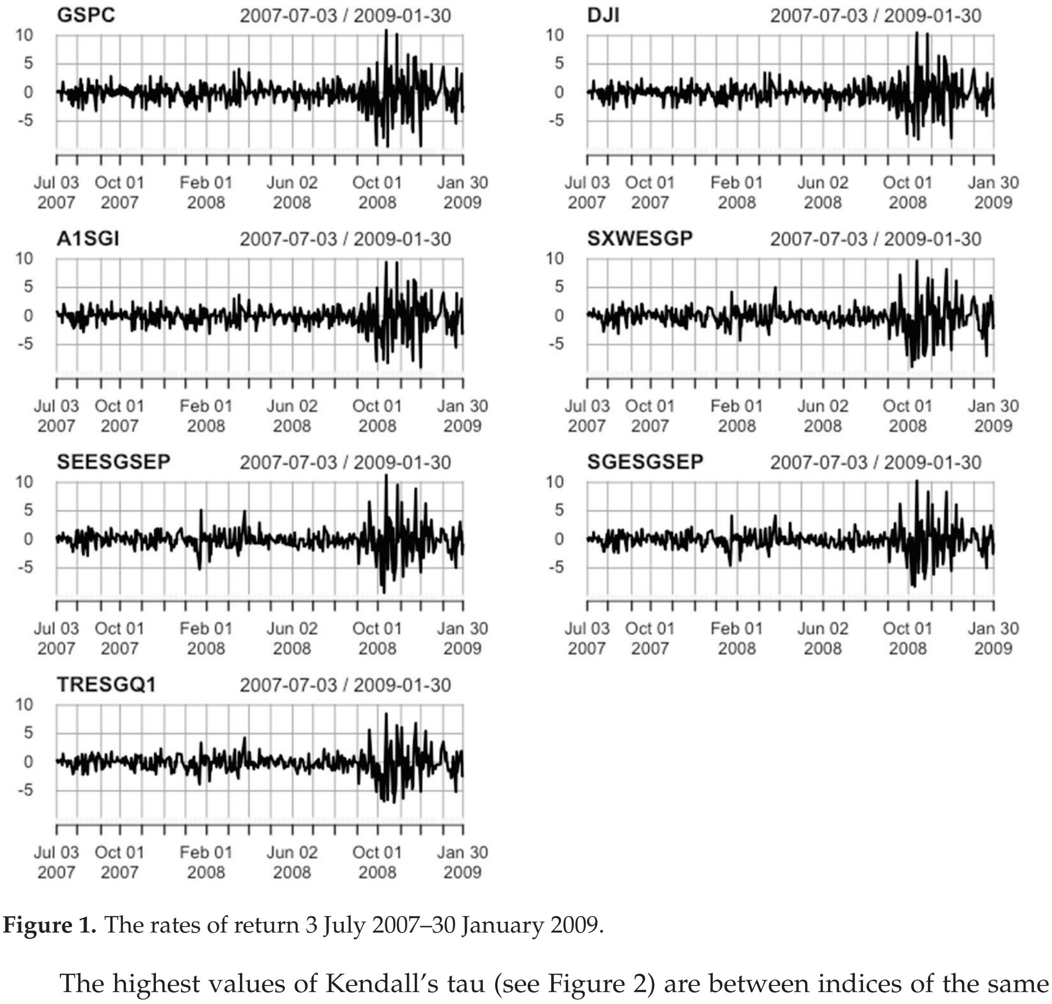


Figure 1. The rates of return 3 July 2007–30 January 2009.

The highest values of Kendall's tau (see Figure 2) are between indices of the same type (S&P 500 and DJI, and SEESGSEP and SGESGSEP). High dependence is observed between S&P 500 and A1SGI and between DJI and A1SGI, low dependence between S&P 500 and SEESGSEP and between DJI and SEESGSEP. For example, for the GSPC–SEESGSEP relationship—the observed copula has the lowest dependency—depends by 34.24% on the upper and lower tail (Student's t-copula was employed). The GSPC–SXWESGP relationship has a dependency of 47.54% on the upper tail, and of 54.41% on the lower tail. It means the interaction has a greater effect in the lower tail (also for GSPC–SGESGSEP, GSPC–TRESGQ1, DJI–SXWESGP, DJI–SGESGSEP, DJI–TRESGQ1).

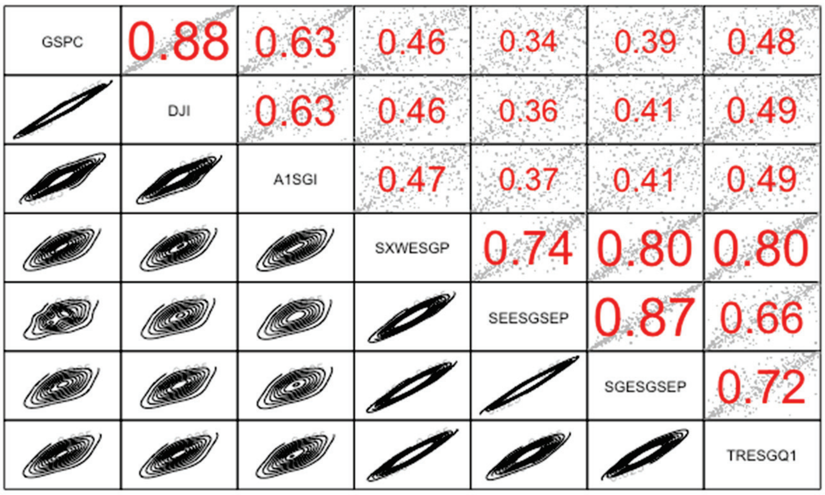


Figure 2. Kendall's τ and copulas 3 July 2007–30 January 2009.

Table 1. Descriptive statistics 3 July 2007–30 January 2009.

	GSPC	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Minimum	−9.470	−8.201	−8.993	−8.927	−9.355	−8.272	−7.069
Maximum	10.957	10.508	9.453	9.712	11.476	10.283	8.504
1. Quartile	−1.190	−1.147	−1.049	−1.120	−1.010	−1.020	−1.008
3. Quartile	0.780	0.744	0.741	0.788	0.736	0.718	0.723
Mean	−0.154	−0.133	−0.153	−0.181	−0.150	−0.153	−0.170
Median	0.024	−0.024	0.011	−0.141	−0.162	−0.093	−0.103
SE Mean	0.116	0.106	0.107	0.106	0.105	0.099	0.094
LCL Mean	−0.382	−0.341	−0.364	−0.390	−0.357	−0.347	−0.356
UCL Mean	0.073	0.074	0.058	0.029	0.056	0.042	0.015
Variance	5.179	4.321	4.479	4.398	4.269	3.796	3.437
Stdev	2.276	2.079	2.116	2.097	2.066	1.948	1.854
Skewness	−0.082	0.181	−0.212	−0.117	0.384	0.240	0.014
Kurtosis (−3)	4.478	4.694	4.268	4.348	6.464	5.731	3.879

Table 2. Dependence structure for the GSPC and ESG indices, 3 July 2007–30 January 2009.

GSPC and ...	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Copula	**t**	**t**	**Joe–Clayton**	**t**	**Joe–Clayton**	**Joe–Clayton**
Par	0.9820	0.8578	1.6436	0.5161	1.4962	1.8196
Par2	2.5474	2.0001	1.1390	2.7057	0.9161	1.1576
Beta			0.4426		0.3921	0.4649
λ_L	0.8673	0.6645	0.5441	0.3424	0.4692	0.5495
λ_U	0.8673	0.6645	0.4754	0.3424	0.4108	0.5363
logLik	668.60	289.87	130.03	79.68	99.50	147.25
AIC	−1333.20	−575.74	−256.07	−155.36	−194.99	−290.51
Indep. (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GoF test	7.3719	4.7832	0.0303	0.5155	0.0472	0.0323
p-value	0.045	0.095	0.89	0.79	0.64	0.875

Indep.—testing for independence for pairs of variables ($H_0: \tau = 0$).

The dependence between the S&P 500 and ESG indices and between the DJI and ESG indices is modeled by Student's t and the Joe–Clayton copula (see Tables 2 and 3). There is one difference—the dependence between S&P 500 and SGESGSEP is described by the Joe–Clayton copula; between DJI and SGESGSEP by the t copula.

Table 3. Dependence structure for DJI and ESG indices, 3 July 2007–30 January 2009.

DJI and . . .	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Copula	**t**	**Joe–Clayton**	**t**	**t**	**Joe–Clayton**
Par	0.8460	1.6701	0.5399	0.6026	1.8628
Par2	2.0001	1.1168	2.5160	2.3969	1.1787
Beta		0.4430			0.4719
λ_L	0.6513	0.5376	0.3706	0.4191	0.5554
λ_U	0.6513	0.4856	0.3706	0.4191	0.5492
logLik	279.78	129.87	87.18	107.47	151.48
AIC	−555.56	−255.74	−170.35	−210.94	−298.95
Indep. (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000
GoF test	4.7128	0.0268	0.1041	1.1651	0.0231
p-value	0.11	0.95	0.95	0.55	0.95

Indep.—testing for independence for pairs of variables ($H_0 : \tau = 0$).

3.3. The Period of Debt Crisis in EU (2 December 2009–30 July 2014)

In the period of debt crisis in the EU, all indices are not normally distributed (mostly high kurtosis, negative asymmetry, as confirmed by the Jarque–Bera test) with positive means and similar standard deviations (lower only for DJI and A1SGI—Table 4). Two periods were characterized by high volatility (visible volatility clustering at all indices in 2009 and 2011—Figure 3). To model the volatility of conventional indices ARMA-EGARCH models (with innovations with skewed GED—sged) were fitted. In the case of ESG indices ARMA-EGARCH models also fitted best, but different distributions for innovation were applied (mostly skewed Student's t-distribution and skewed GED for SEESGSEP). From Table A3 we can see that all the parameters have very small p-values, which shows their statistical significance. Persistence for S&P 500 and DJI is similar (0.9410–0.9411) for ESG indices ranges from 0.9594 to 0.9910, indicating slow movement of variance through time.

The highest values of Kendall's τ (see Figure 4) are between indices of the same type (S&P 500 and DJI, and SEESGSEP and SGESGSEP). High dependence is observed between S&P 500 and A1SGI and between DJI and A1SGI; low dependence between S&P 500 and SEESGSEP and between DJI and SEESGSEP. For example, for the GSPC–SEESGSEP relationship—the observed copula has the lowest dependency (but higher comparing to previous period)—depends by 52.82% on the upper tail and 50.95% on the lower tail. The GSPC-A1SGI relationship has a dependency of 84.33% on the upper tail, and of 89.85% on the lower tail. It means the interaction has a greater effect in the upper tail for the first pair and for DJI–SEESGSEP, DJI–SGESGSEP, and GSPC–SGESGSEP. The greater effect in the lower tail is observed for the second pair and for the other indices.

Table 4. Descriptive statistics, 2 December 2009–30 July 2014.

	GSPC	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Minimum	−6.896	−5.706	−5.794	−5.240	−4.991	−4.376	−5.677
Maximum	4.632	4.153	4.532	5.636	5.797	5.168	4.552
1. Quartile	−0.399	−0.386	−0.385	−0.561	−0.575	−0.449	−0.454
3. Quartile	0.570	0.521	0.539	0.700	0.662	0.559	0.605
Median	0.083	0.064	0.078	0.061	0.090	0.062	0.102
Sum	58.674	48.966	44.397	33.376	16.824	26.214	42.434
SE Mean	0.031	0.028	0.029	0.037	0.035	0.030	0.032
LCL Mean	−0.009	−0.012	−0.017	−0.043	−0.054	−0.036	−0.025
UCL Mean	0.112	0.098	0.095	0.101	0.084	0.082	0.099
Variance	1.078	0.887	0.930	1.536	1.400	1.031	1.138
Stdev	1.038	0.942	0.964	1.239	1.183	1.015	1.067
Skewness	−0.441	−0.412	−0.416	−0.313	−0.352	−0.390	−0.432
Kurtosis (−3)	4.651	4.141	3.932	2.798	2.408	3.049	3.065

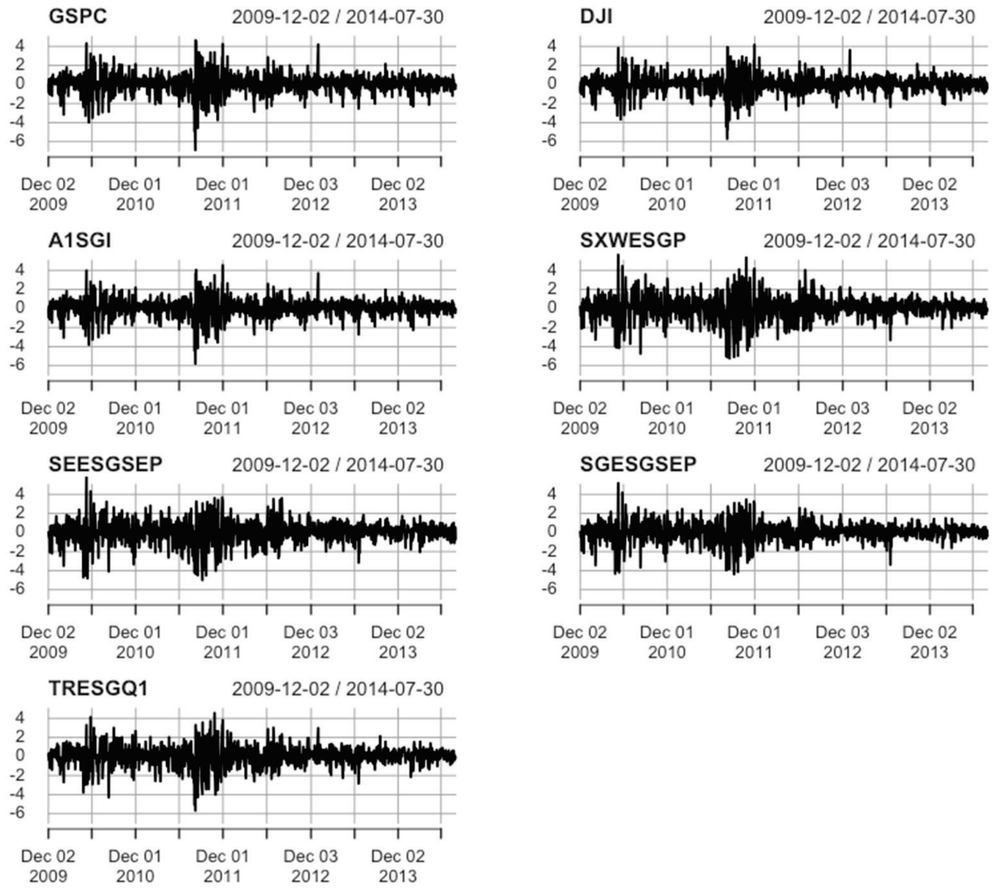


Figure 3. The rates of return, 2 December 2009–30 July 2014.

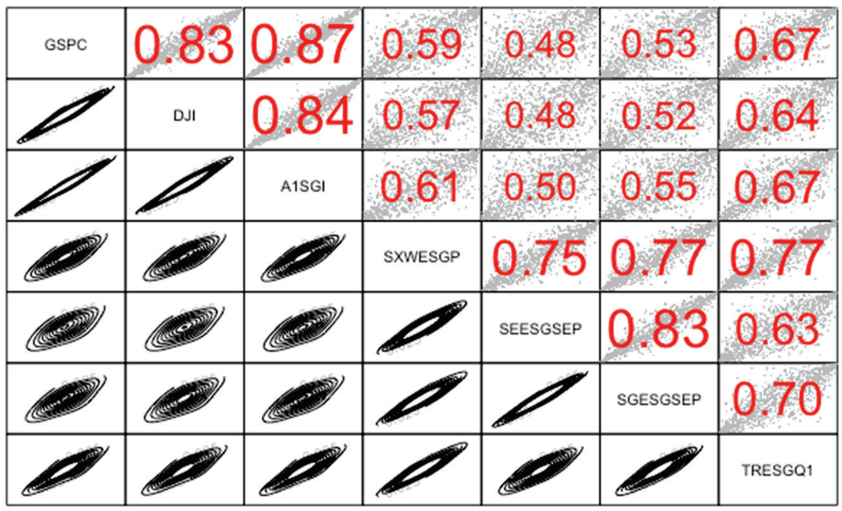


Figure 4. Kendall's τ and copulas, 2 December 2009–30 July 2014.

The dependence between the S&P 500 and ESG indices and between the DJI and ESG indices is modeled by the Clayton–Gumbel and Joe–Clayton copulas (see Tables 5 and 6). The dependences between S&P 500 and SXWESGP, SGESGSEP, and TRESGQ1 comparing to DJI's dependence were modeled differently. For example, GSPC–SGESGSEP relationship is modeled by the Clayton–Gumbel copula, and DJI–SGESGSEP by the Joe–Clayton copula.

Table 5. Dependence structure for GSPC and ESG indices, 2 December 2009–30 July 2014.

GSPC and ...	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Copula	**Clayton–Gumbel**	**Clayton–Gumbel**	**Clayton–Gumbel**	**Joe–Clayton**	**Clayton–Gumbel**	**Clayton–Gumbel**
Par	1.2451	1.3609	0.6552	1.7934	0.5618	0.9411
Par2	3.8209	4.7603	1.8753	1.0280	1.6908	2.0972
Beta	0.8326	0.8701	0.5879	0.4476	0.5276	0.6659
λ_L	0.8644	0.8985	0.5688	0.5095	0.4820	0.7039
λ_U	0.8011	0.8433	0.5528	0.5282	0.4933	0.6083
logLik	1632.44	1919.46	634.47	409.04	493.41	857.63
AIC	−3260.88	−3834.92	−1264.94	−814.07	−982.82	−1711.26
Indep. (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GoF test	0.0462	0.0181	0.0281	0.038	0.033	0.0747
p-value	0.035	0.635	0.86	0.73	0.83	0.115

Indep.—testing for independence for pairs of variables ($H_0: \tau = 0$).

Table 6. Dependence structure for DJI and ESG indices, 2 December 2009–30 July 2014.

DJI and ...	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Copula	**Clayton–Gumbel**	**Joe–Clayton**	**Joe–Clayton**	**Joe–Clayton**	**Survival Joe–Clayton**
Par	1.1574	2.1417	1.8214	1.9686	2.8966
Par2	4.0965	1.5508	1.0006	1.2038	1.6822
Beta	0.8394	0.5301	0.4480	0.4857	0.5951
λ_L	0.8640	0.6396	0.5002	0.5623	0.7296
λ_U	0.8156	0.6178	0.5369	0.5779	0.6623
logLik	1674.90	602.00	411.37	489.18	777.30
AIC	−3345.80	−1200.00	−818.75	−974.35	−1550.61
Indep. (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000
GoF test	0.0359	0.0429	0.0337	0.0226	0.0389
p-value	0.14	0.64	0.795	0.975	0.585

Indep.—testing for independence for pairs of variables ($H_0: \tau = 0$).

3.4. The Russian Financial Crises, Fall in Oil Prices (3 June 2014–31 January 2017)

In the Russian financial crises and fall in oil prices period, all indices are not normally distributed (mostly high kurtosis, negative asymmetry, as confirmed by the Jarque–Bera test) with mostly positive means (for three ESG indices the mean was negative) and similar standard deviations for S&P 500 and DJI. In the case of standard deviation for the ESG indices, large differences were observed—low for A1SGI and high for SEESGSEP (Table 7). In this period, three subperiods characterized by high volatility (visible volatility clustering of all indices in 2015 and the beginning of 2016, and a large drop in the middle of 2016—Figure 5) were present. The volatility of three indices, namely S&P 500, DJI and A1SGI, was modeled by AVGARCH (innovations with skewed Student's t-distribution). There is no consensus in the modeling volatility of the ESG indices. GARCH, EGARCH, and AVGARCH (innovations with Student's t-distribution in all models) were applied. From Table A4, we can see that all the parameters have very small p-values, which shows their statistical significance. Persistence ranges from 0.9655 to 0.9749 for S&P 500 and DJI, and from 0.8972 to 0.9716 for ESG indices.

The highest values of Kendall's τ (see Figure 6) are between indices of the same type (S&P 500 and DJI, and SEESGSEP and SGESGSEP). High dependence is observed between S&P 500 and A1SGI and between DJI and A1SGI; low dependence between S&P 500 and SEESGSEP and between DJI and SEESGSEP. For example, for the DJI–SEESGSEP relationship—the observed copula has the lowest dependency (lower compared to previous periods too)—depends by 21.90% on the upper tail and 21.90% on the lower tail. The GSPC–A1SGI relationship has a dependency of 85.21% on the upper tail, and of 85.21% on the lower tail. It means that the interaction has the same effect in the upper tail and in the lower tail. The greater effects in the lower tail are observed for GSPC–TRESGQ1, DJI–TRESGQ1 and DJI–A1SGI.

Table 7. Descriptive statistics, 3 June 2014–31 January 2017.

	GSPC	DJI	A1SGI	SXWESG	SEESGSE	SGESGSE	TRESGQ
Minimum	−4.021	−3.640	−4.114	−8.160	−9.600	−7.611	−6.608
Maximum	3.829	3.875	3.552	2.998	3.868	2.819	2.970
1. Quartile	−0.355	−0.333	−0.387	−0.513	−0.586	−0.497	−0.392
3. Quartile	0.463	0.463	0.457	0.550	0.550	0.487	0.502
Mean	0.026	0.026	0.018	−0.009	−0.044	−0.029	0.007
Median	0.013	0.035	0.025	0.001	−0.010	−0.034	0.001
SE Mean	0.033	0.033	0.034	0.038	0.042	0.035	0.034
LCL Mean	−0.040	−0.038	−0.049	−0.084	−0.128	−0.098	−0.059
UCL Mean	0.091	0.090	0.085	0.067	0.039	0.040	0.074
Variance	0.743	0.713	0.766	0.982	1.192	0.811	0.761
Stdev	0.862	0.844	0.875	0.991	1.092	0.901	0.872
Skewness	−0.376	−0.312	−0.347	−1.010	−1.031	−0.954	−0.862
Kurtosis (−3)	2.615	2.201	2.265	7.388	9.220	8.011	5.855

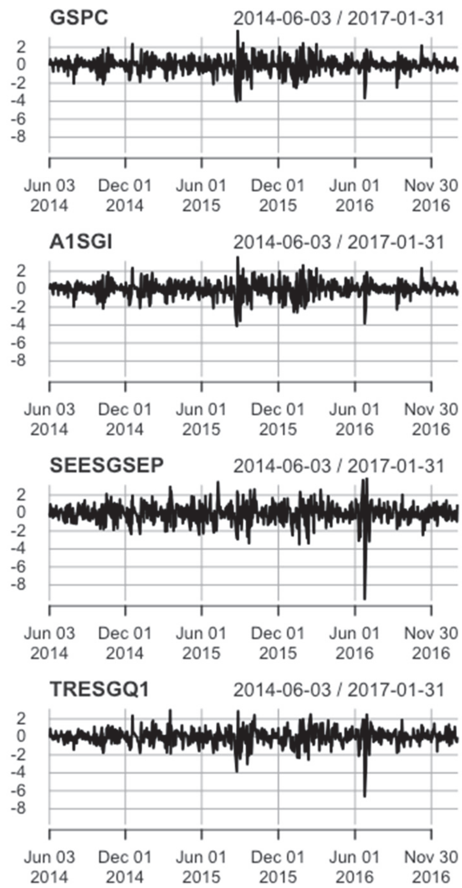


Figure 5. The rates of return, 3 June 2014–31 January 2017.

The dependences between the S&P 500 and ESG indices and between the DJI and ESG indices were modeled by the Clayton–Gumbel and t copulas (see Tables 8 and 9). The dependence between S&P 500 and A1SGI compared to the DJI dependence was modeled differently. For example, the GSPC–A1SGI relationship is modeled by Student's t copula, and DJI–GESGSEP by Clayton–Gumbel copula.

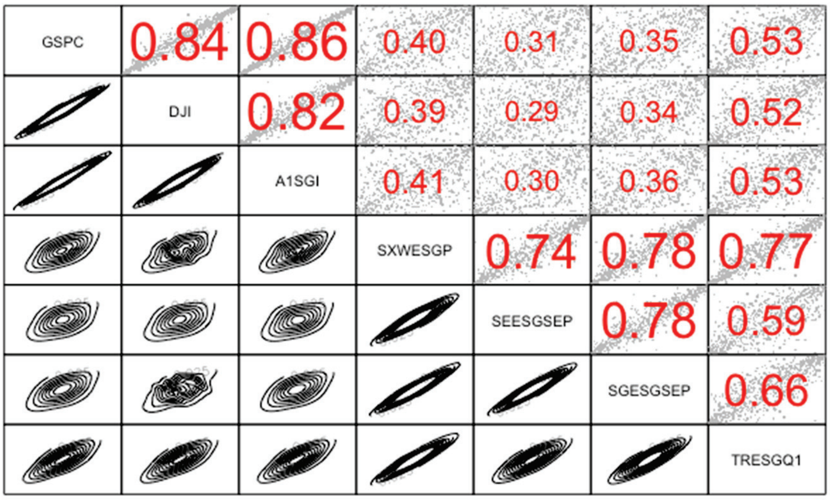


Figure 6. Kendall's τ and copulas, 3 June 2014–31 January 2017.

Table 8. Dependence structure for the GSPC and ESG indices, 3 June 2014–31 January 2017.

GSPC and ...	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Copula	**Clayton–Gumbel**	**t**	**t**	**t**	**t**	**Clayton–Gumbel**
Par	1.0439	0.9777	0.5900	0.4638	0.5336	0.7395
Par2	4.2304	2.5661	3.8015	4.0842	4.0282	1.5687
Beta	0.8388					0.5237
λ_L	0.8547	0.8521	0.3184	0.2296	0.2709	0.5502
λ_U	0.8220	0.8521	0.3184	0.2296	0.2709	0.4444
logLik	963.82	1075.43	161.66	98.89	129.49	282.82
AIC	−1923.64	−2146.7	−319.32	−193.79	−254.97	−561.64
Indep. (*p*-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GoF test	0.0287	2.3726	2.9012	6.674	2.6074	0.0385
p-value	0.285	0.365	0.245	0.03	0.23	0.7

Indep.—testing for independence for pairs of variables ($H_0: \ \tau = 0$).

Table 9. Dependence structure for DJI and ESG indices, 3 June 2014–31 January 2017.

DJI and ...	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Copula	**Clayton–Gumbel**	**t**	**t**	**t**	**Clayton–Gumbel**
Par	1.1266	0.5777	0.4481	0.5130	0.7433
Par2	3.5402	4.0794	4.1390	4.1347	1.5194
Beta	0.8126				0.5093
λ_L	0.8405	0.2954	0.2190	0.2535	0.5413
λ_U	0.7837	0.2954	0.2190	0.2535	0.4219
logLik	868.61	153.33	93.46	120.53	267.18
AIC	−1733.22	−302.65	−182.93	−237.06	−530.36
Indep. (*p*-value)	0.0000	0.0000	0.0000	0.0000	0.0000
GoF test	0.0248	2.8387	7.2067	5.637	0.0421
p-value	0.625	0.265	0.025	0.075	0.675

Indep.—testing for independence for pairs of variables ($H_0: \ \tau = 0$).

3.5. The Stabilization Period (4 January 2017–31 December 2019)

In the stabilization period, all indices are not normally distributed (mostly low kurtosis comparing to previous periods, but higher than for normal distribution, negative asymmetry, as confirmed by the Jarque–Bera test) with positive means and similar standard deviations for S&P 500, DJI, and A1SGI. In the case of standard deviation for other ESG indices, large differences were observed—low for SGESGSEP and high for SEESGSEP (Table 10). In the whole period, two subperiods characterized by high volatility (visible volatility clustering in beginning of 2018 and the beginning of 2019—Figure 7) were observed for S&P 500, DJI, and A1SGI. For other ESG indices not only these clusters were observed (high volatility was during the whole of 2018 and the first half of 2019). Volatility of three indices, namely S&P 500, DJI, and A1SGI, was modeled by AVGARCH (innovations with skewed Student's t and skewed GED distributions). There is no consensus in the modeling volatility of the ESG indices. There were applied GJR-GARCH and EGARCH (innovations with Student's t and skewed normal distributions). From Table A5, we can see that all the parameters have very small *p*-values, which shows their statistical significance. Persistence ranges from 0.9581 to 0.9564 for S&P 500 and DJI, and from 0.9169 to 0.9628 for the ESG indices.

Table 10. Descriptive statistics, 4 January 2017–31 December 2019.

	GSPC	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Minimum	−4.184	−4.714	−4.169	−2.765	−3.297	−2.515	−3.333
Maximum	5.693	5.996	5.563	2.687	3.831	2.477	2.555
1. Quartile	−0.225	−0.266	−0.251	−0.368	−0.436	−0.369	−0.344
3. Quartile	0.451	0.443	0.449	0.454	0.459	0.412	0.448
Mean	0.049	0.049	0.049	0.030	0.011	0.015	0.035
Median	0.069	0.075	0.066	0.046	0.025	0.030	0.049
SE Mean	0.030	0.031	0.029	0.025	0.027	0.023	0.025
LCL Mean	−0.009	−0.011	−0.009	−0.019	−0.042	−0.031	−0.014
UCL Mean	0.108	0.109	0.106	0.078	0.065	0.061	0.084
Variance	0.663	0.702	0.648	0.450	0.554	0.407	0.465
Stdev	0.815	0.838	0.805	0.671	0.744	0.638	0.682
Skewness	−0.502	−0.508	−0.485	−0.273	−0.131	−0.237	−0.420
Kurtosis (−3)	6.620	7.227	6.444	1.214	2.059	1.374	1.696

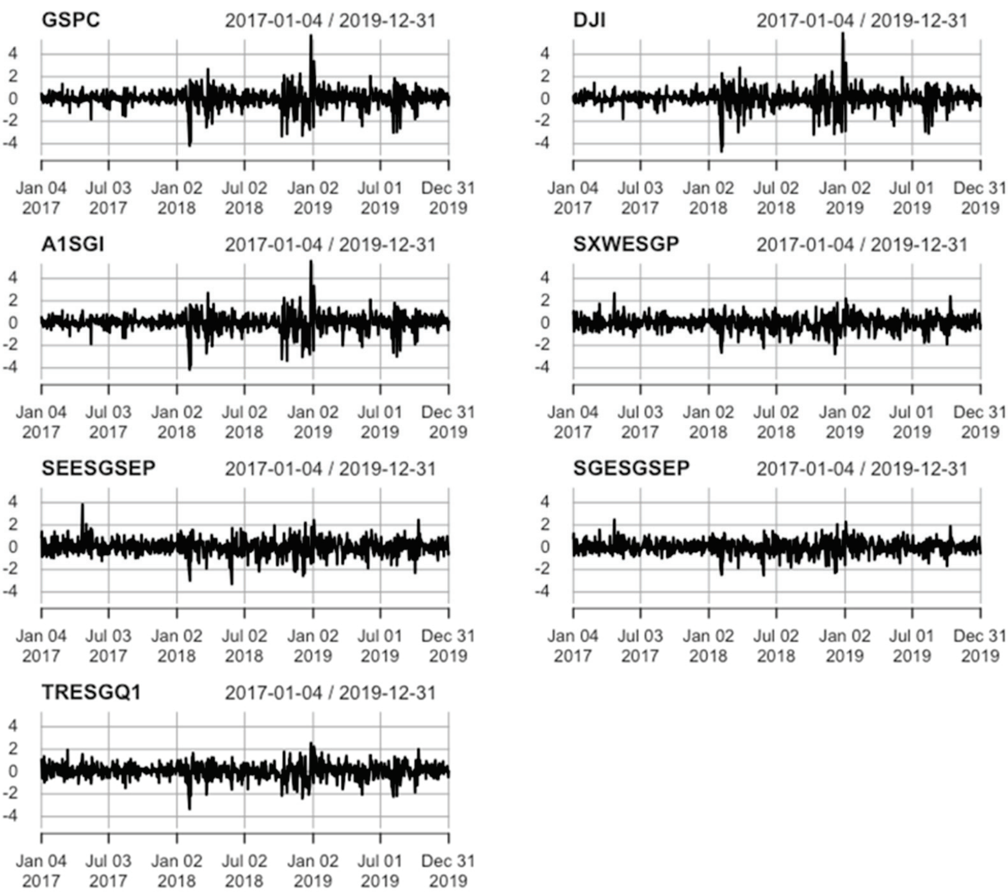


Figure 7. The rates of return, 4 January 2017–31 December 2019.

The highest values of Kendall's τ (see Figure 8) are between indices of the same type (S&P 500 and DJI, and SEESGSEP and SGESGSEP). High dependence is observed between S&P 500 and A1SGI and between DJI and A1SGI, low dependence between S&P 500 and SEESGSEP, and between DJI and SEESGSEP. For example, for the GSPC–SEESGSEP relationship—the observed copula has the lowest dependency—depends by 16.50% on the upper tail and 16.50% on the lower tail (the same dependency). The pair GSPC–A1SGI has a dependency of 88.22% on the lower tail, and of 85.32% on the upper tail. It means the interaction has a greater effect in the lower tail. This interaction was also observed in the pairs DJI–A1SGI, DJI–TRESGQ1, DJI–SXWESGP, and GSPC–TRESGQ1.

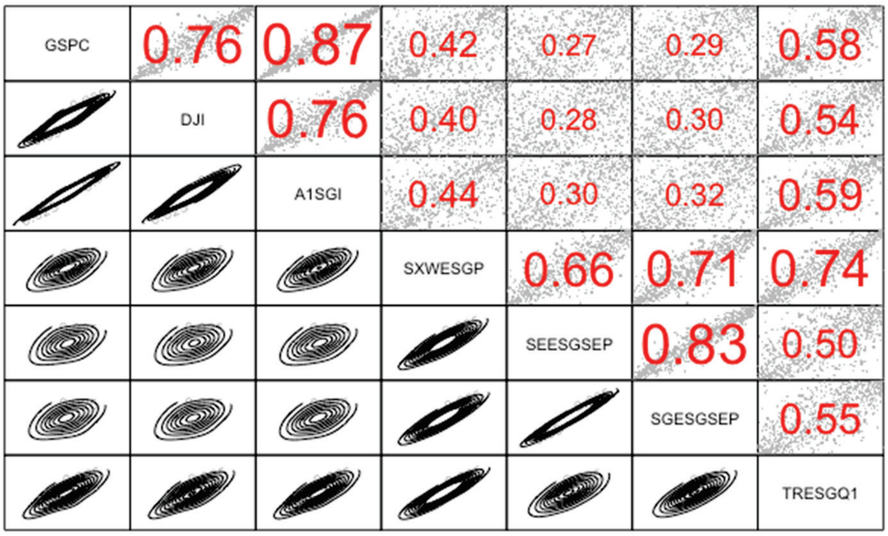


Figure 8. Kendall's τ and copulas, 4 January 2017–31 December 2019.

The dependences between the S&P 500 and ESG indices and between the DJI and ESG indices were modeled by the Clayton–Gumbel, t, and Joe–Clayton copula (see Tables 11 and 12). Dependences between S&P 500 and SEESGSEP, and between S&P 500 and SGESGSEP compared to DJI were modeled by using the t copula. In the other cases, we used other copulas. For example, the GSPC-TRESGQ1 relationship is modeled by the Clayton–Gumbel copula, and DJI–SGESGSEP by the Survival Joe–Clayton copula.

Table 11. Dependence structure for the GSPC and ESG indices, 4 January 2017–31 December 2019.

GSPC and ...	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Copula	**Joe–Clayton**	**Clayton–Gumbel**	**t**	**t**	**t**	**Clayton–Gumbel**
Par	3.9223	1.0932	0.6060	0.4171	0.4443	0.8964
Par2	5.3912	5.0599	4.9383	5.0582	4.9680	1.6639
Beta	0.7145	0.8672				0.5745
λ_L	0.8794	0.8822	0.2733	0.1650	0.1807	0.6283
λ_U	0.8067	0.8532	0.2733	0.1650	0.1807	0.4832
logLik	847.98	1231.71	175.35	80.16	91.18	391.80
AIC	−1691.97	−2459.43	−346.69	−156.33	−178.37	−779.59
Indep. (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
GoF test	0.0167	0.0438	2.4133	0.2038	0.2798	0.1366
p-value	0.96	0.02	0.274	0.926	0.882	0.012

Indep.—testing for independence for pairs of variables ($H_0: \ \tau = 0$).

Table 12. Dependence structure for the DJI and ESG indices, 4 January 2017–31 December 2019.

DJI and ...	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Copula	**Joe–Clayton**	**Survival Clayton–Gumbel**	**t**	**t**	**Survival Joe–Clayton**
Par	3.8520	0.2404	0.4205	0.4489	2.3596
Par2	5.2688	1.4921	5.9725	5.4041	1.0007
Beta	0.7115	0.3935			0.5121
λ_L	0.8767	0.4087	0.1357	0.1665	0.6586
λ_U	0.8028	0.1448	0.1357	0.1665	0.5002
logLik	839.08	167.21	80.10	92.97	353.55
AIC	−1674.17	−330.42	−156.20	−181.93	−703.09
Indep. (p-value)	0.0000	0.0000	0.0000	0.0000	0.0000
GoF test	0.0297	0.0558	1.8649	1.0997	0.0733
p-value	0.685	0.485	0.43	0.525	0.3

Indep.—testing for independence for pairs of variables ($H_0: \tau = 0$).

4. Discussion

Our study does not confirm the outperformance of the ESG indices compared to conventional ones in terms of risk in the considered subperiods. However, generalization of the results is limited by the selection of a market index as a proxy of an investment portfolio. Other empirical studies demonstrate a strong correlation between the lower risk related to sustainability and better financial performance (Whelan et al. 2021). During the 2008 global financial crisis Fernández et al. (2019) found that German green mutual funds had risk-adjusted returns slightly better than their peers (in the noncrisis period, they were equal to conventional funds but better than the SRI funds). Similarly, ESG stock indices performed better and recovered faster after the 2008 global financial crisis (Wu et al. 2017). Other results confirm these findings, as in economic downturns, the high-rated mutual funds outperformed the low-rated funds, based on the Sharpe ratio (Das et al. 2018a, 2018b; Khajenouri and Schmidt 2020). In line are research by Abate et al. (2021) that mutual funds investing in high ESG stocks perform better than investing in low ESG score stocks. Gil-Bazo et al. (2010) confirm that SRI funds perform better than their conventional counterparts.

In the study we evaluate whether conventional stock portfolios, including ESG companies, can effectively decrease portfolio risk, especially in times of financial distress on the market. Generally, we observe in all subperiods almost-weak to high lower-tail dependence between the ESG and conventional indices. Our findings indicate that there is low or symmetric tail dependence between ESG and conventional indices, meaning that if the ESG index decreases, the conventional one will also decrease accordingly. In the two first subperiods (economic downturn periods) lower tail dependence coefficients are higher comparing to the next two periods ('stabilization' periods). We conclude that dependencies exist, can be quantified, and are not negligible, especially in times of crisis.

In risk management of an asset portfolio, we are interested whether the decline of one (or more) assets may influence the behavior of the other assets in a portfolio. Especially, the occurrence of simultaneous extreme events on the market implies that risk diversification breaks down just when it is crucial. In case of extreme events, classical dependence analysis (e.g., linear correlation) fails, a copula approach is used. Some researchers argue that considering ESG practices when creating an equity portfolio (selecting companies with high ESG scores) can act as protection against left-tail risk; therefore, reducing ex-ante expectations of a left-tail event (Shafer and Szado 2020; De and Clayman 2015; Djoutsa Wamba et al. 2020).

The measurement of the tail dependence between ESG and conventional investment based on the copula approach, allows to monitor extreme risks between them. Understanding dependencies and risks is important for setting up adequate risk management as well as construction portfolios—ESG-diversified and resilient to crises. Bax et al. (2021) use the R-vine copula ESG risk model. They estimate all the conditional dependencies among assets as well as specify their interactions as modeled by different copulas families, but they also introduce three ESG risk measures that capture ESG risk, market risk conditionally on the ESG class, as well as an idiosyncratic risk component.

5. Conclusions

As interest in ESG investments grows, it is crucial to better understand the various risks and return tradeoffs between ESG and conventional stocks and the dependence structures between them. We used GARCH family models to estimate conditional volatilities and the copula approach, in particular the (tail) dependence structure between ESG and conventional investments to quantify the overall and lower tail risk between these two investments.

Hypothesis 1, that ESG investments outperform conventional ones in terms of risk was negatively verified—there are indices that underperform the conventional indices in selected subperiods. Hypothesis 2, that asymptotic dependence increases during the crisis on the market (declines of stock market indices), and it stabilizes during non-crisis periods was positively verified.

The findings indicate that there is no significant difference in daily returns of ESG indices and conventional ones.

The parameters are significant among all the estimations and comparisons, showing that GARCH family models may appropriately model the ESG and conventional index data. In most subperiods, the data for the ESG and conventional indices were negatively skewed. The EGARCH models this type of behavior, but in this study, also AVGARCH model was applied. Surprisingly, GJR-GARCH was not often used.

We found relations between the selected ESG indices. A1GSI is related in construction to S&P 500, while three other ESG indices—SXWESGP, SEESGSEP, and SGESGSEP are related to each other. These relations were visible in similar behavior of the rates of return. Volatility clustering observed at S&P 500, DJI, and A1SGI was different from that in other ESG indices (but the scale of these differences depends on the subperiod). In the first subperiod, one cluster was observed for all the indices (there is no difference visible). In the second and third subperiod, two clusters were found for all indices, but the difference between these two groups of indices was present. The most visible differences were observed in the last subperiod. Only in the second period, results of model estimation are consistent for all the indices (ARMA-EGARCH models with innovations sged and sstd).

We cannot confirm that the volatility of conventional indices should be modeled using different models than the volatility of the ESG—this depends on the time period and the market events. There are periods when volatility of all the indices may be modeled by using the same models, when there is a difference between the modeling of conventional and ESG indices, and when there is a difference between the modeling of ESG indices.

The characteristics of the time series indicated the need to apply to the dependence analysis copula approach. To capture tail dependence, Student's t-copula and the combination of the Clayton and Gumbel copulas were fitted best in this study. The choice of an appropriate copula function is crucial. Two features are important regarding the copula selection. The general structure of the chosen copula should coincide with the dependence structure of the real data. If the data show tail dependence than we must apply a copula which comprises tail dependence.

In the periods of economic downturn (declines of stock market indices), the dependencies measured by Kendell's tau coefficients were higher than in the less turbulent periods.

The lower tail dependence and symmetric dependence between ESG and conventional investments were detected. High values of low tail-dependence coefficients were observed in the economic downturn periods; low in stabilization periods. This signifies higher dependence of extreme values in the economic downturn periods and low dependence of extreme values in stabilization periods.

We conclude that when selecting the right model, the preliminary analysis of data is necessary, and the selection of the volatility model should be carried out for subperiods regarding different market-event characteristics. Results show, as in cases of systemic risk, that analysis of volatility and dependence structure should be carried out separately—in the periods of economic downturn and in less turbulent times.

To extend this research for the future, a more detailed analysis, including ESG and non-ESG companies from different markets (developed and developing) would be beneficial. In addition, application of a time-varying model (time-varying copula) would give insights into dependence structure.

Author Contributions: Conceptualization, K.K.; methodology, K.K. and J.G.; software, J.G.; formal analysis, K.K. and J.G.; investigation, K.K. and J.G.; data curation, J.G.; writing—original draft preparation, K.K. and J.G.; writing—review and editing, K.K. and J.G.; visualization, K.K. and J.G.; supervision, K.K.; funding acquisition, K.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Wroclaw University of Economics and Business.

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

Appendix A

Table A1. Description of ESG indices.

Name	Ticker	Description	Market
STOXX GLOBAL ESG LEADERS Index	.SXWESGP	The index offers a representation of the leading global companies in terms of environmental, Social, and governance criteria, based on ESG indicators provided by Sustainalytics	Global
STOXX Europe ESG Leaders Select 30 Price EUR Index	.SEESGSEP	The index captures the performance of stocks with low volatility and high dividends from the STOXX Global ESG Leaders Index.	Europe
STOXX Global ESG Leaders Select 50 Price EUR Index	.SGESGSEP	The index captures the performance of stocks with low volatility and high dividends from the STOXX Global ESG Leaders Index. The component selection process first excludes all stocks whose 3- or 12-month historical volatilities are the highest. Among the remaining stocks, the 50 stocks with the highest 12-month historical dividend yields are selected to be included in the index.	Global
Refinitiv IX Global ESG Equal Weighted Price Only	.TRESGQ1	The index is a benchmark for investors seeking companies that actively invest in and promote ESG values and principles. The index tracks the price return and net total return of publicly traded equities across the world that display relatively high ESG. The constituents' universe is derived from Refinitiv Global Developed Index (the parent index).	Global
Dow Jones Sustainability North America Composite Index	.A1SGI	The index comprises North American sustainability leaders as identified by S&P Global through the Corporate Sustainability Assessment (CSA). It represents the top 20% of the largest 600 North American companies in the S&P Global BMI based on long-term ESG criteria.	North America

Source: own elaboration based on particular indices' websites.

Appendix B

Table A2. Family ARMA-GARCH models, 3 July 2007–30 January 2009.

	GSPC	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Model	**APARCH**	**EGARCH**	**GJR GARCH**	**AVGARCH**	**AVGARCH**	**AVGARCH**	**AVGARCH**
Distribution	**norm**	**sstd**	**norm**	**norm**	**norm**	**norm**	**norm**
ϕ_1	−1.8549 0.0000	−1.8857 0.0000	−1.8615 0.0000		−0.2658 0.0000		
ϕ_2	−0.9319 0.0000	−0.9614 0.0000	−0.8876 0.0000		−1.0050 0.0000		
θ_1	1.7806 0.0000	1.8063 0.0000	1.6787 0.0000		0.2702 0.0000		0.1156 0.0167
θ_2	0.6639 0.0000	0.7189 0.0000	0.3917 0.0000		1.0118 0.0000		−0.0562 0.0285
θ_3	−0.2059 0.0000	−0.1801 0.0000	−0.3394 0.0000				
ω	0.0512 0.0000	0.0001 0.9834	0.0555 0.0000	0.0370 0.0000	0.0454 0.0000	0.0382 0.0000	0.0312 0.0000
α_1	0.0000 1.0000	−0.1579 0.0000	0.0000 1.0000	0.0926 0.0000	0.1398 0.0000	0.1164 0.0000	0.0753 0.0000
α_2	0.1068 0.0000						
β_1	0.8751 0.0000	0.9818 0.0000	0.8774 0.0000	0.8314 0.0000	0.7855 0.0000	0.8196 0.0000	0.8354 0.0000
γ_1	−0.4094 0.9532	0.1037 0.0000	0.1778 0.0001				
γ_2	1.0000 0.0000						
η_{11}				0.4877 0.0000	0.4084 0.0000	0.5056 0.0000	0.6569 0.0000
η_{21}				0.8895 0.0000	0.7126 0.0000	0.6472 0.0000	0.9940 0.0000
δ	1.0751 0.0000						
skew		0.8852 0.0000					
shape		11.4791 0.0642					
Akaike	3.8296	3.7112	3.6759	3.7624	3.6824	3.6065	3.5371
Bayes	3.9521	3.8235	3.7678	3.8135	3.7743	3.6576	3.6085
Ljung–Box test	4.4482 0.9249	6.3567 0.7845	9.9787 0.4424	6.9881 0.7266	5.9840 0.8166	6.6419 0.7588	7.7529 0.6530
Engle Arch test	8.6312 0.7341	15.2820 0.2264	12.3802 0.4156	21.5479 0.0429	16.7602 0.1588	30.0986 0.0027	24.3912 0.0180
Persistence	0.9654	0.9818	0.9663	0.9730	0.9649	0.9693	0.9721

Note: the first row indicates the estimate parameters (or test statistics) and the second row—the *p*-value of the Student's *t*-test (or appropriate test); Ljung–Box and Engle ARCH tests were calculated for standardized innovations.

Table A3. ARMA-EGARCH models, 2 December 2009–30 July 2014.

	GSPC	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Distribution	**sged**	**sged**	**sstd**	**sstd**	**sged**	**sstd**	**sstd**
ϕ_1	−0.5969 0.0000	−0.2909 0.0000		0.0773 0.0000			
θ_1	0.5690 0.0000	0.2501 0.0000					
ω	−0.0042 0.4793	−0.0162 0.0005	−0.0081 0.1138	0.0027 0.1199	0.0023 0.2122	−0.0017 0.6040	0.0024 0.3074
α_1	−0.3983 0.0000	−0.3570 0.0000	−0.3832 0.0000	−0.1023 0.0000	−0.0918 0.0000	−0.1150 0.0000	−0.1117 0.0000
α_2	0.1515 0.0000	0.1371 0.0049	0.1687 0.0000				
β_1	0.9410 0.0000	0.9411 0.0000	0.9594 0.0000	0.9910 0.0000	0.9867 0.0000	0.9816 0.0000	0.9867 0.0000
γ_1	−0.1456 0.0323	−0.0852 0.0003	−0.1791 0.0126	0.0668 0.0000	0.0672 0.0000	0.1182 0.0002	0.0808 0.0019
γ_2	0.2681 0.0001	0.2472 0.0003	0.2805 0.0001				
skew	0.8098 0.0000	0.8356 0.0000	0.7841 0.0000	0.8657 0.0000	0.8563 0.0000	0.8689 0.0000	0.8251 0.0000
shape	1.3919 0.0000	1.4247 0.0000	7.7845 0.0000	8.1811 0.0000	1.5014 0.0000	8.8081 0.0001	6.7950 0.0000
Akaike	2.4940	2.3325	2.3731	2.9187	2.8993	2.5299	2.6496
Bayes	2.5382	2.3767	2.4084	2.9496	2.9258	2.5564	2.6762
Ljung–Box test	7.6686 0.6612	4.6382 0.9140	6.8038 0.7438	5.5808 0.8492	4.8186 0.9030	6.0897 0.8077	7.1695 0.7094
Engle ARCH test	7.8238 0.7987	6.3159 0.8993	7.2004 0.8441	8.4786 0.7467	13.9807 0.3019	17.9072 0.1185	10.1396 0.6037
Persistence	0.9410	0.9411	0.9594	0.9910	0.9867	0.9816	0.9867

Note: the first row indicates the estimate parameters (or test statistics) and the second row—the p-value of the Student's t-test (or appropriate test); Ljung–Box and Engle ARCH tests were calculated for standardized innovations.

Table A4. Family GARCH models, 3 June 2014–31 January 2017.

	GSPC	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Model	**AVGARCH**	**AVGARCH**	**AVGARCH**	**EGARCH**	**EGARCH**	**GARCH**	**AVGARCH**
Distribution	sstd	sstd	sstd	std	std	std	std
ω	0.0361 0.0000	0.0285 0.0000	0.0302 0.0000	−0.0215 0.0560	−0.0147 0.2977	0.0723 0.0052	0.0326 0.0007
α_1	0.2033 0.0000	0.2302 0.0000	0.2782 0.0000	−0.1276 0.0001	−0.1335 0.0007	0.1981 0.0000	0.1807 0.0000
β_1	0.6804 0.0000	0.7186 0.0000	0.7042 0.0000	0.9419 0.0000	0.8972 0.0000	0.7133 0.0000	0.7933 0.0000
γ_1				0.2418 0.0002	0.2775 0.0003		
η_{11}	0.1086 0.1826	−0.1145 0.0000	−0.2498 0.0000				−0.1651 0.0450
η_{21}	1.1870 0.0000	1.1419 0.0000	1.1495 0.0000				0.9274 0.0000

Table A4. *Cont.*

	GSPC	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Model	**AVGARCH**	**AVGARCH**	**AVGARCH**	**EGARCH**	**EGARCH**	**GARCH**	**AVGARCH**
skew	0.8196 0.0000	0.8599 0.0000	0.8440 0.0000				
shape	11.2696 0.0084	7.7111 0.0001	11.4148 0.0037	10.6788 0.0022	8.5965 0.0003	9.6351 0.0018	7.6282 0.0001
Akaike	2.1554	2.1590	2.1985	2.5601	2.7756	2.4024	2.3052
Bayes	2.2029	2.2065	2.2460	2.5940	2.8096	2.4296	2.3458
Ljung–Box test	15.6265 0.1108	12.3524 0.2622	13.1544 0.2152	11.6801 0.3070	7.0749 0.7184	13.3207 0.2063	11.0541 0.3533
Engle ARCH test	5.3622 0.9448	4.8535 0.9627	8.7917 0.7206	9.6174 0.6495	9.4367 0.6653	12.2077 0.4291	12.4372 0.4112
Persistence	0.9655	0.9749	0.9716	0.9419	0.8972	0.9114	0.9668

Note: the first row indicates the estimate parameters (or test statistics) and the second row—the *p*-value of the Student's *t*-test (or appropriate test); Ljung–Box and Engle ARCH tests were calculated for standardized innovations.

Table A5. Family ARMA-GARCH models, 4 January 2017–31 December 2019.

	GSPC	DJI	A1SGI	SXWESGP	SEESGSEP	SGESGSEP	TRESGQ1
Model	**AVGARCH**	**AVGARCH**	**AVGARCH**	**EGARCH**	**EGARCH**	**GJR GARCH**	**EGARCH**
Distribution	**sged**	**sstd**	**sged**	**std**	**std**	**std**	**snorm**
θ_1				0.0766 0.0462			
ω	0.0392 0.0000	0.0406 0.0000	0.0338 0.0000	−0.0707 0.0141	−0.0417 0.0407	0.0208 0.0207	−0.0443 0.0021
α_1	0.1456 0.0000	0.1065 0.0000	0.1321 0.0000	−0.1286 0.0000	−0.0982 0.0036	0.0067 0.7335	−0.1526 0.0000
β_1	0.7806 0.0000	0.8708 0.0000	0.7995 0.0000	0.9169 0.0000	0.9394 0.0000	0.8845 0.0000	0.9402 0.0000
γ_1				0.1402 0.0077	0.1375 0.0009	0.1149 0.0078	0.1793 0.0000
η_{11}	0.5809 0.0000	1.0000 0.0001	0.4573 0.0000				
η_{21}	0.5806 0.0000	0.0786 0.0695	0.6727 0.0000				
skew	0.8846 0.0000	0.8334 0.0000	0.8564 0.0000				0.8620 0.0000
shape	1.2853 0.0000	4.6528 0.0000	1.3843 0.0000	14.0514 0.0145	8.7823 0.0002	12.4842 0.0122	
Akaike	1.8836	1.9762	1.9035	1.9509	2.1482	1.8526	1.9036
Bayes	1.9269	2.0195	1.9468	1.9880	2.1791	1.8835	1.9345
Ljung–Box test	8.6446 0.5661	7.6484 0.6631	9.9759 0.4426	9.0690 0.5256	12.1615 0.2744	9.3624 0.4981	10.8849 0.3666
Engle ARCH test	9.3952 0.6689	12.5922 0.3994	10.9797 0.5307	10.0092 0.6151	20.3661 0.0605	14.3953 0.2762	8.8700 0.7140
Persistence	0.9581	0.9564	0.9628	0.9169	0.9394	0.9487	0.9402

Note: the first row indicates the estimate parameters (or test statistics) and the second row—the *p*-value of the Student's *t*-test (or appropriate test); Ljung–Box and Engle ARCH tests were calculated for standardized innovations.

References

Abate, Guido, Ignazio Basile, and Pierpaolo Ferrari. 2021. The level of sustainability and mutual fund performance in europe: An empirical analysis using ESG ratings. *Corporate Social Responsibility and Environmental Management* 28: 1446–55. [CrossRef]

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

Apergis, Nicholas, Vassilios Babalos, Christina Christou, and Rangan Gupta. 2015. *Identifying Asymmetries between Socially Responsible and Conventional Investments*. Working Papers: 2015-37. Pretoria: Department of Economics, University of Pretoria.

Bax, Karoline, Özge Sahin, Claudia Czado, and Sandra Paterlini. 2021. ESG, Risk, and (Tail) Dependence. *SSRN Electronic Journal*. [CrossRef]

Bollerslev, Tim. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* 31: 307–27. [CrossRef]

Box, George E. P., and Gwilym M. Jenkins. 1983. *Analiza Szeregów Czasowych*. Warszawa: Wydawnictwo PWN.

Brockwell, Peter J., and Richard A. Davis. 1991. *Time Series: Theory and Methods*. New York: Springer. [CrossRef]

Charles, Amélie, Olivier Darné, and Jessica Fouilloux. 2016. The Impact of Screening Strategies on the Performance of ESG Indices. July 12. Available online: https://hal.archives-ouvertes.fr/hal-01344699 (accessed on 1 October 2021).

Das, Nandita, Bernadette Ruf, Swarn Chatterjee, and Aman Sunder. 2018a. Fund characteristics and performances of socially responsible mutual funds: Do ESG ratings play a role? *Journal of Accounting and Finance* 18: 57–69. [CrossRef]

Das, Nandita, Swarn Chatterje, Bernadette Ruf, and Aman Sunder. 2018b. ESG ratings and the performance of socially responsible mutual funds: A panel study journal of finance issues esg ratings and the performance of socially responsible mutual funds: A panel study. *Journal of Finance Issues* 17: 49–57.

De, Indrani, and Michelle R. Clayman. 2015. The benefits of socially responsible investing: An active manager's perspective. *The Journal of Investing* 24: 49–72. [CrossRef]

Ding, Zhuanxin, Clive William John Granger, and Robert F. Engle. 1993. A long memory property of stock market returns and a new model. *Journal of Empirical Finance* 1: 83–106. [CrossRef]

Djoutsa Wamba, Léopold, Jean Michel Sahut, Eric Braune, and Frédéric Teulon. 2020. Does the optimization of a company's environmental performance reduce its systematic risk? New evidence from European listed companies. *Corporate Social Responsibility and Environmental Management* 27: 1677–94. [CrossRef]

Engle, Robert F. 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50: 987–1007. [CrossRef]

Engle, Robert F., and Victor K. Ng. 1993. Measuring and testing the impact of news on volatility. *The Journal of Finance* 48: 1749–78. [CrossRef]

Fama, Eugene F., and Eugene F. Fama. 1965. The behaviour of stock market prices. *Journal of Business* 38: 34–105. [CrossRef]

Fernández, Manuel Salazar, Ahmad Abu-Alkheil, and Ghadeer M. Khartabiel. 2019. Do German green mutual funds perform better than their peers. *Business and Economics Research Journal* 10: 297–312. [CrossRef]

Friede, Gunnar. 2019. Why don't we see more action? A metasynthesis of the investor impediments to integrate environmental, social, and governance factors. *Business Strategy and the Environment* 28: 1260–82. [CrossRef]

Genest, Christian, Jean François Quessy, and Bruno Rémillard. 2006. Goodness-of-Fit procedures for copula models based on the probability integral transformation. *Scandinavian Journal of Statistics* 33: 337–66. [CrossRef]

Giese, Guido, and Linda-Eling Lee. 2019. Weighing the Evidence: ESG and Equity Returns. *MSCI Research Insight*. Available online: https://www.msci.com/www/research-paper/weighing-the-evidence-esg-and/01315636760 (accessed on 24 October 2021).

Gil-Bazo, Javier, Pablo Ruiz-Verdú, and André A. P. Santos. 2010. The performance of socially responsible mutual funds: The role of fees and management companies. *Journal of Business Ethics* 94: 243–63. [CrossRef]

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

HKEX. 2020. Performance of ESG Equity Indices Versus Traditional Equity Indices. Available online: https://www.hkex.com.hk/-/media/HKEX-Market/News/Research-Reports/HKEx-Research-Papers/2020/CCEO_ESGEqIdx_202011_e.pdf (accessed on 24 October 2021).

Jain, Mansi, Gagan Deep Sharma, and Mrinalini Srivastava. 2019. Can sustainable investment yield better financial returns: A comparative study of ESG indices and MSCI indices. *Risks* 7: 15. [CrossRef]

Jondeau, Eric. 2016. Asymmetry in Tail Dependence in Equity Portfolios. *Computational Statistics and Data Analysis* 100: 351–68. [CrossRef]

Khajenouri, Daniel, and Jacob Schmidt. 2020. Standard or sustainable—Which offers better performance for the passive investor. *Journal of Applied Finance & Banking* 11: 61–71. [CrossRef]

Longin, Franqois, and Bruno Solnik. 2001. Extreme correlation of international equity markets. *The Journal of Finance* LVI: 649–76. [CrossRef]

Managi, Shunsuke, Tatsuyoshi Okimoto, and Akimi Matsuda. 2012. Do socially responsible investment indexes outperform conventional indexes? *Applied Financial Economics* 22: 1511–27. [CrossRef]

Mandelbrot, Benoit. 1963. The Variation of Certain Speculative Prices. *The Journal of Business* 36: 394. [CrossRef]

Markowitz, Harry. 1952. Portfolio Selection. *Journal of Finance* 7: 77–91.

Messaoud, Samia Ben, and Chaker Aloui. 2015. Measuring risk of Portfolio: GARCH-Copula model. *Journal of Economic Integration* 30: 172–205. [CrossRef]

Nelsen, Roger B. 2006. *An introduction to Copulas*. New York: Springer. [CrossRef]

Nelson, Daniel B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica* 59: 347. [CrossRef]

Ortas, Eduardo, José M. Moneva, Roger Burritt, and Joanne Tingey-Holyoak. 2014. Does sustainability investment provide adaptive resilience to ethical investors? evidence from Spain. *Journal of Business Ethics* 124: 297–309. [CrossRef]

Ouchen, Abdessamad. 2021. Is the ESG Portfolio Less Turbulent than a Market Benchmark Portfolio? *Risk Management*, 1–33. [CrossRef]

Patton, Andrew J. 2006. *Copula-Based Models for Financial Times Series*. Working Papers: OMI11/07. Oxford: The Oxford-Man Institute, University of Oxford.

Pedersen, Lasse Heje, Shaun Fitzgibbons, and Lukasz Pomorski. 2021. Responsible investing: The ESG-efficient frontier. *Journal of Financial Economics* 142: 572–97. [CrossRef]

Plastun, Alex, Elie Bouri, Rangan Gupta, and Qiang Ji. 2022. Price Effects after one-day abnormal returns in developed and emerging markets: ESG versus traditional indices. *The North American Journal of Economics and Finance* 59: 101572. [CrossRef]

Rydberg, Tina Hviid. 2000. Realistic statistical modelling of financial data. *International Statistical Review/Revue Internationale de Statistique* 68: 233. [CrossRef]

Schwert, G. William. 1990. Stock Volatility and the crash of '87. *Review of Financial Studies* 3: 77–102. [CrossRef]

Shafer, Michael, and Edward Szado. 2020. Environmental, social, and governance practices and perceived tail risk. *Accounting and Finance* 60: 4195–224. [CrossRef]

Sklar, A. 1959. Fonctions de Répartition an Dimensions et Leurs Marges. *Publications de l'Institut de statistique de l'Université de Paris* 8: 229–31.

Stoeber, Jakob, Eike Christian, Benedikt Graeler, Thomas Nagler, Tobias Erhardt, Carlos Almeida, Aleksey Min, Claudia Czado, Mathias Hofmann, Matthias Killiches, and et al. 2018. Package 'VineCopula'. Available online: https://cran.microsoft.com/snapshot/2018-07-26/web/packages/VineCopula/VineCopula.pdf (accessed on 9 August 2020).

Taylor, Stephen J. 1986. *Modelling Financial Time Series: 2nd Edition. Modelling Financial Time Series*. Chichester: John Wiley & Sons.

US SIF. 2020. 2020 Report on US Sustainable, Responsible and Impact Investing Trends. Available online: https://www.ussif.org/currentandpast (accessed on 24 October 2021).

Verheyden, Tim, Robert G. Eccles, and Andreas Feiner. 2016. ESG for all? the impact of ESG screening on return, risk, and diversification. *Journal of Applied Corporate Finance* 28: 47–55. [CrossRef]

Wang, Weijing, and Martin T. Wells. 2000. Model Selection and Semiparametric Inference for Bivariate Failure-Time Data. *Journal of the American Statistical Association* 95: 62–72. [CrossRef]

Weber, Olaf, and Wei Rong Ang. 2016. The performance, volatility, persistence and downside risk characteristics of sustainable investments in emerging market. *ACRN Oxford Journal of Finance and Risk Perspectives* 5: 1–12.

Whelan, Tensie, Ulrich Atz, Tracy Van Holt, and Casey Clark. 2021. ESG and Financial Performance: Uncovering the Relationship by Aggregating Evidence from 1000 Plus Studies Published between 2015–2020. Available online: https://www.stern.nyu.edu/sites/default/files/assets/documents/NYU-RAM_ESG-Paper_2021%20Rev_0.pdf (accessed on 24 October 2021).

Wu, Junjie, George Lodorfos, Aftab Dean, and Georgios Gioulmpaxiotis. 2017. The Market Performance of Socially Responsible Investment during Periods of the Economic Cycle—Illustrated Using the Case of FTSE. *Managerial and Decision Economics* 38: 238–51. [CrossRef]

risks

Article

Exchange Rate Volatility, Currency Misalignment, and Risk of Recession in the Central and Eastern European Countries

Victor Shevchuk [1,*] and Roman Kopych [2]

1 Council of Social Sciences, Cracow University of Technology, 31155 Cracow, Poland
2 Department of International Economic Relations, Lviv National University Named by Ivan Franko, 79000 Lviv, Ukraine; kopych_r@bigmir.net
* Correspondence: vshevchuk@pk.edu.pl

Abstract: This study is aimed at estimation of the exchange rate volatility and its impact on the business cycle fluctuations in four central and eastern European countries (the Czech Republic, Hungary, Poland, and Romania). Exchange rate volatility is estimated with the EGARCH(1,1) model. It is found that exchange rate volatility is affected by the components of the Index of Economic Freedom from the Heritage Foundation, besides inflation and crisis developments. The empirical results using GMM estimation technique and comprehensive robustness checks suggest that exchange rate volatility reduces the risk of recession in the Czech Republic while the opposite effect is found for Hungary and Romania, with a neutrality for Poland. These findings continue to hold after controlling for the fiscal and monetary policy indicators. There is evidence that the RER undervaluation prevents sliding into a recession on a credible basis in Poland only, with a neutral stance for other countries. Except in Romania, higher levels of economic freedom is associated with worsening of the cyclical position of output. Among other results, stabilization policies in the recession imply fiscal tightening for the Czech Republic and Romania, higher money supply for the Czech Republic and Poland, and lower central bank reference rate for Hungary.

Keywords: exchange rate volatility; currency misalignment; business cycle; central and eastern European countries

Citation: Shevchuk, Victor, and Roman Kopych. 2021. Exchange Rate Volatility, Currency Misalignment, and Risk of Recession in the Central and Eastern European Countries. *Risks* 9: 82. https://doi.org/10.3390/risks9050082

Academic Editor: Józef Dziechciarz

Received: 30 March 2021
Accepted: 20 April 2021
Published: 1 May 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Introduction

Exchange rate volatility represents short-term fluctuations of nominal (real) exchange rates about their longer-term trends, while currency misalignment refers to a significant deviation of the actual real exchange rate (RER) from its equilibrium level (Frenkel and Goldstein 1991). Generally, variability of RERs reflects variability of nominal exchange rates. Empirical results on the exchange rate volatility effects on economic growth are rather ambiguous, being sensitive to the choice of estimators, control variables, the time period and the sample taken under consideration, country-specific factors or the sample of countries chosen (for panel studies) (Boar 2010).

Some empirical studies do not find any significant link between exchange rate volatility and economic performance (Ghosh et al. 2003). Other studies are in support of causality between exchange rate volatility and economic growth, with both positive and negative effects being observed. Favorable effects of exchange rate flexibility are found by Edwards and Levy-Yeyati (2005). However, evidence of a negative relationship between exchange rate volatility and economic growth is found for developed and industrial countries (Holland et al. 2013; Jamil et al. 2012; Janus and Riera-Crichton 2015; Papadamou et al. 2016), middle-income countries (Aizenman et al. 2018), as well as developing ones (Dollar 1992). Other studies are rather ambiguous. For example, Bleaney and Greenaway (2001) found for 14 sub-Saharan African countries that volatility exerts negative effects on investment but not on economic growth. Another study of 125 countries finds that exchange rate fluctuation has different effects on economic growth in different countries (Han 2020). Schnabl (2009) highlights the negative effect of volatility on economic growth in several European and Asian countries. Similar results are obtained for the CEE countries by Arratibel et al. (2011); Boar (2010); and Morina et al. (2020).

For developing and emerging countries, volatility seems to be more harmful under a flexible exchange rate regime and financial openness (Barguellil et al. 2018). However, Furceri and Zdzienicka (2012) and Tsangarides (2012) found that countries with a more flexible exchange rate regime tend to experience lower production losses during periods of financial crises. As argued by Schnabl (2009), the mixed results on the link between exchange rate volatility and economic growth can be explained by the country-specific factors such as the level of financial markets development, human capital endowments, as well as institutional features.

Most studies employ conditional volatility measures, for example Antonakakis and Darby (2012); Holland et al. (2013); and Jamil et al. (2012), including studies for the central and eastern European (CEE) countries (Firdmuc and Horváth 2007; Miletić 2015). As alternative measures for exchange rate volatility, standard deviations of monthly exchange rate changes or percent exchange rate changes are used as well (Caporale et al. 2011; Janus and Riera-Crichton 2015; Hau 2002; Morina et al. 2020; Schnabl 2009).

Among controlling variables, the real exchange rate (RER) misalignment and institutional features are worth attention. A number of studies indicate that a more depreciated (undervalued) exchange rate is favorable for economic growth (Béreau et al. 2012; Hausmann et al. 2005), especially for developing and emerging countries (MacDonald and Vieira 2010), though opposite results are not lacking as well (Ahmed et al. 2002; Karadam and Ozmen 2016; Morvillier 2020; Ribero et al. 2020). For nine CEE economies, it is found that exchange rate overvaluation has a negative impact on economic growth, with a stronger effect than undervaluation (Cuestas et al. 2019).

Most of the empirical studies provide evidence that institutions matter for economic growth both in general and in a host of interesting specific contexts (Durlauf 2018). Particularly, there is a positive link between economic freedom and economic performance (De Haan and Sturm 2000), including CEE countries (Uzelac et al. 2020), though not all aspects of economic freedom affect economic growth (Justesen 2008).

This study is aimed at estimation of the exchange rate volatility and its impact on the output fluctuations in four CEE countries (the Czech Republic, Hungary, Poland, and Romania). The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) method is used to model exchange rate volatility and the Generalized Method of Moments (GMM) is used to examine the effect of exchange rate volatility on the business cycle. Our main contribution lies in adopting a more comprehensive approach in studying output effects of exchange rate volatility by integrating volatility and currency misalignment with institutional dimensions.

The rest of the paper is organized as follows. Section 2 presents a survey of relevant literature. In Section 3 data and methods are outlined. Section 4 discusses empirical results and Section 5 concludes.

2. Literature Survey

2.1. Exchange Rate Volatility and Economic Growth

Exchange rate volatility has advantages in the case of adjustment to real asymmetric shocks, as there is no need for slow and costly real exchange rate adjustments to be carried out through relative price and productivity changes under price and wage rigidities, as is the case for a fixed exchange rate regime (Edwards and Levy-Yeyati 2005). For small countries, pegging of a currency may lead to its overvaluation and currency crises, magnifying the effects of asymmetric shocks (Edwards 2011). Besides absorbing properties or resilience to currency crises, a switch to free or managed floating is motivated by high capital mobility and financial innovations (Boar 2010).

Traditional view of the exchange rate as a shock absorber that substitutes for the lack of nominal price flexibility implies a higher volatility in the face of real shocks, especially foreign ones. As argued by Devereux and Engel (2002), exchange rates may be highly volatile because they reflect monetary factors and have little effect on macroeconomic variability:

$$\Delta e_t = \frac{(1+(\sigma/r))(\Delta m_t - \Delta m_t^*) + (\sigma/r)v_t}{[(\sigma/r)+\rho(\theta-(1-\theta^*))]}, \tag{1}$$

where e_t is the nominal exchange rate measured as the price of foreign currency in terms of domestic currency (a rise in e_t means a nominal depreciation), m_t and m_t^* are domestic and foreign money supply, respectively, v_t measures bias in the conditional forecasts of the future exchange rate, ρ is the discount factor, θ and θ^* are shares of profit due to distribution of domestic and foreign products.

Excess volatility is presented as follows:

$$Var_{t-1}(\Delta e_t) = \frac{(1+(\sigma/r))^2 Var_{t-1}(\Delta m_t - \Delta m_t^*)}{\Phi^2\left[1-[\kappa\sigma/r\Phi]^2\right]},$$

$$\Phi = [(\sigma/r)+\rho(\theta-(1-\theta^*))], \tag{2}$$

where κ is the coefficient of proportionality between the conditional variance of the exchange rate and the conditional variance of v_t. Exchange rate volatility depends only on the volatility in relative money supply and becomes much higher than shocks to economic fundamentals, thus creating "disconnect" from the rest of the economy.

Addressing the latest phenomenon of extremely low short-term interest rates, Corsetti et al. (2017) in a tractable model show that a float has advantages in the case of deflationary demand shocks abroad as it is necessary to offset not only the collapse in external demand but deflation abroad as well. Formally, expression for the equilibrium exchange rate presents as follows:

$$e_t = E_t \sum_{k=0}^{\infty} \left(r_{t+k}^* - r_{t+k} + \pi_{H,t+k} - \pi_{t+k}^*\right) + \left(p_{H,t-1} - p_{t-1}^*\right), \tag{3}$$

where r_t and r_t^*, $\pi_{H,t}$ and π_t^* are domestic and foreign nominal interest rates and price inflation, respectively, $p_{H,t}$ and p_t^* are the price indices of domestically produced goods and consumer prices abroad, respectively, and E_t is the expectations operator. Equation (3) implies that present nominal exchange rate depends on the future gaps in real interest rates and the lagged term of trade. If there is a liquidity trap, the domestic currency depreciates in nominal and real terms thus insulating from domestic economy from foreign deflationary pressure.

While the "exchange rate disconnect" implies that the exchange rate volatility depends only on fundamentals, such as the relative money supply, the interest rate and output growth differentials, or terms of trade, is downplayed by the scapegoat theory of exchange rates. It is argued by Bacchetta and Wincoop (2013) that changes in the exchange rate are dependent not only on some observable fundamentals but on changes in expectations of structural parameters as well:

$$\Delta e_t = \mathbf{f}_{t-i}'((1-\lambda)\boldsymbol{\beta}_t + \lambda E_t \boldsymbol{\beta}_t) + (1-\lambda)b_t + \lambda \sum_{i=1}^{T} \mathbf{f}_{t-i}'(E_t \boldsymbol{\beta}_{t-i} - E_{t-1}\boldsymbol{\beta}_{t-i}), \tag{4}$$

where $\boldsymbol{\beta}_t$ is the vector of time-varying true structural parameters, $E_t\boldsymbol{\beta}_t$ is the vector of expected parameters at time t, b_t is the unobserved fundamental, and λ is the discount factor $(0 < \lambda < 1)$. As investors do not know the value of true structural parameters $\boldsymbol{\beta}_t$ and their time variation in the short- to medium term, it becomes a source of exchange rate volatility. Empirical support for the scapegoat theory is claimed by Fratzscher et al. (2015).

As excessive exchange rate fluctuations can be limited by stabilization policies, Corsetti (2006) demonstrates that the exchange rate volatility implied by optimal stabilization rules is inversely related to the import content of consumption. Although a floating exchange rate is indeed associated with higher volatility for both the nominal and the real exchange rate (Mussa 1986; Petracchi 2021), it cannot be a source of concern. As explained by Krugman (1988), the volatility of exchange rates in combination with the substantial sunk costs associated with entering a foreign market made foreign trade prices and volumes unresponsive to exchange rate fluctuations, with the real sector becoming insensitive to exchange rate fluctuations.

Another argument in favor of greater exchange rate flexibility/volatility is the monetary policy autonomy in the presence of strong international capital mobility that offer the possibility of stabilizing the domestic economy (Dornbusch and Giovannini 1990). As implicit in the *Redux* model, money supply shocks can have real effects even in the long run, but at the cost of higher exchange rate volatility (Obstfeld and Rogoff 1995). It is not a problem if exchange rate volatility is expansionary or at least neutral in respect to output, but it can be a problem otherwise.

A fixed exchange rate regime is not without advantages of its own. For small open economies, stable exchange rates provide better insulation against nominal shocks (McKinnon 1963). If exchange rate stability contributes to macroeconomic stability, it creates a favorable environment for investment, consumption, and growth. Straub and Tchakarov (2004) demonstrate that in a model with habit persistence, even non-fundamental exchange rate volatility that generates only small variation in prices and interest rates might induce economically significant welfare changes. In a theoretical model by Devereux and Lane (2003), external debt reduces the efficiency of the exchange rate in responding to external shocks. Also, exchange rate flexibility can be a source of inefficiency due to the presence of speculative actions (Boar 2010).

As surveyed by Arratibel et al. (2011); Demir (2013); Eichler and Littke (2017); and Juhro and Phan (2018), exchange rate volatility has negative effects on economic growth through numerous channels, including lower efficiency of price mechanisms at international level; higher risk premium; greater uncertainty on future consumption and firm revenues; uncertainty about export revenues; increased volatility of business profitability; higher risk for domestic and foreign direct investment, particularly in developing economies; increased inflation uncertainty and higher interest rates along with reduced investment and consumption; obstacles to consumption risk sharing due to the home-bias in portfolio investment; adverse effect of credit constraints on domestic investments; and changes in production cost and increased international transaction risk, or in the relative costs of production with both creative and destructive growth effects. Although exchange rate flexibility reduces the sensitivity of local to base-country central bank policy rates, this feature is weakened by the balance sheet effect because a depreciation of the local currency would raise the cost of servicing and rolling over foreign-currency debt and bank loans (Georgiadis and Zhu 2019). Also, liability dollarization amplifies a negative impact of exchange rate flexibility on growth (Benhima 2012).

Within the framework of a simple monetary growth model, it is demonstrated by Aghion et al. (2009) that for countries with relatively low levels of financial development, (real) exchange rate uncertainty exacerbates the negative investment effects of domestic credit market constraints. An increase in exchange rate volatility may discourage firms from creating jobs as well (Belke and Setzer 2003). Similar to Demir (2013). It is possible to argue that numerous transmission channels imply that growth effects of exchange rate volatility will ultimately depend on the country.

2.2. Currency Misalignments in Open Economies

Another strand of the literature refers to effects of (real) exchange rate misalignment, especially in the context of the business cycle. Under a fixed exchange rate regime, excessive capital inflows may lead to acceleration of inflation and economy overheating, with a risk of recession and costly stabilization efforts to follow. However, flexible rates quite often are not used to play a stabilizing countercyclical role; instead, central banks engineer systemic overvaluation in the context of inflation targeting (Dornbusch 2001).

On the other hand, the fear of exchange rate appreciation under a floating exchange rate regime may be no less damaging than excessive exchange rate stability. As demonstrated by Caballero and Lorenzoni (2014), persistent RER overvaluation can be harmful for the economies with a financially constrained export sector if it is followed by a large exchange rate overshooting once the factors behind the appreciation subside. Such a situation requires exchange rate interventions. Undervaluation can promote growth by stimulating technological progress and knowledge spillovers, but at the cost of negative distributional effects (Ribero et al. 2020). Although a flexible exchange rate regime allows for maintaining a competitive (undervalued) exchange rate in order to boost exports and growth, this kind of policy is difficult to sustain, except for low-income countries, and only in the medium term (Haddad and Pancaro 2010). Such findings are consistent with a more general view that currency misalignments are inefficient and lead to lower welfare (Engel 2011).

The relationship between exchange rate and economic performance can be dependent on institutional factors. Rodrik (2008) argues that there is a trade-off between a currency undervaluation and quality of institutions: a weaker currency is needed in order to compensate for inefficient institutions. As mentioned by Schnabl (2009), the answer to whether flexible exchange rates would reduce the risk of crisis depends on the central bank's response to appreciation pressure. Only if the central bank allows for "uncontrolled appreciation" of the domestic currency does the probability of crisis decline, because the sharp appreciation of the domestic currency deteriorates the economic outlook. The effect of central bank transparency on exchange rate volatility depends on the development of countries (Weber 2017).

2.3. Sources of Exchange Rate Volatility

The exchange rate volatility tends to be lower in more open economies with higher GDP per capita and lower inflation (Bleaney and Francisco 2010) than developing countries with high levels of external debt (Devereux and Lane 2003). Both free floating and more volatile terms-of-trade contribute to higher volatility. Other sources of (real) exchange rate volatility in both industrial and developing countries include highly volatile productivity shocks, sharp oscillations in monetary and fiscal policy shocks, and financial openness (Calderon and Kubota 2018). The link between liquidity and exchange rate volatility depends on the level of financial development (Pham 2018). The exchange rate volatility can be increased by the foreign exchange interventions by central banks (Frenkel et al. 2005) or global economic uncertainty (Juhro and Phan 2018).

Higher exchange rate volatility in the CEE countries is associated with a more flexible exchange rate regime and use of interest rate as a key instrument of monetary policy (Kočenda and Valachy 2006) or low credibility of exchange rate management (Firdmuc and Horváth 2007). Giannellis and Papadopoulos (2011) found that volatility of Polish and Hungarian currencies is caused by the monetary-side of the economy. As established by Lyócsa et al. (2016), the Czech *koruna* and Polish *zloty* appear to be more vulnerable to local and global spillovers than the Hungarian *forint*. The global financial crisis of 2008–2009 increased volatility of the exchange rate in the Czech Republic, Poland, and Romania (Miletić 2015).

Institutional factors can be behind exchange rate volatility as well. For example, Eichler and Littke (2017) propose a theoretical model and provide empirical evidence that better information about monetary policy objectives decreases exchange rate volatility, with a more pronounced effect for countries with a lower level of central bank conservatism. While there is no effect of central bank transparency for developing countries, transparency increases exchange rate fluctuations in developed countries (Weber 2017). Based on the panel estimations for 39 countries from Latin America, Asia, and MENA, it is argued that financial liberalization and integration should be pursued only gradually in emerging countries in order to decrease real exchange rate volatility (Caporale et al. 2011). However, the balance sheet effect implies that it may be optimal for monetary authorities to diminish exchange rate variation in order to decrease the cost of servicing and rolling-over foreign-currency debt (Georgiadis and Zhu 2019).

3. Data and Methods

3.1. Data

For the purpose of this study, we use data for the Czech Republic, Hungary, Poland, and Romania over the period of 2000–2019. All these countries follow free or managed floating exchange rate policies which makes it highly relevant for the study of a nominal exchange rate volatility. Quarterly series of the real gross domestic product (index, 2010 = 100), nominal and real effective exchange rates (index, 2010 = 100), as well as on the central bank reference rate (%), money aggregate M3 (in local currency), and the budget balance (% of GDP) are obtained from the IMF's International Financial Statistics database (www.data.imf.org, accessed on 8 February 2021). As measures of institutional quality, the Index of Economic Freedom from the Heritage Foundation is used (www.heritage.org/index/dowload, accessed on 8 February 2021).

The cyclical components of real output, yc_t (%), as well of the currency misalignment, $rerc_t$ (%), were calculated as a percentage deviation of the current values from the Hodrick-Prescott filtered trend. The use of a nominal effective exchange rate, e_t, is preferred in studies of the exchange rate variability because the RER variability incorporates price fluctuations, which represent another type of uncertainty for private agents (Barguellil et al. 2018). Other studies implement alternative measures of the currency misalignment. For example, Rodrik (2008) and Ribero et al. (2020) define RER misalignment as a difference between exchange rate adjusted for PPP conversion factors, $\ln(RER_t) = \ln(E_t/(PPP_t)$, and the RER obtained from a regression on the log the GDP per capita: $\ln(\widetilde{RER_t}) = \gamma_0 + \gamma_1 \ln(Y_t) + \varepsilon_t$.

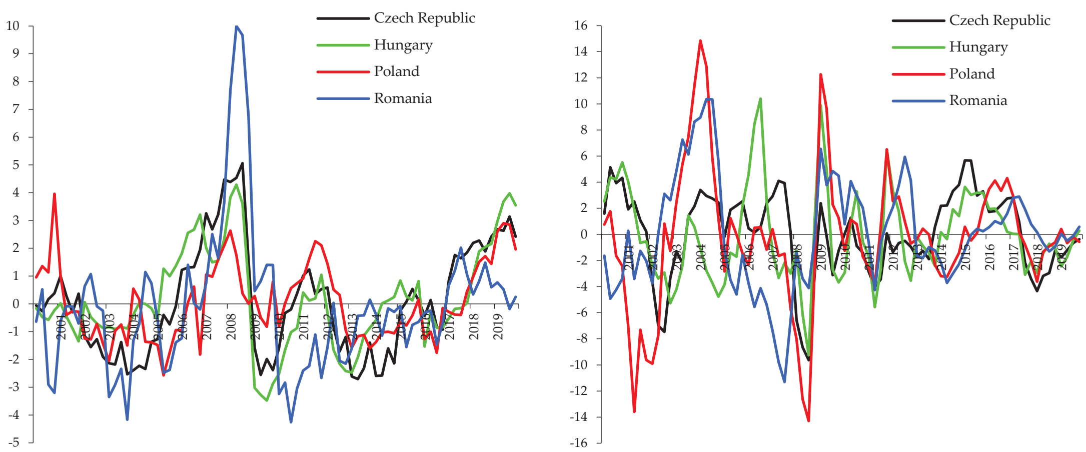
Table 1 shows the descriptive statistics of the main variables that are investigated in our study. Romania is characterized by the largest cyclical peak of its output fluctuations at almost 10%, with the deepest trough at −4.3% as well. The cyclical components of real output reveal less instability in the Czech Republic and Hungary. The business cycle is much smoother in Poland, probably due to a very flat slowdown in 2009. However, Poland reveals the highest level of the currency misalignment, followed by Romania, Hungary, and the Czech Republic. Outcomes are similar for the NEER in first differences. As suggested by the Jarque-Bera statistics, all NEERs show evidence of non-normality. It is not surprising as the NEER is determined by random changes in bilateral exchange rates, including countries with high-risk currencies. Except for Romania, cyclical components of output are normally distributed.

Figure 1 visualizes developments in the cyclical components of output and currency misalignment for the CEE countries. Business cycles of the Czech Republic, Hungary, and Poland seem to be quite synchronized, with cyclical output developments in Romania being somewhat different in terms of timing and amplitude. All countries experienced a boom in 2007–2008 followed by a remarkable slowdown in 2009–2010 (except Poland). After a period of anemic growth in 2013–2016, output dynamics have accelerated, though to a lesser extent in Romania. Currency misalignments had been more substantial in the 2000s, especially for the Polish *zloty* and the Romanian *lei*. On the eve of the world financial crisis there had been a remarkable appreciation of the CEE currencies in 2007–2008, with a steep reverse to follow in 2009. As recently, NEERs have been fluctuating approximately ±5% all around the equilibrium trend.

Table 1. Descriptive statistics.

Country	Mean	Max	Min	STD	Jarque-Bera
		Gross domestic product (yc_t)			
Czech Republic	0.094	5.059	−2.701	1.965	4.345
Hungary	0.148	4.292	−3.474	1.768	2.269
Poland	0.065	3.961	−2.572	1.379	4.105
Romania	−0.091	9.998	−4.259	2.538	140.88 ***
		Nominal effective exchange rate, in first differences (Δe_t)			
Czech Republic	−0.005	0.061	−0.053	0.021	4.809 *
Hungary	0.010	0.128	−0.057	0.032	53.833 ***
Poland	−0.001	0.151	−0.071	0.037	94.879 ***
Romania	0.010	0.105	−0.067	0.030	7.437 **
		Real exchange rate misalignment ($rerc_t$)			
Czech Republic	0.038	5.673	−9.636	3.201	8.508 **
Hungary	−0.092	10.409	−9.257	3.528	6.794 **
Poland	−0.231	14.860	−14.292	5.321	8.987 **
Romania	0.035	10.373	−11.296	4.196	0.968

Notes: ***, **, * denote rejection of the null hypothesis at 1%, 5%, and 10% respectively.



(**a**) the cyclical components of output (yc_t)   (**b**) currency misalignment ($rerc_t$)

Figure 1. Business cycle and currency misalignment, 2000–2019 (in %). Source: own calculations based on data from IMF International Financial Statistics (www.data.imf.org, accessed on 8 February 2021).

The Augmented Dickey-Fuller (ADF) stationarity test indicates that both cyclical components of output and currency misalignment variables are stationary in levels at the 5% significance level (Table 2). As expected, the NEERs are stationary in first differences.

Similar to Borys et al. (2008), to measure the quality of domestic institutions and the progress of market reforms we use the Heritage Foundation database. Besides the composite Index of Economic Freedom ($heritage_t$), we consider nine sub-indices, namely business freedom (1), trade freedom (2), investment freedom (3), financial freedom (4), property rights (5), fiscal health (6), judicial effectiveness (7), labor freedom (8), and monetary freedom (9), ranging from 0 to 10 points. The importance of institutions used to be considered in the context of long-term growth, but it seems to be of the same role in managing short-term output fluctuations. As mentioned by Boar (2010), the key to the macroeconomic success of an emerging economy is not the initial choice of the exchange rate regime but rather the health of the fundamental institutions.

Table 2. Unit root test.

	The Czech Republic		Hungary		Poland		Romania	
	Level	**Δ**	**Level**	**Δ**	**Level**	**Δ**	**Level**	**Δ**
yc_t	−2.428 **	−7.545 ***	−2.696 ***	−6.970 ***	−3.642 ***	−11.849 ***	−2.945 ***	−8.156 ***
e_t	−1.837 *	−5.811 ***	1.066	−8.114 ***	−0.323	−7.392 ***	1.532	−5.340 ***
$rerc_t$	−4.330 ***	−6.986 ***	−6.038 ***	−7.856 ***	−4.553 ***	−7.492 ***	−3.354 ***	−7.577 ***

Notes: ***, **, * denote rejection of the null hypothesis at 1%, 5%, and 10% respectively; Δ is for first differences.

3.2. The Model of Exchange Rate Volatility

The merit of the GARCH model stems from its ability to differentiate and recognize information that generates the exchange rate in a random process. The GARCH model is a robust model that is capable of dealing with the volatility associated with financial data characterized by skewed distribution and the problem of heteroscedasticity. In addition, the GARCH model allows for the differentiation and recognition of information that generates the exchange rate in a random process (Azid et al. 2005). Alternative measures of volatility such as the standard deviation and the coefficient of variation do not take into account the exchange rate uncertainty, which represents the unobserved fraction of exchange rate fluctuations (Barguellil et al. 2018).

In the baseline model, the quarterly exchange rate volatility is specified as follows:

$$\Delta e_t = \mathrm{E}\left(e_t \left|\Omega_{t-1}\right.\right) + \varepsilon_t, \quad \varepsilon_t/\Omega_{t-1} \approx N(0, \sigma_t), \tag{5}$$

$$\ln\left(\sigma_t^2\right) = \omega + \alpha\left|\frac{|\varepsilon_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\sigma_t}}\right| + \gamma\frac{\varepsilon_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \beta\ln\left(\sigma_{t-1}^2\right) + \delta_1\Delta cpi_t + \delta_2 CRISIS_t + \xi_t, \tag{6}$$

where e_t is the nominal effective exchange rate, Ω_{t-1} is the information set available at time $t-1$, ε_t is the stochastic factor, and Δ is the operator of first differences. The expected value of exchange rate $\mathrm{E}\left(e_t \left|\Omega_{t-1}\right.\right)$ is modelled as ARMA(p,q) process.

For the purpose of our study, ARMA(2,2) model is used. In line with Corsetti et al. (2017), the interest rate differential between foreign and domestic rates, as well as the lagged terms of trade, are included as explanatory variables too. Foreign interest rate, r_t^*, is proxied by the 6-month LIBOR. As a measure of the domestic interest rate, r_t , the money market rate is used for Poland and the lending rate is used for other CEE countries. The price index of domestically produced goods, $p_{H,t}$, is proxied by the producer prices, with consumer prices in Germany being used as a measure of foreign prices, p_t^*.

A one-period ahead forecast variance, σ_t, is a function of the mean (ω), the ARCH term (α), the EGARCH term (β), and three explanatory variables. A high value of α means significant impact of stochastic shocks, with a high value of β reflecting persistence in exchange rate volatility. The sum of both coefficients (α and β) indicates the speed of convergence of the forecast of the conditional volatility to a steady state (Kočenda and Valachy 2006). Asymmetry in the standardized shocks to $\ln\left(\sigma_t^2\right)$ exists if $\gamma \neq 0$, why leverage exists if $\gamma < 0$ and $\gamma < \alpha < -\gamma$ (McAleer and Hafner 2014). Similar to Schnabl (2009), we used the consumer price index (CPI) as a proxy for macroeconomic stability. A country-specific dummy $CRISIS_t$ is supposed to control for asymmetric shocks.

In the extended model, we test the link between exchange rate volatility and institutional features, as measured by the composite Index of Economic Freedom ($heritage_t$) and its sub-indices. As data on the Index of Economic Freedom are provided on the annual basis, we used procedure of the Holt-Winters exponential smoothing in order to obtain time series in the quarterly window.

3.3. The Model of Business Cycle

A general representation for the cyclical components of real output model is given as follows:

$$yc_t = a_1 yc_t + a_2 yeuroc_t + a_3 rerc_t + a_4 evar_t^k + a_5 \Delta herit_t + a_6 herit_t + a_7 trade_t + a_8 EU_t + v_t, \tag{7}$$

where $yeuroc_t$ is the cyclical component of real output in the Eurozone (%), $evar_t^k$ are alternative measures of exchange rate volatility ($k = 1, 2$), $herit_t$ is the composite index of economic freedom as provided by the Heritage Foundation, $trade_t$ is the trade balance (% of GDP), EU_t is the dummy for entering the European Union, and v_t is the stochastic factor.

Our regression model incorporates both exchange rate volatility around a constant level and the currency misalignment interpreted as a percentage deviation of the observed RER from the Hodrick-Prescott trend. Other explanatory variables include the Eurozone business cycle, the composite Index of Economic Freedom, the trade balance and the dummy for EU accession. The trade balance accounts for external factors that can affect the cyclical components of real output. A dummy for the EU accession is included as an explanatory variable that accounts for effects of economic integration of the CEE countries. Similar to the country-specific cyclical components of real output, business cycle for the Eurozone is obtained with the Hodrick-Prescott filter.

As suggested by De Haan et al. (2006), a measure of economic freedom is used both in levels and first differences. It is demonstrated that such a specification explains significantly more of the variation in economic growth. While most of empirical studies confirm a positive relationship between all areas of economic freedom and economic growth, for example Doucouliagos and Ulubasoglu (2006) or Emara and Reyes (2020), it is not straightforward whether more of economic freedom is helpful to the same extent in stabilizing cyclical fluctuations of economy. Bjørnskov (2016) finds that crisis risk and duration are not affected by economic freedom, but it has an effect on the peak-to-trough GDP ratios and recovery times of crises.

Two measures of volatility are employed, as presented above. Conditional variance from the baseline model, $evar_t^1$, accounts for domestic consumer prices and crisis developments as external factors. On the other hand, conditional variance from the extended model, $evar_t^2$, reflects impact of the Index of Economic Freedom across its nine sub-indices.

In the extended model, we add the budget balance, $budget_t$ (% of GDP), and two measures of monetary policy, i.e., excess money supply, $moneyc_t$ (%), and the central bank reference rate, rcb_t (%), to the list of explanatory variables. Excess money supply is calculated as a difference between money aggregate M3 and its trend obtained with the Hodrick-Prescott filter. While the use of the central bank reference rate reflects a standard monetary policy tool under a floating exchange rate regime, the excess money supply can control for attempts by the central bank to sterilize capital flows.

4. Results and Discussion

4.1. Determinants of Exchange Rate Volatility

As the Jarque-Bera test implies non-normal error distribution for Δe_t, we use EGARCH(1,1) models with the asymmetric Student's t-distribution. Estimates of the baseline model are presented in Table 3. Among determinants of the mean exchange rate, the interest rate differential is associated with depreciation for three out of four countries, being in line with the logic of Equation (3). However, the relationship for Romania is just the opposite. As for the lagged terms-of-trade, there is no evidence of a direct link between higher domestic prices and depreciation. For the Czech Republic, Poland, and Romania, there is an inverse relationship between both variables. Contrary to predictions of Corsetti et al. (2017), deflation abroad does not lead to depreciation of the exchange rate.

The ARCH term (α) indicates that impact of "surprises" from previous periods is the strongest in the Czech Republic and Poland (higher than one α implies that shocks to exchange rate can destabilize its volatility), with a much weaker effect in Hungary and Romania. Based on the value of the EGARCH term (β), persistence of exchange rate volatility is observed at the statistically significant level in the Czech Republic only. As indicated by the sum of α and β, the speed of convergence of the forecast of the conditional volatility to a steady state is very slow in the Czech Republic and Poland. The standardized shocks to $\ln(\sigma_t^2)$ are symmetrical in Poland, as the value of γ is not statistically different from zero. For the Czech Republic and Romania, the negative value of γ implies that negative news affects the exchange rate volatility more heavily than positive news. It is just the opposite in Hungary. No preconditions for leverage are met in any country.

Table 3. Univariate EGARCH results (baseline model).

	The Czech Republic	Hungary	Poland	Romania
A. Mean equation results				
$r_t^* - r_t$	0.002 ***	0.001 ***	0.005 ***	−0.001 ***
$p_{H,t-1} - p_{t-1}^*$	−0.003 ***	0.008	−0.298 ***	−0.007 ***
B. Variance equation results				
ω	−6.175 ***	−6.687 **	−7.961 ***	−6.671 **
α	1.198 ***	0.660 *	1.172 ***	0.635 ***
γ	−0.380 *	0.491 **	−0.230	−0.316 **
β	0.350 ***	0.212	0.060	0.298
Δcpi_t	−60.059 ***	−11.957	−13.118	14.232 **
$CRISIS_t$	1.647 ***	2.309 **	3.017 ***	1.010 **
Obs	80	80	80	80
AIC	−5.222	−4.497	−4.203	−4.648

Notes: ***, **, * represent statistical significance at 1%, 5%, and 10%, respectively.

Among the control variables, inflation contributes to exchange rate volatility in the Czech Republic and Romania, though with opposite signs. There is no difference between all four CEE countries in that crisis developments are associated with higher exchange rate volatility.

In the extended model (Table 4), a statistically significant direct relationship between the lagged terms-of-trade and mean exchange rate emerges in Hungary. The ARCH effect somewhat decreases in Poland and Romania, with the opposite outcome observed in the Czech Republic and Hungary. As suggested by the value of β, there are no changes to the assessment of persistence in exchange rate volatility in the Czech Republic. However, exchange rate variability becomes more persistent in Hungary. Considering the sum of ARCH and EGARCH coefficients, a control for institutional features implies a significantly slower speed of convergence of the forecast of the conditional volatility to a steady state in Hungary. For other countries, changes are rather marginal. Asymmetry of the standardized shocks to $\ln(\sigma_t^2)$ is confirmed for Hungary and Romania, while the negative coefficient of γ becomes insignificant for the Czech Republic. Among other changes, inflation becomes a factor behind lower exchange rate volatility in Hungary. Except the Czech Republic, the coefficient of $CRISIS_t$ becomes significantly lower for other countries.

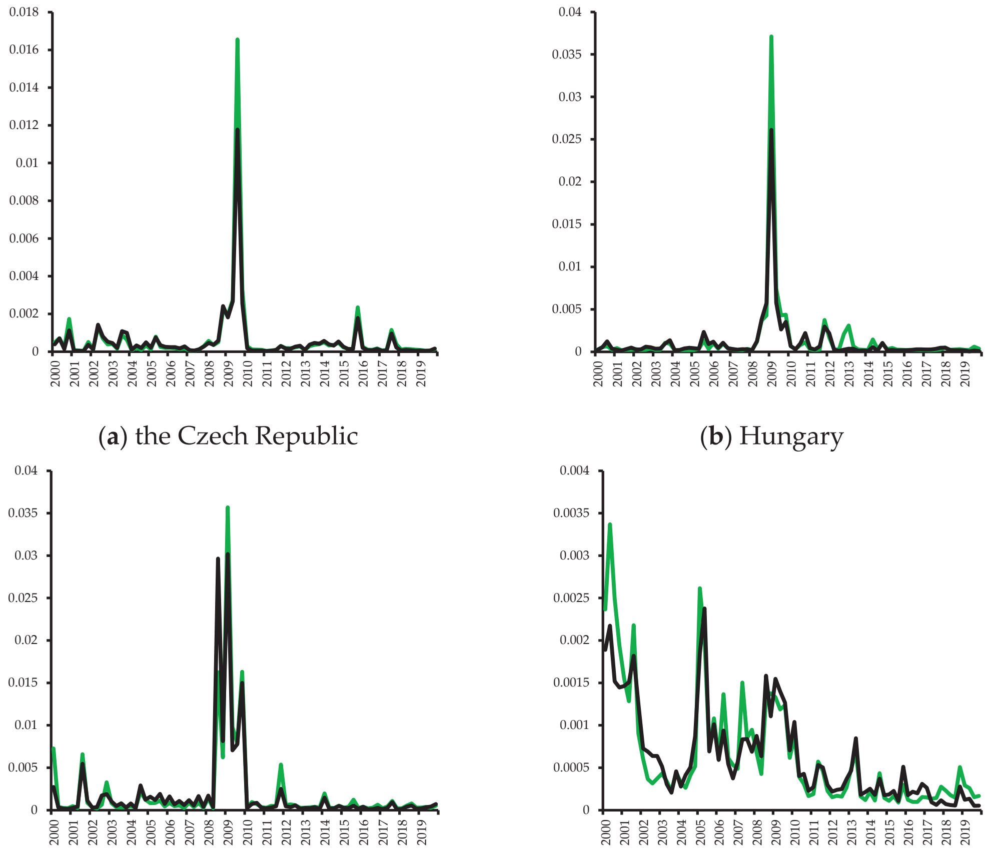
As suggested by the estimated coefficients on sub-indices of economic freedom, in a more liberal environment exchange rate volatility becomes lower. The only exception is Hungary, where monetary freedom brings about a higher exchange rate volatility. On the whole, the effects of economic freedom on exchange rate volatility are country specific. Property rights guarantees decrease exchange rate volatility in Romania. Fiscal health exerts the same effect on volatility in Hungary. Success in anticorruption activities contributes to a lower exchange rate volatility the Czech Republic and Poland. In the extended model, inflation becomes a volatility-decreasing factor in Poland, with a positive coefficient on $libor_t$ changing sign. No changes in the effects of crisis developments are observed.

Figure 2 plots the evolution of quarterly exchange rate volatility by country, with the conditional variation obtained by fitting the baseline and extended EGARCH(1,1) models in green and black colors, respectively. For all countries, periods of low volatility are followed by periods of high volatility which could be associated with periods of global financial crisis of 2008–2009 and/or domestic financial turmoil (the Czech Republic in 2001–2002, Hungary in 2012–2013, Romania in 2004–2005). Differences between two measures of exchange rate volatility seem to be quite small in the Czech Republic and Hungary, while being more pronounced in Poland and Romania. After controlling for the institutional features, volatility becomes somewhat smaller. Except Romania, the exchange rate volatility rose from the beginning of 2007, with a peak during the world financial crisis of 2008–2009. Volatility then subsided in the majority of CEE countries, except Hungary in 2012–2013. For Romania, there is an increase in volatility around 2005 that is followed by a smaller jump in 2008–2009.

Table 4. Univariate EGARCH results (extended model).

	The Czech Republic	**Hungary**	**Poland**	**Romania**
		A. Mean equation results		
$r_t^* - r_t$	0.002 ***	0.002 ***	0.005 ***	−0.001 ***
$p_{H,t-1} - p_{t-1}^*$	−0.016 ***	0.009 ***	−0.174 **	−0.006 ***
		B. Variance equation results		
ω	−3.859 ***	−1.989	−5.156 ***	−5.692 ***
α	1.132 ***	0.735 ***	1.068 ***	0.575 **
γ	−0.310	0.461 ***	−0.148	−0.218 *
β	0.328 **	0.405 **	−0.002	0.198
Δcpi_t	−68.857 ***	−38.708 **	−43.191	15.671 ***
$CRISIS_t$	1.832 ***	1.571 ***	2.965 **	0.584 *
$herit_t^5$	—	—	—	−0.036 **
$herit_t^6$	—	−0.104 ***	—	—
$herit_t^7$	−0.052 **	—	−0.067 ***	—
$herit_t^9$	—	0.059 **	—	—
Obs	80	80	80	80
AIC	−5.215	−4.528	−4.160	−4.640

Notes: ***, **, * represent statistical significance at 1%, 5%, and 10%, respectively.



(**a**) the Czech Republic (**b**) Hungary

(**c**) Poland (**d**) Romania

Figure 2. Conditional variance from EGARCH(1,1) Model.

To summarize, volatility patterns of the CEE countries seem to be similar in respect to both ARCH and EGARCH effects, especially after controlling for the institutional features. However, there are differences in asymmetry of volatility shocks and effects of control variables. In that respect, our results do not reject that volatility in the CEE economies has country-specific features (Kočenda and Valachy 2006).

4.2. Determinants of the Business Cycle

We estimate Equation (3) with the general method of moments (GMM) estimator. Comparing with OLS or IV estimators, the GMM method is preferred for better dealing with problems of simultaneity bias, reverse causality, and omitted variable bias, as well as for obtaining estimates of dummy coefficients (Caporale et al. 2011). Table 5 presents the results for the baseline model of cyclical components of real output.

Table 5. Regression results for cyclical components of output (baseline model).

	The Czech Republic		Hungary		Poland		Romania	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
yc_{t-1}	0.360 ***	0.357 ***	0.640 ***	0.638 ***	0.418 ***	0.430 ***	0.680 ***	0.669 ***
$yeuroc_t$	0.860 ***	0.866 ***	0.318 ***	0.324 ***	0.563 ***	0.524 ***	0.398 **	0.388 **
$evar_t^1$	53.257 ***	—	−91.930 ***	—	30.180	—	−229.22	—
$evar_t^2$	—	72.001 ***	—	−135.27 ***	—	13.939	—	−167.33
$rerc_t$	—	—	0.023	0.023	0.073 ***	0.080 ***	—	—
$rerc_{t-1}$	0.034	0.035	—	—	—	—	0.008	0.014
$\Delta herit_t$	7.773	7.901	−4.697	−3.577	0.967	1.165	7.909	8.169
$herit_t$	−0.196 ***	−0.199 ***	−0.091 *	−0.084 *	−0.088 *	−0.092 *	−0.174	−0.203
$trade_{t-1}$	−0.068 *	−0.067	−0.077 ***	−0.082 ***	−0.039	−0.054 *	−0.076	−0.076
EU_t	1.145 ***	1.149 ***	0.967 ***	0.985 ***	0.285	0.385 *	0.356	0.463
Obs	80	80	80	80	80	80	80	80
Adj. R^2	0.92	0.92	0.87	0.87	0.68	0.67	0.71	0.71

Notes: ***, **, * represent statistical significance at 1%, 5%, and 10%, respectively. Standard errors are corrected for autocorrelation and heteroscedasticity (Newey and West 1994).

For all countries, coefficients on the lagged dependent variable are significant at the 1% level. Inertia of cyclical developments in output is stronger in Hungary and Romania. These two countries are characterized by the lowest correlation with the business cycle of the Eurozone as well. Correlation between national and European business cycles is the strongest in the Czech Republic.

According to the estimates of the baseline model, exchange rate volatility is associated with the risk of recession only in Hungary. An opposite effect is obtained for the Czech Republic. Exchange rate volatility is neutral in respect to the cyclical output developments in Poland and Romania. Such findings can be considered as evidence in favor of the exchange rate disconnect. However, neutrality of output fluctuations in respect to exchange rate volatility does not mean the same lack of reaction to currency misalignment. Exchange rate undervaluation has favorable growth effects in Poland. Assuming that there is exchange rate depreciation in response to such real shock as an increase in the foreign interest rate (Tables 3 and 4), it helps to stabilize the economy. Under such architecture of the exchange rate effects, a switch to free or managed floating looks like a reasonable exchange rate policy. In a wider context, our results imply that the Czech Republic and especially Poland both benefit from exchange rate flexibility and thus may not be interested in joining the Eurozone. However, such benefits are visible for Hungary, as the exchange rate volatility seems to be destabilizing. Among numerous explanations of the inverse relationship between exchange rate volatility and output, several ones are worth attention in the case of Hungary, such as higher risk premium, greater uncertainty about export revenues, higher risk for domestic and foreign direct investment, and adverse effect of credit constraints on domestic investments. Additionally, it is not ruled out that in a country with relatively low levels of financial development (real) exchange rate uncertainty exacerbates the negative investment effects of domestic credit market constraints (Aghion et al. 2009) or reflect disincentives for firms in creating jobs (Belke and Setzer 2003).

There is no evidence of any favorable stabilization effects of economic freedom. A negative effect is the strongest for the Czech Republic, followed by Poland and Hungary. For Romania, a negative coefficient of $herit_t$ is insignificant. It is likely that our findings reflect an excessive level of economic liberalization attained during the period of negotiations with the European Union on the terms of EU accession. While the level of economic freedom is negatively correlated with cyclical fluctuations in output, changes in the level of economic freedom are neutral in respect to yc_t.

Trade deficit is still an important factor behind economic growth in Hungary, Romania, and Poland (to lesser extent). Our results mean that economic recovery depends more on imports, not exports. In this context, traditional supply-side trade channels, as capital accumulation, modernization of industrial structure, and technological and institutional progress, seem to be relevant. As can be seen in the example of the Czech Republic and Poland, statistical significance of the coefficient on $trade_{t-1}$ depends on the choice of the exchange rate variability.

The stimulating effect of the EU accession is the strongest in the Czech Republic, followed by Hungary. For Poland, the coefficient of EU_t is much smaller and statistically significant at the 10% level only in specification with $evar_t^2$. No evidence of any EU accession effects is found for Romania.

After controlling for fiscal and monetary policies (Tables 6 and 7), there are no changes in the assessment of exchange rate effects for Poland and Romania. For the Czech Republic, effects of exchange rate volatility on yc_t are confirmed but the same favorable effect of the RER undervaluation emerges in the specification with the money supply. Similar stimulating effect of the RER undervaluation is found for Hungary, although only in specification with $evar_t^1$. It is confirmed that exchange rate volatility contributes to a recession in Hungary.

A negative link between economic freedom (in levels) and business cycle is very robust for the Czech Republic, while the estimates for other countries are specification dependent. For Hungary, economic freedom becomes neutral in respect to cyclical developments in output in 3 out of 4 specifications. It is just the opposite for Romania, where a statistically significant negative link between $herit_t$ and yc_t emerges in specifications with both $moneyc_t$ and rcb_t. Additionally, Romania emerges as the only CEE country with a statistically significant positive effect of an increase in economic freedom (in first differences) on output. For Poland, a negative relationship between $herit_t$ and yc_t disappears in specification with $moneyc_t$, while being strengthened in the specification with rcb_t.

An excessive money supply, $moneyc_t$, helps to stabilize output in the Czech Republic and Poland. Assuming a link between the money supply and exchange rate volatility (Devereux and Engel 2002), it only strengthens the assumption of shock-absorbing properties of the floating exchange rate regimes for the Czech *koruna* and Polish *zloty*.

Table 6. Regression results for cyclical components of output (extended model-I).

	The Czech Republic		Hungary		Poland		Romania	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
yc_{t-1}	0.240 ***	0.238 ***	0.645 ***	0.619 ***	0.327 ***	0.331 ***	0.670 ***	0.668 ***
$yeuroc_t$	0.936 ***	0.939 ***	0.333 **	0.375 ***	0.521 ***	0.493 ***	0.406 ***	0.381 **
$evar_t^1$	38.360 **	—	−97.045 ***	—	11.991	—	−442.11	—
$evar_t^2$	—	50.948 *	—	−136.54 ***	—	−0.478	—	−517.79
$rerc_t$	—	—	0.034 *	0.031	0.083 ***	0.084 ***	—	—
$rerc_{t-1}$	0.048 **	0.049 **	—	—	—	—	0.057	0.058
$\Delta herit_t$	4.655	4.685	−4.926	−4.460	−4.622	−4.768	11.400	11.059
$herit_t$	−0.113 ***	−0.116 ***	−0.072	−0.088	−0.037	−0.029	−0.277 *	−0.267 *
$trade_{t-1}$	−0.129 **	−0.128 **	−0.072 **	−0.074 **	−0.066 *	−0.078 *	−0.117 **	−0.120 **
EU_t	1.255 ***	1.261 ***	0.836 ***	0.926 ***	0.316	0.350	0.892 *	0.890 *
$moneyc_t$	0.064 *	0.064 *	0.016	0.001	0.066 **	0.071 **	0.033	0.041
$budget_{t-1}$	0.085 ***	0.086 ***	−0.014	−0.021	0.087	0.103	0.160 **	0.163 **
Obs	72	72	76	76	80	80	73	73
Adj. R^2	0.93	0.93	0.87	0.87	0.69	0.70	0.74	0.74

Notes: ***, **, * represent statistical significance at 1%, 5%, and 10%, respectively. Standard errors are corrected for autocorrelation and heteroscedasticity (Newey and West 1994).

Table 7. Regression results for cyclical components of output (extended model-II).

	The Czech Republic		Hungary		Poland		Romania	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
yc_{t-1}	0.304 ***	0.301 ***	0.498 ***	0.514 ***	0.307 ***	0.311 ***	0.668 ***	0.664 ***
$yeuroc_t$	0.830 ***	0.834 ***	0.404 ***	0.392 ***	0.517 ***	0.482 ***	0.437 ***	0.458 **
$evar_t^1$	46.973 ***	—	−59.862 ***	—	31.262 *	—	−83.717	—
$evar_t^2$	—	62.817 ***	—	−91.840 ***	—	21.202	—	−15.943
$rerc_t$	—	—	0.015	0.012	0.091 ***	0.097 ***	—	—
$rerc_{t-1}$	0.022	0.022	—	—	—	—	0.064	0.067
$\Delta herit_t$	3.453	3.605	4.835	5.042	−1.372	−0.556	14.565 *	14.923 *
$herit_t$	−0.164 **	−0.168 **	0.113	0.102	−0.254 **	−0.256 **	−0.419 **	−0.446 **
$trade_{t-1}$	−0.087 *	−0.086 *	−0.147 ***	−0.144 ***	—	—	−0.084	−0.081
EU_t	1.207 ***	1.212 ***	0.783 ***	0.828 ***	0.756 **	0.802 **	1.377 **	1.466 **
rcb_t	0.056	0.058	−0.137 ***	−0.129 ***	0.119 ***	0.129 ***	0.021	0.025
$budget_{t-1}$	0.077 ***	0.077 ***	−0.026	−0.027	0.085	0.102	0.149 **	0.153 **
Obs	80	80	80	80	80	80	68	68
Adj. R^2	0.92	0.92	0.89	0.90	0.70	0.70	0.73	0.73

Notes: ***, **, * represent statistical significance at 1%, 5%, and 10%, respectively. Standard errors are corrected for autocorrelation and heteroscedasticity (Newey and West 1994).

An increase in the central bank reference rate acts in the expected countercyclical manner in Hungary, while counterintuitive proportional link between rcb_t and yc_t is observed in Poland. It is possible to hypothesize that such an outcome results from efforts by the central bank to avoid appreciation of the exchange rate. As a higher central bank rate can tap capital inflows, sterilization policies substitute a stronger currency with a higher excessive money supply that ultimately becomes responsible for an increase in output. For Romania, it is likely that the lack of sensitivity to the central bank policy rate is explained by the balance sheet effect, as argued by Georgiadis and Zhu (2019).

For the Czech Republic and Romania, there is evidence of stabilization properties of the fiscal tightening. The so-called non-Keynesian effects of fiscal policy mean that in the case of recession it is necessary to improve the budget balance, not engage in rounds of fiscal stimuli, as has been the case in many industrial countries since the world financial crisis of 2008–2009.

When the extended dataset is used (using both fiscal and monetary variables), there are several changes to the assessment of trade and EU accession output effects. A negative link between $trade_{t-1}$ and yc_t is confirmed for Hungary and it becomes more stable for the Czech Republic. As for Poland and Romania, a negative impact of the trade balance is observed in specifications with $moneyc_t$, but the effect is lost in specifications with rcb_t. A strong procyclical effect of entering the EU is confirmed for the Czech Republic and Hungary. For Poland, a stimulating effect of similar amplitude emerges in the specification with rcb_t, although it is not observed in the specification with $moneyc_t$.

5. Robustness Check

As suggested by Rodrik (2008), we use the measure of currency misalignment based on the RER adjusted for the level of output. This measure adjusts the relative price of tradables to nontradables for the fact that the relative prices of nontradables tend to rise in line with the higher level of output. Our estimates support the assumption of the RER appreciation for the Czech Republic, Hungary, and Romania (Table 8). However, no link between the level of output and RER is found for Poland.

Table 8. Estimates of the RER adjusted for the level of output.

	The Czech Republic	Hungary	Poland	Romania
γ_0	7.748 ***	5.965 ***	4.450 ***	6.002 ***
γ_1	−0.663 ***	−0.276 ***	0.043	−0.294 ***
Obs	80	80	80	80
Adj. R^2	0.69	0.12	0.01	0.43

Notes: *** represents statistical significance at 1%.

Estimates of the determinants of cyclical components of output are presented in Tables 9 and 10. With the use of the measure of currency misalignment based on the RER adjusted for the level of output, $rerpppc_t$, the architecture of main relationships between exchange rate developments and cyclical changes in output is confirmed. First, exchange rate volatility is a stabilizing factor in the Czech Republic, with an opposite effect in Hungary. For Romania, exchange rate volatility is neutral in respect to the business cycle. As for Poland, a possibility of stimulating effect is offered by specification with $evar_t^1$ and rcb_t, but it is not confirmed by the estimates of other specifications.

Table 9. Regression results for cyclical components of output (extended model-I).

	The Czech Republic		Hungary		Poland		Romania	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
yc_{t-1}	0.187 *	0.185 *	0.586 ***	0.596 ***	0.366 ***	0.373 ***	0.640 ***	0.638 ***
$yeuroc_t$	0.949 ***	0.951 ***	0.355 **	0.356 ***	0.474 ***	0.427 ***	0.303 **	0.278 *
$evar_t^1$	37.831 *	—	−87.808 ***	—	22.424	—	−444.68	—
$evar_t^2$	—	49.362 *	—	−129.40 ***	—	5.169	—	−498.50
$rerpppc_t$	—	—	0.027 *	0.024	0.063 ***	0.063 ***	—	—
$rerpppc_{t-1}$	0.033 *	0.033 *	—	—	—	—	0.040	0.039
$\Delta herit_t$	8.567	8.618	−2.809	−2.327	−6.044	−6.607	11.398	10.853
$herit_t$	−0.184 ***	−0.187 ***	−0.161 **	−0.157 **	−0.140	−0.126	−0.389 **	−0.372 **
$trade_{t-1}$	−0.180 **	−0.179 **	−0.085 **	−0.085 **	−0.099 **	−0.115 ***	−0.160 **	−0.159 **
EU_t	1.805 ***	1.813 ***	1.256 ***	1.244 ***	0.743 **	0.783 **	1.212 **	1.179 **
$moneyc_t$	0.075 **	0.075 *	0.015	0.014	0.045	0.052 *	0.034	0.040
$budget_{t-1}$	0.067 ***	0.067 ***	−0.032	−0.035	0.069	0.093	0.162 **	0.161 **
Obs	72	72	76	76	80	80	73	73
Adj. R^2	0.92	0.92	0.87	0.88	0.67	0.65	0.73	0.73

Notes: ***, **, * represent statistical significance at 1%, 5%, and 10%, respectively. Standard errors are corrected for autocorrelation and heteroscedasticity (Newey and West 1994).

Table 10. Regression results for cyclical components of output (extended model-II).

	The Czech Republic		Hungary		Poland		Romania	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
yc_{t-1}	0.279 ***	0.276 ***	0.505 ***	0.501 ***	0.345 ***	0.354 ***	0.654 ***	0.651 ***
$yeuroc_t$	0.833 ***	0.837 ***	0.410 ***	0.398 ***	0.484 ***	0.437 ***	0.373 **	0.391 *
$evar_t^1$	45.579 **	—	−52.010 ***	—	47.452 ***	—	−157.76	—
$evar_t^2$	—	60.632 **	—	−81.150 ***	—	41.580	—	−82.239
$rerpppc_t$	—	—	−0.014	−0.014	0.059 ***	0.062 ***	—	—
$rerpppc_{t-1}$	0.007	0.007	—	—	—	—	0.024	0.023
$\Delta herit_t$	5.425	5.609	5.398	5.479	−6.651	−6.475	14.545 *	14.455 *
$herit_t$	−0.196 ***	−0.201 ***	0.197	0.186	−0.306 **	−0.303 **	−0.488 **	−0.510 **
$trade_{t-1}$	−0.085	−0.084	−0.147 ***	−0.144 ***	—	—	−0.101	−0.094
EU_t	1.305 ***	1.311 ***	0.528	0.541	0.992 **	1.024 **	1.548 **	1.618 **
rcb_t	0.089	0.091	−0.164 ***	−0.156 ***	0.100 ***	0.108 ***	0.039	0.044
$budget_{t-1}$	0.074 ***	0.074 ***	−0.029	−0.030	0.040	0.061	0.130 *	0.132 *
Obs	78	78	80	80	80	80	68	68
Adj. R^2	0.92	0.92	0.89	0.90	0.67	0.65	0.74	0.74

Notes: ***, **, * represent statistical significance at 1%, 5%, and 10%, respectively. Standard errors are corrected for autocorrelation and heteroscedasticity (Newey and West 1994).

Second, cyclical output effects of currency misalignment are quite similar. Regardless of specifications of regression model or indicators of currency misalignment used, it is confirmed that undervaluation of the Polish *zloty* has a stimulating effect on output. As argued by Rodrik (2008), such an outcome can be explained by the size of the tradable sector (especially industry); however, it is less convincing that undervaluation is aimed at compensating for the institutional weakness and the market failures (information and coordination externalities). Undervaluation of the Czech *koruna* brings about stabilization effects only in the specification with $moneyc_t$. The case is similar with the Hungarian *forint*, but in this case a positive coefficient on $rerpppc_t$ becomes insignificant in the specification with $evar_t^2$.

Higher levels of economic freedom unambiguously destabilize output for the Czech Republic and Romania. However, changes in the level of economic freedom have an opposite effect in the latter, quite similar to the estimates with $rerc_t$ (Table 7). For Hungary, a negative link between $herit_t$ and yc_t becomes statistically significant in the specification with $moneyc_t$. Estimates for Poland do not reveal any differences in respect to output effects of economic freedom.

Using $rerpppc_t$ instead of $rerc_t$ significantly improves assessment of output effects by the EU accession for Poland. An opposite outcome is found for Hungary in specification with rcb_t. If measure currency misalignment by $rerpppc_t$, a negative link between $trade_{t-1}$ and yc_t becomes somewhat stronger in specifications with $moneyc_t$.

6. Conclusions

This study examined the determinants of exchange rate volatility and its impact on the cyclical changes in output for the Czech Republic, Hungary, Poland, and Romania (all of which support a floating exchange rate regime of their currencies). Using quarterly data for the 2000–2019 period, exchange rate volatility was estimated with the EGARCH(1,1) model. As indicated by the ARCH term, the impact of "surprises" from previous periods is the strongest in the Czech Republic and Poland, with a much weaker effect in Hungary and Romania. Based on the value of the EGARCH term, persistence of exchange rate volatility was observed at the statistically significant level in the Czech Republic and Hungary (extended model). For the Czech Republic and Romania, it was found that negative news affects the volatility more heavily than positive news. It is just the opposite in Hungary. Among the control variables, inflation contributes to exchange rate volatility in Romania, while an opposite relationship was found for the Czech Republic and Hungary. There was no difference between all four CEE countries in that crisis developments are associated with higher exchange rate volatility. As suggested by the estimated coefficients on sub-indices of economic freedom, exchange rate volatility becomes lower in a more liberal environment.

The empirical results using GMM estimation technique suggest that exchange rate volatility reduces the risk of recession in the Czech Republic while the opposite effect is found for Hungary and Romania, with a neutrality for Poland. These findings continue to hold after controlling for the fiscal and monetary policy indicators. Ultimately, there is support for an assumption of heterogeneous growth effects of exchange rate volatility across countries. There is evidence that the RER undervaluation prevents sliding into a recession on a credible basis in Poland only (the result is very robust to changes in the specification of regression model), with a neutral stance for other countries. There is no evidence of any favorable stabilization effects of economic freedom. Except in Romania, higher level of economic freedom is associated with worsening of the cyclical position of output. Among other results, stabilization policies in the recession imply fiscal tightening for the Czech Republic and Romania, higher money supply for the Czech Republic and Poland, and lower central bank reference rate for Hungary.

Our research helps to understand potential constrains of exchange rate flexibility as an output stabilization tool in terms of both exchange rate volatility and currency misalignment. However, further research is needed in order to explain significant differences in the output effects of exchange rate volatility across CEE countries.

Author Contributions: Conceptualization, V.S.; methodology, V.S.; software, R.K.; validation, V.S.; formal analysis, V.S.; investigation, R.K.; resources, R.K.; data curation, R.K.; writing—original draft preparation, R.K.; writing—review and editing, V.S.; visualization, R.K.; supervision, V.S. All authors have read and agreed to the published version of the manuscript.

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.

References

Aghion, Philippe, Philippe Bacchetta, Romain Ranciere, and Kenneth Rogoff. 2009. Exchange Rate Volatility and Productivity Growth: The Role of Financial Development. *Journal of Monetary Economics* 56: 494–513. [CrossRef]

Ahmed, Shaghil, Christofer Gust, Steven B. Kamin, and Jonathan Huntley. 2002. *Are Depreciations as Contractionary as Devaluations? A Comparison of Selected Emerging and Industrial Economies*. FRB International Finance Discussion Paper No. 737. Washington, DC: Federal Reserve System.

Aizenman, Joshua, Yothin Jinjarak, Gemma Estrada, and Shu Tian. 2018. Flexibility of adjustment to shocks: Economic growth and volatility of middle-income countries before and after the global financial crisis of 2008. *Emerging Markets Finance and Trade* 54: 1112–31. [CrossRef]

Antonakakis, Nikolaos, and Julia Darby. 2012. *Forecasting Volatility in Developing Countries' Nominal Exchange Returns*. MPRA Paper No. 40875. Munich: University Library of Munich.

Arratibel, Olga, Davide Furceri, Reiner Martin, and Aleksandra Zdzienicka-Durand. 2011. The Effect of Nominal Exchange Rate Volatility on Real Macroeconomic Performance in the CEE Countries. *Economic Systems* 35: 261–77. [CrossRef]

Azid, Toseef, Muhammad Jamil, and Aneela Kousar. 2005. Impact of Exchange Rate Volatility on Growth and Economic Performance: A Case Study of Pakistan, 1973–2003. *Pakistan Development Review* 44: 749–75. [CrossRef]

Bacchetta, Philippe, and Eric van Wincoop. 2013. On the unstable relationship between exchange rates and macroeconomic fundamentals. *Journal of International Economics* 91: 18–26. [CrossRef]

Barguellil, Achouak, Ousama Ben-Salha, and Mourad Zmami. 2018. Exchange Rate Volatility and Economic Growth. *Journal of Economic Integration* 33: 1302–36. [CrossRef]

Belke, Ansgar, and Ralph Setzer. 2003. *Exchange Rate Volatility and Employment Growth: Empirical Evidence from the CEE Economies*. CESifo Working Paper No. 1056. Munich: Center for Economic Studies and Ifo Institute (CESifo).

Benhima, Kenza. 2012. Exchange Rate Volatility and Productivity Growth: The Role of Original Sin. *Open Economies Review* 23: 501–29. [CrossRef]

Béreau, Sophie, Antonia Lopez-Villavicencio, and Valerie Mignon. 2012. Currency Misalignments and Growth: A New Look using Nonlinear Panel Data Methods. *Applied Economics* 44: 3503–11. [CrossRef]

Bjørnskov, Christian. 2016. Economic freedom and economic crises. *European Journal of Political Economy* 45: 11–23. [CrossRef]

Bleaney, Michael, and Manuela Francisco. 2010. What Makes Currencies Volatile? An Empirical Investigation. *Open Economies Review* 21: 731–50. [CrossRef]

Bleaney, Michael, and David Greenaway. 2001. The impact of terms of trade and real exchange rate volatility on investment and growth in sub-Saharan Africa. *Journal of Development Economics* 65: 491–500. [CrossRef]

Boar, Corina Georgeta. 2010. Exchange rate volatility and growth in emerging Europe. *The Review of Finance and Banking* 2: 103–20.

Borys, Magdalena, Éva Polgár, and Andrei Zlate. 2008. *Real Convergence and the Determinants of Growth in EU Candidate and Potential Candidate Countries—A Panel Data Approach*. ECB Occasional Paper No. 86. Frankfurt: European Central Bank.

Caballero, Ricardo J., and Guido Lorenzoni. 2014. Persistent Appreciations and Overshooting: A Normative Analysis. *IMF Economic Review* 62: 1–47. [CrossRef]

Calderon, César, and Megumi Kubota. 2018. Does Higher Openness Cause More Real Exchange Rate Volatility? *Journal of International Economics* 110: 176–204. [CrossRef]

Caporale, Guglielmo, Thouraya Amor, and Christophe Rault. 2011. *Sources of Real Exchange Rate Volatility and International Financial Integration: A Dynamic GMM Panel Approach*. CESifo Working Paper No. 3645. Munich: Center for Economic Studies and Ifo Institute (CESifo).

Corsetti, Giancarlo. 2006. Openness and the case for flexible exchange rates. *Research in Economics* 60: 1–21. [CrossRef]

Corsetti, Giancarlo, Keith Kuester, and Gernot J. Müller. 2017. *Fixed on Flexible Rethinking Exchange Rate Regimes after the Great Recession*. Cambridge Working Papers in Economics No. 1729. Cambridge: University of Cambridge, Available online: http://www.econ.cam.ac.uk/research-files/repec/cam/pdf/cwpe1729.pdf (accessed on 5 February 2021).

Cuestas, Juan Carlos, Mercedes Monfort, and Javier Ordóñez. 2019. *Real Exchange Rates and Competitiveness in Central and Eastern Europe: Have They Fundamentally Changed?* Working Papers 2019/12. Castellón: Universitat Jaume I.

De Haan, Jacob, and Jan-Egbert Sturm. 2000. On the Relationship Between Economic Freedom and Economic Growth. *European Journal of Political Economy* 16: 215–41. [CrossRef]

De Haan, Jacob, Susanna Lundström, and Jan-Egbert Sturm. 2006. Market oriented institutions and policies and economic growth: A critical survey. *Journal of Economic Surveys* 20: 157–91. [CrossRef]

Demir, Firat. 2013. Growth under exchange rate volatility: Does access to foreign or domestic equity markets matter? *Journal of Development Economics* 100: 74–88. [CrossRef]

Devereux, Michael B., and Charles Engel. 2002. Exchange rate pass-through, exchange rate volatility, and exchange rate disconnect. *Journal of Monetary Economics* 49: 913–40. [CrossRef]

Devereux, Michael B., and Philip R. Lane. 2003. Understanding bilateral exchange rate volatility. *Journal of International Economics* 60: 109–32. [CrossRef]

Dollar, David. 1992. Outward-oriented developing economies really do grow more rapidly: Evidence from 95 LDCs, 1976–1985. *Economic Development and Cultural Change* 40: 523–44. [CrossRef]

Dornbusch, Rudiger. 2001. *Fewer Monies, Better Monies*. NBER Working Papers No. 8324. Cambridge: National Bureau of Economic Research.

Dornbusch, Rudiger, and Alberto Giovannini. 1990. Monetary Policy in the Open Economy. In *Handbook of Monetary Economics*. Edited by Benjamin M. Friedman and Frank H. Hahn. Amsterdam: North-Holland Publisher, vol. 2, pp. 1231–303.

Doucouliagos, Chris, and Mehmet Ali Ulubasoglu. 2006. Economic freedom and economic growth: Does specification make a difference? *European Journal of Political Economy* 22: 60–81. [CrossRef]

Durlauf, Steven N. 2018. *Institutions, Development and Growth: Where Does Evidence Stand?* EDI Working Paper Series, 18/04.1; Oxford: Economic Development & Institution.

Edwards, Sebastian. 2011. Exchange-Rate Policies in Emerging Countries: Eleven Empirical Regularities from Latin America and East Asia. *Open Economies Review* 22: 533–63. [CrossRef]

Edwards, Sebastian, and Eduardo Levy-Yeyati. 2005. Flexible exchange rates as shock absorbers. *European Economic Review* 49: 2079–105. [CrossRef]

Eichler, Stefan, and Helge C. Littke. 2017. *Central Bank Transparency and the Volatility of Exchange Rates*. IWH Discussion Papers, 22/2017. Halle: Leibniz-Institut für Wirtschaftsforschung Halle (IWH), Available online: http://nbn-resolving.de/urn:nbn:de:gbv:3:2-80759 (accessed on 5 February 2021).

Emara, Noha, and Loreto Reyes. 2020. *Economic Freedom and Economic Performance: Does Good Governance Matter? The Case of APAC and OECD Countries*. MPRA Paper, No. 103590. Munich: University Library of Munich.

Engel, Charles. 2011. Currency misalignments and optimal monetary policy: A reexamination. *American Economic Review* 101: 2796–822. [CrossRef]

Firdmuc, Jarko, and Roman Horváth. 2007. *Volatility of Exchange Rates in Selected New EU Members: Evidence from Daily Data*. CESifo Working Paper, No. 2107. Munich: Center for Economic Studies and ifo Institute (CESifo).

Fratzscher, Marcel, Dagfinn Rime, Lucio Sarno, and Gabriele Zinna. 2015. The scapegoat theory of exchange rates: The first tests. *Journal of Monetary Economics* 70: 1–21. [CrossRef]

Frenkel, Jacob A., and Morris Goldstein. 1991. Exchange Rate Volatility and Misalignment: Evaluating Some Proposals for Reform. In *Evolution of the International and Regional Monetary Systems*. Edited by Alfred Steinherr and Daniel Weiserbs. London: Palgrave Macmillan, pp. 185–220.

Frenkel, Michael, Christian Pierdzioch, and Georg Stadtmann. 2005. The effects of Japanese foreign exchange market interventions on the yen/U.S. dollar exchange rate volatility. *International Review of Economics & Finance* 14: 27–39. [CrossRef]

Furceri, Davide, and Aleksandra Zdzienicka. 2012. How costly are debt crises? *Journal of International Money and Finance* 31: 726–42. [CrossRef]

Georgiadis, Georgios, and Feng Zhu. 2019. *Monetary Policy Spillovers, Capital Controls and Exchange Rate Flexibility, and the Financial Channel of Exchange Rates*. ECB Working Paper Series; Frankfurt: European Central Bank. [CrossRef]

Ghosh, Atish, Anne-Marie Gulde, and Holger Wolf. 2003. *Exchange Rate Regimes: Choices and Consequences*. Cambridge: MIT Press.

Giannellis, Nikolaos, and Athanasios Papadopoulos. 2011. What causes exchange rate volatility? Evidence from selected EMU members and candidates for EMU membership countries. *Journal of International Money and Finance* 30: 39–61. [CrossRef]

Haddad, Mona, and Cosimo Pancaro. 2010. Can Real Exchange Rate Undervaluation Boost Exports and Growth in Developing Countries? Yes, But Not for Long. In *Economic Premise*. Washington, DC: World Bank, Available online: https://openknowledge.worldbank.org/handle/10986/10178 (accessed on 5 February 2021).

Han, Yujing. 2020. The impact of exchange rate fluctuation on economic growth—Empirical studies based on different countries. *Advances in Economics, Business and Management Research* 146: 29–33.

Hau, Harald. 2002. Real exchange rate volatility and economic openness: Theory and evidence. *Journal of Money, Credit, and Banking* 34: 611–30. Available online: https://www.jstor.org/stable/3270734 (accessed on 5 February 2021).

Hausmann, Ricardo, Lan Pritchett, and Dani Rodrik. 2005. Growth Accelerations. *Journal of Economic Growth* 10: 303–29. [CrossRef]

Holland, Márcio, Flávio Vieira, Cleomar Gomes da Silva, and Luiz Carlos Bottecchia. 2013. Growth and exchange rate volatility: A panel data analysis. *Applied Economics* 45: 3733–41. [CrossRef]

Jamil, Muhammad, Erich W. Streissler, and Robert M. Kunst. 2012. Exchange rate volatility and its impact on industrial production, before and after the introduction of common currency in Europe. *International Journal of Economics and Financial Issues* 2: 85–109.

Janus, Thorste, and Daniel Riera-Crichton. 2015. Real exchange rate volatility, economic growth and the euro. *Journal of Economic Integration* 30: 148–71. [CrossRef]

Juhro, Solikin M., and Dinh Hoang Bach Phan. 2018. Can economic policy uncertainty predict exchange rate and its volatility? Evidence from ASEAN countries. *Bulletin of Monetary Economics and Banking* 21: 251–68. [CrossRef]

Justesen, Mogens K. 2008. The effect of economic freedom on growth revisited: New evidence on causality from a panel of countries 1970–1999. *European Journal of Political Economy* 24: 642–60. [CrossRef]

Karadam, Duygu Yolcu, and Erdal Ozmen. 2016. *Real Exchange Rates and Growth*. ERC Working Papers, No. 1609. Ankara: Middle East Technical University.

Kočenda, Evzen, and Juraj Valachy. 2006. *Exchange Rate Volatility and Regime Change: Visegrad Comparison*. IPC Working Paper Series, No. 7; Ann Abror: University of Michigan.

Krugman, Paul. 1988. *Deindustrialization, Reindustrialization, and the Real Exchange Rate*. NBER Working Papers, No. 2586. Cambridge: National Bureau of Economic Research.

Lyócsa, Štefan, Peter Molnár, and Igor Fedorko. 2016. Forecasting Exchange Rate Volatility: The Case of the Czech Republic, Hungary and Poland. *Czech Journal of Economics and Finance* 66: 453–75.

MacDonald, Ronald, and Flávio Vieira. 2010. *A Panel Data Investigation of Real Exchange Rate Misalignment and Growth*. CESifo Working Paper, No. 3061. Munich: Center for Economic Studies and ifo Institute (CESifo).

McAleer, Michael, and Christian M. Hafner. 2014. A one line derivation of EGARCH. *Econometrics* 2: 92–97. [CrossRef]

McKinnon, Ronald. 1963. Optimum Currency Areas. *The American Economic Review* 53: 717–25.

Miletić, Siniša. 2015. Modeling exchange rate volatility in CEEC countries: Impact of global financial and European sovereign debt crisis. *Megatrend Review* 12: 105–22. [CrossRef]

Morina, Fatbardha, Eglantina Hysa, Uğur Ergün, Mirela Panait, and Marian Voica. 2020. The effect of exchange rate volatility on economic growth: Case of the CEE countries. *Journal of Risk and Financial Management* 13: 177. [CrossRef]

Morvillier, Florian. 2020. Do currency undervaluations affect the impact of inflation on growth? *Economic Modelling* 84: 275–92. [CrossRef]

Mussa, Michael. 1986. Nominal exchange rate regimes and the behaviour of real exchange rates: Evidence and implications. *Carnegie-Rochester Conference Series on Public Policy* 25: 117–214. [CrossRef]

Newey, Whithney K., and Kenneth D. West. 1994. Automatic Lag Selection in Covariance Matrix Estimation. *The Review of Economic Studies* 61: 631–53. [CrossRef]

Obstfeld, Maurice, and Kenneth Rogoff. 1995. Exchange Rate Dynamics Redux. *The Journal of Political Economy* 103: 624–60. [CrossRef]

Papadamou, Stephanos, Moïse Sidiropoulos, and Eleftherios Spyromitros. 2016. Central bank transparency and exchange rate volatility effects on inflation-output volatility. *Economics and Business Letters* 5: 25–133. [CrossRef]

Petracchi, Cosimo. 2021. *The Mussa Puzzle: A Generalization*. BU Working Papers, 2021-001. Providence: Brown University.

Pham, Thi Hong Hanh. 2018. *Liquidity and Exchange Rate Volatility*. LEMNA Working Paper, EA 4272. Nantes: University of Nantes.

Ribero, Rafael S. M., John S. L. McCombie, and Gilbeto Tadeu Lima. 2020. Does real exchange undervaluation really promote economic growth? *Structural Change and Economic Dynamics* 52: 408–17. [CrossRef]

Rodrik, Dani. 2008. The real exchange rate and economic growth. *Brookings Papers on Economic Activity* 39: 365–412. [CrossRef]

Schnabl, Gunther. 2009. Exchange rate volatility and growth in emerging Europe and East Asia. *Open Economies Review* 20: 565–87. [CrossRef]

Straub, Roland, and Ivan Tchakarov. 2004. *Non-Fundamental Exchange Rate Volatility and Welfare*. ECB Working Paper Series, 328; Frakfurt: European Central Bank.

Tsangarides, Charalambos G. 2012. Crisis and recovery: Role of the exchange rate regime in emerging market economies. *Journal of Macroeconomics* 34: 470–88. [CrossRef]

Uzelac, Ozren, Milivoje Davidovic, and Marijana Dukic Mijatovic. 2020. Milivoje Davidovic, and Marijana Dukic Mijatovic. 2020. Legal Framework, Political Environment and Economic Freedom in Central and Eastern Europe: Do They Matter for Economic Growth? *Post-Communist Economies* 32: 697–725. [CrossRef]

Weber, Christoph S. 2017. *The Effect of Central Bank Transparency on Exchange Rate Volatility*. BGPE Discussion Paper, 174. Nürnberg: Friedrich-Alexander-Universität Erlangen-Nürnberg.

risks

Article

Systemic Illiquidity Noise-Based Measure—A Solution for Systemic Liquidity Monitoring in Frontier and Emerging Markets

Ewa Dziwok [1] and Marta A. Karaś [2,*]

1 Department of Applied Mathematics, University of Economics in Katowice, 40-287 Katowice, Poland; ewa.dziwok@ue.katowice.pl
2 Department of Financial Investments and Risk Management, Wroclaw University of Economics and Business, 53-345 Wrocław, Poland
* Correspondence: marta.karas@ue.wroc.pl

Abstract: The paper presents an alternative approach to measuring systemic illiquidity applicable to countries with frontier and emerging financial markets, where other existing methods are not applicable. We develop a novel Systemic Illiquidity Noise (SIN)-based measure, using the Nelson–Siegel–Svensson methodology in which we utilize the curve-fitting error as an indicator of financial system illiquidity. We empirically apply our method to a set of 10 divergent Central and Eastern Europe countries—Bulgaria, Croatia, Czechia, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, and Slovakia—in the period of 2006–2020. The results show three periods of increased risk in the sample period: the global financial crisis, the European public debt crisis, and the COVID-19 pandemic. They also allow us to identify three divergent sets of countries with different systemic liquidity risk characteristics. The analysis also illustrates the impact of the introduction of the euro on systemic illiquidity risk. The proposed methodology may be of consequence for financial system regulators and macroprudential bodies: it allows for contemporaneous monitoring of discussed risk at a minimal cost using well-known models and easily accessible data.

Keywords: systemic risk; systemic illiquidity; liquidity crisis; parametric models; quantitative methods; emerging markets; frontier markets; CEE

JEL Classification: G12; G28; G32; C58; E44

Citation: Dziwok, Ewa, and Marta A. Karaś. 2021. Systemic Illiquidity Noise-Based Measure—A Solution for Systemic Liquidity Monitoring in Frontier and Emerging Markets. *Risks* 9: 124. https://doi.org/10.3390/risks9070124

Academic Editor: Paolo Giudici

Received: 1 April 2021
Accepted: 4 June 2021
Published: 1 July 2021

1. Introduction

In the last decade, with the global financial crisis, followed by the European debt crisis, the subsequent economic stagnation, and the current pandemic, many shortcomings have been highlighted in systemic risk monitoring, including the underestimation of illiquidity risk. Since then, many papers have proposed various methods of liquidity risk measurement at the macroscale. However, these methods were developed for and applied to advanced economies with mature financial systems.

Unfortunately, when one wants to quantify liquidity risk in frontier and emerging financial markets[1], the task is not that simple. The specificity and scarcity of data available in such systems render the mentioned methods unusable. To this end, this study aims to fill the existing research gap in systemic liquidity analysis by proposing a novel approach to the task. We developed and applied a measure based on the well-known Nelson–Siegel–Svensson methodology; however, we utilized data from the curve-fitting error to obtain, after technical modifications, a daily indicator of systemic illiquidity that is applicable not only to emerging markets, but also to frontier markets.

For the sample, we selected a set of 10 divergent Central and Eastern European (CEE) countries, with seven frontier (Bulgaria, Croatia, Estonia, Latvia, Lithuania, Romania, and Slovakia) and three emerging financial markets (Czechia, Hungary, and Poland).

Our analysis showed increased systemic liquidity risk in the three periods of global disturbance: the global financial system crisis, the European public debt crisis and the COVID-19 pandemic. However, several recorded risk peaks corresponded in time to events that were only locally significant for financial stability. Additionally, the results allowed us to identify three divergent sets of countries with different systemic liquidity risk characteristics. Our analysis also illustrated the impact of the introduction of the euro on systemic illiquidity risk.

The paper layout is as follows. In Section 2, we discuss the role of liquidity and its imbalances in systemic risk materialization, and then give an overview of the results of systemic-illiquidity-focused theoretical research, as well as the models and methods proposed to measure this phenomenon. The studies of liquidity effects are categorized by focus and the sectors of the financial system. This is supplemented with an overview of the empirical papers studying the discussed illiquidity effects. We also present a comprehensive overview of liquidity risk indicators and discuss 13 different methods of measuring systemic illiquidity. We then discuss why these methods are inapplicable to the CEE region. We devote Section 3 to parametric models and their application in systemic liquidity analysis. We also describe our proposition to adopt Nelson–Siegel–Svensson methodology for systemic liquidity measurement, using the curve-fitting error as an indicator of financial system illiquidity. Section 4 presents the empirical results obtained using our Systemic Illiquidity Noise (SIN)-based measure for 10 selected CEE countries, using interbank market data and the information embedded in the interest rate term structure. Section 5 provides the conclusion.

2. Liquidity in Systemic Risk

Market turbulence and liquidity in the financial system are very closely related. When analyzing systemic liquidity, one should consider the level of market liquidity and its resilience. Both of these aspects decide how (and with what consequences) the financial system will withstand a possible liquidity shock. When liquidity is low, it also tends to change in a volatile manner and is prone to sudden drops. In such circumstances, the prices become less informative, diverging from fundamentals and increasing market volatility further. In extreme circumstances, this leads to systemic outcomes.

A high level of market liquidity means "the ability to rapidly execute sizable transactions at a low cost and with a limited price impact" (IMF 2015, p. 49). Since liquidity influences the efficiency of fund transfers from savers to borrowers, stable and adequate liquidity in the financial system fosters economic growth. Even more importantly, resilient market liquidity is crucial for the ability to dilute instances of instability, as "it is less prone to sharp declines in response to shocks" (IMF 2015, p. 49). Importantly, even seemingly ample market liquidity may be fragile if its sources are undiversified (IMF 2015, pp. 49–87); for instance, if the main source of liquidity are several banks with a similar risk profile. This concern is particularly important for the frontier and emerging financial markets, in which the financial system structure is still not that diversified.

In general, market liquidity is likely to be high, if (IMF 2015, p. 50):

- Market infrastructures are efficient and transparent, leading to low search and transactions costs;
- Market participants have easy access to funding;
- Risk appetite is abundant;
- A diverse investor base ensures that factors affecting individual investors do not translate into broader price volatility.

However, all of the mentioned conditions evaporate from the financial system when it is faced with a crisis. Search frictions related to the lack of liquidity may include information asymmetry between dealers and traders, communication breakdowns, uncertainty about the counterparty's ability to carry out the trade, and dealer failures. These frictions are especially significant in extreme situations, when they "may lead to considerable market illiquidity, even when funding liquidity is high" (IMF 2015, p. 51). Furthermore, because financial systems are elaborate networks, liquidity effects tend to be self-reinforcing, which creates a range of multiple equilibria with different liquidity characteristics (Buiter 2008).

A shortage of liquidity has obvious negative consequences. However, benign cyclical conditions may mask liquidity risks (Bessembinder et al. 2011). Moreover, ample market liquidity driven by cyclical factors may promote excessive risk-taking (Clementi 2001). It may also lead financial institutions to build up unsustainable leverage, with negative consequences for financial stability (Geanakoplos 2010). Similarly, irrational overconfidence in highly liquid markets favors trading frenzies, amplifying asset price bubbles (Scheinkman and Xiong 2003; Brunnermeier 2008). This situation appeared after the crisis in 2007–2009, when an increase of control, and lower rates moved the lending industry towards the nonbanking industry. Furthermore, the COVID-19 crisis revealed the scale of leverage in nonbanking investment (Duffie 2020; Vivar et al. 2020; Vassallo et al. 2020). This rapid growth of the nonbanking sector, however, has rendered traditional monetary policy tools, such as increasing the money supply to banks and accepting broader collateral, insufficient.

2.1. Systemic Illiquidity: Research and Existing Measures

Multiple illiquidity-related effects lead to systemic risk amplification. Table 1 sums up the more prominent literature contributions focused on such liquidity effects and their impacts on systemic risk. We specify these effects by categorizing them in relation to phenomena typical for systemic risk and the financial system sector in which they occurred in the cited studies. We also indicate other sectors that may potentially be affected by the described effects.

The empirical studies on the effects presented above are presented in the papers by Coval and Stafford (2007), Loutskina and Strahan (2009), Aragon and Strahan (2009), and Boyson et al. (2010). Among more recent papers, one may find the study by Banerjee and Mio (2014), who researched the empirical impacts of new liquidity regulation on the banking sector, using the UK as an example. The paper by Chan-Lau et al. (2009) and the IMF's (2009) Global Financial Stability Review contained two network models of interbank exposures, allowing them to assess the network externalities of bank failures using institutional data. In a similar framework, Sapra (2008) found that mark-to-market accounting creates an illiquidity contagion, unlike historical cost accounting. Boss et al. (2004) and Gofman (2015) used network models based on empirical data from the interbank market to model contagion signal transmission in the banking sector.

Table 1. The studies of illiquidity effects categorized by the focus and sector of the financial system.

Systemic Risk Occurrence	Liquidity Effects	Primary Sector of Occurrence	Other Sectors Possibly Affected by the Effect	Authors
Illiquidity exposure	Correlated exposures to illiquidity, free-riding	Banking sector	Shadow banking	Bhattacharya and Gale (1987)
	Maturity rat-race and excessive short-term debt [1]			Brunnermeier and Oehmke (2013)
Illiquidity contagion	Fire sales and their effect on prices	Financial assets markets	Banking sector, shadow banking, investment funds, SIFIs	Shleifer and Vishny (1992)
	Market incompleteness and effects of illiquidity on prices			Allen and Gale (1994, 2000a, 2000b)
	Snowball effect, in which the loss spiral interacts with a margin spiral [1]			Brunnermeier and Pedersen (2009)
	Market illiquidity contagion			Cespa and Foucault (2014)

Table 1. *Cont.*

Systemic Risk Occurrence	Liquidity Effects	Primary Sector of Occurrence	Other Sectors Possibly Affected by the Effect	Authors
Illiquidity-driven crises	Constraints to arbitrage adding to illiquidity	Financial assets markets	-	Shleifer and Vishny (1997)
	Arbitrage affecting liquidity both ways			Gromb and Vayanos (2002)
	Runs caused by mark-to-market accounting		Banking sector, shadow banking, investment funds, SIFIs	Cifuentes et al. (2005)
	Bank runs triggering illiquidity, which triggers further bank runs	Banking sector		Diamond and Rajan (2005)
	Leverage, illiquidity spirals, and financial frictions			Brunnermeier et al. (2013)
				Brunnermeier and Sannikov (2014)
Informationally driven market freezes	Interbank market fragility due to fear of adverse selection	Banking sector	-	Flannery (1996)
	Lack of information about the counterparty risk causes the banks to stop lending to each other upon large shocks			Caballero and Simsek (2013)
	Interbank market freezes caused by information asymmetry			Heider et al. (2015)
	Information asymmetry as a source of repo markets collapse	Financial assets markets	Banking sector, financial markets, shadow banking, investment funds,	Acharya et al. (2011)
	Collateral value vs. its price			Gorton and Ordonez (2014)

The table categorizes papers investigating liquidity risk effects that are relevant for systemic risk. The reported effects are grouped into four types of systemic risk triggers and arranged according to financial system sector or segment. [1] A loss spiral occurs when the losses on a few assets induce the market participants to reduce their positions in many other assets. Then these sales depress market prices, prompting further losses; a margin spiral occurs when market participants apply higher margin requirements because of the reduced market liquidity. Both effects reinforce each other, increasing the pressure to sell more assets (Brunnermeier and Pedersen 2009).

More recent publications related to systemic risk treat illiquidity as the necessary condition for fragility accumulation or for contagion. In relation to market freezes, Afonso et al. (2011) revealed how interbank loans in the US became more sensitive to borrower characteristics during the crisis. Still, they reported no evidence of liquidity hoarding, in contrast to the predictions in the theoretical model by Allen et al. (2009) and to the empirical findings from the interbank markets in the UK (Acharya and Merrouche 2013), and in the euro area (Gabrieli and Georg 2014). On a similar note, Morris and Shin (2012) analyzed toxic asset market freezes caused by the breakdown of common knowledge about maximum losses. In turn, banking panics have been empirically studied by Iyer and Peydro (2011) and Iyer and Puri (2012), among others. Finally, Schrimpf et al. (2020) pointed out the consequences of the leverage and margin spiral that amplified liquidity risk in the euro area during the COVID-19 crisis, which was also emphasized in the recent Financial Stability Review (ECB 2020). All mentioned phenomena have liquidity problems at their core.

2.2. Measures of Systemic Illiquidity—Overview

We will now discuss and categorize the measures proposed by other authors to measure systemic liquidity. They form two vast sets: simple indicators and much more complex—often multifaceted—models.

Indicators are structurally simple constructs built of a few readily observable variables that allow for straightforward interpretation. By virtue, they are most often related to a specific segment of the financial system; therefore, they are not cross-sectional. Among financial soundness indicators (FSIs), one may distinguish current and forward-looking indicators. The first group allows the analysis of the current developments in the financial system, while the second one allows inferences to be drawn about possible future outcomes (see: Berg and Pattillo 1999 or Kumar and Persaud 2001). Sometimes, the same indicator may serve both purposes if analyzed vis-à-vis its historical path (trend) or distribution (quantile).

Nelson and Perli (2007, p. 350) state that the US Federal Reserve was using more than 100 different indicators at the time of their publication. They discussed, for instance, indicators of market liquidity, including bid–ask spreads and volumes (e.g., on bonds, bills, and various derivatives, such as swaps), credit default swap (CDS) spreads, and liquidity premiums (yield on less-liquid security minus yield on highly liquid (benchmark) security). Indicators used by others include interbank market rates, interbank market traffic, and the demand changes for central bank facilities (Afonso et al. 2011).

In relation to the banking sector, there is the basic liquidity ratio (short-term resources vs. short-term liabilities) and other similar ratios, such as quick assets to assets or client deposit ratios (Gersl and Heřmánek 2007). The ECB uses a broad set of indicators to analyze financial soundness, such as the ratio of liquid assets to short-term liabilities (see, e.g., ECB 20). Finally, basic composite indicators are available for advanced financial markets. These include volatility indices, such as the VIX.

A complete list of liquidity-focused indicators is very extensive. However, Jobst (2012) selected the indicators most useful from the systemic risk perspective (Table 2).

Table 2. Liquidity risk indicators.

	Quantity-Based Indicators	**Price-Based Indicators**
Monetary liquidity	Base money and broader monetary aggregates	Policy and money-market interest rates
	Access to central bank liquidity facility (e.g., bidding volume)	Monetary conditions indices
	Foreign exchange reserves	
Funding liquidity	Bank liquidity ratios	Unsecured interbank lending (Libor–OIS spreads)
		Secured interbank lending (repo rates)
	Bank net cash flow estimates	Margins and haircuts on repo collateral
		FX swap basis
	Maturity mismatch measures	Violation of arbitrage conditions (bond–CDS basis, covered interest rate parity)
	Commercial paper market volumes	Spreads between assets with similar credit characteristics
		Qualitative surveys of funding conditions
Market liquidity	Transaction volumes	Bid–ask spreads on selected global assets
		Qualitative fund manager surveys

The table presents existing liquidity risk indicator types categorized in relation to the type of liquidity and the numerical base of the indicator. Indicators are limited to those used in systemic risk analysis. Source: Jobst (2012, p. 13).

The systemic risk perspective requires a broader view that goes beyond a set of individual indicators for individual institutions or markets and allows for a system-level analysis. For this reason, multiple complex measures focused on systemic liquidity have been developed in recent years (see Appendix B). These measures significantly differ in terms of the data requirements and the output they produce. Some models relate to the whole financial system, while others have a narrower focus. There are methods that use the data from a given market segment to capture the liquidity crisis in that same segment. Others use data from one segment to shed light on another one. Finally, there are cross-sectional proposals. For some of the overviewed measures, the link between the measurement method and liquidity is direct (e.g., SRL in Jobst 2014). For others, it is indirect and comes from the theoretical justification of a given measure, rather than from the data per se.

2.3. Measures of Systemic Illiquidity—Empirical Application Possibilities

For any risk measure to be effective, the theoretical assumptions necessary for its use must be fulfilled. In this study, the systemic illiquidity measure must be in line with the fact that during the sample period, the Central and Eastern European financial systems were characterized by:

- Developing (frontier or emerging) markets in terms of the structure (banking sector dominance, with traditional banking products), maturity (affecting data availability and historical data span), and depth (including the limited variety of markets, the size of the stock market, and the numbers and types of existing financial instruments);
- Relatively well-developed economies in terms of the stability of prices (relatively low and stable inflation), currency, capital flows, and monetary policy targets and tools.

Furthermore, timing is critical in systemic risk monitoring. An adequate liquidity risk measure should produce at least a daily frequency time series, because liquidity may evaporate very fast. For the same reason, the input data should also be minimally affected by lags—any data reporting and preprocessing time must be minimal. Finally, the data should also represent all financial institutions that are systemically important (SIFIs) in the given system.

After analyzing almost 60 systemic risk measures found in the literature, we identified only 13 measures focused on liquidity-related turbulence, despite the unargued impact of illiquidity on systemic risk. These are the approaches proposed by Getmansky et al. (2004); Chan et al. (2006); Perotti and Suarez (2011); Khandani and Lo (2011); Severo (2012); Brunnermeier et al. (2014); Jobst (2014); Greenwood et al. (2015); Karkowska (2015); and Duarte and Eisenbach (2019). We analyzed all of them in terms of applicability to the studied CEE region. We describe this process below and illustrate it in Table 3 afterward. We also provide details about each of these measures in Table A1 (Appendix B).

The first step of elimination involved the practical aspects, such as data availability and dependability. For each country in our study, we asked whether solid data required for the calculation of a given measure existed. Several approaches required the data from market segments that were not sufficiently developed in the CEE region. More specifically, they were based on data regarding instruments or indices that were not quoted regularly (or at all) in frontier markets. For emerging markets, even though the data existed, it was too scarce to draw solid conclusions about systemic liquidity. Given the factors discussed above, we eliminated the measures based on hedge fund data (Getmansky et al. 2004; Chan et al. 2006) and the method utilizing derivatives (Severo 2012).

Another question that we asked regarded the facilitation of daily risk monitoring. Market liquidity can evaporate from the markets very fast. Thus, to be useful for systemic risk analysis, a liquidity measure must provide information on a daily basis. Unfortunately, the existing methods focused on the banking sector could not be used to obtain a daily time series. They included the liquidity risk charges proposal by Perotti and Suarez (2011), Liquidity Mismatch Index (Brunnermeier et al. 2014), Jobst's (2014) Systemic Risk-Adjusted Liquidity Model, the Cumulative Distance to Default by Karkowska (2015), and the measure of systemicness by Greenwood et al. (2015) and its expansion by Duarte and Eisenbach (2019). They were incompatible with the goal of creating a daily systemic illiquidity monitoring tool, even though the institutional focus of these measures was proper for the frontier and emerging markets in which banks are the main providers of systemic liquidity.

Frontier stock markets are shallow and the data are scarce, which significantly limits the potential of measures based solely on stock-market data to indicate system-wide liquidity in the CEE region. Therefore, in our empirical analysis, we could not use liquidity-focused measures such as the liquidity factor (Pastor and Stambaugh 2003), the contrarian strategy and price-impact liquidity measures (Khandani and Lo 2011), or the liquidity noise measure by Hu et al. (2013).

In effect, we were unable to identify any ready-made daily frequency systemic illiquidity measure that could be successfully applied in frontier and emerging markets. Therefore, we developed a new measure to fill the existing gap.

Table 3. Analysis of systemic illiquidity measures for applicability to CEE.

Measure	Authors	Is the Application Possible? (Data Limitations)	Is Contemporaneous Measurement Possible? (Issues of Lags and Frequency)	Does it Facilitate Systemic Risk Analysis? (Coverage/Proxying the Whole Financial System)
Liquidity factor	Pastor and Stambaugh (2003)	YES	YES	NO
A set of interpretable parameters	Getmansky et al. (2004)	NO	x	x
Broader hedge-fund-based systemic risk measures	Chan et al. (2006)	NO	x	x
A system of liquidity risk charges (LRCs)	Perotti and Suarez (2011)	YES	NO	x
Contrarian strategy liquidity measure (CSL)	Khandani and Lo (2011)	YES	YES	NO
Price-impact liquidity measure (PIL)	Khandani and Lo (2011)	YES	YES	NO
Systemic Liquidity Risk Index (SLRI)	Severo (2012)	NO	x	x
Daily liquidity noise measure	Hu et al. (2013)	NO	x	x
Liquidity Mismatch Index (LMI)	Brunnermeier et al. (2014)	YES	NO	x
Systemic risk-adjusted liquidity (SRL) model	Jobst (2014)	YES	NO	x
Systemicness	Greenwood et al. (2015)	NO	x	x
Cumulative Distance to Default (CDD)	Karkowska (2015)	YES	NO	x
Aggregate vulnerability (AV) and illiquidity concentration	Duarte and Eisenbach (2019)	NO	x	x

The table presents the step-by-step process used to find a systemic liquidity risk measure by answering three questions (YES/NO). A negative answer to a given question eliminated the measure from further analysis (x).

3. Parametric Models and Their Potential in Systemic Liquidity Analysis

Financial market participants find multiple applications for the estimated yield curve. The first application of the Nelson–Siegel–Svensson methodology took place in the late 1980s, when Nelson and Siegel (1987) described their fitting technique for the first time. They used the estimated yield curve to predict the price of a long-term US Treasury bond. However, the application possibilities were much broader, including modeling the demand functions, testing theories regarding the term structure of the interest rates, and graphic display for informative purposes.

The forward rate, a solution to the differential equation that generates spot rates that are applicable as a forecast, was a main driver for the parsimonious models' exploration and their future popularity. After the introduction of Svensson's (1994, 1995, 1999) extension to the Nelson–Siegel model, in which forward rates are used to indicate market expectations of future interest rates, the model started to be widely used by central banks to estimate market expectations of future rates, as well as depreciation rates.

The reports published by BIS (2005) and ECB (Nymand-Andersen 2018) indicated that the Nelson–Siegel–Svensson model had become the most popular tool used to estimate the term structure of interest rates and market expectations. Additionally, the relatively recent appearance of negative rates called for a revision of term-structure estimation models, rendering various modern approaches inapplicable. However, as Garcia and Carvalho (2019) noticed, despite the negative rates observed in 20 countries, the Nelson–Siegel–Svensson model maintains good prognostic features and seems to be a good option for monetary policy institutions and market players. Other common uses for the structural models include marking-to-market, interest-rate modeling, and portfolio risk-management methods (see, e.g., Martellini et al. 2003 and Choudhry 2018).

Structural models are also utilized for the calculation of systemic risk buffers in the insurance sector. In particular, the latest solvency requirements for economic and regulatory capital purposes suggest using the Nelson–Siegel–Svensson model to determine the ultimate long-term forward rate (UFR) (EIOPA 2017). This change resulted from the study by Zigraiova and Jakubik (2017), which emphasized the benefits of the Nelson–Siegel methodology (EIOPA 2016).

Furthermore, parametric models also have been used in liquidity risk measurement. A good example is the study by Hu et al. (2013), which used the Svensson model on hedge fund returns and currency-carry trade data to create a measure of dispersion (a so-called "noise measure"). They constructed the measure of market noise by calculating the root mean square error between the market and theoretical yields, and applied it as a liquidity risk factor in portfolio risk modeling. Noise in the Treasury market informs about liquidity in the broad market because the Treasury market has low intrinsic noise, high liquidity, and low credit risk; i.e., the noise becomes high when liquidity drops. This particular application shows the potential to use structural models in liquidity risk measurement.

Our idea consists of applying the structural models to measure the liquidity risk of the financial system as a whole. In particular, we used the information about how market yields deviate from the theoretically expected yields (modeled in different ways) in response to market frictions. We postulate that this phenomenon results from the liquidity shortage that manifests in response to systemic events. The main two channels of risk transmission here are information asymmetry and behavioral effects. To obtain information about systemic liquidity in the banking-based financial systems (such as CEE), we applied the measure to the interbank market. Therefore, we used the interbank market data and the information embedded in the interest-rate-term structure.

The term structure of interest rates has informational value for systemic risk analysis. It reacts to the expectations of the market participants, especially in the short term. It changes with changing expected risk premiums for liquidity and default risk, and it depends on risk-aversion characteristics and preferences of the market participants. It also reacts to central banks' activities (as proven inter alia by Lucas 1978; Cox et al. 1981; Shiller and McCulloch 1990; and Mehra 1995). Therefore, it is an essential source of information about the stability of the financial market, and in a broader sense, the financial system affected by this market.

The money market is a component of the financial market of assets with a maturity not exceeding one year, and by definition, it is a wholesale market with its core in interbank transactions. The interest rate on loans in the developed interbank market is a reference system for determining fixed-asset prices, as well as for loan contracts in the entire economy. Therefore, a well-functioning interbank market plays a key role in the transmission of monetary policy and the redistribution of liquid assets (Schmitz 2011).

Central banks are interested in constructing interbank market yields mainly because of the information about the forward rates embedded in them. In fact, many financial instruments' parameters in the CEE region are based directly on the interbank rates (Interbank Offered Rates—"IBOR"). Successful monetary-policy transmission involves a linkage between the banks' operating target and the interbank lending rate. Thus, the conditions in the interbank lending market have significant effects on monetary-policy transmission. The weakening of this link creates a significant challenge for central banks and is one of the factors that motivated the creation of extraordinary liquidity and credit facilities.

The importance of the money market in maturity transformation was relatively small before 1980. However, in recent decades banks have increasingly replaced government-guaranteed individual deposits with uninsured wholesale deposits from the interbank money market. For example, their value in the US had increased by 160% by the year 2000 (Feldman and Schmidt 2001). At the same time, the loans granted to other banks in many countries have had a growing share in assets. For instance, at the end of 2005, interbank loans accounted for 29% of Swiss and 25% of German banks' assets (Upper 2007). By the end of 2006, the interbank assets exceeded their shares in five out of eight developed countries. In many European banks, interbank assets accounted for five times or more than the equity (Upper 2011). Moreover, during COVID-19, the balance sheets for the biggest central bank have increased by 50%, making interbank loans a potential contagion channel.

Indeed, one of the most characteristic symptoms of the global financial crisis was the increase in interbank market tensions, which manifested through a decrease in the turnover and a sharp increase in interest rates and spreads. Explaining this mechanism, Lubiński (2013, p. 22) articulated that "the contribution of the interbank money market to the stability of the system boils down to facilitating banks' liquidity management."

Banks' resilience to liquidity shocks and their ability to lend to each other is crucial for macroeconomic stability. The tensions in the interbank market limit this ability. Nonetheless, interbank loans are generally not included in the macroprudential regulations against overexposure and concentration, especially when groups of banks are concerned. Due to high flows in currencies and derivatives, mutual exposure of financial institutions is treated as an element of the sector's specificity, and the resulting exposure to direct contagion is considered its attribute (Blåvarg and Nimander 2002). In addition, due to the lack of appropriate regulation, information on interbank exposures is usually not available, and market participants only have an approximate idea of the actual scale of dependence. For this reason, they do not know which banks have claims against bankruptcy, which may lead to a general undermining of trust (Schoenmaker 1996).

As uninsured money-market instruments are associated with higher risk, they react to changes more quickly. Thus, their interest rates are more variable than the interest on regular deposits (Mishkin 2007). This market is also most sensitive to the loss of confidence that accompanies turbulence. This is usually immediately reflected in widening spreads, lowering numbers of transactions, and the shortening of their maturity. The market may also be ineffective due to the asymmetry of information, its incompleteness, or the market power of some entities (especially SIFIs). During turbulence, solvent banks' liquidity problems may lead to insolvency because such banks cannot obtain sufficient interbank loans, and they must sell long-term assets below their fundamental value.

Regardless of the nature of the adverse stimulus, interbank loans may contribute to contagion through an associated flow of information and the linking of portfolios and balance sheets. In the first case, the contagion results from passing information from more liquid markets or markets in which prices are previously disclosed to others. Based on unfavorable information about one institution, business entities draw conclusions about the threat to others (which may be correct or not) (Kiyotaki and Moore 2002). Additionally, unfavorable interpretation arises from the observation that individual institutions' portfolios and balance sheets are connected, while assets and liabilities must be equal.

There are several methods proposed by various authors that use the interbank market as a source of information about systemic risk. Among these, one may find the aforementioned paper by Hu et al. (2013), but also the network model proposed by Elsinger et al. (2006) or the PA–CA–BA measure developed by Drehmann and Tarashev (2011). Among the most interesting empirical studies of the interbank market in terms of systemic risk is the publication by Allen and Gale (2000a, 2000b), who found that the interbank market's susceptibility to adverse liquidity shocks depends on its structure.

4. Empirical Application of the Systemic Illiquidity Noise-Based Measure

In a preliminary phase of this research, we successfully applied the proposed Systemic Illiquidity Noise-based measure, SIN, to the Polish interbank market (Dziwok 2017). This small study showed that the Polish market is sufficiently sensitive to new information inflow to apply a "noise-type" liquidity measure based on parametric models. Using daily WIBOR data and applying the Nelson–Siegel–Svensson models to limited time horizons, Dziwok (2017, pp. 34–35) confirmed that the model was suitable for analyzing systemic liquidity. The measure detected increased illiquidity-driven volatility in the Polish financial system around the global financial crisis. Karaś (2019) confirmed these results in a longer horizon study (for the years 2006–2018).

This method is advantageous for contemporaneous liquidity measurement. For instance, the Basel III liquidity criteria (LCR and NSFR measures) are based on the asset–liability position of the banking sector, and therefore they are prone to a time lag, because the data needs to be gathered, recalculated, and delivered (published) before the measures can be calculated. SIN depicts the current condition of the interbank almost instantly. This makes it a better indicator of financial system liquidity for systemic risk analysis.

4.1. Methodology

Let us assume that τ is the point in time when the curve is constructed. Then, the value of a zero-coupon instrument at maturity is equal to one: $P_t(\tau, t) = 1$, where t is maturity and capital growth takes a continuous form. A spot rate could be described as the average of instantaneous forward rates:

$$i(\tau, t) = \frac{1}{t - \tau} \int_{\tau}^{t} f_{\tau}(s) ds. \tag{1}$$

The value of a zero-coupon instrument at the moment τ when the curve is constructed $P_{\tau}(\tau, t)$ is equal to the discount factor $\delta(\tau, t)$ and follows the formula (de La Grandville 2001):

$$P_{\tau}(\tau, t) = \delta(\tau, t) = e^{-i(\tau,t)\cdot(t-\tau)} = e^{-\int_{\tau}^{t} f_{\tau}(s)ds}. \tag{2}$$

In a special case, when the moment of the rate's construction is $\tau = 0$, and assuming that:

$$P_{\tau}(\tau, t) = P_0(0, t) \equiv P(t),\ \delta(\tau, t) = \delta(0, t) \equiv \delta(t),\ f_{\tau}(s) = f_0(s) \equiv f(s), \tag{3}$$

we may simplify Formula (2) into the following form:

$$P(t) = \delta(t) = e^{-i(0,t)\cdot t} = e^{-\int_0^t f(s)ds}. \tag{4}$$

As outlined above, we can model the yield curve by constructing a continuous function based on existing discrete market data, using the functional relationship between the discount factor, the spot rate, and the instantaneous forward rate (2).

The existing interrelation among a discounting factor $\delta(t)$, a spot rate $i(0,t)$, and an implied forward rate $f(s)$ enables us to search for only one of them. When one rate is established, the level of the others is received through equation (James and Weber 2000).

We divided the yield-curve construction process into several phases, including selecting the data, building the cash flow matrix, defining the theoretical price vector, and establishing the estimation criteria (to fit the curve to real data).

Phase 1: data selection. For the moment $\tau = 0$ a set of k zero-coupon assets with different maturities is chosen, for which the present values are P_l for $l = 1, 2, \ldots, k$, while the face value equals 1.

Phase 2: building of the cash-flow matrix. For the collected zero-coupon data, a diagonal cash-flow matrix $\mathbf{C}$ is constructed, for which the elements correspond to the payments.

Phase 3: a vector of theoretical prices. A vector of theoretical prices $\overline{\mathbf{P_l}} = \{\overline{P_l}\}_{l=1,2,\ldots,k}$ is described as the product of the cash-flow matrix and the estimators of discount factors (interrelated with parameters through Formula (4)):

$$\begin{bmatrix} \overline{P_1} \\ \overline{P_2} \\ \vdots \\ \overline{P_k} \end{bmatrix} = \mathbf{C} \cdot \left[\overline{\delta}(t_1),\ \overline{\delta}(t_2), \cdots, \overline{\delta}(t_k)\right]^T \tag{5}$$

Phase 4: the fitting criteria. The parameters are found by minimizing the mean square error (MSE) between theoretical and market data. The measure could involve either prices or yields that allow the function $\Psi(\cdot)$ to be minimized, such as:

$$\Psi(P) = \sum_{l=1}^{k} \left(P_l - \overline{P_l}\right)^2 \rightarrow \min, \tag{6}$$

or

$$\Psi(Y) = \sum_{l=1}^{k} \left(i_l - \overline{i_l}\right)^2 \rightarrow \min. \tag{7}$$

One of the main reasons for the extensive use of the parametric model for yield-curve modeling is its plainness and a limited number of estimated parameters. The Nelson–Siegel–Svensson model shows the instantaneous forward rate as a function of six parameters, β_0, β_1, β_2, β_3, v_1, v_2, such that:

$$f(s) = \beta_0 + \beta_1 \cdot e^{-\frac{s}{v_1}} + \beta_2 \cdot \frac{s}{v_1} \cdot e^{-\frac{s}{v_1}} + \beta_3 \cdot \frac{s}{v_2} \cdot e^{-\frac{s}{v_2}} \tag{8}$$

The spot rate $i(0,t)$ received through Formula (1) has the following form:

$$i(0,t) = \beta_0 + (\beta_1 + \beta_2) \frac{1 - e^{-\frac{t}{v_1}}}{\frac{t}{v_1}} - \beta_2 \cdot e^{-\frac{t}{v_1}} + \beta_3 \cdot \left(\frac{1 - e^{-\frac{t}{v_2}}}{\frac{t}{v_2}} - e^{-\frac{t}{v_2}} \right) \tag{9}$$

Through the description of the discount factor $\delta(t) = e^{-i(0,t) \cdot t}$ (Formula (4)), the spot rate (Formula (9)), and the theoretical vector of prices (Formula (5)), the estimation process (Formula (6)) aims to find parameters that minimize the function $\Psi(\cdot)$, which involves imposing a set of specified initial conditions during the estimation process on the parameter vector. For each point of the estimated curve, the error value (i.e., the noise) reflects the degree of deviation between the theoretical and the market rates, regardless of the length of the transaction.

In the final step, we introduce a modification. For short-term instruments, their prices are similar, despite the significant differences in yields. This relation results from the nonlinear relationship between the price and the yield to maturity, which shows that for short terms, the asset price goes to unity (face value) (Schich 1997). To maximize the potential of the error function $\Psi(Y)$ to serve as a noise-based illiquidity measure, we gave higher weights to errors in the prices of instruments with a shorter maturity. To improve the quality of $\Psi(P)$ in this way, we used the concept of duration (Fabozzi 2007).

Correcting the price-error function via the inverse of the duration allows the quality of matching to be increased for instruments with shorter maturities. After this modification, the yield-curve estimation requires finding the parameters that minimize the following function:

$$\Psi(P/D) = \sum_{l=1}^{k} \left(\frac{P_l - \overline{P_l}}{D_l} \right)^2 \rightarrow \min \tag{10}$$

In such a form, the noise-based measure better signals these deviations from the theoretical curve that are informative of sudden changes in the interbank market systemic liquidity position. This characteristic makes SIN even more useful for systemic risk measurement.

4.2. Data and Empirical Results

We applied the presented computational methodology to selected Central and Eastern European (CEE) countries, including Bulgaria, Croatia, Czechia, Estonia, Hungary, Latvia, Lithuania, Poland, Romania, and Slovakia. We used interbank data in the form of the Interbank Offered Rates. Data span encompasses years 2006 to 2020.

Typically, in the CEE region, one can build a term structure of the spot interest rates for the interbank market (interbank deposit rates, Treasury bonds, and bills) and forward interest rates (interest-rate-based derivatives). The interbank deposit market is characterized by the ease of conducting transactions, their growing volume in the studied period, and the domination of short-term maturities (between one day (overnight, O/N) and one year). However, these characteristics relate only to the "IBOR" reference rates in the region. In several countries, continuous data for derivatives (e.g., FRAs, swaps) do not exist. Hence, to keep the estimations comparable, we limited the data used in yield estimations to IBORs for all the sampled countries.

Figures 1–10 present the results. Generally, we can say that in all the studied cases, the SIN measure signaled increased risk in two periods between 2007 and early 2010, as well as between late 2010 and 2013. This observation corresponds to the unfolding of the global financial crisis and the sovereign debt crisis in Europe, validating the sensitivity of the SIN measure to the clear-cut financially driven systemic crises in the study period. We also observed a period of increased liquidity risk in the euro area during the ongoing COVID-19 pandemic.

We also observed for all the analyzed markets that following the global financial crisis, liquidity volatility and illiquidity risk gradually fell to comparatively low levels around 2015. At that time, the macroprudential regulations aimed at the elimination of systemic problems from the interbank market were also introduced. The fact that these tools were, to a point, successful is visible in Figures 1–10. Although the risk was not eliminated in total, the markets entered the economic crisis caused by the pandemic in a completely different state of liquidity than was the case for the global financial crisis and the public debt crisis. The markets were characterized by a high shock-absorption capacity this time around.

The SIN measure indicated a set of characteristic differences between the studied countries. The differences corresponded to the scale of risk and the specific timing of it. The results also pointed to country-specific periods of higher risk, which seemed to be driven by local events. Below we discuss the results in more detail, separately for the emerging markets and for the frontier markets.

In this study, we analyzed three markets that were classified as emerging during the study period, namely the Czech, Hungarian, and Polish markets. These markets differed quite significantly. Poland had a much bigger market than the other two countries, while the currency risk issues were most pronounced in the Hungarian market, which is in a closer geopolitical proximity to the frontier markets of Bulgaria and Romania. Furthermore, Hungary and Poland had a bigger market share of domestic institutions in the banking sector than Czechia. All of these characteristics were significant for materialization of systemic illiquidity risk. Figures 1–3 present the results for the emerging markets.

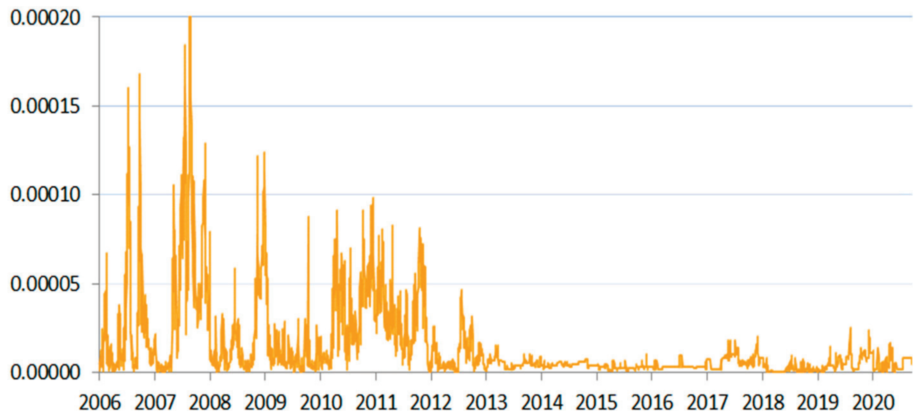


Figure 1. SIN measure for Czechia between 2006 and 2020.

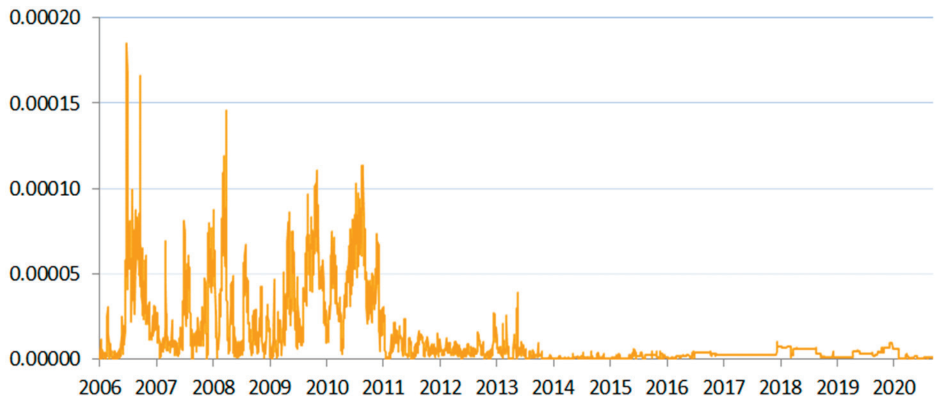


Figure 2. SIN measure for Poland between 2006 and 2020.

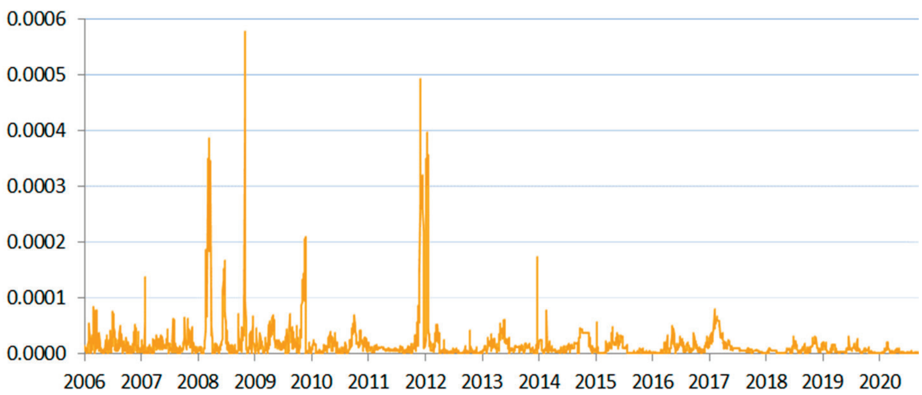


Figure 3. SIN measure for Hungary between 2006 and 2020.

Poland and Czechia showed very similar risk profiles as far as systemic liquidity was concerned. Given the difference in size of their domestic interbank markets, such a similarity may have resulted from the high convergence with the financial sector of the developed European Union countries. Adequately, we observed high risk in 2010 and 2011, around the time of the European debt crisis, when neither Poland's nor Czechia's budgets were significantly hit by it[2]. We observed a kind of contagion by association with the western EU states. It should be noted that the presence of foreign-owned and foreign-parented banks in the CEE region was a significant contagion channel in such circumstances.

This observation was also in line with the risk peaks in both countries before the global financial crisis (2007), coinciding with the intensification of the liquidity pressures in foreign markets. Czech National Bank (CNB 2008, p. 35) concluded that "market participants behavior [was] over-sensitive or herd-like", and that market participants were keen to "assign an unhealthy higher weight to market liquidity" in risk assessments, which resulted in severely increased volatility. The composite liquidity indicator calculated by the central bank for that period was at the lowest level since 2000 (see Box 3.1, CNB 2007). On a similar note, Czech National Bank explained the risk in the period between 2010–2011 as caused by the "persisting increased nervousness in the markets" (CNB 2010–2011, p. 40). At that time, the money market was still significantly affected by the recent crisis, as financial institutions were hoarding liquidity in the face of prevailing uncertainty. This constatation also held for Poland and Hungary.

For Czechia, we saw another significant risk peak in 2006. At that time, commercial banks were intensely speculating on the interbank market in the expectation of the increase in the base rates. In response, CNB decreased the frequency of the repo transactions from daily to three weeks. This resulted in even higher volatility that lasted for about four months. Admittedly, the central bank confirmed that its action affected market liquidity negatively (CNB 2007, p. 28).

The Polish interbank market, similar to the other studied markets, was characterized by a decreasing volume of trading, falling volatility of the spreads, and declining differentiation of the interest rates in the period after the financial crisis. However, for Poland, the years 2015–2017 were exceptional in this respect—almost zero volatility and no differentiation of daily quotations of banks participating in fixing were observed. This period was also the time of maintaining constant interest rates by the Monetary Policy Council (Kapuściński and Stanisławska 2017).

In Hungary, we observed several exceptionally high peaks in systemic illiquidity risk. While the increased risk in 2006 and 2014 corresponded to the events described above that were common for all three emerging markets analyzed in this work, it was significantly amplified by domestic events. Among those, we may enumerate currency instability and related problems with entering the euro area, slowdown of the economic growth that stopped the convergence towards the EU's developed economies, and fiscal problems resulting in fiscal emergency measures (Valentinyi 2012).

Negative market reactions were also observed in relation to various banking tax changes that were introduced between 2014 and 2021, as all of these changes put a strain on the sector, which was much bigger than in other CEE countries. The year 2014 was especially hard for the Hungarian banking sector, as 10 bank collapses took place then. There were media suggestions that this resulted from political decisions related to longer-term vision of the renationalization of the banking sector (Balogh 2015). The most significant closures of banks took place in January and in December of that year.

A vital difference appeared between the Visegrad Group and the rest of the analyzed countries. In Poland, Czechia, Hungary, and Slovakia (Figures 1–3 and 10), the scale of SIN was one order of magnitude smaller than in other countries. The described scale of risk also corresponded to the relative strength of each local currency. Poland had the most stable local currency among the studied countries, which corresponded to the smallest (other than euro area's) registered scale of the SIN measure (for comparison, see Figure A1 in Appendix A).

In our study, we also analyzed seven frontier markets. Three of them were developing economies with relatively weak currencies (compared to other European countries): Croatia, Bulgaria and Romania. They were the most aggressively economically developing countries in the sample with the least stable flow of funds from abroad. In addition, for these countries, several domestic crisis-like events may be enumerated (see Kubinschi and Barnea 2016; Andrieş et al. 2018; Barkauskaite et al. 2018; and Karaś and Szczepaniak 2019). It was for these countries that we observed the largest volatility of the SIN measure. Figures 4–6 show the results of systemic illiquidity measurements for these frontier markets.

When analyzing the results obtained for Bulgaria (Figure 4), we noticed several periods of systemic illiquidity other than the biggest peak at the time of the global financial crisis. Of special significance was the bank run that took place in 2014. Corporate Commercial Bank AD, which was affected by this run, was the 4th-largest bank in Bulgaria at that time. It had problems already in the late 2012, and it had a negative audit results in mid-2013. The bank's assets were frozen in June 2014, but no resolution mechanism was put in place immediately. At that time, Bulgaria was in a political crisis, having had five government changes between 2013 and 2014, noting a rise in the deficit from 1.2% to 3.4% and an episode of deflation (BNB 2014). To avoid a bigger financial market run, Bulgaria rescued First Investment Bank, which was reputationally affected by the problems of Corporate Commercial Bank, but was otherwise in a relatively sound condition. This action likely stopped the further contagion effect of the loss of confidence in the banking sector and allowed the restoration of liquidity over time.

Another—much smaller, but still quite significant—peak in liquidity risk was recorded for Bulgaria in 2016. The peak coincided with the increase of conservative prudential liquidity measures for the banking sector. In the financial stability assessment of Bulgaria (IMF 2017, p. 10), we may read that in June 2016, Bulgarian banks had a liquid-assets ratio at a level of 31% (11% above the prudential requirement). In 2020, the liquidity-coverage ratio in the Bulgarian banking sector was at a level of over 260%, much above the regulatory requirements (Radev 2020, p. 1).

The almost flat shape of the SIN measure recorded for the period of the COVID-19 pandemic corresponded to the liquidity-providing operations of the government, including several rescue packages for businesses, as well as the moratorium on deferral of loan repayments possible on the basis of the EU-wide regulatory framework established by the European Banking Authority (EBA) in its guidelines on legislative and nonlegislative moratoria on loan repayments applied in the light of the COVID-19 crisis (EBA 2020).

This EU-wide solution, which affected all the countries analyzed in this paper, has had a positive risk effect in the short term, but is likely to cause delayed negative effects for the financial system in the medium term, when the moratorium expires and loan-performance deterioration will intensify in the region.

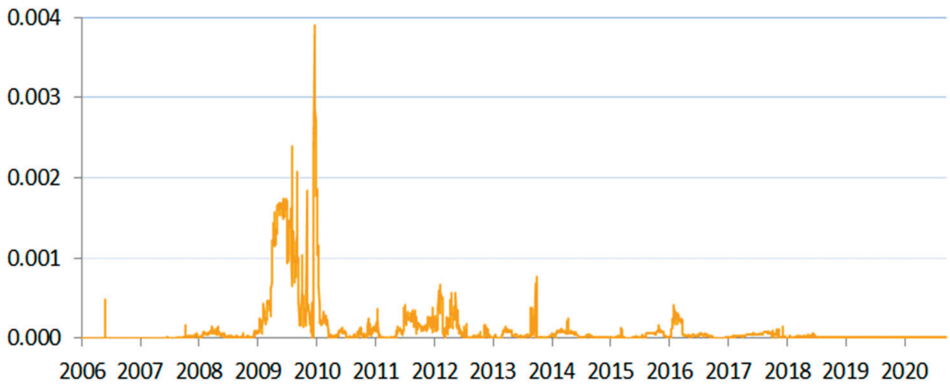


Figure 4. SIN measure for Bulgaria between 2006 and 2020.

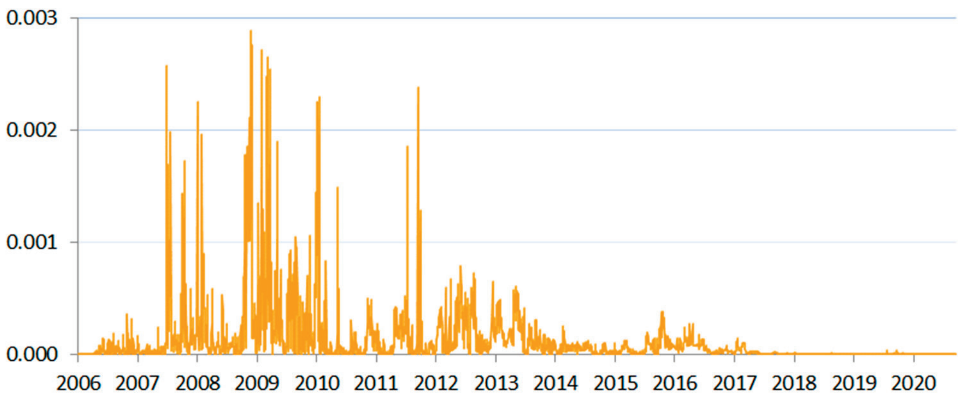


Figure 5. SIN measure for Croatia between 2006 and 2020.

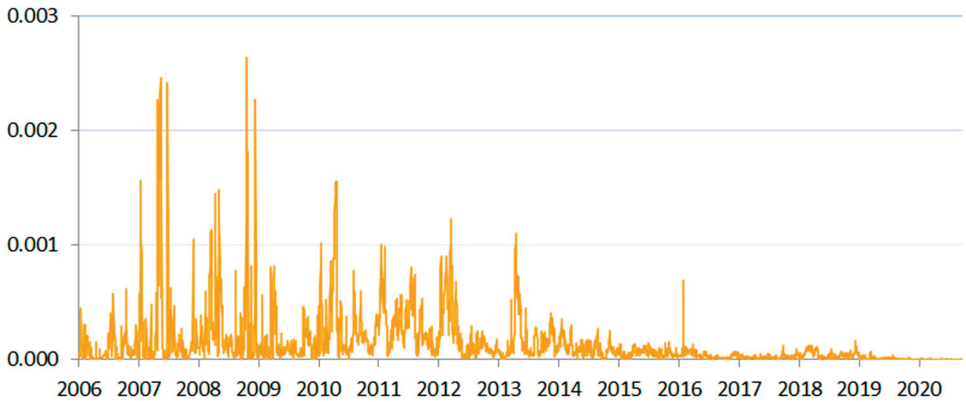


Figure 6. SIN measure for Romania between 2006 and 2020.

Croatia (Figure 5) and Romania (Figure 6) were characterized by high interbank market volatility. One of the major risk factors here was the exchange rate. For instance, Croatia had negative GDP growth in 2010 and 2012, and a major increase in the cost of euro. This coincided with the decrease of the foreign capital inflow, putting further strain on the exchange rate and forcing the Croatian National Bank to take action several times. At that time, the euro interbank market also recorded increased volatility; e.g., the EONIA rate was prone to sudden drops and peaks. Big European banks were sharply deleveraging, while their CDS spreads raised significantly (see Chart 5 in CNB 2012, p. 12). The matter concerned several European banks that had a key presence, inter alia, in Croatia. The coincidence of all of these risk factors was likely responsible for the sizeable peaks in Figure 5. Since 2015, the Croatian banking sector has been noting an improving liquidity position (in 2016, it was the best since 2004—see CNB 2016, p. 48), and this corresponds to the decrease in the levels of the SIN measure.

As far as Romania is concerned, we saw a similar level of volatility that was also driven by, among other factors, the exchange-rate variability. It was further amplified by the fact that many systemically important Romanian banks rely heavily on European parent financing. We also saw the reactions to the same international factors as described for other frontier markets (see, e.g., NBR 2007, p. 7) and the reaction to the European Union accession that put a significant competitive pressure on domestic banks (NBR 2008, p. 72). The intensity of the impact of the external risk factors corresponded to very low liquidity of the interbank market. For instance, in the beginning of the study period, the Romanian interbank market recorded only 200 transactions between October 2005 and February 2007 (NBR 2007, p. 47). Such low liquidity in that time period was a typical trait of all the frontier European markets included in this study. Among more specific Romanian risk factors, we may count a mild liquidity decrease in the first quarter of 2019, when the relatively big bank Bancpost (assets corresponding to 3.3% of the banking sector) was merged into a systemically important bank in Romania—Banca Transilvania S.A. As was the case for other analyzed countries, in Romania we also observed increasing liquidity of banks since 2015. The good liquidity position of the banks was reflected in the low SIN levels in that period.

The final four frontier market countries analyzed in the paper (Estonia, Lithuania, Latvia, and Slovakia) were characterized by more-developed economies and closer economic ties with the European Union. Liquidity risk peaks corresponded in their case to various events in the developed countries that composed the external financial market environment for them (Western EU countries and the Nordic countries). For instance, we could see risk peaks for the Baltics in 2010 that corresponded to liquidity-tightening in the markets in the mentioned countries.

In the periods preceding the euro area entrance, the SIN measure reacted to several domestic events. For instance, in case of Lithuania, we saw a sudden peak by the end of 2008, when the banking sector suffered losses, its loans portfolio was shrinking, and a higher withdrawal of deposits occurred. The peak subsided immediately because "parent banks fully covered liquidity shortage in the market by additional lending to their subsidiaries" (BL 2009, p. 5). Such an occurrence was typical for the Baltic countries, where—in the study period—Northern European parent banks were most willing to intervene to help their subsidiaries.

In Estonia, we saw a big drop in volatility in 2010 in the period when financial institutions were slowly withdrawing from the Estonian interbank market and anticipating Estonia joining the euro area in 2011 (BE 2011). Such a preceding effect also was visible for the other three countries.

Regarding Latvia, in February 2007, Standard & Poor's lowered the country's rating forecast from neutral to negative, which hit its currency significantly. The central bank intervened, but the exchange rate remained affected by this event until mid-March. These events also affected the stock exchange and the money market. "The average weighted interest rate on overnight transactions in lats amounted to 4.97% in 2007 (by 177 basis points in excess of the 2006 level). Lats money market rates on longer term transactions also went up, with 6-month RIGIBOR picking up 450 basis points (to 8.98% on average on an annual basis) and 12-month RIGIBOR growing by 453 basis points (to 9.11%) and reaching the maximum in October 2007" (BoL 2007, str. 15). The peaks in Estonia for that period also were related to the turbulence in the Latvian and euro-based interbank markets (BE 2007, p. 47).

On an opposite note, on 23 February 2018, following the money-laundering accusations made by the Financial Crimes Enforcement Network of the United States Department of the Treasury (2018, Federal Register 83/33/Friday), the Financial and Capital Market Commission made a decision on unavailability of deposits at the Latvian ABLV Bank, which was one of the systemically important banks in Latvia at that time. The bank decided on voluntary liquidation. Prompt actions of Latvian regulators allowed the country to avoid a systemic crisis. At the same time, thanks to the fact that Latvia was already in the European interbank market, this domestic event did not affect systemic liquidity. The depth of the European market was large enough to dissipate any related shocks (BoL 2019) A similar case involved the Estonian branch of Danske Bank, which entered into liquidation in August 2018, after a money-laundering[3] scandal in early 2018. In this case, no significant market freeze was reported as well. The SIN measure reacted accordingly in both cases—no significant peaks were recorded.

Of special note was the shape of the time series indicative of systemic liquidity in the Baltics and Slovakia. Here, we saw sharp but relatively short liquidity declines. For the rest of the time, the indicator remained very low, suggesting stability. We also saw significant overlap in systemic conditions for the Baltic countries, showing that the region was the most convergent among the analyzed states. This pointed to a higher risk of illiquidity-driven contagion in the Baltics. Such observations were in line with the fact that Estonia, Lithuania, and Latvia only have a small number of systemically important banks, all of which strongly depend on the same financing from Sweden and Norway.

The Baltic countries and Slovakia adopted the euro currency in the study period (2009, 2011, 2014, and 2015, respectively), which changed their foreign-exchange-rate risk profile. Figures 7–10 show how the accession changed the depth of their interbank market, increasing liquidity and lowering volatility very significantly, as these countries turned from small-scale, local-rate-based wholesale funding to euro-area funding.

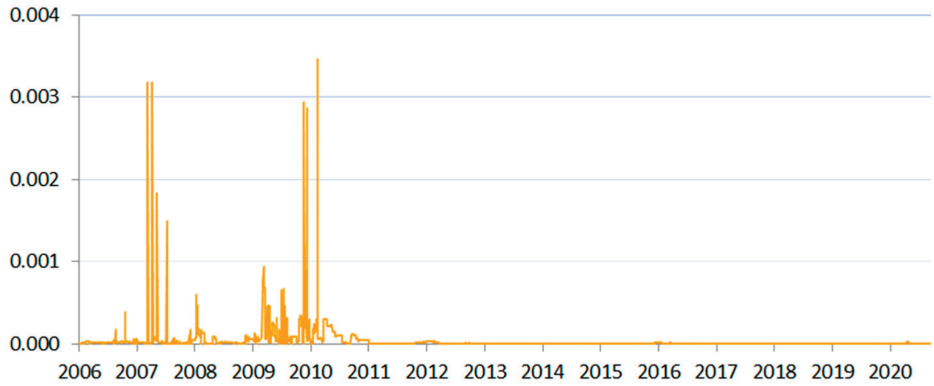


Figure 7. SIN measure for Estonia between 2006 and 2020.

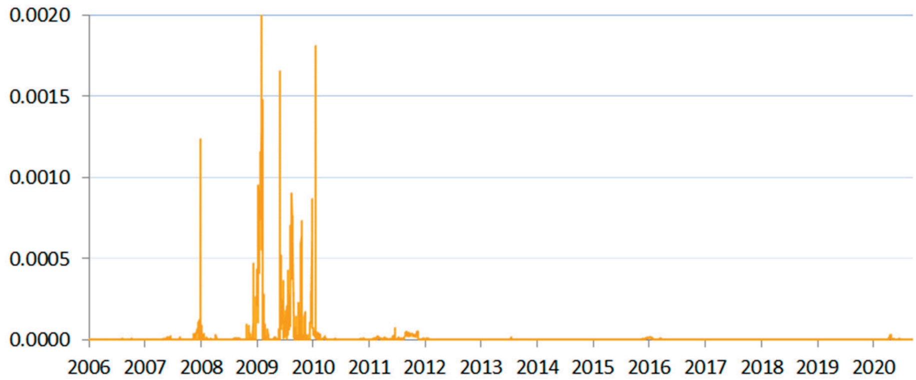


Figure 8. SIN measure for Lithuania between 2006 and 2020.

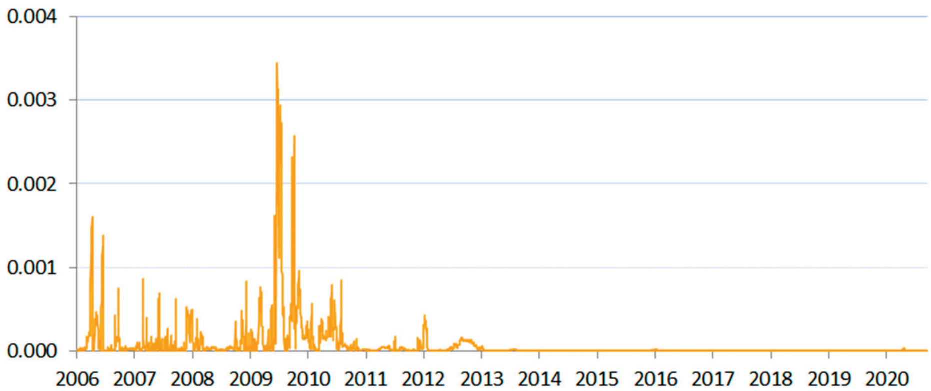


Figure 9. SIN measure for Latvia between 2006 and 2020.

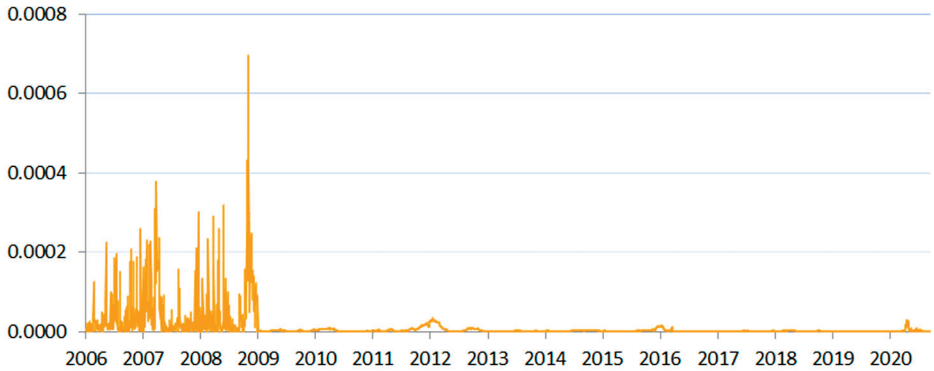


Figure 10. SIN measure for Slovakia between 2006 and 2020.

4.3. COVID-19 Pandemic

In the postcrisis period and before the COVID-19 pandemic, when ample liquidity still remained in the financial system—a remnant of the quantitative easing and prolonging negative interest rates—systemic illiquidity risk was minimal. The results suggested that the scale of risk in the interbank market has gone down. Nevertheless, one cannot be certain that this change is a permanent one. In fact, one may also argue that the risk was not mitigated, but simply shifted to the shadow-banking sector (ECB 2020).

The direction of the diffusion of the COVID-19 crisis varied from the systemic shocks observed before. Compared to the crisis in 2007–2010, when the shock originated in the financial sector and then spilled over to the real economy, the COVID-19 pandemic spread oppositely: first, the real economy was affected, then the spillover to the financial sector followed (BIS 2020).

For less-developed countries of the EME region, the COVID-19 pandemic mainly caused shocks in capital flows that influenced currency exchange rates, causing the depreciation of local currencies (Financial Stability Board FSB). Central banks started to offer foreign-exchange operations to stabilize the exchange-rate volatility, and immediately announced liquidity support to protect the financial system against any disruptions (IOSC 2020). In the CEE region, the central banks of Hungary, Poland, and Romania purchased government securities in secondary markets to restore their liquidity and strengthen the mechanism of monetary-policy transmission (Cantú et al. 2021, p. 15). Among other actions, the central banks implemented reserve policy changes to quickly free up liquidity and established nontargeted lending operations in the first months of the pandemic (Cantú et al. 2021, p. 11). The immediate intervention of the central banks, which supported the process of market running, enabled them to minimize negative consequences of the pandemic for the financial sector in the study period. This was depicted by the low levels of the SIN measure.

In the euro area, the situation in the interbank market was quite different. "While the core of the financial system—including major banks and financial infrastructures—entered the crisis more resilient than in the run-up to the global financial crisis, the COVID-19 shock led to severe liquidity stress in the system" (BIS 2021, p. 3). Described liquidity shocks were recorded by the SIN measure (Figure 11). They were short and took place at the beginning of the COVID-19 pandemic.

In March 2020, asset markets froze in many countries. The growing cash needs resulted in the widening of spreads on fixed-income instruments, the yields (mainly long-term) of which increased significantly (Schrimpf et al. 2020; Hördahl and Shim 2020). The situation was observed mainly in the developed markets in Europe—especially in the euro area, where the outflow from money market funds reflected sudden liquidity needs (IOSC 2020). Investors tried to move from the more-liquid, but also riskier, sector (various asset classes) into a less-sensitive one (sovereign bonds).

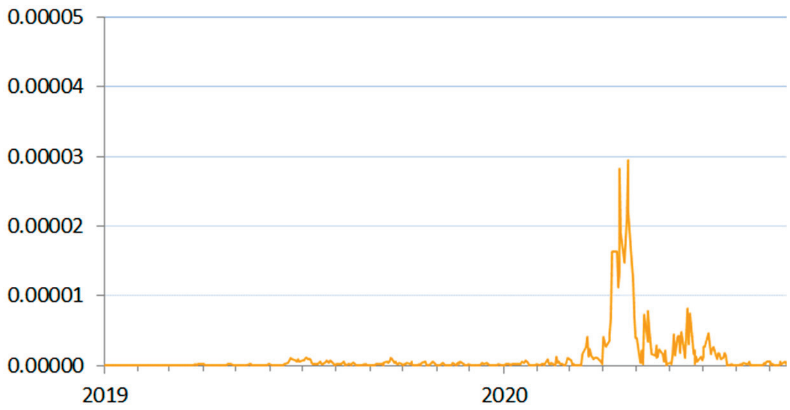


Figure 11. SIN measure for Estonia, Latvia, Lithuania, and Slovakia (euro area) between 2019 and 2020. To compare the scale of risk with previous years, see Appendix A.

The situation stabilized very quickly (by the end of March in the euro market) thanks to the immediate and synchronized reaction of the central banks. ECB announced and implemented several supportive actions: the temporary capital and operational relief in reaction to coronavirus (12 March 2020), the temporary pandemic emergency purchase program (18 March 2020) and a package of collateral easing measures (7 April and 22 April 2020) (EBA 2020). Our results confirmed that for the time being, these measures alleviated systemic risk in the financial system.

One should bear in mind that the illiquidity and the effects of the COVID-19 pandemic may be delayed in time because of the financial help from euro-area governments that basically poured liquidity directly into the system, a series of actions that in this respect resembles quantitative easing (see Christensen and Gillan 2014). It will thus certainly be interesting to see how the SIN measure behaves in the upcoming months and in a longer period of two or more years, when the economic downturn that started with the pandemic will put most of the strain on systemic risk. The longer-term effects of the help packages will most definitely be significant, and it is difficult to say with certainty what adverse effects will follow.

Nonetheless, at least three facts are certain. Governments are running unprecedented deficits and public debt has been building up fast, also in the face of the locked-down economy. In the past, the public debt buildup resulted in the increase of systemic illiquidity risk, among other things. This time, the inflation effects are also uncertain, as the monetary policy interest-rate channel working through the interbank market is almost nonexistent, with historically low base rates across the region.

Secondly, the issues of moral hazard are clear. The large-scale support measures introduced recently “may induce moral hazard, causing investors to underestimate market risk [and] infer that liquidity support will always be provided” (BIS 2021, p.18), which will cause systematic mispricing of the market liquidity risk.

Thirdly, many central banks have had deteriorated balance sheets ever since the global financial crisis, which was the effect of the previous unprecedented activities to stabilize the financial system, such as direct quantitative easing and buying back bad collateral from the banking sector. Now, “The need to intervene in such a substantial way has meant that central banks had to take on material financial risk” (BIS 2021, p. 2). It is impossible to say how this will affect their stability in the longer run and what the exit strategy is at this point.

From a global perspective, we are facing unprecedented uncertainty in any economic and financial aspect that exists. The effects of the current events are so unpredictable for one reason: namely, the entire global economy and the entire global financial sector is being affected by the same crisis at once. It is no longer the case of one market problem spilling over to the rest of the world—this time the shock is simultaneous everywhere. There are questions regarding whether the rescue measures will be able to outlast the pandemic until it subsides for good, and what will happen when the support measures start to be phased out.

5. Conclusions

In the course of this paper, we discussed the role of liquidity and its imbalances in systemic risk materialization, and we overviewed the results of existing systemic illiquidity-focused theoretical and empirical research. This literature review showed how important liquidity measurement and monitoring are in systemic risk analysis. Next, we discussed the applicability of the existing methods of systemic liquidity measurement to the CEE region. We concluded that these methods were inapplicable to the frontier and emerging financial markets under our analysis, and therefore a new approach is necessary to measure systemic liquidity risk for the given set of countries.

This conclusion also held for many other less-developed financial markets in the world that are characterized by the same specificity as the countries analyzed by us. In this way, we found an existing research gap in systemic liquidity analysis that relates to countries with frontier and emerging markets.

To fill the gap, we developed a new approach to illiquidity risk measurement using the interbank market data and Nelson–Siegel–Svensson methodology. Our measure—the Systemic Illiquidity Noise (SIN)-based measure—was empirically applied to a selected set of 10 CEE countries. In this way, we obtained the results of systemic illiquidity analysis for seven frontier and three emerging financial markets, for which such analysis was impossible before.

The empirical results displayed a successful application of the proposed method. The SIN measure proved to be sensitive to the global liquidity breakdown that took place during the global financial crisis and the European debt crisis. Similarly, the measure was reactive at times of locally important systemic events, such as runs on banks or periods of significant currency depreciation in different countries. Moreover, SIN facilitated identifying three divergent sets of countries with different systemic liquidity risk characteristics.

The results also captured the impact of introducing the euro currency on systemic liquidity risk. For the example of the euro area, we also showed that the SIN measure reacted to a liquidity shock caused by the current pandemic. This was despite the fact that the financial sector was very liquid before the shock. In effect, we may conclude that the SIN measure was sensitive to systemic liquidity shocks of different origins and magnitudes, regardless of the prevailing levels of liquidity per se.

There were at least two significant advantages to the methodology developed in the course of this study. First, the SIN measure allowed for contemporaneous monitoring of systemic liquidity at a minimal cost—using well-known models and easily accessible data. This is of potential high value to any frontier or emerging market regulator and supervisor, as such monitoring may be easily introduced and sustained, giving macroprudential bodies a better chance to react to future financial crises and to mitigate potential costs of such crises.

Second, although the study encompassed only 10 selected countries, given their diversity and the stability of our results, there is a potential for the successful application of our method to other markets, as long as there is a continuous interbank market there. Although the Interbank Offered Rates were used in this study, other similar types of rates also would be feasible, regardless of their fixing methodology. Such a potential of broad application of our measure creates an opportunity for a better and increased understanding of systemic liquidity disturbances on an international scale, regardless of the level of financial market development in each specific country.

Finally, this study is of pragmatic value not only to regulators, but also to other participants in the financial system, especially banks, because they may also use the SIN measure to monitor the systemic liquidity risk that affects them so significantly. Better-informed regulators and financial market participants would be better-equipped to make better risk management decisions, which in the long run might add to lowering systemic risk in the financial system as a whole.

Author Contributions: Conceptualization, M.A.K. and E.D.; Data curation, E.D.; Formal analysis, M.A.K. and E.D.; Funding acquisition, M.A.K.; Investigation, M.A.K. and E.D.; Methodology, M.A.K.

and E.D.; Project administration, M.A.K.; Software, E.D.; Validation, M.A.K. and E.D.; Visualization, E.D.; Writing—original draft, M.A.K. and E.D.; Writing—review & editing, M.A.K. and E.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research project was funded by the National Science Centre under Agreement UMO-2018/29/N/HS4/02783.

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

Appendix A

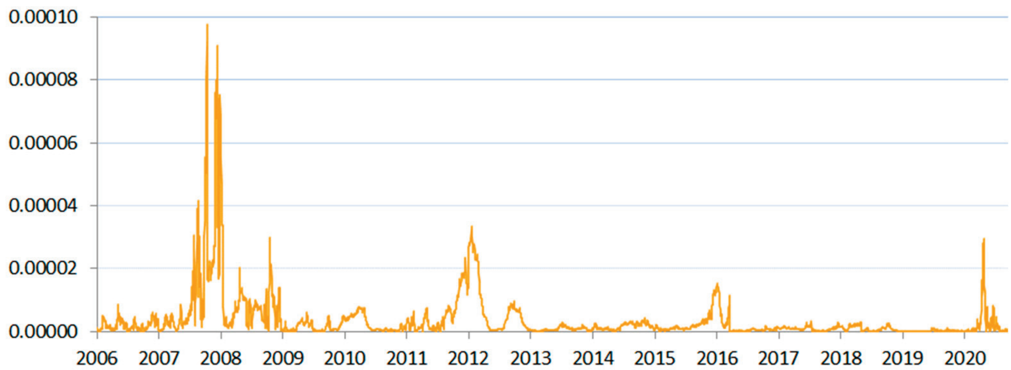


Figure A1. Systemic Illiquidity Noise-based measure calculated for the euro area.

Appendix B

Table A1. Measures focused on illiquidity applicable to systemic-scale analysis.

Measurement Output	Authors	Short Description
Liquidity factor	Pastor and Stambaugh (2003)	A measure of market liquidity computed as the equally weighted average of the liquidity measures of individual stocks, using daily data. Specifically, the liquidity **measure for a stock is the ordinary least squares regressed function of quantities** of the daily returns on this stock in a given month, its volume, and the value-weighted market return. The measure relies on the principle that order flow induces greater return reversals when liquidity is lower, viewing volume-related return reversals as arising from liquidity effects.
A set of interpretable parameters	Getmansky et al. (2004)	The proposal to use autocorrelation of returns of hedge funds as a proxy of their liquidity; **the first-, second-, and third-order autocorrelations** for each hedge fund's returns are computed using an econometric model of return smoothing coefficients and used as **a proxy for quantifying illiquidity exposure**—the less liquid the fund, the more serial correlation is observed.

Table A1. *Cont.*

Measurement Output	Authors	Short Description
Broader hedge-fund-based systemic risk measures	Chan et al. (2006)	A set of three measures quantifying the **hedge funds' impact on systemic risk** by examining the risk/return profiles of hedge funds, using returns and sizes data, at the individual and aggregate levels in relation to the **investment risk** they bear: **autocorrelation**-based measure of illiquidity exposures, a **liquidation probability**-based measure, and the regime-switching-based model quantifying the **aggregate distress** level in the hedge fund sector.
Five measures of contagion potential	Billio et al. (2012)	A structured approach to measure systemic risk with indicators based on **illiquidity** (quantified by autocorrelation) and **correlation**, using principal component analysis (indicating the degree of assets commonality), regimeswitching models, Granger causality tests (indicating the direction of propagation of systemic triggers), and network diagrams (visualizing the connectedness via directional networks), focused on detecting of **interdependence between banks, brokers, insurers, and hedge funds**, based on statistical relations among their market returns. This way, the authors quantify the potential contagion effects in the analyzed financial system.
A system of liquidity risk charges (LRCs)	Perotti and Suarez (2011)	**Pigouvian charges** are calculated per unit of refinancing risk-weighted liabilities based on a vector of additional systemic factors (such as size and interconnectedness) in a given period. The weighting function is decreasing and smooth to avoid regulatory arbitrage, which could distort market rates. The model is aimed at **making banks internalize negative systemic effects of fragile funding strategies**, but the computed size of charges may be used as a tool for quantifying liquidity risk showing which institutions generate more risk for the financial system.
Contrarian strategy liquidity measure (CSL)	Khandani and Lo (2011)	A proposal to apply **mean-reversion equity market strategy** (buying losers and selling winners over 5 to 60 min lagged returns) **to proxy the market-making** (i.e., liquidity-provisioning) profits and to obtain equity market liquidity measure by observing the performance of this trading strategy. The authors showed that when it does very well, there is less liquidity in the market, and vice versa.
Price-impact liquidity measure (PIL)		An **inverse proxy of liquidity**, in which liquidity is measured with a linear-regression estimate of the volume required to move the price of a security by one dollar; i.e., higher values of lambda imply lower liquidity and market depth. The aggregate measure of market liquidity (PIL) is computed as the **daily cross-sectional average** of the estimated price-impact coefficients.
Systemic Liquidity Risk Index (SLRI)	Severo (2012)	The SLRI is calculated by integrating the deviations of the following **basis spreads**: covered interest parity, the on-the-run versus the off-the-run interest-rate spread on government bonds, and the interest-rate spread between the overnight index swap (OIS) and short-term government bonds and the CDS basis spread, to represent **the degree of their comovement** first component score from a principal component analysis (based on historical time-series data) is used.

Table A1. *Cont.*

Measurement Output	Authors	Short Description
Liquidity Mismatch Index (LMI)	Brunnermeier et al. (2014)	Measures the difference between the cash-equivalent future values of the assets and liabilities of a bank; it utilizes the **cash-equivalent value**, which is the product of the asset or liability current value, multiplied by the liquidity weight (positive for assets, negative for liabilities), which depends on an assumed stress scenario, **Value-at-Liquidity-Risk**, defined as the quantile of worst losses (e.g., 5%), and the **Expected Liquidity Loss**, which corresponds to the average of the liquidity losses beyond this threshold. The authors proposed to use LMI to identify the most systemically important financial institutions.
Systemic risk-adjusted liquidity (SRL) model	Jobst (2014)	Estimates **the probability and severity of joint liquidity events**; i.e., instances of **banks jointly breaching their Net Stable Funding Ratios**. Estimation process: 1. The components of the NSFR are valued at market prices in order to generate a time-varying measure of funding risk relative to prudential liquidity standards. 2. Aggregate cash flow implications of changes to liquidity risk are modeled as a put option to estimate losses expected from insufficient stable funding. 3. Individually estimated liquidity risk net exposures are aggregated via a multivariate distribution to determine the probabilistic measure of joint liquidity shortfalls on a system-wide level.
Systemicness	Greenwood et al. (2015)	A linear **model of fire-sale-induced liquidity crises**, computing banks' equity shock exposures to system-wide deleveraging and to spillovers induced by individual banks; systemicness is a (quantity) measure of a bank's contribution to financial sector fragility, proportional to its **size**, **leverage**, and **connectedness** (owning large and illiquid asset classes to which other banks are also highly exposed).The key assumption is that banks target a given level of leverage, and this implies asset sales when leverage grows beyond the target. It allows the measurement of how the distribution of banks' leverage and risk exposures contributes to systemic risk.
Cumulative Distance to Default (CDD)	Karkowska (2015)	The distance-to-default measure is **a market-based measure of credit risk** based on Merton's model, in which the equity of a firm is modeled as a call option on the value of its assets. The exercise price is equal to the value of the liabilities (the firm defaults when its assets' value falls below its debt face value). For implementation, the face value of debt is assumed to be equal to the sum of short-term liabilities and half the long-term liabilities from the balance-sheet data. The model is calibrated using the analyzed institution's market value and its equity price volatility. Karkowska used this method to derive the DD value for each institution forming the studied banking system and **aggregated the data** to obtain a systemic risk measure equal to the **total probability of default** of all the studied institutions.

Table A1. *Cont.*

Measurement Output	Authors	Short Description
Aggregate vulnerability (AV) and illiquidity concentration	Duarte and Eisenbach (2019)	An extension of the systemicness measure that includes the panel analysis tracking vulnerabilities over time. It takes banks' leverage, asset holdings, asset liquidation behavior, and the price impact of liquidating assets in the secondary market as given, and models banks' responses to negative liquidity shocks (fire-sale spillovers); using information embedded in repo haircuts to account for changes in asset-specific liquidity and flow-of-funds data, it allows to measure **aggregate liquidity**, defined as the **sum of all the second-round spillover losses** (not the initial direct losses) as a share of the total equity capital in the system; the factors' decomposition applied produces a new component of AV, namely **illiquidity concentration**. The authors showed that the measure Granger-causes most other systemic risk measures.

The table presents all the complex measures applicable to systemic risk analysis focused on illiquidity considered in the study. For each method or measure, we provide a short description of the mechanism behind the measurement output.

Notes

1. Countries were classified according to the criteria of the S&P DJI's Global Benchmark Index for the study period.
2. Poland instigated the emergency mechanism to limit public debt in 2014, when the debt was at 56% of GDP.
3. In that period, several cases of monely laundering were reported in the CEE region, including ABLV bank (Latvia), Danske Bank (Estonia), Versobank (Estonia), and other smaller banks in the Baltics.

References

Acharya, Viral V., and Ouarda Merrouche. 2013. Precautionary Hoarding of Liquidity and Interbank Markets: Evidence from the Subprime Crisis. *Review of Finance* 17: 107–60. [CrossRef]

Acharya, Viral V., Douglas Gale, and Tanju Yorulmazer. 2011. Rollover Risk and Market Freezes. *The Journal of Finance* 66: 1177–209. [CrossRef]

Afonso, Gara, Anna Kovner, and Antoinette Schoar. 2011. *Stressed, Not Frozen: The Federal Funds Market in the Financial Crisis*. Staff Report 437. New York: Federal Reserve Bank.

Allen, Franklin, and Douglas Gale. 1994. Limited Market Participation and Volatility of Asset Prices. *The American Economic Review* 84: 933–55.

Allen, Franklin, and Douglas Gale. 2000a. Bubbles and Crises. *The Economic Journal* 110: 236–55. [CrossRef]

Allen, Franklin, and Douglas Gale. 2000b. Financial Contagion. *Journal of Political Economy* 108: 1–33. [CrossRef]

Allen, Franklin, Elena Carletti, and Douglas Gale. 2009. Interbank market liquidity and central bank intervention. *Journal of Monetary Economics* 56: 639–52. [CrossRef]

Andrieş, Alin Marius, Simona Nistor, and Nicu Sprincean. 2018. The impact of central bank transparency on systemic risk—Evidence from Central and Eastern Europe. *Research in International Business and Finance* 51: 100921. [CrossRef]

Aragon, George, and Philip Strahan. 2009. *Hedge Funds as Liquidity Providers: Evidence from the Lehman Bankruptcy*. Working Paper 15336. Cambridge: National Bureau of Economic Research.

Balogh, Eva S. 2015. Ten Hungarian Banks Failed Within One Year. *Hungarian Spectrum*, March 4.

Banerjee, Ryan N., and Hitoshi Mio. 2014. *The Impact of Liquidity Regulation on Banks*. BIS Working Papers 470. Basel: Bank for International Settlements.

Bank for International Settlements (BIS). 2005. *Zero-Coupon Yield Curves: Technical Documentation*. BIS Paper No 25. Basel: Monetary and Economic Department, Bank for International Settlements.

Bank for International Settlements (BIS). 2020. *A Global Sudden Stop*. Chapter 1 and Chapter 2, Annual Economic Report 2020. Basel: Bank for International Settlements.

Bank for International Settlements (BIS). 2021. COVID-19 Support Measures. Extending, Amending and Ending. April 6. Available online: https://www.fsb.org/wp-content/uploads/P060421-2.pdf (accessed on 23 March 2021).

Bank of Estonia (BE). 2007. Financial Stability Review 2/2007. Available online: https://www.eestipank.ee/en/publications/series/financial-stability-review (accessed on 23 March 2021).

Bank of Estonia (BE). 2011. Financial Stability Review 2/2011. Available online: https://www.eestipank.ee/en/publications/series/financial-stability-review (accessed on 23 March 2021).

Bank of Latvia (BoL). 2007. Financial Stability Report. Available online: https://www.bank.lv/en/publications-r/financial-stability-report (accessed on 23 March 2021).

Bank of Latvia (BoL). 2019. Financial Stability Report. Available online: https://www.bank.lv/en/publications-r/financial-stability-report (accessed on 23 March 2021).

Bank of Lithuania (BL). 2009. Financial Stability Review. Available online: https://www.lb.lt/en/ch-publications/category.39/series.169#group-2 (accessed on 23 March 2021).

Barkauskaite, Aida, Ausrine Lakstutiene, and Justyna Witkowska. 2018. Measurement of Systemic Risk in a Common European Union Risk-Based Deposit Insurance System: Formal Necessity or Value-Adding Process? *Risks* 6: 137. [CrossRef]

Berg, Andrew, and Catherine Pattillo. 1999. Are Currency Crises Predictable? A Test. *IMF Staff Papers* 46: 108–38.

Bessembinder, Hendrik, Jia Hao, and Michael L. Lemmon. 2011. *Why Designate Market Makers? Affirmative Obligations and Market Quality*. Working Paper. Washington, DC: International Monetary Fund.

Bhattacharya, Sudipto, and Douglas Gale. 1987. Preference shocks, liquidity and central bank policy. In *New Approaches to Monetary Economics*. Edited by William Barnett and Kenneth Singleton. New York: Cambridge University Press, pp. 69–88.

Billio, Monica, Mila Getmansky, Andrew W. Lo, and Loriana Pelizzon. 2012. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics* 104: 535–59. [CrossRef]

Blåvarg, Martin, and Patrick Nimander. 2002. Interbank Exposures and Systemic Risk. *Sveriges Riksbank's Economic Review* 2: 19–45.

Boss, Michael, Helmut Elsinger, Martin Summer, and Stefan Thurner. 2004. Network topology of the interbank market. *Quantitative Finance* 4: 677–84. [CrossRef]

Boyson, Nicole M., Christof W. Stahel, and Rene M. Stulz. 2010. Hedge Fund Contagion and Liquidity Shocks. *Journal of Finance* 65: 1789–816. [CrossRef]

Brunnermeier, Markus K. 2008. Bubbles. In *New Palgrave Dictionary of Economics*. Edited by Lawrence E. Blume and Steven Durlauf. London: Palgrave Macmillan.

Brunnermeier, Markus K., and Lasse H. Pedersen. 2009. Market Liquidity and Funding Liquidity. *Review of Financial Studies* 22: 2201–38. [CrossRef]

Brunnermeier, Markus K., and Martin Oehmke. 2013. The Maturity Rat Race. *The Journal of Finance* 68: 483–521. [CrossRef]

Brunnermeier, Markus K., and Yuliy Sannikov. 2014. A Macroeconomic Model with a Financial Sector. *The American Economic Review* 104: 379–421. [CrossRef]

Brunnermeier, Markus K., Gary Gorton, and Arvind Krishnamurthy. 2014. Liquidity Mismatch Measurement. In *Risk Topography: Systemic Risk and Macro Modeling*. NBER Chapters. Cambridge: National Bureau of Economic Research, pp. 99–112.

Brunnermeier, Markus K., Thomas M. Eisenbach, and Yuliy Sannikov. 2013. Macroeconomics with Financial Frictions: A Survey. In *Advances in Economics and Econometrics, Tenth World Congress of the Econometric Society*. New York: Cambridge University Press, vol. 2, pp. 3–96.

Buiter, Willem Hendrik. 2008. *Central Banks and Financial Crises*. London School of Economics and Political Science Discussion Paper 619. Financial Markets Group. London: London School of Economics and Political Science.

Bulgarian National Bank (BNB). 2014. Economic Review. April. Available online: https://www.bnb.bg/bnbweb/groups/public/documents/bnb_publication/pub_ec_r_2014_04_en.pdf (accessed on 23 March 2021).

Caballero, Ricardo, and Alp Simsek. 2013. Fire Sales in a Model of Complexity. *The Journal of Finance* 68: 2549–87. [CrossRef]

Cantú, Carlos, Paolo Cavallino, Fiorella De Fiore, and James Yetman. 2021. *A Global Database on Central Banks' Monetary Responses to Covid-19*. BIS Working Papers 934. Basel: Bank for International Settlements.

Cespa, Giovanni, and Thierry Foucault. 2014. Illiquidity Contagion and Liquidity Crashes. *Review of Financial Studies* 27: 1615–60. [CrossRef]

Chan, Nicholas, Mila Getmansky, Shane M. Haas, and Andrew W. Lo. 2006. Do Hedge Funds Increase Systemic Risk? *Federal Reserve Bank of Atlanta Economic Review* 91: 49–80.

Chan-Lau, Jorge A., Marco Espinosa, and Juan Sole. 2009. *On the Use of Network Analysis to Assess Systemic Financial Linkages*. Global Financial Stability Report. Washington, DC: International Monetary Fund.

Choudhry, Moorad. 2018. *An Introduction to Banking: Principles, Strategy and Risk Management (Securities Institute)*, 2nd ed. Chichester: John Wiley & Sons, Ltd.

Christensen, Jens H., and James M. Gillan. 2014. Does Quantitative Easing Affect Market Liquidity? Federal Reserve Bank of San Francisco Working Paper 2013-26. Available online: https://www.frbsf.org/economic-research/publications/working-papers/2013/26/ (accessed on 12 February 2021).

Cifuentes, Rodrigo, Hyun Shin, and G. ianluigi Ferrucci. 2005. Liquidity Risk and Contagion. *Journal of the European Economic Association* 3: 556–66. [CrossRef]

Clementi, David. 2001. Financial Markets: Implications for Financial Stability. *Balance Sheet* 9: 7–12. [CrossRef]

Coval, Joshua D., and Erik Stafford. 2007. Asset Fire Sales (and Purchases) in Equity Markets. *Journal of Financial Economics* 86: 479–512. [CrossRef]

Cox, John C., Jonathan Ingersoll, and Stephen Ross. 1981. The relation between forward prices and futures prices. *Journal of Financial Economics* 9: 321–46. [CrossRef]

Czech National Bank (CNB). 2007. Financial Stability Report. Available online: https://www.cnb.cz/en/financial-stability/fs-reports/ (accessed on 23 March 2021).

Czech National Bank (CNB). 2008. Financial Stability Report. Available online: https://www.cnb.cz/en/financial-stability/fs-reports/ (accessed on 23 March 2021).
Czech National Bank (CNB). 2010–2011. Financial Stability Report. Available online: https://www.cnb.cz/en/financial-stability/fs-reports/ (accessed on 23 March 2021).
Czech National Bank (CNB). 2012. Financial Stability Report. Available online: https://www.cnb.cz/en/financial-stability/fs-reports/ (accessed on 23 March 2021).
Croatian National Bank (CNB). 2016. Financial Stability, No. 17. Available online: https://www.hnb.hr/en/-/financial-stability-17 (accessed on 23 March 2021).
de La Grandville, Olivier. 2001. *Bond Pricing and Portfolio Analysis*. Cambridge and London: MIT Press.
Diamond, Douglas W., and Raghuram G. Rajan. 2005. Liquidity Shortages and Banking Crises. *Journal of Finance* 60: 615–47. [CrossRef]
Drehmann, Mathias, and Nikola Tarashev. 2011. Systemic Importance: Some Simple Indicators. BIS Quarterly Review. Available online: https://ssrn.com/abstract=1785264 (accessed on 12 February 2021).
Duarte, Fernando, and Thomas Eisenbach. 2019. *Fire-Sale Spillovers and Systemic Risk*. Staff Report 645. New York: Federal Reserve Bank.
Duffie, Darrell. 2020. *Still the World's Safe Haven? Redesigning the U.S. Treasury Market after the COVID-19 Crisis*. Hutchins Center Working Papers No 62. June. Available online: https://www.brookings.edu/research/still-the-worlds-safe-haven/ (accessed on 25 March 2021).
Dziwok, Ewa. 2017. Chosen measures for pricing of liquidity. In *Contemporary Trends and Challenges in Finance*. Edited by Krzysztof Jajuga, Lucjan Orlowski and Karsten Staehr. Springer Proceedings in Business and Economics. Cham: Springer, pp. 3–9.
EIOPA. 2016. *Financial Stability Report, EIOPA-FSI-16-016*. Frankfurt am Main: European Insurance and Occupational Pensions Authority.
EIOPA. 2017. *Technical Documentation of the Methodology to Derive EIOPA's Risk-Free Interest Rate Term Structures*. EIOPA-BoS-15/035. January 30, Frankfurt am Main: European Insurance and Occupational Pensions Authority.
Elsinger, Helmut, Alfred Lehar, and Martin Summer. 2006. Risk Assessment for Banking Systems. *Management Science* 52: 1301–14. [CrossRef]
European Banking Authority (EBA). 2020. Guidelines on Legislative and Non-Legislative Moratoria on Loan Repayments Applied in the Light of the COVID-19 Crisis, Consolidated Version, December 2. Available online: https://www.eba.europa.eu/regulation-and-policy/credit-risk/guidelines-legislative-and-non-legislative-moratoria-loan-repayments-applied-light-covid-19-crisis (accessed on 25 March 2021).
European Central Bank (ECB). 2020. *Financial Stability Review*. Frankfurt am Main: European Central Bank.
Fabozzi, Frank J. 2007. *Fixed Income Analysis*, 2nd ed. Hoboken: John Wiley & Sons, Inc.
Feldman, Ron, and Jason Schmidt. 2001. *Increased Use of Uninsured Deposits: Implications for Market Discipline*. Minneapolis: Federal Reserve Bank of Minneapolis-Fed Gazette, pp. 18–19.
Financial Stability Board (FSB). 2020. *Holistic Review of the March Market Turmoil*. November 17, Basel: Financial Stability Board.
Flannery, Mark J. 1996. Financial Crises, Payment System Problems, and Discount Window Lending. *Journal of Money, Credit and Banking* 28: 804–24. [CrossRef]
Gabrieli, Silvia, and Co-Pierre Georg. 2014. *A Network View on Interbank Market Freezes*. Discussion Paper 44/2014. Frankfurt am Main: Deutsche Bundesbank.
Garcia, Maria Teresa Medeiros, and Vítor Hugo Ferreira Carvalho. 2019. A cross-sectional application of the NelsonSiegel-Svensson model to several negative yield cases. *Cogent Economics & Finance* 7: 1. [CrossRef]
Geanakoplos, John. 2010. Solving the present crisis and managing the leverage cycle. *Economic Policy Review* 16: 101–31. [CrossRef]
Gersl, Adam, and Jaroslav Heřmánek. 2007. Financial Stability Indicators: Advantages and Disadvantages of their Use in the Assessment of Financial System Stability. In *CNB Financial Stability Report*. Thematic Article 2. Prague: Research Department, Czech National Bank, pp. 69–79.
Getmansky, Mila, Andrew Lo, and Igor Makarov. 2004. An econometric model of serial correlation and illiquidity in hedge fund returns. *Journal of Financial Economics* 74: 529–609. [CrossRef]
Gofman, Michael. 2015. *Efficiency and Stability of a Financial Architecture with Too-Interconnected-To-Fail Institutions*. External Seminar Paper. Washington, DC: International Monetary Fund.
Gorton, Gary, and G. uillermo Ordonez. 2014. Collateral Crises. *The American Economic Review* 104: 343–78. [CrossRef]
Greenwood, Robin, Augustin Landier, and David Thesmar. 2015. Vulnerable banks. *Journal of Financial Economics* 115: 471–85. [CrossRef]
Gromb, Denis, and Dimitri Vayanos. 2002. Equilibrium and welfare in markets with financially constrained arbitrageurs. *Journal of Financial Economics* 66: 361–407. [CrossRef]
Heider, Florian, Marie Hoerova, and Cornelia Holthausen. 2015. Liquidity hoarding and interbank market spreads: The role of counterparty risk. *Journal of Financial Economics* 118: 336–54. [CrossRef]
Hördahl, Peter, and Ilhyock Shim. 2020. EME bond portfolio flows and long-term interest rates during the Covid-19 pandemic. *BIS Bulletin 18*. May 20. Available online: https://www.bis.org/publ/bisbull18.htm (accessed on 25 March 2021).
Hu, Grace Xing, Jun Pan, and Jiang Wang. 2013. Noise as Information for Illiquidity. *The Journal of Finance* 68: 2341–82. [CrossRef]

International Monetary Fund (IMF). 2009. Assessing the Systemic Implications of Financial Linkages. In *Global Financial Stability Review*. Washington, DC: International Monetary Fund, pp. 73–110.
International Monetary Fund (IMF). 2015. Vulnerabilities, Legacies, and Policy Challenges. In *Risks Rotating to Emerging Markets, Global Financial Stability Report*. Washington, DC: International Monetary Fund.
International Monetary Fund (IMF). 2017. *Bulgaria. Financial System Stability Assessment*. IMF Country Report, no. 17/132. Washington, DC: International Monetary Fund.
International Organization of Securities Commissions (IOSC). 2020. Money Market Funds during the March-April Episode, Thematic Note, OR03/2020. November. Available online: https://www.iosco.org/library/pubdocs/pdf/IOSCOPD666.pdf (accessed on 23 March 2021).
Iyer, Rajkamal, and Jose-Luis Peydro. 2011. Interbank Contagion at Work: Evidence from a Natural Experiment. *Review of Financial Studies* 24: 1337–77. [CrossRef]
Iyer, Rajkamal, and Manju Puri. 2012. Understanding Bank Runs: The Importance of Depositor-Bank Relationships and Networks. *American Economic Review* 102: 1414–45. [CrossRef]
James, Jessica, and Nick Weber. 2000. *Interest Rate Modeling*. Hoboken: John Wiley & Sons, Inc.
Jobst, Andreas A. 2012. *Measuring Systemic Risk-Adjusted Liquidity (SRL)*. A Model Approach, International Monetary Fund Working Paper, WP/12/209. Washington, DC: International Monetary Fund.
Jobst, Andreas A. 2014. Measuring systemic risk-adjusted liquidity (SRL)—A model approach. *Journal of Banking and Finance* 45: 270–87. [CrossRef]
Kapuściński, Mariusz, and Ewa Stanisławska. 2017. Measuring bank funding costs in the analysis of interest rate pass-through: Evidence from Poland. *Economic Modelling* 70: 288–300. [CrossRef]
Karaś, Marta. 2019. Measuring Financial System Stability—Analysis and Applications for Poland. Ph.D. thesis, Wrocław University of Economics, Wrocław, Poland, February.
Karaś, Marta, and Witold Szczepaniak. 2019. Three ways to improve systemic risk analysis of the CEE region using SRISK and CoVaR. Paper presented at the International Risk Management Conference, Working Paper, Milan, Italy, June 17–18.
Karkowska, Renata. 2015. *Ryzyko Systemowe. Charakter i Źródła Indywidualizacji w Sektorze Bankowym*. Kraków: Wolters Kluwer.
Khandani, Amir E., and Andrew W. Lo. 2011. What Happened to the Quants in August 2007? Evidence from Factors and Transactions Data. *Journal of Financial Markets* 14: 1–46. [CrossRef]
Kiyotaki, Nobuhiro, and John Moore. 2002. Evil Is the Root of All Money. *American Economic Review* 92: 62–66. [CrossRef]
Kubinschi, Matei, and Dinu Barnea. 2016. Systemic risk impact on economic growth—The case of the cee countries. *Romanian Journal of Economic Forecasting* 16: 79–94.
Kumar, Manmohan P., and Avinash Persaud. 2001. *Pure Contagion and Investors Shifting Risk Appetite; Analytical Issues and Empirical Evidence*. IMF Working Papers WP/01/134. Washington, DC: International Monetary Fund.
Loutskina, Elena, and Philip E. Strahan. 2009. Securitization and the Declining Impact of Bank Finance on Loan Supply: Evidence from Mortgage Originations. *Journal of Finance* 64: 861–89. [CrossRef]
Lubiński, Marek. 2013. Międzynarodowy rynek pieniężny i zarażenie, Gospodarka Narodowa. *The Polish Journal of Economics* 264: 19–41.
Lucas, Robert E. 1978. Asset Prices in an Exchange Economy. *Econometrica* 46: 1429–45. [CrossRef]
Martellini, Lionel, Philippe Priaulet, and Stéphane Priaulet. 2003. *Fixed-Income Securities: Valuation, Risk Management, and Portfolio Strategies*. Chichester: John Wiley & Sons Ltd.
Mehra, Yash P. 1995. Some Key Empirical Determinants of Short-Term Nominal Interest Rates. Federal Reserve Bank of Richmond. *Economic Quarterly* 81: 33–51.
Mishkin, Frederic. 2007. Systemic Risk and the International Lender of Last Resort, Board of Governors of the Federal Reserve. Paper presented at Tenth Annual International Banking Conference, Chicago, IL, USA, September 28.
Morris, Stephen, and Hyun Song Shin. 2012. Contagious Adverse Selection. *American Economic Journal: Macroeconomics* 4: 1–21. [CrossRef]
National Bank of Romania (NBR). 2007. Financial Stability Report. Available online: https://www.bnr.ro/Regular-publications-2503.aspx (accessed on 23 March 2021).
National Bank of Romania (NBR). 2008. Financial Stability Report. Available online: https://www.bnr.ro/Regular-publications-2504.aspx (accessed on 23 March 2021).
Nelson, Charles R., and Andrew F. Siegel. 1987. Parsimonious Modelling of Yield Curves. *Journal of Business* 60: 473–89. [CrossRef]
Nelson, William R., and Roberto Perli. 2007. Selected Indicators of Financial Stability. In *Risk Measurement and Systemic Risk*. Frankfurt am Main: ECB, pp. 343–72.
Nymand-Andersen, Per. 2018. *Yield Curve Modelling and a Conceptual Framework for Estimating Yield Curves: Evidence from the European Central Bank's Yield Curves*. Statistics Paper Series, No 27; Frankfurt am Main: European Central Bank.
Pastor, Lubos, and Robert Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111: 642–85. [CrossRef]
Perotti, Enrico C., and Javier Suarez. 2011. A Pigovian Approach to Liquidity Regulation. *International Journal of Central Banking* 7: 3–41. [CrossRef]

Radev, Dimitar. 2020. "The Bulgarian Banking Sector in 2020—How Have We Coped and What Is Forthcoming. Address of the Governor of the Bulgarian National Bank", Bulletin of the Association of Banks in Bulgaria, Issue 64. December. Available online: https://www.bis.org/review/r201209a.htm (accessed on 23 March 2021).

Sapra, Haresh. 2008. Do accounting measurement regimes matter? A discussion of mark-to-market accounting and liquidity pricing. *Journal of Accounting and Economics* 45: 379–87. [CrossRef]

Scheinkman, Jose A., and Wei Xiong. 2003. Overconfidence and Speculative Bubbles. *Journal of Political Economy* 111: 1183–219. [CrossRef]

Schich, Sebastian T. 1997. *Estimating the German Term Structure*. Discussion Paper 4/97. Frankfurt: Economic Research Group of the Deutsche Bundesbank.

Schmitz, Hendrik. 2011. Direct evidence of risk aversion as a source of advantageous selection in health insurance. *Economics Letters* 113: 180–82. [CrossRef]

Schoenmaker, Dirk. 1996. *Contagion Risk in Banking*. LSE Financial Markets Group. London: London School of Economics and Political Science, pp. 86–104.

Schrimpf, Andreas, Hyun Song Shin, and Vladyslav Sushko. 2020. *Leverage and Margin Spirals in Fixed Income Markets during the Covid-19 Crisis*. BIS Bulletin, No 2. April 2. Basel: Bank for International Settlements.

Severo, Tiago. 2012. *Measuring Systemic Liquidity Risk and the Cost of Liquidity Insurance*. IMF Working Paper WP/12/194. Washington, DC: International Monetary Fund.

Shiller, Robert, and J. Huston McCulloch. 1990. The term structure of interest rates. In *Handbook of Monetary Economics*. Edited by B. M. Friedman and F. H. Hahn. Amsterdam: Elsevier, pp. 627–722.

Shleifer, Andrei, and Robert W. Vishny. 1992. Liquidation Values and Debt Capacity: A Market Equilibrium Approach. *The Journal of Finance* 47: 1343–66. [CrossRef]

Shleifer, Andrei, and Robert W. Vishny. 1997. The Limits of Arbitrage. *The Journal of Finance* 52: 35–55. [CrossRef]

Svensson, Lars E. 1994. *Estimating and Interpreting forward Interest Rates: Sweden 1992–94*. NBER Working Paper Series 4871; Cambridge: National Bureau of Economic Research.

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

Svensson, Lars E. 1999. *Price Stability as a Target for Monetary Policy: Defining and Maintaining Price Stability*. NBER Working Paper 91. Cambridge: National Bureau of Economic Research.

United States Department of the Treasury. 2018. Federal Register, 83/33/Friday, Financial Crimes Enforcement Network Notification. Available online: https://www.fincen.gov/sites/default/files/federal_register_notices/2018-02-16/2018-03214.pdf (accessed on 25 March 2021).

Upper, Christian. 2007. *Using Counterfactual Simulations to Assess the Danger of Contagion in Interbank Markets*. BIS Working Paper 234. Basel: Bank for International Settlements.

Upper, Christian. 2011. *Simulation Methods to assess the Danger of Contagion in Interbank Markets*. BIS Working Paper 234. Basel: Bank for International Settlements.

Valentinyi, Akos. 2012. The Hungarian Crisis, Vox EU. March 19. Available online: https://voxeu.org/article/hungarian-crisis (accessed on 25 March 2021).

Vassallo, Danilo, Hermans Lieven, and Thomas Kostka. 2020. *Volatility-Targeting Strategies and the Market Sell-Off*. Financial Stability Review. Frankfurt am Main: European Central Bank.

Vivar, Molestina Luis, Michael Wedow, and Christian Weistroffer. 2020. *Burned by Leverage? Flows and Fragility in Bond Mutual Funds*. Working Paper Series No 2413; Frankfurt am Main: European Central Bank.

Zigraiova, Diana, and Petr Jakubik. 2017. *Updating the Long Term Rate in Time: A Possible Approach*. IES Working Paper No. 03/2017. Prague: Charles University in Prague, Institute of Economic Studies (IES).

risks

MDPI

Article

Measurement of Systemic Risk in the Colombian Banking Sector

Orlando Rivera-Escobar [1], John Willmer Escobar [2,*] and Diego Fernando Manotas [3]

1 Department of Economic Sciences, Universidad del Valle, Cali 760001, Colombia; orlando.rivera@correounivalle.edu.co
2 Department of Accounting and Finance, Universidad del Valle, Cali 760001, Colombia
3 School of Industrial Engineering, Universidad del Valle, Cali 760001, Colombia; diego.manotas@correounivalle.edu.co
* Correspondence: john.wilmer.escobar@correounivalle.edu.co

Abstract: This paper uses three methodologies for measuring the existence of systemic risk in the Colombian banking system. The determination of its existence is based on implementing three systemic risk measures widely referenced in academic works after the subprime crisis, known as CoVaR, MES and SRISK. Together, the three methodologies were implemented for the case of Colombian Banks during the 2008–2017 period. The findings allow us to establish that the Colombian banking sector did not present a systemic risk scenario, despite having suffered economic losses due to external shocks, mainly due to the subprime crisis. The results and findings show the efficiency of the systemic risk measures implemented in this study as an alternative to measure systemic risk in banking systems.

Keywords: systemic risk; banking sector; DCoVaR; MES; SRISK; quantile regression; EGARCH; DCC; value at risk

Citation: Rivera-Escobar, Orlando, John Willmer Escobar, and Diego Fernando Manotas. 2022. Measurement of Systemic Risk in the Colombian Banking Sector. *Risks* 10: 22. https://doi.org/10.3390/risks10010022

Academic Editors: Krzysztof Jajuga and Józef Dziechciarz

Received: 26 October 2021
Accepted: 2 December 2021
Published: 13 January 2022

1. Introduction

After the financial crisis of 2007–2008, the debate on the systemic risk that materialized took center stage. This circumstance led both the financial regulatory institutions and the academy to take part in the different debates generated to know, in a better way, the circumstances and facts that would have provided the necessary inputs for the realization and materialization of this type of risk.

The results of this crisis produced severe consequences and high costs for the economy and society. Although many of the costs involved in the development and subsequent outcome are not directly quantifiable, it is estimated that banks' costs would be around 2.3 trillion dollars (International Monetary Fund 2010). On the other hand, both direct and indirect impacts on the real sector materialized in economic losses and weakened confidence in the markets and the economy. These events provided the necessary inputs for the appearance of major adverse events such as deflation and economic stagnation (Cáceres 2009). The above results made systemic risk one of the central research topics, thus generating many contributions encouraged by providing answers to the questions generated during and after the subprime crisis.

Within the expected results of the various contributions made to the debate on systemic risk, there is a tendency to suggest that a determining factor would be integration between markets. Over time, financial crises have tended to broaden and deepen more and more, to the extent that the financial sector's participation within economies has increased significantly. This increase goes hand in hand with technological advances in the computing and information sectors, which, together with other factors, have consolidated a continuous and solid integration between global and local financial markets (Guerra et al. 2016; Silva et al. 2017). In this way, the above events would support the necessary mechanisms for the spread and accentuation of systemic risk events between economies and within them, regardless of their growth and development levels (Grilli et al. 2015).

On the other hand, the financial sector's participation within economies would be a key factor when evaluating the impacts and spread. The financial crisis of 2008, from now on subprime crisis, also evidenced that the inefficient participation of financial regulation and the prevalence of moral hazard in decision-making by credit institutions and risk evaluators contributed to the creation and extension of systemic risk, which as a result produced the bankruptcy of the largest banks and mortgage institutions in the United States and Europe during 2008 (Cáceres 2009).

The subprime crisis caused an indecision environment due to the lack of consensus regarding its definition and measure. The systemic risk concept had already been addressed in studies before the subprime crisis; however, these did not represent a point of collective interest, and thus remained underdeveloped (Drakos and Kouretas 2015; Silva et al. 2017). As a result of the previous, more information and consensus regarding its consequences than its causes has been fostered (Guerra et al. 2016).

For this article's purpose, the definition of systemic risk proposed in the Report on Financial Consolidation for the year 2001 proposed by the Bank for International Settlements BIS is used. Systemic risk is defined as the risk induced by exogenous shocks that trigger losses in the economic value of a significant portion of the financial system, and with these, the imposition of adverse effects on the real economy. Based on the findings, the definition proposed here adequately captures the Colombian economy's role during the subprime crisis. However, the economic losses experienced by the banking sector were not significant enough to materialize in a scenario of systemic crisis. Furthermore, the negative impact imposed on the real sector by the banking sector was carried out mainly through the restriction of the supply of credit by them, as a measure to limit their exposure to credit risk (Cáceres 2009).

The vulnerability of financial systems, both national and international, to a possible realization of a new systemic crisis has been a latent concern due to the lack of dynamism on the part of regulatory policies; despite showing the dynamics present in the financial systems, they remained static without corresponding to these dynamics. This idea is commonly found in the works carried out by Drakos and Kouretas (2015); López-Espinosa et al. (2012) and Castelao et al. (2013), among others. By the above, it is necessary to implement strategies or policies to neutralize or reduce the transmission mechanisms by which systemic risk empowers the system's stability. In this way, this paper proposes a methodology for measuring systemic risk based on using three risk measures that capture each banking institution's contribution to global risk. Likewise, the identification of the characteristics conducive to systemic risk in banking entities is determined. In this way, it is possible to evaluate the incidence and the individualization of the determining factors for the generation of systemic risk in the Colombian banking system.

The proposed methodology is based on the implementation of various systemic risk measures. The choice to use different measures to contribute to systemic risk is based on the difficulty of capturing the multiple facets of this type of risk with just one measure (Kleinow et al. 2017). Thus, in this paper, measures based on the DCoVaR (Tobias and Brunnermeier 2016), Marginal Espected Shofoll (Acharya et al. 2017) and SRISK (Brownlees and Engle 2016) have been adapted. The measures have been integrated into a methodological scheme for measuring systemic risk. Tobias and Brunnermeier (2016) developed a systemic risk measure called DCoVaR, which measures the increase in financial risk, defined as the change in the Value at Risk of the system, caused by an institution in trouble. For their part, Acharya et al. (2017) structure an economic model in which they accommodate the systemic risk measure called Marginal Espected Shofoll (MES), which represents the marginal change in the expected losses in the tail of the distribution of the returns of the system, before equally marginal changes in the participation of each institution that makes it up. Finally, Brownlees and Engle (2016) developed a systemic risk index called

SRISK, which measures the contribution to systemic risk, the capital deficit of an institution conditioned to a significant decline in the market.

These measures provide different ways of understanding the dynamics surrounding a systemic risk event and provide, in this way, a more global look at the characteristics involved in the development of this type of event. This paper makes it possible to complement each risk measure's results, considering banks' exposure, contribution and vulnerability to potentially triggering episodes of systemic risk. This paper uses banks' characteristics to evaluate their operational structure and thus outline the transmission mechanisms by which a bank becomes systemically essential to evaluate the determining factors contributing to systemic risk by a banking institution. The use of characteristics identified by Bostandzic and Weiß (2018) proposed that the methodology was used to choose variables and link them with systemic risk measures to evaluate the explanatory power of the above characteristics on the risk measures implemented in this study. The paper's main contribution is the comparison of three measures of Systemic Risk and the empirical confirmation of the study on the Colombian Banking Sector. Finally, in addition to increasing state of the art related to systemic risk for the Colombian economy, this work aims to contribute to international evidence on the effects and consequences of the spread of systemic risk among economies. Unlike the published works, we have extrapolated data to make a fair comparison with the monthly information.

The work has been organized as follows: Section 2 presents and discusses the literature referring to systemic risk and its link with the proposed methodology. In Section 3, the proposed methodology is developed and the data used. Section 4 presents the obtained results. Section 5 discusses the results and highlights the relevant aspects of the proposed methodology. Finally, Section 5 presents the conclusions and future remarks.

2. Literature Review

In the economic literature following the subprime crisis, there is a tendency to identify this event as systemic because it substantially affected agents' confidence in the financial system, thus deepening the deterioration of economic growth perceived worldwide (Aparicio et al. 2012). The causes that had a significant impact, from the point of view of financial regulation, were associated with incompatibility of the microprudential policies of that time to adequately incorporate scenarios or episodes of systemic risk (Castelao et al. 2013; Drakos and Kouretas 2015; López-Espinosa et al. 2012).

Castelao et al. (2013) present three limitations in microprudential regulation: (i) it ignores the correlations between financial institutions and the concentration of risk between them when analyzing institutions in isolation without considering the endogeneity of risks; (ii) it ignores the systemic importance of certain factors such as size, leverage and interrelations with the rest of the system, factors widely recognized today by the Basel III agreement as criteria when determining systemically important banks and (iii) it does not allow us to see the variation in the risks taken by institutions throughout the economic cycle.

In the contributions from academia, there is a notable absence of a collective consensus regarding the definition of systemic risk, leading to the development of several lines of approach to the problem that structurally mutate in terms of the scope and objectives proposed by the researcher (Acharya 2009; Cabrera-Rodríguez et al. 2014; Aparicio et al. 2012). Aparicio et al. (2012) state that in the literature concerning the analysis of systemic risk, two lines of approach are generically addressed; the first is associated with the calculation of the probability of occurrence of generalized crises in the financial system, and the second one addresses the identification of the underlying causes that give rise to this systemic event.

Drakos and Kouretas (2015) distinguished, like Aparicio et al. (2012), two ways of analyzing systemic risk. The first explores the channels through which risk is transmitted from one financial institution to another, commonly known in the literature as contagion. The second focuses more on the quantification of systemic risk with the use of high-frequency time series. Of these, different approaches have been proposed based on the information involved in the analysis.

Furthermore, there are different perspectives on research focuses. Estrada and Osorio-Rodríguez (2006) state that the recent literature can be classified into three groups that address different analysis perspectives. The first focuses on considering systemic risk as a natural result of considering a possible bank run. The second is based on the analysis of the individual behavior of a financial institution through liquidity risk and how this transforms into systemic risk, i.e., when the existence of explicit links between institutions allows the failure of one, or a small group of these, to be transmitted to others. Finally, an analytical perspective includes the conception of problematic banks that induce disruptions in the financial market. In this way, the market becomes the scenario where the effects of the malfunctioning of one bank are transmitted to others due to the negative changes in the positions of the other banks in the system.

Finally, Furfine (2003) considers two types of systemic risk. The first is associated with the risk caused by financial shocks that simultaneously cause a set of institutions to fall into an inefficiency function. This definition follows the same approach presented by Group of Ten (2001), who define systemic risk as the risk that exogenous shocks produce on the financial system that trigger economic losses in a significant portion of the system and that, in turn, adversely affect the real economy. The second type of systemic risk contemplated by Furfine (2003) is the risk resulting from the failure of one or a small number of institutions that could be transmitted to others in the presence of interconnecting links between institutions.

Berger et al. (2021) show that supervision enforcement actions (EAs)—the primary tools of supervisors—affect systemic risk. The authors empirically investigate relations between EAs and banks' contributions to systemic risk. The results show that the primary channel behind this relation is reduced leverage, but lower portfolio risk also plays a role. Meuleman and Vennet (2020) investigate the effectiveness of macroprudential policy to determine systemic risk in the short and long run. The systemic risk criteria are decomposed into an individual bank risk component and a systemic linkage component. The results show that the announcements of macroprudential policy actions generally have a downward effect on bank systemic risk. On average, all banks benefit from macroprudential tools in terms of their risk.

Duan et al. (2021) conducted the first broad-based international study of the effect of the COVID-19 pandemic on bank systemic risk by considering 1584 listed banks from 64 countries. The authors find that the pandemic has increased systemic risk across countries. The effect operates through government policy response and bank default risk channels. Besides, the authors suggest that the adverse effect on systemic stability is more pronounced for large, highly leveraged, riskier, high loan-to-asset, undercapitalized, and low network centrality banks.

In conclusion, the presence of different definitions of the concept of systemic risk, as well as the existence of a significant variety of measures of this, are the result of the representation of different aspects of this complex phenomenon that manifests through a wide range of different characteristics and whose results materialize in the affectation of both the financial system and the real economy through spillover effects (Vogl 2015).

The previous ways of addressing and understanding the problem of systemic risk have led to vast literature based on these approaches. For this study, the primary focus is addressing the problem from the perspective of the second approach proposed by Drakos and Kouretas (2015), complemented in turn by the second group proposed by Estrada and Osorio-Rodríguez (2006). In this way, the quantification of systemic risk is carried out from the analysis of time series, from the perspective of an individual risk analysis of institutions and impacts on systemic risk through the presence of links with the system.

As explained by Drakos and Kouretas (2015), in the existing literature after a subprime crisis, approaches that quantify systemic risk predominate; among these, as explained by Cabrera-Rodríguez et al. (2014), a series of approaches that use quantitative methods are appreciated for the development of a ranking based on the systemic importance of financial institutions, using indicators such as asset size, connectivity and substitutability; these are considered a proxy of systemic importance. Authors such as León and Machado (2011) and Laverde and Gutiérrez-Rueda (2012) are commonly referenced authors who address this type of interest for the Colombian case.

Studies have significantly contributed to this phenomenon in the literature regarding systemic risk through its authors' different interpretations. However, among this large number of contributions, the relatively constant reference to CoVaR (Tobias and Brunnermeier 2016), expected marginal shortfall (MES) (Acharya et al. 2017) and SRISK (Brownlees and Engle 2016) methodologies are notable. These methodologies stand out for being relatively easy to develop; they can be built with public information and are easily accessible. The objectives of these methods focus on measuring the contribution of each financial institution to systemic risk.

In a work developed by Tobias and Brunnermeier (2016), the systemic risk measure known as CoVaR is introduced; this measure corresponds to the value at risk for the financial system conditioned in the scenario in which an institution is at risk. The difference is taken between the previously defined CoVaR and the CoVaR for a financial system conditioned on the normal functioning of the same institution to capture the marginal contribution of a particular institution to systemic risk. With the above methodology, the development of a systemic risk measure is sought, as are the variables of financial institutions that can predict systemic events.

Unlike Tobias and Brunnermeier (2016); Acharya et al. (2017) introduce an economic model to formalize and measure a financial institution's contribution to systemic risk. For this, the authors develop an economic model called "systemic expected shortfall" (SES), interpreted as the propensity of an institution to be undercapitalized when the system is undercapitalized. The results of the application of this methodology are materialized in a useful tool for the development of policies with important practical utility for regulating systemic risk.

Luciano and Wihlborg (2018) analyze banks' choice of organizational structures theoretically in branches, subsidiaries, or standalone banks, in the presence of public bailouts and default costs. The authors consider the highest risk-taking as measured by leverage and expected loss. Zedda and Cannas (2020) propose an analysis of the systemic risk and contagion determinants by considering the effect of excluding one bank from the general financial system. This work defines the contribution of individual banks to systemic risk as to the sum of the standalone bank risk and the contagion risk.

A work considering the relationship between bank competition and systemic risk has been proposed by Silva-Buston (2019). The risk is split into a component driven by banks' commonality with the market and a component arising from other sources of interbank commonality. This relationship is robust for informationally opaque banks, financed with a larger share of uninsured sources and in countries with lower deposit insurance coverage. Davydov et al. (2021) consider the relationship between bank liquidity creation and systemic risk. The proposed approach is applied to data of US banks showing that liquidity creation decreases systemic risk at the individual bank level after controlling for bank size, asset risk and other bank-specific attributes. Besides, the paper shows that the riskiness of individual banks is negatively linked to liquidity creation.

Finally, the SRISK methodology developed by Brownlees and Engle (2016) provides a tool that captures a financial institution's contribution to systemic risk; as a measure of contribution to systemic risk, the methodology uses the expected capital deficit of an institution conditioned on a prolonged market decline. The inputs for the development of this methodology are size, leverage and expected loss of capital conditioned on the fall of the market, which the authors of this methodology call the long-run marginal expected shortfall (LRMES).

In addition to the previous methodologies, this work considers a series of characteristics identified by different contributions that address systemic risk; within this literature, it is relatively common to find associations between characteristics such as size, fund structure and business model of credit institutions, with channels through which systemic risk emerges under certain conditions. Bostandzic and Weiß (2018) proposed that the methodology was used to choose variables and link them with systemic risk measures to evaluate the explanatory power of the above characteristics on the risk measures implemented in this study.

Among the works with a research proposal similar to that for this study are the contributions by Lin et al. (2018) and Bostandzic and Weiß (2018), which also use three measures of systemic risk, previously exposed, to determine the financial institutions with a more significant contribution to systemic risk and determine the underlying factors that were conducive to such contribution. However, the two studies mentioned above use geopolitically different samples. Lin et al. (2018) use a sample of Taiwanese financial institutions; within this configuration, factors such as size, leverage ratios and price/book value influence the contribution to systemic risk in a transversal dimension. Bostandzic and Weiß (2018) use a sample of North American and European banks, finding that, on average, European banks contribute more to global risk, mainly due to a riskier portfolio and greater interconnection with the system.

As of the time of writing this manuscript, no similar studies were found in which the three risk measures were used to investigate the Colombian economy. Thus, this work presents the first approach to determine the overall performance of these systemic risk measures implemented for the Colombian banking sector.

3. Proposed Methodology

The following provides a summary of the basic inputs for constructing the systemic risk measures *CoVaR* and *MES*. Value at Risk (*VaR*) and expected shortfall (ES) are considered standard risk measures in financial institutions. *VaR* allows estimating the maximum expected loss of a risky asset or portfolio in a defined time frame and under a confidence level α, given $(1-\alpha)$ corresponds to the probability of a loss greater than the established level. For the case presented here, α equals 95%. The representation of this concept is shown below:

$$\Pr\Big(R < -VaR_{(1-\alpha)}\Big) = (1-\alpha) \tag{1}$$

where R is a random variable that represents the profit or loss of a given portfolio. *ES* is the expected shortcoming conditioned on the losses, being more significant than that of the *VaR*, that is:

$$ES_{(1-\alpha)} = -E\Big[R\Big|R \leq -VaR_{(1-\alpha)}\Big] \tag{2}$$

Indeed, the expected shortcomings are the average returns in the months in which the portfolio losses exceed the *VaR* limit.

3.1. CoVaR

It is important to note that downside risk statistics such as *VaR* are customary to present the outcomes in positive values (i.e., $-VaR$) as in Equation (1) since it is implicitly understood that these refer to a loss. However, when addressing the CoVaR methodology, the definition of VaR presented in Equation (3) shows that it does not follow this convention, in the author's own words: "In practice, the sign is often switched, a sign convention we will not follow" (Tobias and Brunnermeier 2011, p. 7).

$$P\Big(X^i \leq VaR^i_{(1-\alpha)}\Big) = (1-\alpha) \tag{3}$$

where X^i corresponds to the loss of the bank i for which the $VaR^i_{(1-\alpha)}$ is defined. Considering the previous definition, X^i corresponds to the negative variation rates of the market value of the total assets of bank i, which are covered in the data and variables section.

When performing the analysis of the risk measure of *CoVaR*, a confidence interval equal to 95% ($\alpha = 95\%$) is assumed for both the calculation of *CoVaR* and *VaR*. This confidence interval is commonly used for the development of this methodology, unless otherwise specified.

Following the interpretation of the *CoVaR* made by López-Espinosa et al. (2012), the CoVaR is defined as the maximum expected loss for a specific portfolio (in this case, a representative portfolio of the entire banking market) for a given confidence level and time horizon, conditional on the maximum expected loss on the part of one of the institutions that make up the portfolio (in this case, bank), at a specific level of trust and time horizon. According to Tobias and Brunnermeier (2016), $CoVaR^{j|i}_{(1-\alpha)}$ distinguishes the $VaR\ _{(1-\alpha)}$ system *j* conditioned on $VaR\ _{(1-\alpha)}$ of entity i, which is:

$$P\left(X^j \leq CoVaR^{j|i}_{(1-\alpha)} \middle| X^i = VaR^i_{(1-\alpha)}\right) = (1-\alpha) \tag{4}$$

where X^j is the variable of the institution for which $CoVaR^{j|i}_{(1-\alpha)}$ is defined. This value corresponds to the negative weighted variation rate of the market value of the total assets banking system for institution *i*. While Equation (3) describes the definition of *VaR*, Equation (4) has a conditioned event attached to the definition of *VaR*; this implicitly defines the *CoVaR* of the banking system conditioned on bank *i* being at a level of $(1-\alpha)\%$ of *VaR*. To calculate the entity *i* to the systemic risk of the system *j* (banking system for this study), Tobias and Brunnermeier (2016) suggest the following equation:

$$DCoVaR^{j|i}_{(1-\alpha),t} = CoVaR^{j|X^i_t = VaR^i_{(1-\alpha)}}_{(1-\alpha),t} - CoVaR^{j|X^i_t = Median}_{(1-\alpha),t} \tag{5}$$

Equation (5) presents the contribution of bank *i* to the systemic risk of the banking sector as the difference between the *CoVaR* of the system conditioned on bank i being at a level (1 − α)% of the *VaR* and the *CoVaR* of the system conditioned on bank *i* being in its "normal" state (its median). This methodology uses quantile regression to calculate *CoVaR* because of its simplicity and efficiency (Tobias and Brunnermeier 2016).

After the presentation of the methodology, some considerations will be presented below when interpreting the results. We note the lack of a causal relationship within the measure because it does not distinguish whether the contribution to systemic risk is causal or derived by common factors. Although indeed, the authors do not explicitly address the issue of causality, they mitigate the presence of common factors with the use of state variables because these fulfill the function of capturing the risk variation, which is not directly related to the risk exposure of the banking system (Tobias and Brunnermeier 2016). The econometric details of this tool are exposed in Appendix B.

The choice of *CoVaR* as a tool to characterize systemic risk has, for this study, three important considerations, according to López-Espinosa et al. (2012). The first consideration is related to the possibility of deleveraging due to greater exposure to market risk in an environment of financial stress. The second consideration is related to the possibility of monitoring the dynamics presented by the systemic contribution of a particular bank to the system. The third consideration is the adaptability of configurations to condition nonlinear patterns and other relevant effects that account for the contributions of large banks to the banking system.

3.2. MES

MES arises from the following argumentative structure developed by Acharya et al. (2017). Banks must break down the losses of the entire company into contributions of individual groups or negotiation tables (investments); in this way, the return of bank R can be decomposed into the sum of the returns obtained in each investment r_i as follows: $R = \sum_i y_i r_i$, where y_i is the investment share i in the portfolio. From the definition of ES, we have:

$$ES_{(1-a)} = -\sum_i y_i E\left[r_i \middle| R \leq -VaR_{(1-\alpha)}\right] \quad (6)$$

From this equation, the sensitivity of the general risk to exposure y_i to each investment i is given by:

$$\frac{\partial ES_{(1-\alpha)}}{\partial y_i} = -E\left[r_i \middle| R \leq -VaR_{(1-\alpha)}\right] \equiv MES^i_{(1-\alpha)} \quad (7)$$

From the above equation, MES^i is the marginal expected shortcoming of investment i and is interpreted as the risk taken by investment i and how this attaches risk to the general return for the bank. The previous measure is derived from the perspective of a single institution, and its investment instruments can be approximated to a system comprising different firms (banks). In this way, the *ES* of the banking system can be established by considering R as the aggregate return of the banking sector. Under this configuration, the *MES* corresponds to the partial derivative of the *ES* of the banking system concerning bank participation when the banking system is in an adverse scenario (Boucher et al. 2013). In this way, each bank's contribution to system risk can be measured through its *MES*; the more significant the MES of a bank, the more outstanding the bank's contribution to systemic banking risk. Finally, the construction of the *MES* methodology presents, in a reduced form, a measure of systemic risk that is a function of observable data and statistical techniques similar to the models (Acharya et al. 2017).

3.3. SRISK

SRISK, proposed by Brownlees and Engle (2016), indicates a financial institution's contribution to systemic risk; the methodology is defined as the expected capital deficit conditioned on a significant market decline. *SRISK* is a function of institution size, leverage and the expected loss in share price conditioned on the market decline. *SRISK* is derived from the stock market and accounting data to construct a measure based on market behavior and the size and degree of leverage of firms; its purpose is to measure the expected undercapitalization of a firm conditioned on a systemic event (Brownlees and Engle 2016).

The variable of interest is capital deficit, which, according to Brownlees and Engle (2016), is obtained taking into account the minimum capital reserve that the firm needs (for regulatory reasons) minus its market capitalization; in this way, the capital deficit of firm i in month t is given by CS_{it} (capital shortfall), which corresponds to:

$$CS_{i,t} = kA_{i,t} - W_{i,t} = k(D_{i,t} + W_{i,t}) - W_{i,t} \quad (8)$$

where W_{it} is market capitalization, D_{it} is the book value of the debt, A_{it} is the value of the quasi asset and k is the prudential capital requirement. For this study, the prudential capital requirement of 8% was taken as a reference. If the capital deficit determined with Equation (8) is negative, the result is interpreted as excess capital, a result that would indicate an "appropriate" functioning of the institution; if capital deficit is positive, the institution is in a state in which its main activities are at risk.

We assume that the systemic event corresponds to a sufficiently extreme scenario, understood as a decline in the market below threshold C during time horizon h to increase the utility of *SRISK*; the justification of the above lies in the model developed by Acharya et al. (2017), where the capital deficit of a firm generates negative externalities if it occurs when the system is truly at risk; for this, they denote a return of the multiperiod arithmetic market between the period $t+1$ and $t+h$ as $R_{mt+1:t+h}$ and the systemic event as $\{R_{mt+1:t+h} < C\}$. For this study, a horizon of 6 months and a threshold of 40% are considered. A representation of this concept is presented below:

$$SRISK_{i,t} = E_t(CS_{i,t+h} | R_{m,t+1:t+h} < C) = kE_t(D_{i,t+h} | R_{m,t+1:t+h} < C) - (1-k)E_t(W_{i,t+h} | R_{m,t+1:t+h} < C) \quad (9)$$

In order to calculate expectations, the authors assume (as in this work) that in the case of a systemic event, the debt cannot be renegotiated; from the above, it is necessary to:

$$E_t(D_{i,t+h}|R_{m,t+1:t+h} < C) = D_{it} \quad (10)$$

Based on the above mentioned:

$$SRISK_{i,t} = kD_{i,t} - (1-k)W_{i,t}(1 - LRMES_{i,t}) = W_{i,t}[kLVG_{i,t} + (1-k)LRMES_{i,t} - 1] \quad (11)$$

where LVG_{it} corresponds to the ratio of quasi-leverage $(D_{it} + W_{it})/W_{it}$ and $LRMES_{it}$ to expected long-term marginal losses. The arithmetic multiperiod expected return of the firm conditioned to a systemic event, is defined below:

$$LRMES_{i,t} = -E_t(R_{i,t+1:t+h}|R_{m,t+1:t+h} < C) \quad (12)$$

where $R_{it+1:t+h}$ is the multiperiod arithmetic return of the firm between periods $t+1$ and $t+h$. Based on the above, *SRISK* is a function of the size of the firm, its degree of leverage and the expected devaluation of stock conditioned on a fall in the market. In this way, *SRISK* is higher for firms that are larger, have the most leverage and have greater sensitivity to market movements (Brownlees and Engle 2016). For simplicity, the prudential ratio k (8%), the threshold C (40%) and the time horizon h (6 months) are implicit in the SRISK notation.

Brownlees and Engle (2016) use the *SRISK* of institutions to build a systemic measure of financial stress; in this way, the total systemic risk in the financial system is measured by:

$$SRISK_t = \sum_{i=1}^{N} (SRISK_{i,t})_+ \quad (13)$$

where $(SRISK_{it})_+$ denotes $max(SRISK_{it}, 0)$. $SRISK_t$ is interpreted as the approximate total amount of capital that the government should provide to rescue the financial system conditioned on the systemic event. Similarly, only the positive contributions of *SRISK* are considered, and the negative contributions are ignored because these correspond to capital surplus scenarios.

In a crisis, it is unlikely that much of the excess capital will be mobilized into loans; therefore, it is not necessarily available to support the firms affected by the systemic event (Brownlees and Engle 2016). For example, Cáceres (2009) presents evidence for the Colombian case, where a substantial reduction in lending activity was observed during 2008, mainly caused by the supply side. These facts support the idea that the excess capital of banking entities during a financial crisis does not represent a risk environment for these institutions, at least in the short term, but that it does not tend to mitigate the spread and accentuation of the systemic event significantly.

The methodology seems to be simple; however, within its structure, a relative complexity is evident when considering the *LRMES* approximation, although there are different specifications and estimation techniques to obtain it. The alternative proposed by the volatility laboratory (V-Lab) is used. The *LRMES* is constructed as follows:

$$LRMES_{i,t} = 1 - \exp(\log(1-h)x\ \beta_{i,t}) \quad (14)$$

where:

$$\beta_{i,t} = \rho_{i,t}\frac{\sigma_{i,t}}{\sigma_{m,t}} \quad (15)$$

Expressions (14) and (15) show that the *LRMES* is a function of the time horizon h, in this case, a threshold of six months for the market return to decrease by 40%, and the beta coefficient of the firm, constructed from the dynamic correlations (ρ_{it}) and conditional volatilities $(\sigma_{it}, \sigma_{mt})$ of bank i and system variable m (V-Lab n.d.). This study utilized the EGARCH model developed by Nelson (1991) for variance and a standard DCC model developed by Engle (2002) for the correlations. For more details, please see Appendix C.

The choice of the EGARCH model for estimating conditional volatilities is the result of multiple tests performed with different models of the GARCH family of models. Indeed, the EGARCH model presents the best behavior fit of volatilities on the series of the rate variations of the value in the market of the total assets of each bank i. The criteria used to determine the above were based on the minimization of information criteria such as Bayesian and Akaike.

4. Computational Results

The sample for this study comprises banks present in the Colombian market and listed on the Colombian stock exchange. It is common to find similar studies that investigate the same topic of interest as this study, but they consider many financial institutions, including insurance companies, pension funds and stockbrokers. The difficulty of accessing these financial sector segments lies in the fact that many of them are not listed on the stock market, thus preventing the proper application of the methodologies described above. Thus, only commercial banks are considered; these banks tend to be classified as the most systemically important and the most impacted by a shock from the system. This fact is due to the specialty of its business and its activities and access to liquidity configuring its actions within the system (León et al. 2011).

The sample consists of the following commercial banks: Banco de Bogotá SA, Banco Popular SA, Banco de Occidente SA, Banco Comercial AV Villas SA, Banco Bilbao Vizcaya Argentaria SA and Bancolombia SA. Data were collected monthly from 2008 to 30 June 2017; stock prices, financial statements and the number of shares in circulation for the banks were collected. The information on stock prices and financial statements is in Colombian pesos and was obtained from the National Registry of Securities and Issuers—RNVE. The period for the analysis of systemic risk measures corresponds to the availability of the information required for this study.

The three primary sources of information in this study present different frequencies in their publication: daily, monthly and quarterly, for the respective prices of shares, financial statements and annexes. This temporary mismatch requires analyzing each of the sources to determine the frequency of the result of the risk measures are presented, resulting in the choice of a monthly frequency that implies making "transformations" to the series that are not in that frequency. In the case of share prices, the monthly average of these was taken as a reference. Likewise, the number of outstanding shares, both ordinary and preferred, was transformed monthly by keeping them constant during the quarter following the cut-off date.

For the construction of this study, it is essential to carry out the analysis of the systemic risk measures with a monthly frequency instead of daily or weekly, as tended to be applied in previous studies. In this way, there is a tendency to give greater prominence to movements in the stock market, considering the behavior of the rest of the information constant with a lower frequency than that of share prices. Based on the above, the monthly frequency of the financial statements was selected as the reference for constructing risk measures. This configuration allows the results of the development of the institution's activities to be counteracted with the changes in their valuation in the stock markets. Lastly, the adoption of a monthly frequency reduced the presence of seasonal discontinuities that tend to be observed when implementing a weekly or daily frequency. The share of assets in the banking sector for the sample considered is shown in Figure 1.

As of 31 January 2008, the banking sector had 16 institutions recognized as banking establishments; however, at the end of the study period (30 June 2017), a total of 26 banking establishments were registered, representing a growth of 62.5% in the number of establishments.

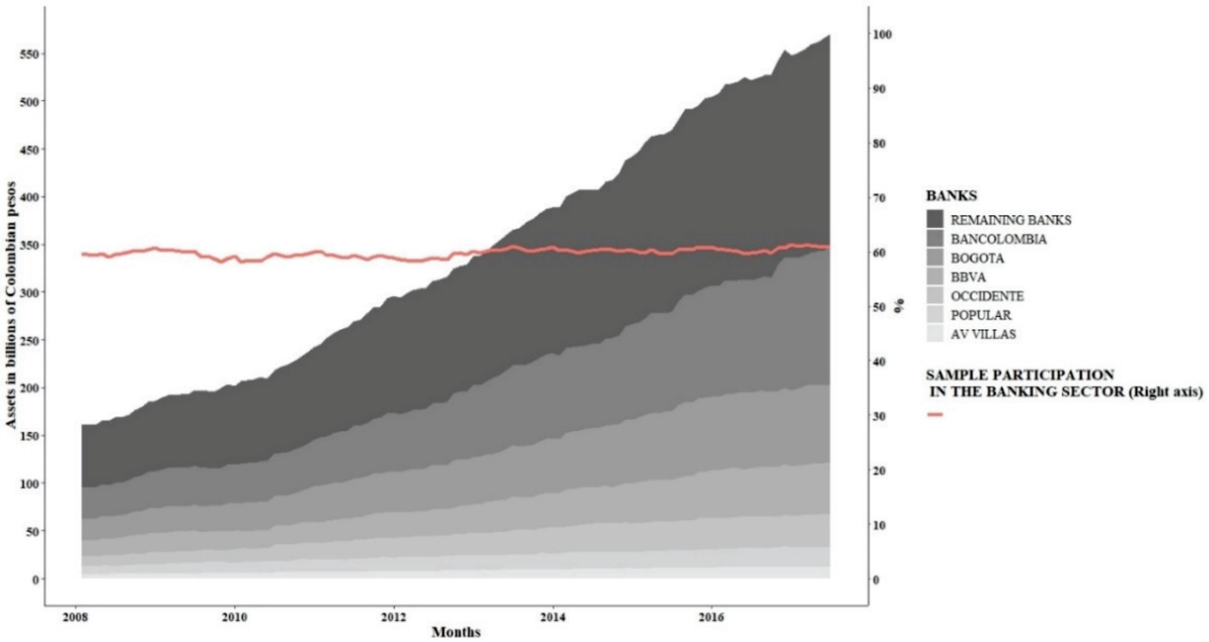


Figure 1. Participation in the level of assets of the selected sample on the sector. Source: Owner.

For the construction of the *DCoVaR*, in addition to the accounting and stock market information for the banks, information concerning state variables was required. It was decided to use the state variables cited in the academic literature that addresses the *DCoVaR* methodology. These variables are used under the argument that the interconnection between the global and local financial systems would allow these variables, coming from the US economy, to capture the temporal variation in the conditional moments of the returns of shares in large parts of the economies. Along with these, it was decided to attach the performance of relevant variables for the Colombian economy, such as the conditional volatilities of the exchange rate and indices of the Colombian stock market and the TES; for more information, see Appendix A.

According to Arias et al. (2010), we have calculated the state variables with the use of principal components. This methodology allows us to avoid multicollinearity between variables and capture 80% of the volatility of the standardized variables. The results matrix of this methodology is incorporated into the quantile regressions to capture time variation in conditional moments of asset returns. Based on the above, when incorporating the state variables into Equations (4) and (5), they would be as follows:

$$P\left(X^i \leq VaR^i_{(1-\alpha)}, M\right) = (1-\alpha) \tag{16}$$

$$P\left(X^j \leq CoVaR^{j|i}_{(1-\alpha)}, M \middle| X^i = VaR^i_{(1-\alpha)}, M\right) = (1-\alpha) \tag{17}$$

where M represents the matrix that contributes to the model 80% of the volatilities presented by the state variables, and their effect within the quantile regression is interpreted as the effect they exert on the systemic risk of the banking sector in a specific quantile q, given the risk contributed by the banking institution.

4.1. Characteristics That Contribute to Systemic Risk

To determine the characteristics of banks that contribute to systemic risk in the banking sector, the works of Bostandzic and Weiß (2018), Laeven et al. (2016), Laverde and Gutiérrez-Rueda (2012) and Lin et al. (2018) identify three characteristics that tend to generate the channels through which a bank becomes systemically important, i.e., the size of the institution, the fund structure of the institution and the business model of the institution. For more information, see Appendix A. Each of these characteristics is described below.

- Size of the Institution: Evidence indicates that size is correlated with the complexity of the institution and the interconnection with the rest of the system (Bostandzic and Weiß 2018). Large institutions are generally less substitutable, a circumstance that tends to present a greater risk to the system's integrity. The natural logarithm of an institution's total assets is taken as a proxy for the above characteristics.
- Fund Structure: For the structure of banks' funds, the leverage behavior of the institution and the fragility and composition of the funds are used. To capture the leverage behavior of an institution, the leverage variable proposed by Acharya et al. (2017), defined as the quasi-value in the market of the assets divided by the market capitalization of the institution under consideration, is used. The long-term financing indicator developed by the Financial Superintendence of Colombia and the ratio between deposits and total obligations are used to capture the structure and fragility of the funds.
- Business Model: Three variables are used that can adequately characterize this characteristic in banks to determine the structure of the business model: income other than interest, the portfolio as a percentage of assets and the portfolio provision's natural logarithm.

4.2. Market Value of Total Assets

Tobias and Brunnermeier (2011) analyze the VaR and DCoVaR based on the growth rate of the market value of the total assets of each institution; based on these, the market value of total assets is related to the supply of credit to the real economy. For this, the following transformation is performed:

$$X_t^i = \frac{ME_t^i\left(\frac{BA_t^i}{BE_t^i}\right) - ME_{t-1}^i\left(\frac{BA_{t-1}^i}{BE_{t-1}^i}\right)}{ME_{t-1}^i\left(\frac{BA_{t-1}^i}{BE_{t-1}^i}\right)} = \frac{LEV_t^i ME_t^i - LEV_{t-1}^i ME_{t-1}^i}{LEV_{t-1}^i ME_{t-1}^i} = \frac{MA_t^i - MA_{t-1}^i}{MA_{t-1}^i} \tag{18}$$

where:

$BA_t^i =$ Corresponds to the book value of the total assets of bank i at time t.
$BE_t^i =$ Corresponds to the book value of the total shares of bank i at time t.
$ME_t^i =$ Corresponds to the market value of the total shares of bank i at time t. This variable considers both ordinary and preferred shares issued by the institution under consideration.
$LEV_t^i = \frac{BA_t^i}{BE_t^i} =$ Corresponds to the ratio between total assets and the book value of the shares of bank i at time t.
$MA_t^i =$ Corresponds to the market value of the total financial assets of the bank i at time t.

The above is conducted for each institution included in the analysis. To determine the systemic risk in the market, the authors consider the weighted average of the growth rate of total assets at market price for all financial institutions; in this way, the returns of the representative systemic portfolio for bank i are configured according to:

$$X_t^{S,i} = \sum_{j=1, j\neq i}^{n} \omega_{j,t} X_t^j \tag{19}$$

$$\omega_{j,t} = W_t^j \left(\sum_{j=1, j\neq i}^{n} W_t^j\right)^{-1} \quad \textit{and} \quad 0 \leq \omega_{j,t} \leq 1 \tag{20}$$

where W_t^j is the variable used to perform the weighting. Based on the findings by López-Espinosa et al. (2012), the lagged value of the total assets of the institution under consideration is used.

$X_t^{S,i}$ corresponds to the global systemic portfolio for each institution; this is a weighted average of the returns for all banks except bank i. With the objective of generating results for three risk measures that are "comparable", it was decided to use the previously established variables for the construction of the CoVaR, MES and SRISK. Finally, Table 1 presents the descriptive statistics of the variables presented in Equations (18) and (19).

Table 1. Descriptive statistics of the market value growth rate of total assets and the system variable.

Descriptive Statistics	N	Banks											
		AV VILLAS		BANCOLOMBIA		BBVA		BOGOTÁ		OCCIDENTE		POPULAR	
		Bank	System	Bank	System	Bank	System	Bank	System	Bank	System	Bank	System
Mín	113	−0.428	−0.336	−0.514	−0.178	−0.302	−0.236	−0.142	−0.212	−0.095	−0.314	−0.197	−0.267
1st Qu.	113	−0.049	−0.038	−0.056	−0.031	−0.059	−0.023	−0.02	−0.023	−0.016	−0.029	−0.027	−0.033
Median	113	0.01	0.028	0.051	0.01	−0.01	0.017	0.006	0.022	0.002	0.026	0.003	0.024
Average	113	0.029	0.027	0.03	0.017	0.014	0.021	0.01	0.021	0.009	0.024	0.004	0.022
3rd Qu.	113	0.073	0.104	0.117	0.069	0.056	0.073	0.038	0.075	0.023	0.087	0.039	0.07
Max	113	1776	0.362	0.594	0.284	0.341	0.24	0.157	0.298	0.199	0.326	0.252	0.258

Source: Owner.

Table 1 shows two variables of interest for each Bank. The first column labeled with the name of the "Bank" corresponds to the result of Equation (18) that represents the rate of change in the market value of the total assets of a particular bank. On the other hand, the columns identified as the "System" name correspond to Equation (19). The Table 1 also reports the respective values of minimum, first quartile, median, average, third quartile and maximum for each of the previous variables. The sample period is between 28 February 2008 and 31 June 2017, with 1356 data.

5. Discussion and Managerial Insights

Tables 2 and 3 provide the EGARCH, DCC and quantile regression results for each bank and their respective proxy of the systemic variable. Likewise, the results of different tests performed to determine the suitability of the selected variables are presented. These tests include only the EGARCH model estimation that includes an ARMA model for the mean. The values in parentheses correspond to the standard error for the coefficients of the EGARCH model and the respective lags of the ARCH effects test. Finally, ***, **, * denote the significance of the coefficients, tests, or statistics at levels of 1, 5 and 10%, respectively.

The behavior of the different risk measures was assessed to address the issue related to the realization of a systemic crisis in the Colombian banking sector; the results are presented in Table 4 and Figures 2–6.

Table 2. ARMA and EGARCH results for the sample data.

Specification		AV VILLAS		BANCOLOMBIA		BBVA		BOGOTÁ		OCCIDENTE		POPULAR	
		System	Bank	System	Bank	System	Bank	System	Bank	System	Bank	System	Bank
ARMA	AR1	−0.034479 *** (0.000002)	−0.246957 *** (0.000079)		−0.049502 *** (0.000090	−0.214773 *** (0.000130)	0.302472 *** (0.000307)	−1.07579 *** (0.000134)		−1.026864 *** (0.001008)	−0.404966 *** (0.000017)	−0.103862 *** (0.000009)	
	AR2	0.842407 *** (0.000113)	0.151093 *** (0.000241)		−0.195470 *** (0.000173)	0.824432 *** (0.000828)	−0.726260 *** (0.002171)	−0.94786 *** (0.000556)		−1.095858 *** (0.001594)	−0.731353 *** (0.000040)		
	AR3				0.041250 *** (0.000139)					−0.144043 *** (0.000095)			
	MA1	0.275944 *** (0.000001)		0.224597 *** (0.000269)		0.470256 *** (0.000408)	−0.133322 *** (0.000054)	1.05083 *** (0.003667)	−0.25046 *** (0.000009)	1.125084 *** (0.000669)	0.533917 *** (0.000044)		−0.046343 *** (0.000009)
	MA2	−1.095661 *** (0.000231)				−0.991695 *** (0.000063)	0.538468 *** (0.000342)	0.86165 *** (0.001624)		1.234945 *** (0.000874)	0.686710 *** (0.000009)		−0.076016 *** (0.000005)
	MA3	−0.240641 *** (0.000104)				−0.188052 *** (0.000150)	0.278034 *** (0.000166)			0.292615 *** (0.000109)	0.229717 *** (0.000012)		−0.073302 *** (0.000001)
	MA4	0.355981 *** (0.000014)				0.264615 *** (0.000076)	−0.447289 *** (0.001104)				−0.184071 *** (0.000318)		0.051772 *** (0.000028)
EGARCH	ARCH1	0.799366 *** (0.000303)	0.012710 *** (0.000031)	0.246493 *** (0.000754)	−0.576275 *** (0.000133)	0.605561 *** (0.000035)	−0.069722 *** (0.000166)	0.33965 *** (0.001264)	0.44333 *** (0.000105)	0.537437 *** (0.000114)	0.422964 *** (0.001229)	0.011183 *** (0.000047)	0.149429 *** (0.000013)
	ARCH2	−0.579092 *** (0.000059)	0.323211 *** (0.000908)	−0.384483 *** (0.001015)	0.140552 *** (0.000034)	−0.789385 *** (0.000009)				0.132865 *** (0.000190)		0.154661 *** (0.000050)	0.157589 *** (0.000014)
	ARCH3	−0.451457 *** (0.000105)	−0.853260 *** (0.000486)		0.702936 *** (0.000578)	0.219627 *** (0.000737)						−0.104234 *** (0.000032)	
	ARCH4	0.471059 *** (0.000138)			−0.376674 *** (0.000009)								
	GARCH1	0.644798 *** (0.000292)	0.726499 *** (0.000853)	0.997655 *** (0.009199)	0.036037 *** (0.000044)	0.540945 *** (0.000161)	0.961643 *** (0.003489)	0.35768 *** (0.001203)	0.41782 *** (0.000115)	−0.042545 *** (0.000041)	0.815780 *** (0.000302)	−0.125789 *** (0.000047)	−0.227812 *** (0.000061)
	GARCH2		−0.107191 *** (0.000058)	−0.022221 *** (0.000009)	0.534309 *** (0.000690)	−0.490466 *** (0.000137)		−0.21096 *** (0.000147)	−0.77307 *** (0.000482)	−0.328719 *** (0.000051)	−1.000000 *** (0.000075)	0.973641 *** (0.001867)	−0.009186 *** (0.000001)
	GARCH3		0.787467 *** (0.001844)		0.296964 *** (0.000203)	0.688145 *** (0.000609)		0.70607 *** (0.000028)	0.53256 *** (0.000298)	0.493064 *** (0.000072)	0.712506 *** (0.000023)	0.036081 *** (0.000036)	0.734146 *** (0.000337)
	GARCH4		−0.748643 *** (0.001662)					−0.62262 *** (0.001066)	−0.86264 *** (0.000495)	0.037023 *** (0.000091)	−0.066404 *** (0.000157)	−0.682825 *** (0.000912)	−0.129795 *** (0.000016)
	GAMMA1	−0.885067 *** (0.000101)	0.996195 *** (0.000574)	0.115575 *** (0.000175)	0.632487 *** (0.000820)	−0.420413 *** (0.000873)	−0.246882 *** (0.001143)	1.15198 *** (0.032347)	1.21458 *** (0.000511)	0.721123 *** (0.000233)	0.930880 *** (0.000041)	1.589955 *** (0.001813)	0.459014 *** (0.000084)
	GAMMA2	0.029432 *** (0.000030)	−0.943253 *** (0.000887)	−0.331180 *** (0.001007)	−1.079320 *** (0.000993)	−0.417899 *** (0.000043)				1.408392 *** (0.000516)		0.076924 *** (0.000026)	1.140554 *** (0.000129)
	GAMMA3	1.135463 *** (0.000306)	−0.448216 *** (0.000036)		−1.413676 *** (0.000585)	−0.140913 *** (0.000193)						−0.888913 *** (0.001216)	
	GAMMA4	−1.427526 *** (0.0000499			1.201497 *** (0.002484)								

Source: Owner.

Table 3. Estimated results for the data sample.

Specifications		AV VILLAS		BANCOLOMBIA		BBVA		BOGOTÁ		OCCIDENTE		POPULAR	
		System	**Bank**	**System**	**Bank**	**System**	**Bank**	**System**	**Bank**	**System**	**Bank**	**System**	**Bank**
Statistic of Weighted ARCH LM Tests	LAG	3.562 * (8)	3.006 * (6)	2.149 (5)	0.2935 (8)	0.3971 (7)	0.8164 (3)	0.1252 (6)	0.006104 (6)	2.378 (7)	0.05129 (6)	0.01764 (8)	0.004192 (7)
	LAG	5.027 (10)	4.306 (8)	2.688 (7)	1.3683 (10)	1.5182 (9)	0.9351 (5)	0.4744 (8)	0.108698 (8)	2.679 (9)	0.34029 (8)	1.50173 (10)	0.008476 (9)
	LAG	6.327 (12)	4.751 (10)	3.553 (9)	1.8888 (12)	2.8858 (11)	1.656 (7)	1.0434 (10)	0.211909 (10)	4.512 (11)	3.94416 (10)	1.72761 (12)	0.013002 (11)
Statistic of Weighted Ljung-Box Test on Standardized Residuals	LAG(1)	0.345	0.225	0.5031	0.01174	0.4766	0.3104	0.9113	0.6687	0.003	0.1306	0.7995	0.07855
	LAG [2 x (p + q) + (p + q) − 1]	9.145	1.497	1.1633	1.43577	9.1705	5.7734	5.6035	2.1882	7.931	6.9093	1.3296	3.1782
	LAG [4 x (p + q) + (p + q) − 1]	15.927	2.522	2.2855	3.18303	18.0695	15.5513	8.8423	3.8606	13.376	11.9768	2.4611	6.2216
t-value of Sign Bias Test	Sign Bias	0.2625	1.92338 *	0.474	1.0172	1.1945	0.1973	1.0692	0.6155	0.4014	0.5628	0.7137	0.9539
	−Sign Bias	0.3557	0.31216	0.01592	0.21837	0.1144	0.4438	0.4976	0.061	0.6078	0.152	0.783	0.2512
	+Sign Bias	1.289	0.05067	0.24486	0.02748	1.1768	0.1897	1.6055	0.2599	0.6237	1.532	1.476	1.2071
	Joint Effect	2.043	6.30388 *	0.33806	3.16204	2.0518	0.2438	3.2451	1.4817	1.0996	2.3709	3.0869	7.3373 *
Statistic of Adjusted Pearson Goodness-of-Fit Test	Group 20	25.23	24.52	24.52	23.11	16.73	10.01	13.9	14.61	18.86	13.55	17.44	19.92
	Group 30	32.75	30.1	29.57	26.38	36.47	20.01	22.66	29.04	27.97	22.13	27.44	20.54
	Group 40	42.4	48.77	40.27	31.78	36.73	27.53	28.95	30.37	45.23	33.19	37.44	36.73
	Group 50	47.62	58.24	37.88	48.5	42.7	37	44.08	31.2	47.62	39.65	47.62	37.88
Dynamic Conditional Correlation (DCC)	p1	0.051755 ** (0.025315)		0.039008 ** (0.015635)		0.086556 *** (0.004175)		0.25804 *** (0.059624)		0.000000 ** (0.000000)		0.063595 *** (0.021299)	
	q1	0.486054 *** (0.185008)		0.860154** (0.380315)		0.461762 ** (0.214375)		0.37346 *** (0.131889)		0.902351 *** (0.115806)		0.553310 *** (0.165288)	
Normal distribution		Yes		Yes		Yes		Yes		Yes		Yes	
Quantile Regression (τ = 0.05)	Intercept	−0.13094 *** (0.01403)		−0.103357 *** (0.009594)		−0.10624 *** (0.01127)		−0.10761 *** (0.01062)		−0.13298 *** (0.01459)		−0.10531 *** (0.01152)	
	Beta	−0.03467 (0.22383)		0.189388 *** (0.055078)		0.20543 ** (0.07904)		0.73401 *** (0.17231)		0.73588 *** (0.18341)		−0.11914 (0.16436)	

Banks: AV VILLAS, BANCOLOMBIA, BBVA, BOGOTÁ, OCCIDENTE, POPULAR.

Table 4. Descriptive statistics of systemic risk measures.

Banks	Descriptive Statistics	N	SRISK	%DCoVaR	%MES
AV VILLAS	Mín	113	454.30	–2.38	–6.26
	1st Qu.	113	787.00	–0.96	–1.58
	Median	113	1263.00	–0.79	–1.02
	Average	113	1191.00	–0.81	–1.13
	3rd Qu.	113	1496.00	–0.55	–0.55
	Max	113	1891.00	–0.24	1.80
BANCOLOMBIA	Mín	113	3593.00	1.93	0.60
	1st Qu.	113	7959.00	3.66	2.44
	Median	113	10,307.00	4.34	3.64
	Average	113	9675.00	4.47	3.76
	3rd Qu.	113	11,528.00	5.11	4.48
	Max	113	13,455.00	0.80	10.98
BBVA	Mín	113	966.00	1.78	0.94
	1st Qu.	113	2298.00	2.27	4.03
	Median	113	2981.90	2.54	4.77
	Average	113	2888.20	2.51	4.81
	3rd Qu.	113	3460.10	2.71	5.68
	Max	113	4348.40	3.15	7.45
BOGOTA	Mín	113	3480.00	3.25	–3.26
	1st Qu.	113	8466.00	5.23	0.17
	Median	113	12,759.00	6.27	1.20
	Average	113	12,024.00	6.29	1.27
	3rd Qu.	113	15,956.00	6.96	2.39
	Max	113	19,400.00	14.01	7.44
OCCIDENTE	Mín	113	1835.00	2.46	0.69
	1st Qu.	113	3680.00	3.92	1.13
	Median	113	4393.00	4.97	1.44
	Average	113	4110.00	5.22	1.52
	3rd Qu.	113	4988.00	5.89	1.72
	Max	113	5624.00	13.36	3.95
POPULAR	Mín	113	1530.00	–1.98	–1.78
	1st Qu.	113	2160.00	–1.25	0.05
	Median	113	2973.00	–1.07	0.38
	Average	113	2766.00	–1.13	0.41
	3rd Qu.	113	3182.00	–0.93	0.82
	Max	113	4395.00	–0.58	3.11
SAMPLE	Mín	113	12,044.00	2.96	0.00
	1st Qu.	113	26,491.00	3.66	1.89
	Median	113	35,914.00	4.09	2.39
	Average	113	32,653.00	4.15	2.60
	3rd Qu.	113	39,109.00	4.56	3.21
	Max	113	44,470.00	6.26	5.62

Source: owner.

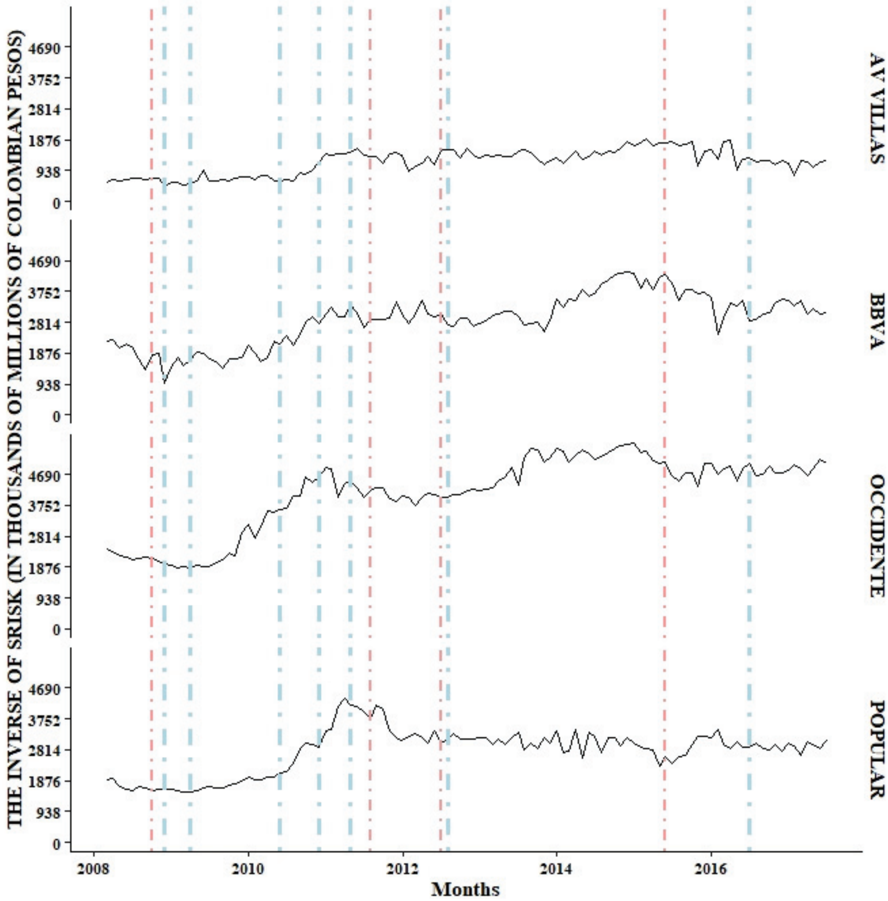


Figure 2. Monthly behavior of systemic risk measures. Source: owner.

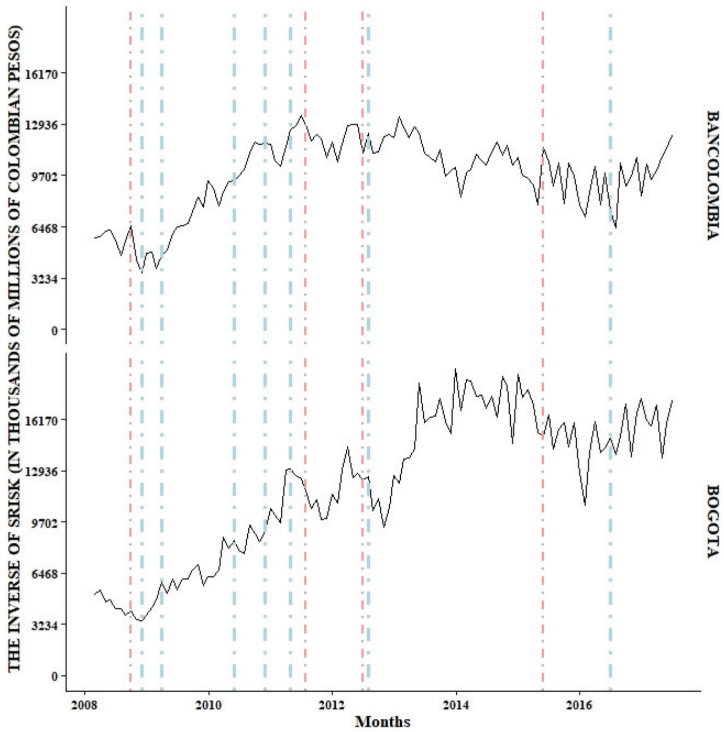


Figure 3. Monthly behavior of the SRISK systemic risk measure for BANCOLOMBIA and BOGOTÁ banks. Source: owner.

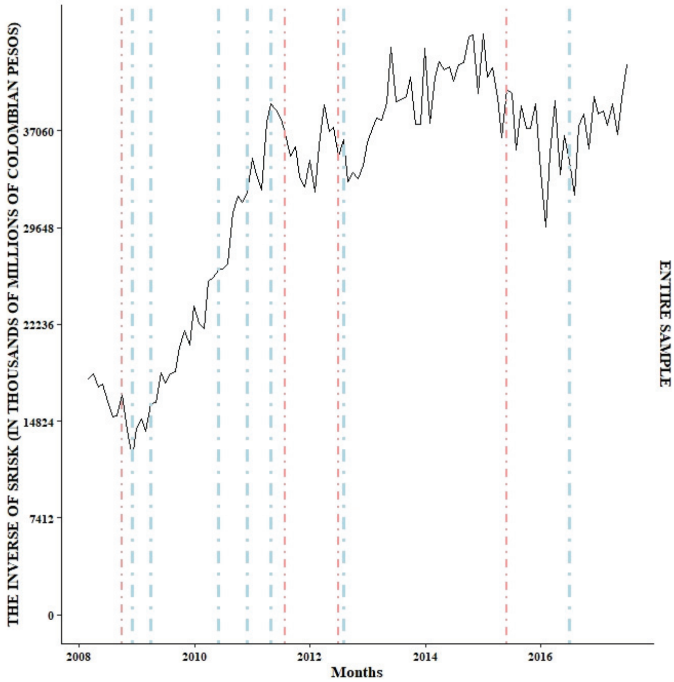


Figure 4. Monthly behavior of the SRISK systemic risk measure for the entire sample. Source: owner.

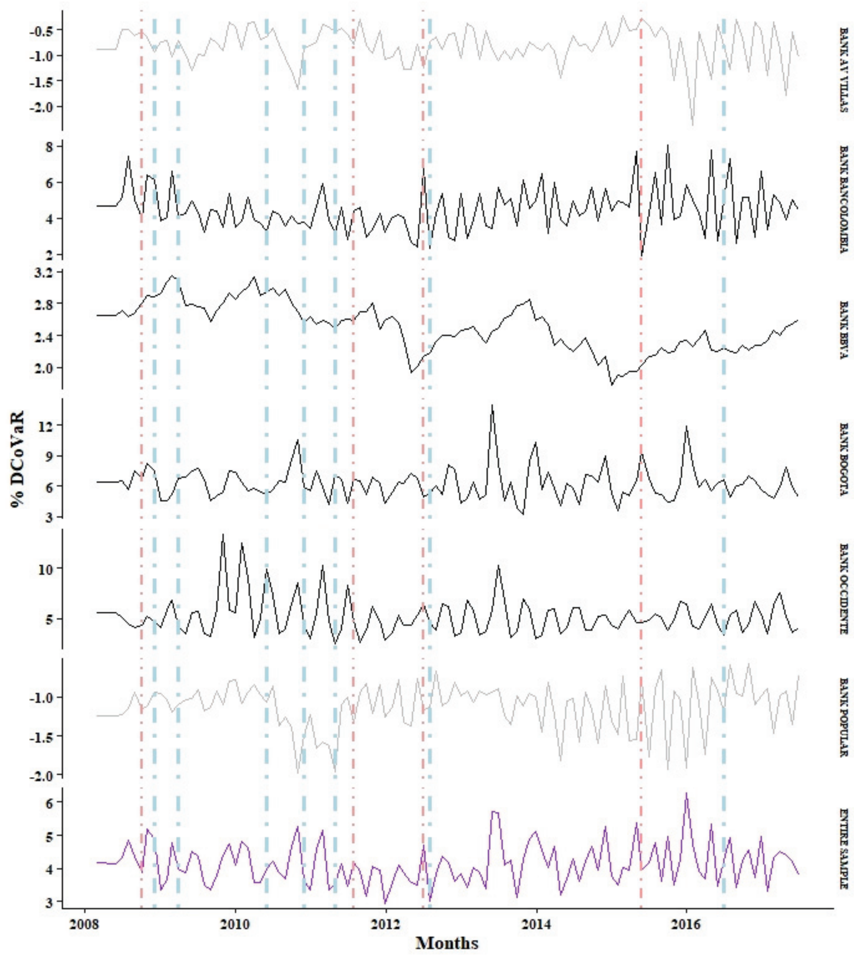


Figure 5. Monthly behavior of the DCoVaR systemic risk measure for both banks and the entire sample. Source: owner.

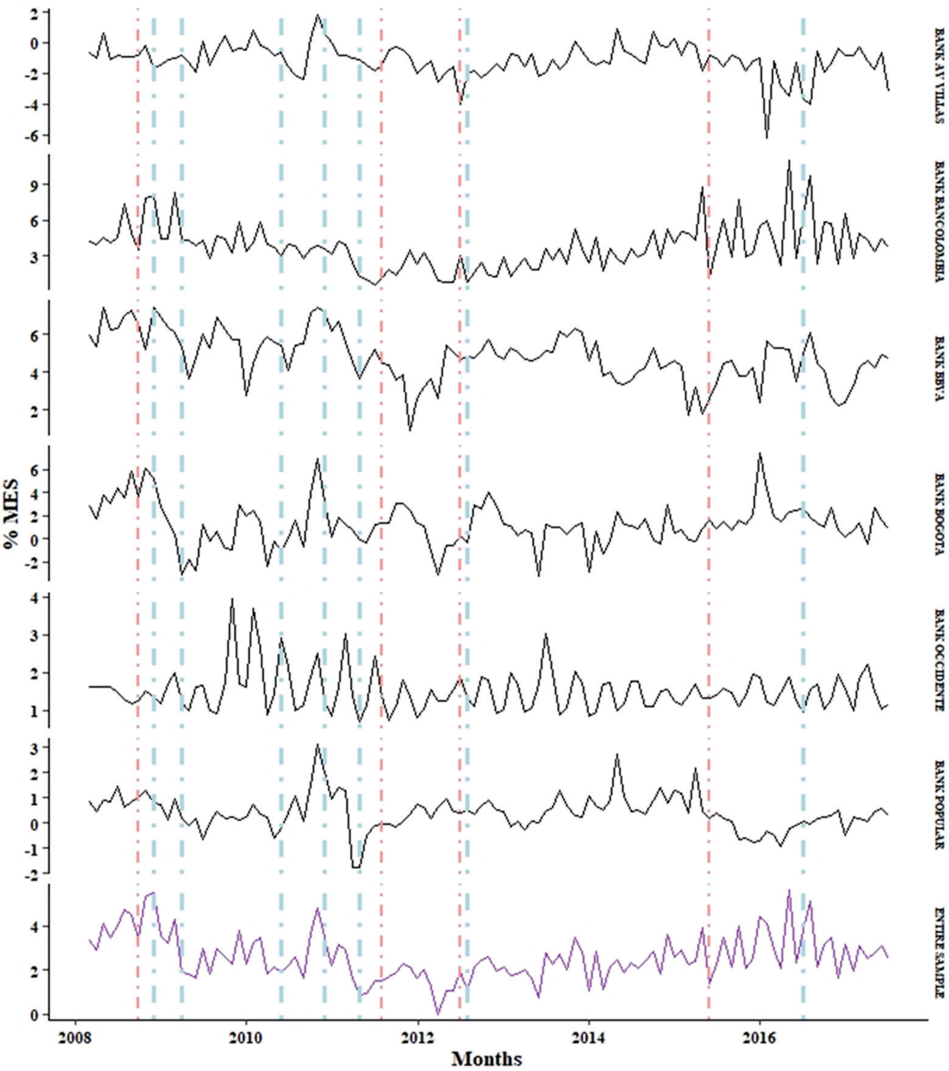


Figure 6. Monthly behavior of the MES systemic risk measure for both banks and the entire sample. Source: Owner.

For the DCoVaR estimation, no evidence showed that the lagged state variables capture the variation in tail risk not directly related to the financial system risk exposure. This finding suggests that using the rate of variation of the value in the market of total assets with a monthly frequency significantly impacted the prominence of share prices on the behavior of the time series from stocks. We developed the methodology in search of a relative equilibrium on the information sources that make up the market value of the total assets for each institution and its respective systemic variable. Thus, the results of this risk measure correspond to dynamic behaviors without state variables.

Table 5 provides the descriptive statistics for the SRISK, DCoVaR and MES risk measures for both the banks included in the study and the respective proxy for the sample; for the SRISK indicator, the results are the sum of the individual results for the banks following the methodology described by Brownlees and Engle (2016), and for the DCoVaR and MES risk measures, the results are the weighted average of assets.

Table 5. Results of the explanatory power of the characteristics of the banks.

	Dependent Variables		
Independent Variables	**Inverse of SRISK**	**DCoVaR**	**MES**
Log-natural assets		−0.00123 (0.00440)	0.00625 (0.00630)
Leverage		−0.00170 (0.00153)	−0.00023 (0.00110)
Long-term financing index	−0.00140 (0.00390)	−0.00250 *** (0.00088)	−0.00060 (0.00210)
Deposits	−0.00986 (0.00626)	−0.00054 (0.00081)	−0.00250 (0.00270)
Other incomes different from the interest	−0.00080 (0.00340)	−0.00045 (0.00036)	−0.00100 (0.00100)
Portfolio	−0.00170 (0.00620)	0.00011 (0.00071)	−0.00080 (0.00100)
Provisions Portfolio	−0.00490 (0.02000)	−0.00057 (0.00470)	0.00630 (0.00850)
Fixed effects—banks	Yes	Yes	Yes
Fixed effects—time	Yes	Yes	Yes
Observations	672	452	565
R^2	0.0006	0.0170	0.0420
Adjusted R^2	−0.2190	−0.3500	−0.2252

Source: owner.

The SRISK systemic risk measure is presented separately to implement the same range for the Y-axis. However, we decided to separate the sample according to the size of each bank. The banks were separated into two samples. According to the sample, Figure 2 corresponds to the relatively small banks, these being the AV VILLAS, OCCIDENTE, POPULAR and BBVA banks. Figure 3 corresponds to the two largest banks in the Colombian banking sector, Banco Bogotá and Bancolombia. Finally, the SRISK measurement for the sample is found in Figure 4. On the other hand, in the case of the DCoVaR and MES systemic risk measures, no way was found to divide the figure based on some factor that differentiates them.

In Figure 5, the DCoVaR values for the AV Villas and Popular banks are shown in a lighter tonality than those for the other banks, a fact that is explained by the lack of evidence (statistically speaking) that extreme events (VaR) of these banks have some effect on the extreme events (VaR) of the sector. In addition, the observed results indicate an inverse relationship to that expected for this risk measure. Furthermore, this lack of agreement also transcends the MES for these banks.

The monthly behavior of the systemic risk measures (INVERSE OF SRISK (Panel A), DCoVaR (Panel B) and MES (Panel C)) is presented for each bank and the entire sample. The lines formed by dots and vertical dashes represent the dates in which there were shocks (positive and negative) with importance in the making of investment decisions: (−) September 2008, the bankruptcy of Lehman Brothers; (+) November 2008, implementation of the Quantitative Easing policy; (+) March 2009, the central banks of large countries came together to face the crisis; (+) May 2010, the first bailout was provided (Greece, 110,000 million euros); (+) November 2010, 67,500 million euros to Ireland; (+) April 2011, 78,000 million euros to Portugal; (−) July 2011, 109,000 million euros to Greece (second bailout); (−) June 2012, Spain requested economic aid from the European Union; (+) July 2012, the European central bank declared its willingness to do everything necessary to sustain the euro; (−) May 2015, 86,000 million euros to Greece (third bailout); (+) June 2016, recovery began in Spain (Reig 2017).

There is evidence of increases in the DCoVaR and MES risk measures during periods of greater volatility in global markets; however, should these results be interpreted as realizations of systemic events in the banking sector, or are they simply the results of market shocks? The above question can be clarified when considering the estimation results for the systemic risk index of the inverse SRISK, that is, the behavior of the capital surplus in the face of significant declines in the market, providing evidence that indicates that the Colombian banking sector did not experience a systemic crisis in at least 60% of the assets managed in the sector, thus avoiding compromising the stability of the system.

Causes that could have influenced the risk measure results are the different impacts that occurred in both developed and emerging economies; the latter tend to show impacts on risk measures that are essentially due to a slowdown or reversal in capital flows that are managed through the banking sector and that, in turn, would be influenced by external factors, regardless of the economic fundamentals of the (emerging) country under consideration (Foggitt et al. 2017).

Likewise, Coleman et al. (2018) provide a critical reinterpretation of the systemic risk indicator SRISK, i.e., the different ways in which the systemic risk event was propagated during the subprime crisis and how economies were linked to this event determined the suitability and interpretation of SRISK, and therefore this indicator should be interpreted as the propensity to face losses during a crisis.

To complete this section of this research and complement the previous results, Table 5 presents the regression results using panel data for the respective risk measures, and explanatory variables addressed in the methodology. The objective of this regression is to find evidence that the explanatory variables related to characteristics such as the institution's size, the structure of funds and the banks' business model explain the contribution of different measures to systemic risk.

The regression using panel data considers fixed effects both for the banks and for time, and the respective standard errors provided in Table 5 were controlled for both heteroscedasticity and the serial correlation from the robust estimation of the covariance matrix. In addition, to avoid the presence of a unit root in the SRISK series, the growth rate (in logarithms) of this indicator is used. Finally, the explanatory variables were lagged to mitigate that both the dependent and independent variables were determined simultaneously (Bostandzic and Weiß 2018). Likewise, the independent variables were standardized to have a mean of zero and a standard deviation of one, thus facilitating the results' interpretation.

In Table 5, the rows show the explanatory variables, and columns 2, 3 and 4 show the dependent variable in the regression. For the inverse SRISK, the natural logarithm of the assets and leverage were not considered because these variables are inputs for the construction of the SRISK. Below the estimated coefficients are their respective standard errors. Finally, ***, **, and * represent statistical significance at the levels of 1, 5 and 10%, respectively.

The results in Table 5 indicate that most of these variables do not significantly influence the behavior of the risk measures implemented, except for the long-term financing index, which has a significant influence on the DCoVaR risk measure. Considering the above, increasing one standard deviation in the long-term financing index decreases the DCoVaR risk measure by 25 basis points. Despite having statistical significance, the above result economically presents an incongruence between the logic of this variable and its relationship with the risk measure.

In general, the three configurations of the panel data regressions show a poor fit that the R^2 and the R^2 adjusted evidence; however, the most critical finding is that the variables present in the regressions, for the case studied herein, do not explain the results observed in the measures of systemic risk, evidence that would support a systemic scenario in the Colombian banking sector through the channels identified as promoters of systemic contributions.

6. Concluding Remarks

This paper proposes a methodology for measuring systemic risk in the banking system. The methodology estimates three systemic risk measures widely referenced in academic papers after the subprime crisis, known as DCoVaR, MES and SRISK systemic risk index. These measures individually tend to capture characteristics of systemic risk events. Therefore, the combined use would better understand and identify the causes or triggers of systemic risk in the Colombian banking sector. The proposed methodology has been tested in the Colombian banking system to determine if this sector presented a systemic crisis between February 2008 and June 2017. Similarly, the banks' characteristics would have to support the contribution to the sector's systemic risk.

We found evidence that a systemic event would not have materialized in the Colombian banking sector. The conclusion was provided by the SRISK systemic risk index results, which capture a particular bank's undercapitalization in the face of the prolonged market downturn. The results showed that none of the banks considered presented a scenario of undercapitalization, implying that the economic losses imposed, mainly due to the subprime crisis, failed to endanger the stability of the Colombian banking system.

Finally, the research results that seek to determine the explanatory power of the variables used as proxies of the characteristics identified as causing the systemic importance of an institution showed that these would not be explaining the behavior of the risk measures. This result would reinforce the conclusion of the absence of a systemic risk scenario in the Colombian banking sector and that the observed results in the risk averages were the product of the external impacts to which the sector was exposed.

For methodological terms, it is essential to point out that the configuration of a monthly frequency for the analysis of systemic risk measures could be generating a Loss of explanatory power by the state variables evaluated for the DCoVaR modeling. On the other hand, we found that the mechanisms identified as drivers of systemic risk did not explain the behavior of the risk measures obtained in this work. The need to exhaustively evaluate the mechanisms by which both direct and indirect impacts, coming from the crisis suppresses, interact with the financial and regulatory systems of the Colombian economy by proposing a topic of interest to be developed. However, we consider that the preceding does not distort the results of the risk measurements implemented in this study, but rather, it is a reflection of the complexity that sustains this type of risk for both local and international financial systems.

Determining that the presence of systemic risk in the Colombian banking sector was not configured during the sample does not imply that these results should be interpreted in a wrong way that leads to thinking that the Colombian banking sector is prepared to face any risk arising from the international context. On the contrary, the subprime crisis scenario reveals the growing dynamism of financial systems, which poses new challenges for the administration and management of risk both at the institutional level and for financial regulatory institutions and the central bank. For this reason, it is necessary to carry out the consolidation of joint work for the appropriate schematization to address the mechanisms that trigger instability within the national financial sector, in search of safeguarding not only its stability and solidity, but also the productive sector of the country that is today more exposed to the dynamics presented by the financial systems, mainly national.

The results observed through the SRISK systemic risk index consistently respond to the events observed during and after the subprime crisis. In this way, this index is postulated with a tool that could be implemented to base or complement an early warning indicator for the Colombian economy, since its implementation in this research adequately captured the periods in which the sector was more affected by the events of greater relevance in the international context.

Author Contributions: Conceptualization, O.R.-E. and J.W.E.; methodology, O.R.-E.; software, O.R.-E.; validation, O.R.-E., J.W.E. and D.F.M.; formal analysis, J.W.E.; investigation, O.R.-E.; resources, O.R.-E.; data curation, O.R.-E.; writing—original draft preparation, J.W.E.; writing—review and editing, J.W.E.; visualization, O.R.-E., supervision, J.W.E.; project administration, J.W.E. and D.F.M.; funding acquisition, D.F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Appendix A

Variable Name	Definition	Data Source
VIX	VIX measures market expectation of near-term volatility conveyed by stock index option prices	Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org (accessed on 27 January 2020)
LIQSPR	Short-term liquidity margin. Difference between three-month repo rate and three-month treasury bill rate.	Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org (accessed on 27 January 2020)
TBR3M	Change in the three-month treasury bill rate.	Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org (accessed on 27 January 2020)
YIESPR	Change in the slope of the returns curve. Spread between ten-year and three-month treasury bill rate.	Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org (accessed on 27 January 2020)
CRESPR	Change in the credit spread between BAA-rated bonds and treasury bill rate (both with a maturity of ten years).	Federal Reserve Bank of St. Louis, https://fred.stlouisfed.org (accessed on 27 January 2020)
VOLTRM	Conditional volatility of the representative foreign exchange rate returns of the Colombian foreign exchange market. Obtained from an ARMA-EGARCH (3.4–7.2) model.	Own calculations with data obtained from the Banco de la república de Colombia.
VOLCOLG	Conditional volatility of the Colombia’s stock market index returns. Obtained from an ARMA-EGARCH (5.3–7.1) model.	Own calculations with data obtained from the Banco de la república de Colombia.
VOLTES	Conditional volatility of the returns of Colombia’s treasury bills index (IDXTES). Obtained from an ARMA-EGARCH (3.4–6.3) model.	Own calculations with data obtained from the Banco de la república de Colombia.
Banks’ characteristics		
Total Assets	Natural logarithm of the asset’s book value.	Own calculations with data obtained from the managerial indicators published by the Superintendencia Financiera de Colombia.

Variable Name	Definition	Data Source
Leverage	It corresponds to the quasi-market value of assets divided by the market value of equity, where the quasi-market value of assets is the book value of assets minus book value of equity + market value of equity (Acharya et al. 2017).	Own calculations with data obtained from the managerial indicators published by the Superintendencia Financiera de Colombia.
Operating income different from interests	Operating income different from interest divided by total interest income.	Own calculations with data obtained from the managerial indicators published by the Superintendencia Financiera de Colombia.
Portfolio	Participation of gross portfolio in total assets. This variable is calculated as the division of the gross portfolio by total assets.	Own calculations with data obtained from the managerial indicators published by the Superintendencia Financiera de Colombia.
Portfolio provision	Natural logarithm of the portfolio provision.	Own calculations with data obtained from the managerial indicators published by the Superintendencia Financiera de Colombia.
Long-term financing index	Indicator that captures the need to finance short-term debts with long-term resources (assets). It is obtained by dividing the subtraction of the obligations and short-term assets by the long-term assets.	Own calculations with data obtained from the managerial indicators published by the Superintendencia Financiera de Colombia.
Deposits	Corresponds to the division of total deposits by total liabilities.	Own calculations with data obtained from the managerial indicators published by the Superintendencia Financiera de Colombia.

Source: Owner.

Appendix B

Quantile Regression

The econometric tool used to capture the codependency between institutions and the system is the Quantile Regression. This is considered adequate according to Arias et al. (2010), since:

This methodology provides a more extensive analysis than ordinary least squares because it estimates the relationship between random variables considering different quantiles. (. . .) Furthermore, this is a methodology that can be easily estimated for a large number of independent variables (p. 4).

This Regression method also approximates the different risk scenarios since it enables the evaluation of specific quantiles, thus capturing a large part of the states of nature captured by the sample used. A low quantile of the distribution is taken into account to estimate the CoVaR since it is in these small quantiles that financial stress episodes materialize; for this reason, the Regression by quantiles is convenient and applicable.

In general, estimating a Regression by quantiles consists of minimizing the sum of the residuals, weighted asymmetrically by a function that depends on the analyzed quantile τ.

That is, the τ Quantile Regression, where $0 < \tau < 1$. The above can be represented as a solution for the following expression:

$$min_{\beta} \sum_{t} \rho\tau(y_t - f(x_t, \beta)) \tag{A1}$$

where y_t is the dependent variable, $f(x_t, \beta)$ represents a linear function of the parameters and variables used to explain the behavior of y_t and $\rho\tau$ represents the weight assigned to each observation, depending on the quantile τ analyzed. In the methodology proposed by Koenker and Bassett (1978), they propose the following representation for Equation (A1):

$$min_{\beta} \left[\sum_{t \in \{t: y_t \geq f(x_t, \beta)\}} \tau |y_t - f(x_t, \beta)| + \sum_{t \in \{t: y_t < f(x_t, \beta)\}} (1 - \tau)|y_t - f(x_t, \beta)| \right] \tag{A2}$$

Appendix C

Specifications for the Modeling of Variance and Conditioned Dynamic Correlations (DCC)

The equation to capture the time variation of volatility follows the structure of the exponential GARCH model (EGACH) developed by Nelson (1991), which presents a solution to the problems associated with pessimistic estimates of the variance parameters. This specification makes it possible to capture the differential effect observed in shocks from both the "bad" and the "good" news on the behavior and magnitude of volatility, facts not captured by the standard ARCH/GARCH models because in these the conditional variance is not affected by the sign of the errors of past periods.

The equation of the EGARCH model for the volatility dynamics follows the following structure:

$$\ln\left(\sigma_t^2\right) = \omega + \sum_{i=1}^{q} \left[\alpha_i \epsilon_{t-i}^2 + \gamma_i \left(\left|\epsilon_{t-i}^2\right| - E(|\epsilon_{t-i}|)\right)\right] + \sum_{i=1}^{p} \beta_i \ln\left(\sigma_{t-i}^2\right) \tag{A3}$$

From Equation (A3), we have that the shocks of positive returns ("Good news") have $\alpha_i + \gamma_i$ impacts on the volatility of the return, while a negative shock in the returns ("bad news") has a shock of $\alpha_i - \gamma_i$ on the volatility of the return. The DCC conditioned dynamic correlation model specification uses volatility-adjusted returns as follows:

$$\in_{it} = \frac{r_{it}}{\sigma_{it}}, \quad \in_{mt} = \frac{r_{mt}}{\sigma_{mt}} \tag{A4}$$

$$Cor\begin{pmatrix} \in_{it} \\ \in_{mt} \end{pmatrix} = R_t = \begin{bmatrix} 1 & \rho_{it} \\ \rho_{it} & 1 \end{bmatrix} = diag(Q_{it})^{-\frac{1}{2}} diag(Q_{it})^{-\frac{1}{2}} \tag{A5}$$

where Q_{it} is also called the pseudo correlation matrix, in this way, the DCC model then specifies the dynamics of the Q_{it} pseudo-correlation matrix as:

$$Q_{it} = (1 - \alpha c_i - \beta c_i)S_i + \alpha c_i \begin{bmatrix} \in_{it-1} \\ \in_{mt-1} \end{bmatrix} \begin{bmatrix} \in_{it-1} \\ \in_{mt-1} \end{bmatrix}' + \beta c_i Q_{it-1} \tag{A6}$$

where S_i is the matrix of unconditional correlations of the adjusted returns of the firm and the market, the model is estimated in two steps through the quasi maximum likelihood estimation. For more details on the estimation of this model, refer to Engle (2002).

References

Acharya, Viral V. 2009. A theory of systemic risk and design of prudential bank regulation. *Journal of Financial Stability* 5: 224–55. [CrossRef]

Acharya, Viral V., Lasse H. Pedersen, Thomas Philippon, and Matthew Richardson. 2017. Measuring systemic risk. *The Review of Financial Studies* 30: 2–47. [CrossRef]

Aparicio, Mónica, Alexandra Heredia, Camilo José Hernández, Juan Carlos Quintero, and Olga Esperanza Serna. 2012. *Crisis Financieras Sistémicas en Colombia y Contraste con el Escenario Actual*. Documento de Investigación. Colombia: Fogafin.
Arias, Mauricio, Juan Carlos Mendoza-Gutiérrez, and David Pérez-Reyna. 2010. Applying covar to measure systemic market risk: The colombian case. *Temas de Estabilidad Financiera* 47: 1–16.
Berger, Allen N., Jin Cai, Raluca A. Roman, and John Sedunov. 2021. Supervisory enforcement actions against banks and systemic risk. *Journal of Banking & Finance*. [CrossRef]
Bostandzic, Denefa, and Gregor N. F. Weiß. 2018. Why do some banks contribute more to global systemic risk? *Journal of Financial Intermediation* 35: 17–40. [CrossRef]
Boucher, Christophe M., Patrick S. Kouontchou, Bertrand B. Maillet, and O. Scaillet. 2013. The Co-CoVaR and Some Other Fair Systemic Risk Measures with Model Risk Corrections. Available online: www.systemic-riskhub.org/papers/selectedworks/BoucherKouontchouMailletScaillet_2013_fullpaper.pdf (accessed on 27 January 2020).
Brownlees, Christian, and Robert F. Engle. 2016. SRISK: A conditional capital shortfall measure of systemic risk. *The Review of Financial Studies* 30: 48–79. [CrossRef]
Cabrera-Rodríguez, Wilmar Alexander, Jorge Luis Hurtado-Guarín, Miguel Morales, and Juan Sebastián Rojas-Bohórquez. 2014. *A Composite Indicator of Systemic Stress (CISS) for Colombia*. Borradores de Economía; No. 826. Colombia: Banco de República.
Cáceres, José Antonio. 2009. Colombia ante la crísis financiera global. *Revista Escuela de Administración de Negocios* 65: 5–30. [CrossRef]
Castelao, Silvina, Sofía Palmigiani, and Patricia Lampes. 2013. *Riesgo Sistémico: Una Aproximación Para el Sistema Bancario Uruguayo*. No. 2013003. Montevideo: BCU.
Coleman, Thomas F., Alex LaPlante, and Alexey Rubtsov. 2018. Analysis of the SRISK measure and its application to the Canadian banking and insurance industries. *Annals of Finance* 14: 547–70. [CrossRef]
Davydov, Denis, Sami Vähämaa, and Sara Yasar. 2021. Bank liquidity creation and systemic risk. *Journal of Banking & Finance* 123: 106031.
Drakos, Anastassios A., and Georgios P. Kouretas. 2015. Bank ownership, financial segments and the measurement of systemic risk: An application of CoVaR. *International Review of Economics & Finance* 40: 127–40.
Duan, Yuejiao, Sadok El Ghoul, Omrane Guedhami, Haoran Li, and Xinming Li. 2021. Bank systemic risk around COVID-19: A cross-country analysis. *Journal of Banking & Finance* 133: 106299.
Engle, Robert. 2002. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics* 20: 339–50.
Estrada, Dairo Ayiber, and Daniel Esteban Osorio-Rodríguez. 2006. A market risk approach to liquidity risk and financial contagion. *Revista Ensayos Sobre Política Económica* 24: 242–71. [CrossRef]
Foggitt, Gregory M., Andre Heymans, Gary W. van Vuuren, and Anmar Pretorius. 2017. Measuring the systemic risk in the South African banking sector. *South African Journal of Economic and Management Sciences* 20: 1–9. [CrossRef]
Furfine, Craig H. 2003. Interbank exposures: Quantifying the risk of contagion. *Journal of Money, Credit and Banking* 35: 111–28. [CrossRef]
Grilli, Ruggero, Gabriele Tedeschi, and Mauro Gallegati. 2015. Markets connectivity and financial contagion. *Journal of Economic Interaction and Coordination* 10: 287–304. [CrossRef]
Group of Ten. 2001. Report on Consolidation in the Financial Sector. Available online: https://www.imf.org/external/np/g10/2001/01/Eng/index.htm (accessed on 27 January 2020).
Guerra, Solange Maria, Thiago Christiano Silva, Benjamin Miranda Tabak, Rodrigo Andrés de Souza Penaloza, and Rodrigo César de Castro Miranda. 2016. Systemic risk measures. *Physica A: Statistical Mechanics and its Applications* 442: 329–42. [CrossRef]
International Monetary Fund. 2010. Annual Report 2010 Supporting a Balanced Global Recovery. Available online: https://www.imf.org/~{}/media/Websites/IMF/imported-flagship-issues/external/pubs/ft/ar/2010/eng/pdf/_ar10engpdf.ashx (accessed on 1 September 2021).
Kleinow, Jacob, Fernando Moreira, Sascha Strobl, and Sami Vähämaa. 2017. Measuring systemic risk: A comparison of alternative market-based approaches. *Finance Research Letters* 21: 40–46. [CrossRef]
Koenker, Roger, and Gilbert Bassett Jr. 1978. Regression quantiles. *Econometrica: Journal of the Econometric Society* 1: 33–50. [CrossRef]
Laeven, Luc, Lev Ratnovski, and Hui Tong. 2016. Bank size, capital, and systemic risk: Some international evidence. *Journal of Banking & Finance* 69: S25–S34.
Laverde, Mariana, and Javier Gutiérrez-Rueda. 2012. *¿Cómo Caracterizar Entidades Sistémicas? Medidas de Impacto Sistémico Para Colombia*. Temas de Estabilidad Financiera, No. 65. Colombia: Banco de República.
León, Carlos, and Clara Machado. 2011. *Designing an Expert Knowledge-Based Systemic Importance Index for Financial Institutions*. Borradores de Economía, No. 669. Colombia: Banco de República.
León, Carlos, Clara Machado, Miguel Sarmiento, Freddy Cepeda, Orlando Chipatecua, and Jorge Cely. 2011. Riesgo sistémico y estabilidad del sistema de pagos de alto valor en Colombia: Análisis bajo topología de redes y simulación de pagos. *Revista Ensayos Sobre Política Económica (ESPE)* 29: 109–75. [CrossRef]
Lin, Edward MH, Edward W. Sun, and Min-Teh Yu. 2018. Systemic risk, financial markets, and performance of financial institutions. *Annals of Operations Research* 262: 579–603. [CrossRef]
López-Espinosa, Germán, Antonio Moreno, Antonio Rubia, and Laura Valderrama. 2012. Short-term wholesale funding and systemic risk: A global CoVaR approach. *Journal of Banking & Finance* 36: 3150–62.

Luciano, Elisa, and Clas Wihlborg. 2018. Financial synergies and systemic risk in the organization of bank affiliates. *Journal of Banking & Finance* 88: 208–24.
Meuleman, Elien, and Rudi Vander Vennet. 2020. Macroprudential policy and bank systemic risk. *Journal of Financial Stability* 47: 100724. [CrossRef]
Nelson, Daniel B. 1991. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica: Journal of the Econometric Society* 59: 347–70. [CrossRef]
Reig, Paula. 2017. Diez Años de Crisis en 24 Acontecimientos Clave. Available online: https://www.bbva.com/es/diez-anos-crisis-24-acontecimientos-clave/ (accessed on 23 March 2020).
Silva, Walmir, Herbert Kimura, and Vinicius Amorim Sobreiro. 2017. An analysis of the literature on systemic financial risk: A survey. *Journal of Financial Stability* 28: 91–114. [CrossRef]
Silva-Buston, Consuelo. 2019. Systemic risk and competition revisited. *Journal of Banking & Finance* 101: 188–205.
Tobias, Adrian, and Markus K. Brunnermeier. 2011. *CoVaR*. No 17454, NBER Working Papers. Cambridge: National Bureau of Economic Research, Inc., Available online: https://econpapers.repec.org/repec:nbr:nberwo:17454 (accessed on 23 April 2020).
Tobias, Adrian, and Markus K. Brunnermeier. 2016. CoVaR. *The American Economic Review* 106: 1705.
V-Lab. n.d. Dynamic MES. Available online: https://vlab.stern.nyu.edu/docs/srisk/MES (accessed on 24 April 2020).
Vogl, Christian. 2015. *Systemic Risk Measurement in the Eurozone—A Multivariate GARCH Estimation of CoVaR*. Lund: Lund University.
Zedda, Stefano, and Giuseppina Cannas. 2020. Analysis of banks' systemic risk contribution and contagion determinants through the leave-one-out approach. *Journal of Banking & Finance* 112: 105160.

risks

MDPI

Article

New Definition of Default—Recalibration of Credit Risk Models Using Bayesian Approach

Aneta Ptak-Chmielewska [1,*] and Paweł Kopciuszewski [2]

1 Institute of Statistics and Demography, Warsaw School of Economics, 02-554 Warsaw, Poland
2 Faculty of Art, Technique and Communication, Vistula University of Warsaw, 02-787 Warsaw, Poland; p.kopciuszewski@vistula.edu.pl or pawel.kopciuszewski@ing.com
* Correspondence: aptak@sgh.waw.pl or aneta.ptak-chmielewska@ing.com; Tel.: +48-501-71-23-71

Abstract: After the financial crisis, the European Banking Authority (EBA) has established tighter standards around the definition of default (Capital Requirements Regulation CRR Article 178, EBA/GL/2017/16) to increase the degree of comparability and consistency in credit risk measurement and capital frameworks across banks and financial institutions. Requirements of the new definition of default (DoD) concern how banks recognize credit defaults for prudential purposes and include quantitative impact analysis and new rules of materiality. In this approach, the number and timing of defaults affect the validity of currently used risk models and processes. The recommendation presented in this paper is to address current gaps by considering a Bayesian approach for PD recalibration based on insights derived from both simulated and empirical data (e.g., a priori and a posteriori distributions). A Bayesian approach was used in two steps: to calculate the Long Run Average (LRA) on both simulated and empirical data and for the final model calibration to the posterior LRA. The Bayesian approach result for the PD LRA was slightly lower than the one calculated based on classical logistic regression. It also decreased for the historically observed LRA that included the most recent empirical data. The Bayesian methodology was used to make the LRA more objective, but it also helps to better align the LRA not only with the empirical data but also with the most recent ones.

Keywords: new definition of default; credit risk models; Bayesian approach

Citation: Ptak-Chmielewska, Aneta, and Paweł Kopciuszewski. 2022. New Definition of Default—Recalibration of Credit Risk Models Using Bayesian Approach. *Risks* 10: 16. https://doi.org/10.3390/risks10010016

Academic Editors: Krzysztof Jajuga and Józef Dziechciarz

Received: 18 November 2021
Accepted: 6 January 2022
Published: 9 January 2022

1. Introduction

Following the financial crisis, EBA has established tighter standards around the definition of default (Capital Requirements Regulation—CRR Article 178, EBA/GL/2017/16) (EBA 2017) to achieve a higher comparability and consistency in models used for credit risk measurement and procedures and capital frameworks across banks and financial institutions. These requirements were supposed to be implemented by the end of 2020.

The initial deadline for the implementation (for all banks using IRB approach) was the 1st of January 2021. These deadlines were under discussion with European Central Bank (ECB) by many banks, and now they are under further revision by ECB due to COVID—19 circumstances (EBA 2016b).

Banks were allowed to choose either a one-step, or a two-step approach:

- One-step approach—the introduction of the new definition of default (DoD) and recalibration of all relevant models in one step;
- Two-step approach—first the introduction of the new DoD and then recalibration of relevant models.

The two-step approach was introduced because ECB realized that one-step approach would most likely automatically trigger a material change of all models within a bank.

The new DoD concern how banks recognize credit defaults for prudential purposes and also include quantitative impact analysis and new rules of materiality. In this approach, the number and timing of defaults determine whether existing models and processes are valid. Banks need to update their risk management practices and support pricing and accounting decisions related not only to the expected credit loss methodology (according to International Financial Reporting Standards—IFRS 9) (EU 2016) but also those related to capital requirements models (according to Internal Ratings Based Approach -IRB and Internal Capital Adequacy Assessment Process—ICAAP) (ICAAP 2018).

These extensive and very detailed guidelines often challenge information technology (IT) infrastructure, processes, data engines and data analytics, commonly used model risk platforms, model implementation and execution and automated solutions (BCBS 2013). In addition, these new standards have a significant impact on risk governance and management, frameworks and methodologies, data quality process assessments and reviews, as well as model recalibration needs and internal management policies and approval processes (Basel 2010, 2014).

In recent research on the credit risk parameters modeling, different approaches incorporate small samples problem incorporation (Zięba 2017), unbalanced samples or unresolved cases for LGD. Different techniques are proposed starting with traditional logistic or linear regression approach, two-stage modeling, ensemble models (Papouskova and Hajek 2019), and new machine learning methods or non-parametric approach (Peláez Suárez et al. 2021). In the literature of the PD modeling mostly a frequentist approach is applied (Bellotti and Crook 2009; Crone and Finlay 2012; Lessmann et al. 2015; Wang et al. 2020), while only limited use of a Bayesian approach is present (Bijak and Matuszyk 2017; Bijak and Thomas 2015). Among the papers using a Bayesian approach, many use non-informative priors on the assumption of the likelihood normally asymptotically distributed with large data (Bijak and Thomas 2015). Some papers emphasize the importance of expert information and use that to gain informative priors (Jacobs and Kiefer 2010; Kiefer 2009) or use coefficient estimates of past data as priors for current data (Bijak and Matuszyk 2017). However, we find that effective use of informative priors is still a gap to be covered.

The main contribution of this paper is to propose an approach to complement the scarce observational data post redefinition of DoD with simulated data. The Bayesian method was applied to anticipate this simulated and empirical data for a basic credit risk measure: probability of default (PD) re-calibration. Such example of Bayesian approach utilization in modeling and calibration of credit risk parameters is promising for trying to incorporate such approach in calibration of PD, using parameters estimated on empirical data for new DoD as prior information. For other parameters such as loss given default (LGD) and exposure of default (EAD) only a theoretical approach was proposed.

The paper is organized as follows: Section 1 provides an overview of the main implications of regulatory requirements impacting banks from a risk management perspective, with a focus on credit risk models. In this section a literature review focused on Bayesian approach is presented. Section 2 describes current challenges from a modeling perspective and discusses a methodological proposal to address these gaps. Section 3 provides an empirical example of application on real data for retail customers. Final section contains concluding remarks and suggestions for future research.

2. Implications of the New Definition of Default—Literature Review

This section provides a summary of the main changes implied by the new DoD and its impacts for credit risk modeling with specific emphasis for the institution under study in this article (one of European commercial banks). The literature review is limited to regulatory background and to Bayesian approach in research as the main focus of the paper.

The regulation contains specific requirements addressing specific event identification and threshold calculations. As a high-level overview, the main directions of the introduced changes can be split into specific focus areas: Days Past Due (DpD) calculation, Unlikeliness to Pay Criteria (UTP), Return to Non-Default Status (probation period), and Other Significant Changes (EBA/GL/2017/16, EBA/RTS/2016/06) (EBA 2014, 2016a, 2016b, 2017, 2020).

Materiality thresholds for days past due (DpD) includes not only absolute but also relative materiality thresholds for counting DpD until the event of default at 90 DpD (for retail exposures: 1% relative and 100 euros absolute, for non-retail exposures: 1% relative and 500 euros absolute).

The amounts past due represent the sum of all amounts past due, including all fees, interests, and principal. For the relative threshold calculation, this amount is divided by the total on-balance and off-balance exposure. If the principal is not repaid when an interest only loan expires, counting DpD starts from that date despite the fact that the obligor continues to pay interest.

Unlikeness to pay criteria (UTPs) will be recognized if the credit obligation gets a non-accrued status according to the accounting framework. Specific credit risk adjustments (SCRA) can be as follows: (i) sale of client's credit obligation recognized as defaulted if economic loss exceeds 5%, (ii) distressed restructuring if the net present value of the obligation decreases by more than 1% the obligation is considered defaulted, (iii) bankruptcy, (iv) additional indications of UTP (including fraud, significant increase in obligor leverage, individual voluntary arrangements, significant delays in payment to other creditors, impaired credit history indicators, expecting worst status).

A minimum probation period of three months is required for all defaults. The exemption stands for distressed restructurings where applies a 1-year minimum probation period. It is required to monitor the effectiveness of the cure policy on a regular basis, including also impact on cure rates and impact on multiple defaults.

Indeed, to implement the new DoD it is important to consider the holistic view of all processes impacted. In this respect, the key aspect to consider is a robust control framework from a risk management perspective. This can be disentangled into aspects related to external data, application of the definition of default from a broader banking perspective and specific features linked to retail exposures.

The implemented new definition of default (DoD), with above mentioned changes comparing to previous DoD, had a significant impact on existing rating systems and credit risk models. Along with days past due (DpD) calculations, changes to relative and absolute thresholds and default triggers were proposed. The implemented changes imposed changes to A-IRB models, their discriminatory power and calibration accuracy. As a consequence, some changes to IFRS 9 models will be required.

All A-IRB models with material change must be recalibrated and redeveloped (ECB 2018). This exercise will require recalibration and/or re-development and validation of all existing rating systems. This will be executed in the following steps:

- Data sources for modeling acquisition;
- Simulation of data according to the new definition of default;
- Back-test of all A-IRB models: PD, LGD, EAD;
- Recalibration of all models that showed material change during back-test;
- Recalibration of all IFRS 9 models as a result of A-IRB models recalibration and redevelopment;
- New Margin of Conservatism (MoC) calculation for existing models and the new DoD;
- Assessment of Risk Weighted Assets (RWA) impact of this change;
- Additional validation of rating systems.

As a consequence of PD, LGD and EAD models re-calibration all IFRS 9 models based on IRB parameters should be re-calibrated. Lifetime adjustment for parameters PD and LGD could change due to different starting points. Finally, all changes require independent validation. In the case of multiple models and portfolios it will be a time and resources critical process in bigger financial institutions as the one under study in this article.

The literature on new DoD is rather limited. We focused on available examples of incorporation of a Bayesian approach in credit risk parameters modeling, mostly for PD and LGD parameters. A methodology for credit default estimates applying Bayesian mixture models was presented in Simonian (2011). The author proposed robust models taking parameter uncertainty into account to generate a new model. In the context of credit risk parameters modeling, robust models are beneficial to practitioners when estimating default probabilities.

Much less frequent are examples of Loss Given Default (LGD) estimations using a Bayesian approach. One of the examples is for LGD for unsecured retail loans as often found difficult to model. The typical is two-step approach, two separate regression models are estimated independently. This approach can be potentially problematic because it must be combined to make the final predictions about LGD. LGD can be than modeled using Bayesian methods (Bijak and Thomas 2015). In this approach only a single hierarchical model can be built instead of two separate models. It makes this a more appropriate approach. Authors used Bayesian methods, and alternatively the frequentist approach, and applied to the data on personal loans provided by a large UK bank. The posterior estimates of means of parameters that have been calculated using the Bayesian approach were very similar to the ones calculated in frequentist approach. An advantage of the Bayesian model was an individual predictive distribution of LGD for each loan. According to regulatory requirements applications of such distributions include also the downturn LGD calculations and the so called stressed LGD calculations.

The lack of data is typical for a low default portfolio (LDP). Probability of default (PD) calibration in such situation is limited to add conservative add-ons that should cover the gap of information due to scarce default event data. As described in the article (Surzhko 2017), a PD calibration framework proposes Bayesian inference. The main idea proposed is to calibrate prior using a "closest" available portfolio with reliable default statistics. Author proposed the form of the prior, criteria for a "closest" portfolio selection and application of the approach to real life data and artificial portfolios. The advantage of the approach proposed in the article is avoidance of the subjective level of conservatism assumption. The author also proposed an approach that could be used for stress-testing purposes.

Bayesian informative prior selection method is also proposed for including additional information to credit risk modeling, specifically for PD, and to improve model performance (Wang et al. 2018). Authors used logistic regression to model the probability of default of mortgage loans; they applied the Bayesian approach with various priors and the frequentist approach for comparison. The authors proposed for the Bayesian informative prior selection method the coefficients in the PD model as time series variables. They built ARIMA models to prognose the coefficient values in future time periods and used these prognoses as Bayesian informative priors. According to their results the Bayesian models using this prior selection method outperformed in accuracy both approaches: frequentist models and Bayesian models with other priors.

Based on U.S. mortgage loan data, the probability of default at account level using discrete time hazard analysis was analyzed (Wang et al. 2020). Authors employed the frequentist and Bayesian methods in estimation of the parameter, and also the default rate (DR) stress testing. By applying the Bayesian parameter posterior distribution to simulating the DR distribution, they reduced the estimation risk coming from usage of point estimates in stress testing. As estimation risk was addressed in this approach, they obtained more prudential forecasts of credit losses. The simulated DR distribution obtained using the Bayesian approach with the parameter posterior distribution had a standard deviation more than 10 times as large as the standard deviation from a frequentist approach with parameter mean estimates. The same observation was found for VaR (Value at Risk) estimates.

Such examples of Bayesian approach utilization in modeling and calibration of credit risk parameters are promising for trying to incorporate such an approach in calibration of PD using parameters estimated on empirical data for new DoD as prior information. From this perspective this approach is unique in the research.

3. Proposed Recalibration and Re-Development Methods for Credit Risk Parameters

The introduction of the new DoD entails backtesting of the old models built on the earlier version of DoD and taking appropriate remediation actions. The easiest way to fit a model to the current definition of default is to recalibrate the old model using new simulated and new empirical data. If the recalibration fails, a new model must be built. Many different recalibration methods can be considered—from the easiest based on any scaling approach to more sophisticated ones. Greater attention should be paid to understanding data before selecting an appropriate approach for recalibration or redevelopment.

Data used in the process of recalibration and re-development of models can be simulated, empirical and mix type (see Table 1). Figure 1 presents how these types of data can be used. Before implementation of a new DoD flag in the system only simulated data on new DoD are available. After a sufficient DpD period post implementation there is new DoD empirical data available in the system. Both types of data can be used in re-calibration and re-development. Re-development on real empirical data will be only possible after two years period post implementation. Re-calibration must be undertaken on mixed data simulated and empirical. The only limitation is for LGD re-development because collection data are available only empirically; there is no simulation data for collections.

Steps of recalibration/re-development of models for credit risk parameters:

- Simulation of data concerning the new definition of default for different portfolios.
- Construction of Analytical Based Tables (ABTs) for model's re-calibration process.
- Construction of codes for recalibration process automatization—model development.
- Construction of codes for validation process automatization—model validation.
- Re-calibration of LGD/CR, EAD/CCF parameters.
    - Scaling factors.
    - Regression models adjusting the old risk parameter to the new one.
- Re-development of models in case of negative back-test after re-calibration of models.
    - Models including a-priori distributions (Bayesian approach).
    - Parametric models.
    - Non-parametric models.
- Validation of the model's re-calibration/re-development process.

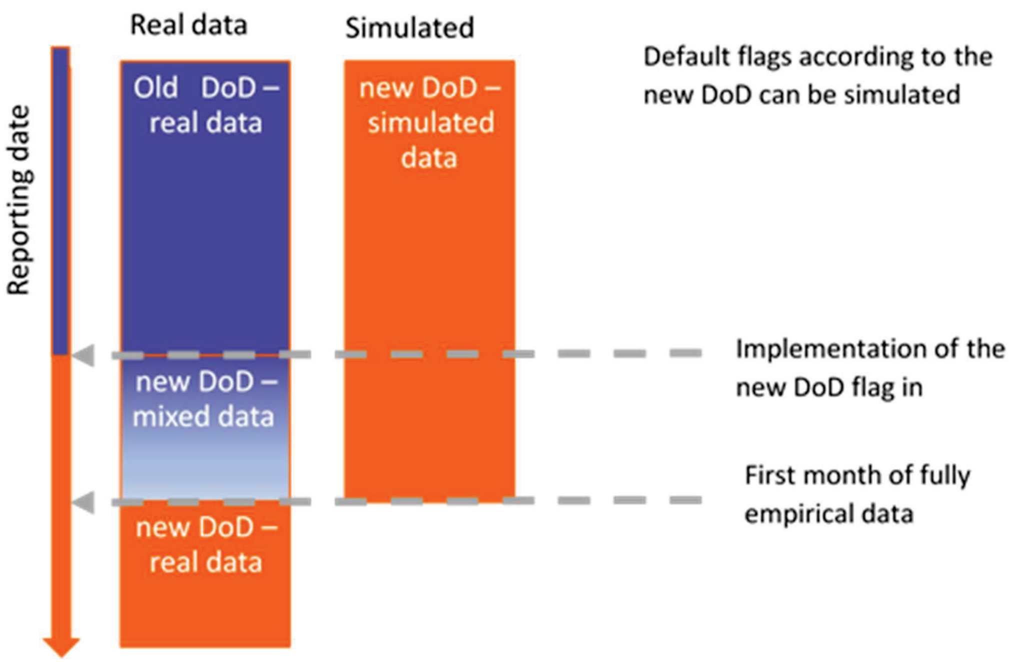


Figure 1. Scheme of data used in recalibration. Source: own elaboration.

Table 1. Types of data utilized.

Model	Re-calibration of IRB models (end of year 2021)	Re-development of models using new data (years 2022/2023)
PD	Default data—simulated Additional data—empirical	Simulated data (option of immediate redevelopment) Empirical data (redevelopment possible after 2 years of data collection)
LGD	Default data—simulated Collection data—real empirical data after default date	Empirical/mixed data, No simulated data

3.1. Probability of Default PD—Model Recalibration

There are two approaches for PD model recalibration—one based on scaling factors and the alternative one based on model estimation.

The first approach is based on scaling factors as conservative add-on. We have two or more different scaling factors proposed in this approach. A basic scaling factor assumes that new default rate can be re-scaled proportionally for particular risk grades, segments by proportional change of *PD* using the following formula, assuming Default Rate (*DR*) as empirical realization of *PD*:

$$PD_{new} = PD_{old} \cdot f, \quad f = \frac{DR_{new}}{DR_{old}}, \tag{1}$$

where: *f*—scaling factor, *PD*—probability of default, *DR*—default rate.

When scaled probability takes a value out of the accepted range, then another approach is needed. One of them can be a logit scaling factor given by the following formula:

$$PD_{new} = \frac{f \cdot PD_{old}}{1 + (f-1) \cdot PD_{old}}, \quad f = \frac{\frac{1-DR_{old}}{DR_{old}}}{\frac{1-DR_{new}}{DR_{new}}}, \tag{2}$$

Of course, other scaling factors are also possible. They can be, for example, based on relations between neighboring rating grades.

Quite a different approach for calibration is based on the estimation of a new model. There are, of course, some options such as logit regression or the Bayesian approach. As an alternative to regression the non-parametric model can be estimated. In that case we need non-parametric model based on PD curve on rating scale grades.

For the redevelopment approach a Bayesian approach could be considered that would require prior information on simulated or mixed data, and likelihood based on real data for the new DoD.

For built models it is not recommended to have calibration separate for rating grades, rather for the model level. In such situation the lack of PD monotonicity is observed, and additional algorithms are required to maintain monotonicity. To check the necessity of model re-calibration or re-development the backtest is performed. The consistency between model probability of default PD comparing to empirical default rate DR is backtested. As the first step, a backtest on simulated data is performed (see Figure 2) and if it is positive (not rejected), then MoC calculation is undertaken to adjust the new DoD. Only when it is positive and MoC is applied, a backtest on real data is performed and, if positive, the implementation phase is launched. If the backtest on real data is negative (rejected), recalibration on real data is performed. If the backtest on simulated data is negative, an additional step of recalibration on simulated data is performed.

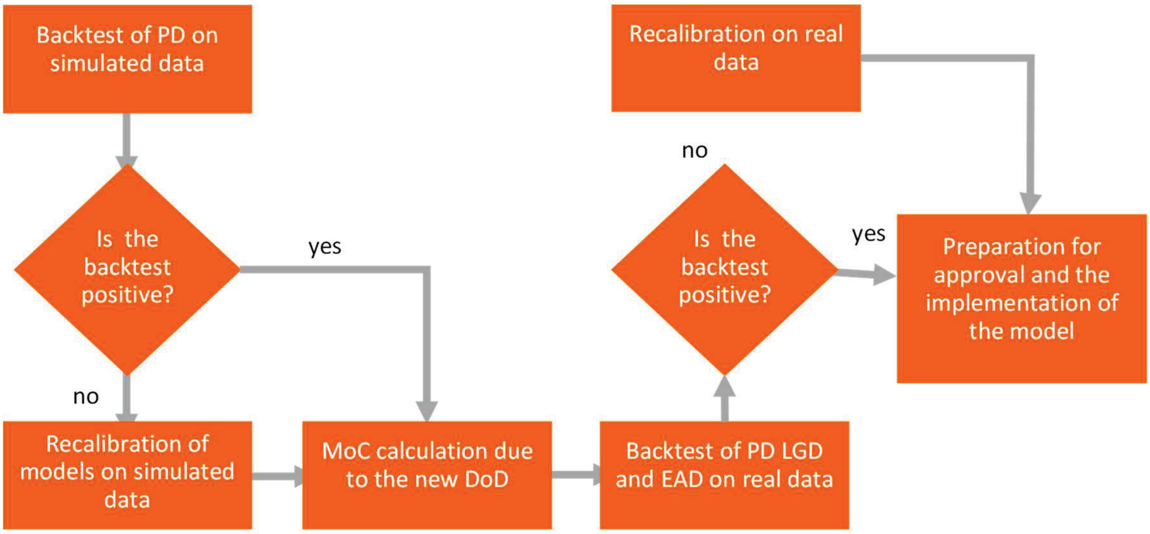


Figure 2. PD model recalibration scheme. Source: own elaboration.

3.2. Loss Given Default LGD—Model Recalibration

Backtesting, recalibration or redevelopment processes are different for LGD due to the limited size of population for modeling and more problematic use of historical simulated data. The LGD formula may contain many components that are estimated separately but the final formula depends on the specific debt collection process. The basic formula used for LGD is presented below.

$$\mathrm{LGD} = (1 - \mathrm{CR}) \cdot (1 - \mathrm{SRR}) \cdot (1 - \mathrm{URR}) + \mathrm{CR} \cdot \mathrm{L}_{\mathrm{cure}}, \tag{3}$$

where:

CR—cure rate,
SRR—secured recovery rate,
URR—unsecured recovery rate in relation to the reduced EAD with secured recoveries,
$\mathrm{L}_{\mathrm{cure}}$—the economic loss associated with cured cases expressed as a percentage of EAD.

Calibration of basic components of this LGD formula is based on:

1. For the cure rate parameter (CR):
   - ○ scaling factor (including no-loss),
   - ○ logistic regression model,
   - ○ full recalibration using the Bayesian approach,
   - ○ recalibration using survival methods (with censored observations).
2. For the secured and unsecured recovery rate (SRR, URR):
   - ○ calibration using the scaling factor,
   - ○ linear or non-linear regression model,
   - ○ full calibration using regression models including simultaneous calibration of both parameters.

The recalibration of LGD depends on the type of the new period of data available—if it is a downturn period or not (see Figure 3).

If the period cannot be qualified as downturn, a backtest approach with a scaling factor is applied and if it is positive (not rejected), it is the end of recalibration. If it is not positive (rejected), a full recalibration based on the Bayesian approach is applied. More sophisticated, hierarchical Bayesian models for LGD are applied instead of two combined models. The first model is used to separate all positive values from zeroes and the second model is than used to estimate only positive values (Bijak and Thomas 2015; Papouskova and Hajek 2019). If the backtest is not positive, the conservative MoC is needed. If the period is qualified as downturn, a backtest is applied on real data for LGD downturn. If it is positive, the recalibration ends. If the backtest is not positive (rejected), a conservatism option for LGD downturn is applied.

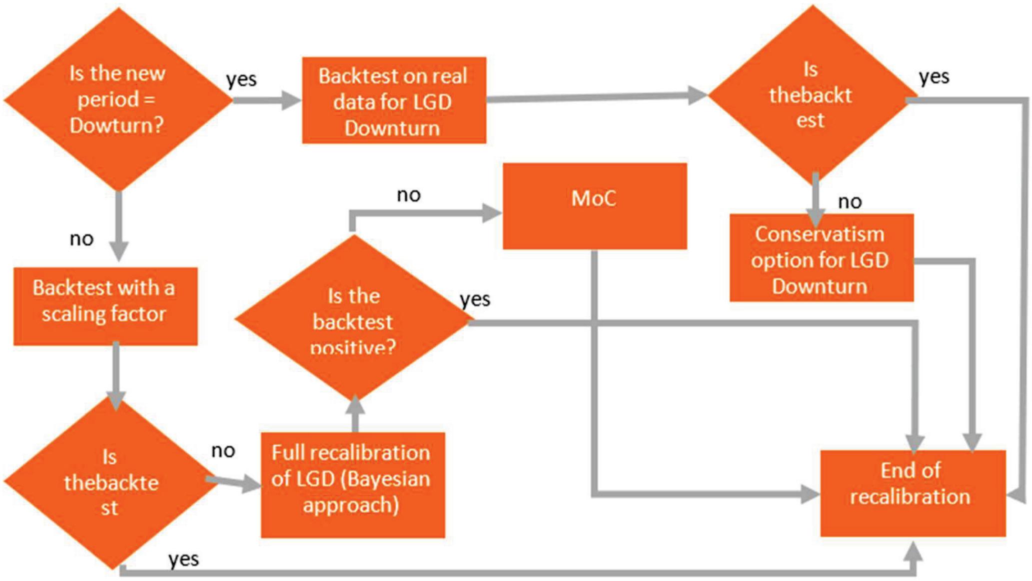


Figure 3. LGD model recalibration scheme. Source: own elaboration.

Understanding the data is crucial in the case of LGD recalibration. As shown in Figure 4, the data period for the new DoD may be different than for the old DoD in terms of its length and start date. The worst-case scenario is when the new default date is several months after the old default date. This is because all recoveries and other LGD components have been observed from the old default date, but the modeling or recalibration process requires them to be used right after the new default date. Since default entry and exit dates can be simulated on historical data this does not apply to the recovery process. The challenge for a modeler is to determine which data should be used—the old observer recoveries that are inconsistent with the new default date or only those that are observed right after the new default date. The answer is not unambiguous but in the case of significant differences between the old and new recoveries the former should be taken into account. The choice of using observed recoveries beyond the old default date may be justified as the recovery process does not change in most cases after the introduction of the new DoD.

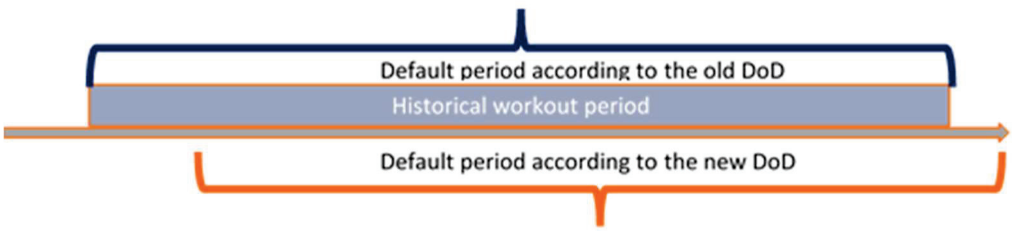


Figure 4. LGD recalibration—differences between the old and new default period. Source: own elaboration.

3.3. Exposure at Default EAD—Model Recalibration

The EAD recalibration is based on constant. We have four possible scenarios as indicated below:

Scenario	Parameters Description
1	Exposure below limit
2	Exposure equal limit or above (>0)
3	Exposure positive, limit 0
4	Exposure and limit equal 0

For the above scenarios adjustment coefficients are defined. Those coefficients equal the relation between the previous and new parameter such as:

- credit conversion factor (CCF) in the first case;
- the ratio of limit in the observation date for a facility to all limits for all facilities in a sample in the second case;
- the ratio of exposure at observation date for a facility to all exposures for all facilities in a sample in the third case;
- constant value estimated on the sample in the fourth case.

Those factors usually equal average factors for a portfolio, as the following formula indicates in the first case:

$$adj_{factor} = \frac{CCF_{new}}{CCF_{old}}, \tag{4}$$

where CCF_{new}, CCF_{old} are conversion factors based on the new and old definition of default.

The process of EAD recalibration is shown in Figure 5. If scaling with constant gives a positive (not-rejected) backtest result, we end recalibration. If the result is negative (rejected), a full recalibration is applied. When the full recalibration still gives a negative backtest (rejected), then conservative MoC is applied.

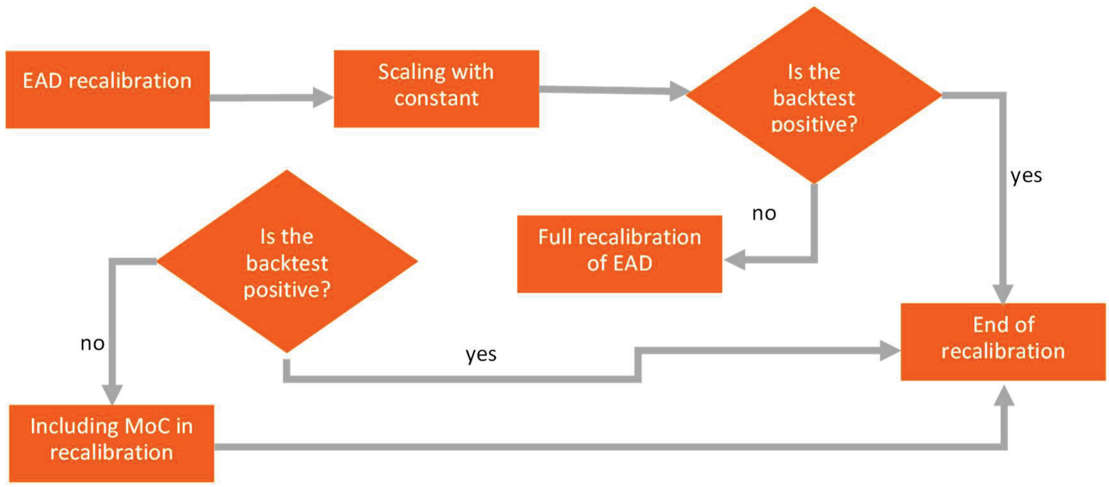


Figure 5. EAD model recalibration scheme. Source: own elaboration.

3.4. The Bayesian Approach in PD Recalibration

When the backtest fails on the recalibrated model, it is necessary to rebuild a new model. The redevelopment should take into account both simulated and empirical data. There is no need to use more sophisticated models in terms of advanced analytics, but rather to ensure clear data preparation and use models that take into account all information available when rebuilding the model. At the early stage of collecting data for new defaults, empirical data is not large enough to build a model on its own. A suitable methodology for both simulated and empirical data is the Bayesian approach, which can be used for different data sources, even if some data sets are small (Surzhko 2017; Zięba 2017). An additional advantage of Bayesian approach is the possibility of using expert knowledge not contained directly in the data (Simonian 2011). When the PD model is rebuilt, the simulated data is combined with the empirical data. From the purer statistical point of view the final

PD is based on posterior distribution which is the product of the prior information and the likelihood.

First some basic assumptions must be introduced:

- PD_{sim}—PD estimated on simulated data;
- PD_{emp}—PD estimated on empirical data, at start this population is much smaller than the simulated one;
- prior information comes from simulated data;
- likelihood comes from empirical data;
- posterior distribution is a combination of prior distribution and likelihood;
- final PD is based on posterior distribution characteristics;

In general, the Bayesian approach can be applied when simulated PD_{sim} does not pass the alignment on empirical data, which means that there is no match between simulated PD_{sim} and empirical PD_{emp}.

The estimated parameter is PD_{sim} on simulated data with additional information from empirical data:

$$\pi(PD_{sim}|PD_{emp}) \propto \pi(PD_{sim}) \cdot f(PD_{emp}|PD_{sim}), \tag{5}$$

where:

$\pi(PD_{sim}|PD_{emp})$—is the posterior distribution under condition of PD_{emp},
$\pi(PD_{sim})$—is the prior distribution,
$f(PD_{emp}|PD_{sim})$—likelihood—conditional distribution of PD_{emp} parameter under the condition of PD_{sim}.

All simulations can be performed using Markov Chain Monte Carlo (MCMC) methods. The main advantages of this approach are as follows (Robert 1994):

1. The Bayesian approach fits estimate parameters very well with the use of knowledge from different data sources, both internal and external. However, before applying the Bayesian technique the following problems should be solved. The choice of prior distribution for PD simulated. The prior should be based on portfolio characteristics and the uncertainty in data;
2. The choice of MCMC technique, such as Gibbs sampling, Hastings–Metropolis or other when explicit posterior cannot be found;
3. The possibility of including other prior information that is not derived from the data such as future macroeconomic conditions in some selected industries;
4. The possibility of calculating confidence intervals for PD and using a more conservative estimator.

4. PD Recalibration—Application on Real Data

Application on real data sample was based on selected portfolio of retail mortgages loans for one of European banks. Time frame for the sample was restricted to years 2008–2020. The development sample for the PD model includes around 15K observations, where 4K observations come from the period of simulated defaults and 11K from the period of empirical defaults. This sample was drawn from the population including around 132K observations by selecting all defaults and drawing non-default cases with probability less than 0.1. Hence, the development sample was reweighted. PD model was built on the reweighted population because the total population includes too many “good” customers for modeling purpose, that is the inclusion of too many observations could lead to numerical problems in estimation and is more time consuming. Another reason to weigh population in processes of model building is the better learning of the risk profile by increasing the share of “bad” customers. The default rate in the entire population before drawing observations for the modeling process was 1.75% and 1.52% for simulated and empirical cases, respectively. The stability of both the default rate and the arrears amount (maximum and the average arrears for the facility) is presented in the Table 2. The arrears amount means the total outstanding of the customer after default moment.

Table 2. Default rate and arrears amount.

Year	Count	Default Rate	Arrears Max [Thousand EUR]	Arrears Average [Thousand EUR]
2008	6939	0.72%	10.59	6.36
2009	8973	0.57%	15.93	3.32
2010	7502	0.72%	14.29	2.64
2011	6889	1.32%	19.29	4.82
2012	7860	2.01%	19.21	4.19
2013	8993	2.62%	26.60	7.72
2014	9774	2.34%	32.68	7.00
2015	10,725	2.45%	28.64	8.52
2016	12,545	2.18%	22.42	6.40
2017	16,832	1.73%	36.84	20.93
2018	3290	1.52%	19.81	6.21

Simulated defaults were selected within the period 2008–2018 but empirical defaults were more relevant after 2019. The sample was split into two independent subsamples for simulated and empirical defaults. While the model was built on the whole sample, including both the simulated and empirical defaults, the recalibration was mainly performed on simulated data with the additional use of empirical defaults according to the Bayesian methodology.

The process of recalibration includes the following steps:

1. Building the new PD model on the joint population for simulated and empirical defaults, i.e., a mixed population. The model is built with use of logistic regression;
2. Calculation of Long Run Average (LRA) on the simulated data. LRA is the average of default rates calculated within the given period of time;
3. Adjusting LRA through the Bayesian methodology, which combines both the simulated and empirical data. The role of empirical data is to adjust the LRA calculated on the simulated data;
4. Final recalibration of the PD parameter estimated in the 1st step at the facility level according to the posterior mean calculated at the step 3. The final step is also performed with Bayesian approach.

Summarizing the above algorithm, the Bayesian approach was used both to find the posterior estimators and final recalibration of the model but was applied to two different models.

The PD model was built on a population of around 15K observations and a default rate of 0.017424. The entire data set contained around 400 risk drivers that had been previously selected due to their business intuitiveness and high data quality. The model estimation procedure was based on logistic regression with a stepwise selection method assuming that the p-value for entering the variable into the model was 0.05 and the p-value for remaining in the data was 0.001. The final model was based on around 10 variables with the p-value for the test of significancy of the coefficient at the variable not higher than 0.004. The estimation procedure was performed in SAS. Ultimately, PD was explained by some transformations of the following variables:

- Absolute breach in the past;
- Relative breach in the past;
- Maximum DpD in the past;
- Amount of maximum arrears in the past;
- Total obligations;
- The age of the customer;
- Account balance.

The quality of the model measured with Area Under ROC Curve (AUC) is 0.9. The model was built on the reweighted population to increase the default rate. Both simulated and empirical data were included in the population, but the defaults were underrepresented by the reduction in the population of "good" customers. The abovementioned reasons require a recalibration process, but the latter require more sophisticated methods such as Bayesian approach.

The next step after building the model was the choice of the calibration methodology. The starting point in this process was the calculation of Long Run Average of DR (LRA) on the simulated data, which was 0.017499, while the standard deviation of its estimator was 0.131123. An attempt was made to include empirical data that turned out to be too small to estimate risk parameters and the Bayesian methodology was used. The Bayesian approach can be considered as an improvement of LRA using empirical data that was initially calculated on the simulated data. In general, the PD calculated on simulated data is prior information, but the PD calculated on empirical data provides real information that is not biased by the simulation process. Nevertheless, the empirical population is too small to recalibrate the model. According to the Bayesian methodology, the following assumptions (6) and (7) were made and incorporated into the MCMC SAS procedure:

$$PD_{sim} \sim N(E(LRA_{sim}), std(LRA_{sim})), \tag{6}$$

where: PD_{sim} is a random variable normally distributed around $E(LRA_{sim})$.

$E(LRA_{sim})$ and $std(LRA_{sim})$ are the expected value of LRA and standard deviation of LRA, respectively, calculated on the simulated data.

The above assumptions related to the distribution of PD_{sim} is directly derived from the historical data. The distribution of the variable PD_{sim} is considered as prior distribution. This is the case where prior distribution is more objective as it is based on historical data. We used many different prior distributions which are allowed by SAS MCMC procedure and the final posterior results were very similar which further proves robustness of the posterior estimators.

The likelihood which is the distribution of the PD_{emp} parameter on the empirical data is the binomial distribution with the number of trials that equals the number of observations n_{emp} in the empirical data and the probability of success equals the PD_{sim} calculated on the simulated data. Hence, the likelihood can be viewed as a conditional distribution of the PD parameter on empirical data given the PD_{sim} calculated on the simulated data is defined as follows:

$$PD_{emp} \middle| PD_{sim} \sim B\left(n_{emp},\ PD_{sim}\right), \tag{7}$$

where: PD_{emp} is a random variable binomially distributed (Joseph 2021) with parameters n_{emp} and PD_{sim}.

Based on the above two distributions of prior and likelihood, the characteristics of posterior density were calculated using the MCMC procedure and the following results for the posterior expected value and posterior standard deviation were obtained:

$$PD_{sim|emp} = E\left[PD_{sim}\middle|PD_{emp}\right] = 0.0155$$
$$s\left[PD_{sim}\middle|PD_{emp}\right] = 0.0021$$

The obtained posterior mean $PD_{sim|emp}$ of 0.0155 is the new calibration purpose.

The following table (see Table 3) summarizes all considered PD estimators obtained on the simulated data, empirical data, development data and the Bayesian calculation as a posterior mean.

Table 3. PD estimator results.

Simulated Data (Prior)	Development Data	Empirical Data	Posterior
0.017499	0.017424	0.015198	0.0155

The final recalibration of the PD parameter was based on the new recalibration purpose using the logistic Bayesian model. The default flag shows binary distribution with the expected value that equals the unknown calibrated PD (PD_{cal}) depending on the simulated PD (PD_{sim}) and two additional parameters a and b as follows:

$$\begin{gathered} D \sim Bin(f(PD_{sim}, a, b) \big| PD_{emp}) \\ PD_{cal} = f(PD_{sim}, a, b) = f(a \cdot \ln\left(\tfrac{PD_{sim}}{1-PD_{sim}}\right) + b) \end{gathered} \quad (8)$$

where f is the logistic function.

Parameters a and b are hyperparameters normally distributed with the following mean and standard deviation:

$$\begin{gathered} a \sim N(1, \sigma_a) \\ b \sim N\left(\ln\left(\tfrac{PD_{sim|emp}}{1-PD_{sim|emp}}\right) - \ln\left(\tfrac{DR_{mod}}{1-DR_{mod}}\right), \sigma_b\right) \end{gathered} \quad (9)$$

where DR_{mod} = 0.017424 is the default rate on the modeled data set.

The concept of the expected value for parameter b is that it should be positive when $PD_{sim|emp}$ is higher than DR_{mod} but otherwise negative bearing in mind that this is an additive part to the log odds. The formula for b parameter is the main point in this Bayesian model to include the information about the posterior mean to which to calibrate the model.

Both standard deviations σ_a and σ_b have standard deviation uniformly distributed over the interval (0,1). The upper value of the interval is based on previous experiments with data. An alternative approach to find estimators a and b is to use simple non-Bayesian regression logistic. The Bayesian approach is key in the previous step to find the posterior mean, but here it only serves as a consistent methodology and is not needed for finding the final recalibration formula. On the other hand, it can be easily extended by providing a greater external knowledge of the estimated parameters. Therefore, it can be taken as a pattern for the further calculations on other data, which is very flexible. Ultimately, posterior averages for a and b were calculated, respectively (see Table 4).

Table 4. Posterior parameters for the Bayesian approach.

Parameter	Mean	STD	95% HPD Interval	
a	0.88	0.0896	0.7229	1.0791
b	−5.4135	0.2399	−5.8676	−4.9243
sigma_a	0.4013	0.298	0.00168	0.9242
sigma_b	0.9423	0.0535	0.8326	1

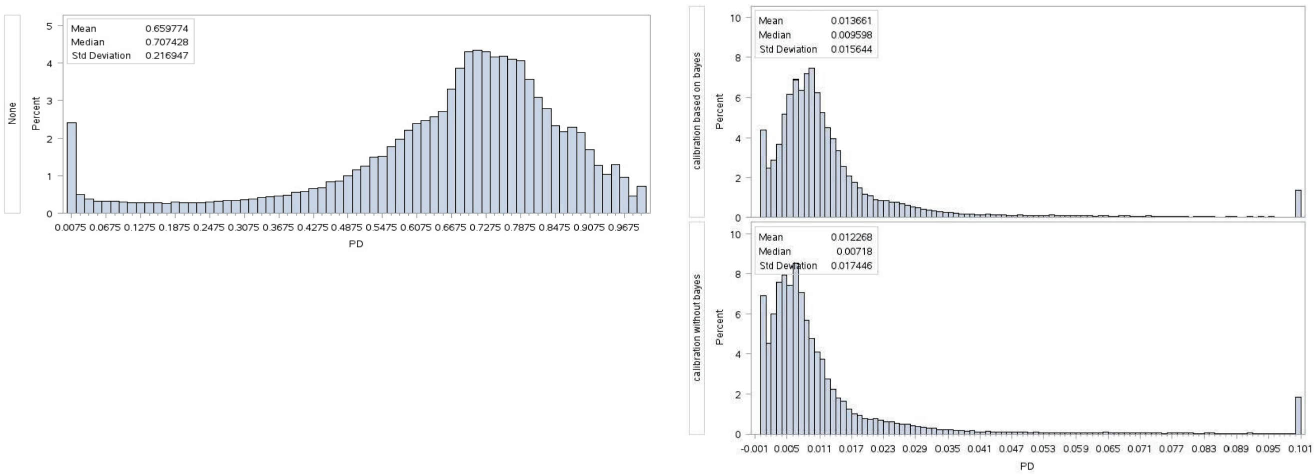
In order to obtain the final PD estimation formula, the logistic regression Equation (8) for the entire PD was applied. Additionally, the PD calibrated without the Bayesian method was calculated for comparison (see Table 5). It is worth mentioning that the simple logistic regression model based on the Equation (8) but without taking into account parameter distributions can be built with use of the following weights for all observations to adjust the mean default rate to posterior mean. In case of Bayesian approach weights were not needed as b parameter in the model plays the role of such adjustment.

$$w = \frac{PD_{sim|emp}}{1 - PD_{sim|emp}} \cdot \frac{1 - DR_{mod}}{DR_{mod}} \quad (10)$$

Table 5. Parameters calculated without the Bayesian approach.

Parameter	Mean	STD	95% HPD Interval	
a	1.1087	0.0192	1.0736	1.1498
b	−5.9081	0.0512	−6.0129	−5.8132

All PD's distributions for the raw PD obtained from the model built on the reweighted sample, PD calibrated without the Bayesian methodology and PD using the Bayesian approach are presented below in Figure 6a,b.



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

Figure 6. PD distribution: (**a**) Raw PD; (**b**) PD calibrated with and without the Bayesian approach. Source: own elaboration using SAS Enterprise Guide.

Summarizing the classic approach to building the logistic regression model as the final calibration formula is possible only with introducing weights for observations; the Bayesian approach does not require weights but an appropriate definition of distributions for regression parameters.

5. Conclusions and Discussion

Recent changes in economy caused by COVID pandemic (Batool et al. 2020) had an impact on sharing economy but also a significant impact on the banking sector. Recent changes in Industry 4.0 were also not neglectable for Banking 4.0 (Mehdiabadi et al. 2020). Significant regulatory changes imposed by regulatory authorities followed those changes.

The regulatory requirements related to the change in the definition of default have significant impacts from a banking perspective, with the most material one on credit risk modeling. From a methodological point of view, all provisioning IFRS9 and capital IRB regulatory models need recalibration. At the same time, empirical evidence that is obtained when the new definition is applied to real portfolios is still being collected. In this context, the paper provides an overview of the main implications deriving from the regulation in different risk management areas (e.g., data, default definition, implementation, identification) and discusses a methodological proposal for IRB modeling recalibration to address the current challenges. The idea is to leverage the Bayesian approach for PD recalibration by retrieving information from both simulated and empirical data (e.g., a priory and a posteriori distributions). As discussed in the methodological section, this mathematical approach seems to be a promising solution to building a robust framework in the current phase and addressing the gaps. In our plans for future research, we foresee an empirical study to test how the proposed methodology can perform in different credit risk modeling contexts.

Finally, we used the Bayesian approach in two steps: the first basic approach is to calculate the LRA on both simulated and empirical data. In addition, we used the same approach for the final model calibration to the posterior LRA. In summary, the Bayesian approach result for the PD LRA was slightly lower than the one calculated based on classical logistic regression (Figure 6). It also decreased for the historically observed LRA (Table 3) that included the most recent empirical data. The Bayesian methodology was used to make the LRA more objective, but it also helps to better align the LRA not only with the empirical data but also with the most recent ones. It allows us to consider the LRA as a random variable, where its variance tells us more about the significance of the point estimation. The greater the variance, the more empirical data need to be included in the calculation of the end value. When comparing the standard deviation of the LRA calculated on the simulated data, which is 0.131123 with the mean of 0.017499, there is still room to improve this estimator with less volatile data, such as empirical with the mean LRA of 0.0155 and much lower standard deviation of 0.00214. To sum up, it is a unique approach to statistical modeling that can combine different information, even expertise, not covered by historical data. Moreover, it can be applied to both LGD and EAD, but PD is preferred as the starting point. Promising results for PD were also obtained by (Wang et al. 2018) with ARIMA model results as prior for Bayesian approach confirming outperformance comparing to frequentist approach.

From a practical point of view, using the Bayesian approach can significantly decrease the capital requirements, thus making savings for organization from managerial perspective. Close to the empirical values, the calculations of capital adequacy and provisions are as more precise, keeping of course regulatory requirements fulfilled.

A weakness of the proposed approach, however mitigated, is due to normal distribution assumption. Some ideas for future research rely on further investigation of the Bayesian approach, but perhaps for small samples or other parameters as well. This approach seems also promising for IFRS 9 lifetime parameter estimation.

Author Contributions: Conceptualization, A.P.-C. and P.K.; methodology, A.P.-C. and P.K.; software, P.K.; validation, A.P.-C.; formal analysis, P.K.; investigation, A.P.-C.; resources, A.P.-C.; data curation, P.K.; writing—original draft preparation, A.P.-C. and P.K.; writing—review and editing, A.P.-C.; visualization, P.K.; supervision, A.P.-C.; project administration, A.P.-C.; funding acquisition, A.P.-C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Restrictions apply to the availability of these data. Data was obtained from 3rd Party and are not available. MDPI Research Data Policies.

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

References

Basel. 2010. *Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems*. Basel: Basel Committee on banking Supervision, Bank for International Settlements, December. Available online: https://www.bis.org/publ/bcbs189.pdf (accessed on 12 November 2021).

Basel. 2014. *Basel III: The Net Stable Funding Ratio*. Basel: Basel Committee on banking Supervision, Bank for International Settlements, October. Available online: https://www.bis.org/bcbs/publ/d295.htm (accessed on 12 November 2021).

Batool, Maryam, Huma Ghulam, Muhammad Azmat Hayat, Muhammad Zahid Naeem, Abdullah Ejaz, Zulfiqar Ali Imran, Cristi Spulbar, Ramona Birau Icon, and Tiberiu Horațiu Gorun. 2020. How COVID-19 has shaken the sharing economy? An analysis using Google trends data. *Economic Research-Ekonomska Istraživanja* 34: 2374–86. [CrossRef]

BCBS. 2013. *BCBS 239—Principles for Effective Risk Data Aggregation and Risk Reporting*. Basel: Basel Committee on Banking Supervision, Bank for International Settlements, January. Available online: https://www.bis.org/publ/bcbs239.pdf (accessed on 12 November 2021).

Bellotti, Tony, and Jonathan Crook. 2009. Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications* 36: 3302–8. [CrossRef]

Bijak, Katarzyna, and Anna Matuszyk. 2017. Bayesian models of car lease frauds. Paper presented at the Credit Scoring and Credit Control XV, Edinburgh, UK, August 30–September 1.

Bijak, Katarzyna, and Lyn Thomas. 2015. Modelling LGD for unsecured retail loans using Bayesian methods. *Journal of the Operational Research Society* 66: 342–52. [CrossRef]

Crone, Sven F., and Steven Finlay. 2012. Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting* 28: 224–38. [CrossRef]

EBA. 2014. *EBA Report Results from the 2014 Low Default Portfolio (LDP) Exercise*. Paris: EBA European Banking Authority. Available online: https://www.eba.europa.eu/sites/default/documents/files/documents/10180/950548/b3adc4f7-653d-408d-b950-da2e4229e294/EBA%20results%20from%20the%202014%20Low%20Default%20portfolio%20%28LDP%29%20exercise.pdf?retry=1 (accessed on 12 November 2021).

EBA. 2016a. *Draft Regulatory Technical Standards on the Materiality Threshold for Credit Obligations Past due under Article 178 of Regulation (EU) No 575/2013*. EBA/RTS/2016/06. 28/09/2016. Available online: https://www.eba.europa.eu/regulation-and-policy/credit-risk/regulatory-technical-standards-on-materiality-threshold-of-credit-obligation-past-due (accessed on 12 November 2021).

EBA. 2016b. *Guidelines on the Application of the Definition of Default under Article 178 of Regulation (EU) No 575/2013*. EBA/GL/2016/07. Available online: https://www.eba.europa.eu/regulation-and-policy/credit-risk/guidelines-on-the-application-of-the-definition-of-default (accessed on 12 November 2021).

EBA. 2017. *Guidelines on PD Estimation, LGD Estimation and the Treatment of Defaulted Exposures*. EBA/GL/2017/16. Available online: https://www.eba.europa.eu/regulation-and-policy/model-validation/guidelines-on-pd-lgd-estimation-and-treatment-of-defaulted-assets (accessed on 12 November 2021).

EBA. 2020. *Guidelines on Credit Risk Mitigation for Institutions Applying the IRB Approach with Own Estimates of LGDs*. EBA/GL/2020/05. Available online: https://www.eba.europa.eu/regulation-and-policy/model-validation/guidelines-on-credit-risk-mitigation-for-institutions-applying-the-irb-approach-with-own-estimates-of-lgds (accessed on 12 November 2021).

ECB. 2018. *ECB Guide to Internal Models Risk-Type-Specific Chapters*; Brussels: ECB Banking Supervision, September. Available online: https://www.bankingsupervision.europa.eu/legalframework/publiccons/pdf/internal_models_risk_type_chapters/ssm.guide_to_internal_models_risk_type_chapters_201809.en.pdf (accessed on 12 November 2021).

EU. 2016. *Commission Regulation (EU) 2016/2067 of 22 November 2016 Amending Regulation (EC) No 1126/2008 Adopting Certain International Accounting Standards in Accordance with Regulation (EC) No 1606/2002 of the European Parliament and of the Council as Regards International Financial Reporting Standard 9*. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R2067 (accessed on 12 November 2021).

ICAAP. 2018. *ECB Guide to the Internal Capital Adequacy Assessment Process (ICAAP)*. Brussels: ECB Banking Supervision, November. Available online: https://www.bankingsupervision.europa.eu/ecb/pub/pdf/ssm.icaap_guide_201811.en.pdf (accessed on 12 November 2021).

Jacobs, Michael, and Nicholas M. Kiefer. 2010. *The Bayesian Approach to Default Risk: A Guide*. CAE Working Paper No. 10-01. Ithaca: Cornell University, Center for Analytic Economics.

Joseph, Dominic. 2021. Estimating credit default probabilities using stochastic optimization. *Data Science in Finance and Economics* 1: 255–71. [CrossRef]

Kiefer, Nicholas M. 2009. Default estimation for low-default portfolios. *Journal of Empirical Finance* 16: 164–73. [CrossRef]

Lessmann, Stefan, Bart Baesens, Hsin-Vonn Seow, and Lyn C. Thomas. 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research* 247: 124–36. [CrossRef]

Mehdiabadi, Amir, Mariyeh Tabatabeinasab, Cristi Spulbar, Amir Karbassi Yazdi, and Ramona Birau. 2020. Are We Ready for the Challenge of Banks 4.0? Designing a Roadmap for Banking Systems in Industry 4.0. *International Journal of Financial Studies* 8: 32. [CrossRef]

Papouskova, Monika, and Petr Hajek. 2019. Two-stage consumer credit risk modelling using heterogeneous ensemble learning. *Decision Support Systems* 118: 33–45. [CrossRef]

Peláez Suárez, Rebeca, Ricardo Cao Abad, and Juan M. Vilar Fernández. 2021. Probability of default estimation in credit risk using a nonparametric approach. *TEST* 30: 383–405. [CrossRef]

Robert, Christian P. 1994. *The Bayesian Choice*. Paris: Springer.

Simonian, Joseph. 2011. A Bayesian approach to building robust structural credit default models. *Applied Economic Letters*. [CrossRef]

Surzhko, Denis. 2017. Bayesian approach to PD Calibration and stress testing in low default portfolios. *Journal of Banking and Finance* 7: 83–98.

Wang, Zheqi, Jonathan N. Crook, and Galina Andreeva. 2018. Improving Model Predictive Accuracy Using a Bayesian Approach: Application to PD Modelling of Mortgage Loans. *SSRN Electronic Journal*. [CrossRef]

Wang, Zheqi, Jonathan N. Crook, and Galina Andreeva. 2020. Reducing Estimation Risk Using a Bayesian Posterior Distribution Approach: Application to Stress Testing Mortgage Loan Default. *European Journal of Operational Research* 287: 725–38. [CrossRef]

Zięba, Paweł. 2017. Methods of Extension of Databases Used to Estimate LGD Parameter. *Studia i Prace Kolegium Zarzadzania i Finansów* 150: 31–55.

risks

Article

Bankruptcy Prediction with a Doubly Stochastic Poisson Forward Intensity Model and Low-Quality Data

Tomasz Berent * and Radosław Rejman

Capital Markets Department, Warsaw School of Economics, 02-554 Warszawa, Poland; rrejman@sgh.waw.pl
* Correspondence: tberent@sgh.waw.pl

Abstract: With the record high leverage across all segments of the (global) economy, default prediction has never been more important. The excess cash illusion created in the context of COVID-19 may disappear just as quickly as the pandemic entered our world in 2020. In this paper, instead of using any scoring device to discriminate between healthy companies and potential defaulters, we model default probability using a doubly stochastic Poisson process. Our paper is unique in that it uses a large dataset of non-public companies with low-quality reporting standards and very patchy data. We believe this is the first attempt to apply the Duffie–Duan formulation to emerging markets at such a scale. Our results are comparable, if not more robust, than those obtained for public companies in developed countries. The out-of-sample accuracy ratios range from 85% to 76%, one and three years prior to default, respectively. What we lose in (data) quality, we regain in (data) quantity; the power of our tests benefits from the size of the sample: 15,122 non-financial companies from 2007 to 2017, unique in this research area. Our results are also robust to model specification (with different macro and company-specific covariates used) and statistically significant at the 1% level.

Keywords: default; bankruptcy risk; Poisson process; doubly stochastic assumption; ROC curve; accuracy ratio; leverage

Citation: Berent, Tomasz, and Radosław Rejman. 2021. Bankruptcy Prediction with a Doubly Stochastic Poisson Forward Intensity Model and Low-Quality Data. *Risks* 9: 217. https://doi.org/10.3390/risks9120217

Academic Editor: Mercedes Ayuso

Received: 14 October 2021
Accepted: 22 November 2021
Published: 2 December 2021

1. Introduction

In the world ravaged by the pandemic, with a fragile macro-economic outlook, low interest rates, and record high leverage, any academic research on bankruptcy is welcome. As the authors of this paper, we are neither professionally prepared nor interested in the debate on the degree of macro fragility or the effectiveness of the unprecedented stimulus packages adopted worldwide. Mentioning the unparalleled global leverage positions or the global health crisis, we merely acknowledge the emergence of the almost unmatched global uncertainty. In contrast to more optimistic views, exemplified by the stock exchange post-COVID-19 valuations, we believe the uncertainty is the only certain thing around us these days. Understanding the process of going down (in these circumstances) seems to us more than critical. Throughout this paper, bankruptcy and default are used interchangeably.

Below are some universally available leverage statistics (Altman 2020). Global non-financial corporate debt increased from the pre-Global Financial Crisis level of $42 trillion to $74 trillion in 2019. The government debt position more than doubled from $33 trillion in 2007 to $69 trillion in 2019. Even the financial sector increased its leverage from the record high pre-crisis level to $62 trillion. Households increased their debt globally from $34 trillion to $48 trillion. With the exception of the financial sector, debt also grew in relation to global GDP. It increased to 93% for non-financials (up from 77%), to 88% for governments (up from 58%), and to 60% for households (up from 57%). Despite this, as Altman (2020) notes, the corporate high-yield bond default rate was surprisingly low at 2.9% in 2019, below a 3.3% historic average, the recovery rate of 43.5% was quite in line with the historic average of 46%, with the high-yield spreads lagging behind historic averages, too. Consequently, Altman believes that the pre-COVID-19 debt market, in contrast to the current state, was still at a benign cycle stage. However, the very levels of debt globally, coupled with the increased appeal of a very long end of the yield curve and a massive increase in the BBB issuance, makes the debt markets and the global economy quite vulnerable, even without the health crisis. The unconventional monetary policies and the low interest environment also lead to the proliferation of "zombie" firms. Regardless of the precise definition, these companies are kept alive rather artificially thanks to the availability of cheap debt. Banerjee and Hofmann (2018) estimate as much as 16% of US listed firms may have "zombie" status—eight times more than in 1990. Acharya et al. (2020) estimate that 8% of all loans may also be infected with the "zombie" virus. Needless to say, COVID-19 and the resultant generous governmental relief packages do not help mitigate the problem.

Bankruptcy research, in all its guises, has been truly impressive and has produced many insightful results for some decades now. Such results include the classical structural models of Merton (1974), Fischer et al. (1989), and Leland (1994), numerous reduced-form models ranging from the simple scoring methods of Beaver (1966, 1968) and Altman (1968), qualitative response models, such as the logit of Ohlson (1980) and probit of Zmijewski (1984), to the third generation of the reduced form, the duration-type models of, for e.g., Shumway (2001), Kavvathas (2000), Chava and Jarrow (2004), and Hillegeist et al. (2004). They all propose various and divergent econometric methods and methodological approaches, with an impressive sectorial and geographic empirical coverage (see Berent et al. 2017).

To discriminate healthy from unhealthy firms is one challenge; to predict the (multi-period) bankruptcy probabilities is another. One way to address the problem is to model default as a random counting process. A Poisson process is such an example. In the bankruptcy literature, it is the Poisson process with stochastic intensities that is frequently used. In the doubly stochastic setting, the stochastic intensity depends on some state variables which may be firm-specific or macroeconomic, also called "internal" or "external" in the works on the statistical analysis of the failure time data (Lancaster 1990; Kalbfleisch and Prentice 2002). We adopt the Duffie–Duan model, as described in Duan et al. (2012), who, with their forward intensity approach and the maximum pseudo-likelihood analysis, follow in the footsteps of Duffie et al. (2007). In 2007, Duffie et al. (2007) first formulated a doubly stochastic Poisson multi-period model with time-varying covariates and Gaussian vector autoregressions. Duan et al. (2012) resolve some specification and estimation challenges inherent in Duffie et al. (2007). With their forward intensity concept, Duan et al. (2012) no longer need a high-dimension state variable process to be assumed, but instead use the data known at the time of making predictions. Both Duan et al. (2012) and Duffie et al. (2007) are well grounded in the doubly stochastic hypothesis literature debated in, for e.g., Collin-Dufresne and Goldstein (2001), Giesecke (2004), Jarrow and Yu (2001), and Schoenbucher (2003).

Duffie et al. (2007) apply their model to US-listed industrial firms. Duan et al. (2012) use also US public companies traded on NYSE, AMEX, and Nasdaq. Other researchers use the Duffie–Duan model to assess the default risk of public firms and/or in the context of developed (e.g., Caporale et al. 2017), or emerging markets (Duan et al. 2018). In this paper, we demonstrate that the Duffie–Duan model not only successfully describes a default process for public companies from developed countries with well-functioning capital markets, but is also equally successful in the context of privately owned equity markets with frequently patchy, low-quality data, operating in an emerging market characterized by lower transparency and governance standards (Aluchna et al. 2019). Compared to Duan et al. (2012) and Duffie et al. (2007), we apply the model to a significantly larger dataset of over 15,000 firms. As it is the applicability of the model rather than the discrimination itself that is our priority, we are not optimizing any cut-off point to maximize the accuracy of discrimination (typically made within an in-sample estimation context), but we make use of the out-of-sample accuracy measure calculated across all cut-off points, as is done in the context of the ROC analysis.

First, we collect a unique dataset for as many as 15,122 non-financial companies in Poland over the period 2007–2017. We make a huge effort to cross-check and cleanse the data so that the intricate estimation procedures could be run on this (initially) patchy input. Then, we document the performance differential (in the form of financial ratios) between the healthy and the (future) bankrupt firms one, two, and three years before default. In the next stage, company-specific variables (liquidity, profitability, leverage, rotation, and size) and macroeconomic variables (GDP growth, inflation, and interest rates) are used as state variables to estimate the default forward intensity employed in the doubly stochastic Poisson formulation. Partly due to the size of the dataset, we believe, we are able to exploit the differences between the attributes of the two groups. What we lose in the (data) quality, we seem to regain in (data) quantity. Our results surpassed our expectations. Not only are the estimated covariate parameters in line with the expectations and the literature, but the out-of-sample accuracy ratios produced—85% one year before default, 81% two years before default, and 76% three years before default—are at least as high, if not better, than those obtained for the high-quality public companies from developed countries. All our results are statistically significant and robust to the (state variable) model specification.

We hasten to repeat that this paper is not about searching for the determinants of default, neither is it about the maximization of the discrimination power between the two groups at any optimal cut-off point. In particular, we are not interested in artificially lifting up the in-sample fit. The main objective is to prove that the doubly stochastic Poisson model can be successfully used in the context of low-quality data for non-public companies from emerging markets. We believe this objective has been fully achieved. We are not aware of any similar effort in this area.

The rest of the paper is organized as follows. Below, still within the introduction section, we briefly introduce Poland's bankruptcy law, an important ingredient, given the recent overhaul of the legislation framework. In the Materials and Methods section, we describe our unique dataset, introduce the model and the micro and macro covariates, and then we define the accuracy ratio, our preferred goodness-of-fit measure, the ROC curves, and the statistical tests applied. In the Results section, we first produce and comment on the descriptive statistics, separately for survivors and defaulters, one, two, and three years prior to bankruptcy. We then analyze the estimated covariate parameters and accuracy ratios, in- and out-of-sample. The critical discussion of our results, in the context of the literature, follows in the Discussion section. We conclude with some proposals for future research in the Conclusions section.

Poland's Bankruptcy Law

With regard to Poland's bankruptcy law, it was not until 2003, nearly one and a half decades after the end of communism in Poland, that the new legislation came into force. The new Bankruptcy and Rehabilitation Act replaced the pre-war ordinance of the President of the Republic of Poland, dated as far back as 1934. The new law was universally praised for bringing together, under one umbrella, two separate bankruptcy and restructuring (composition) proceedings, hitherto governed by the two separate legal acts. It is paradoxical that the essence of the latest changes in Poland's bankruptcy law consisted in the carving out of the rehabilitation part into once again a separate Restructuring Act, which came into force in 2016. Apart from some substantial changes to the proceedings (e.g., the extension of both the time as well as the list of persons entitled/obliged to file for bankruptcy), the new law gave a debtor the option to choose, depending on the severity of insolvency, between four distinct ways to reach an agreement with the creditors. Given the discontinuity/alteration of the default definition brought upon the changes in the legal frameworks, care must be taken while conducting research in the bankruptcy field in Poland.

The motivation for the new 2016 law was clear, as the number of liquidation proceedings dwarfed the restructurings by the ratio of 5:1. As Figure 1 illustrates, the effort was worth making, as the number of restructuring proceedings has significantly improved ever since. In 2020, restructuring proceedings outnumbered liquidations, partly due to COVID-19-driven regulations. The trend is generally assumed to remain even after the pandemic.

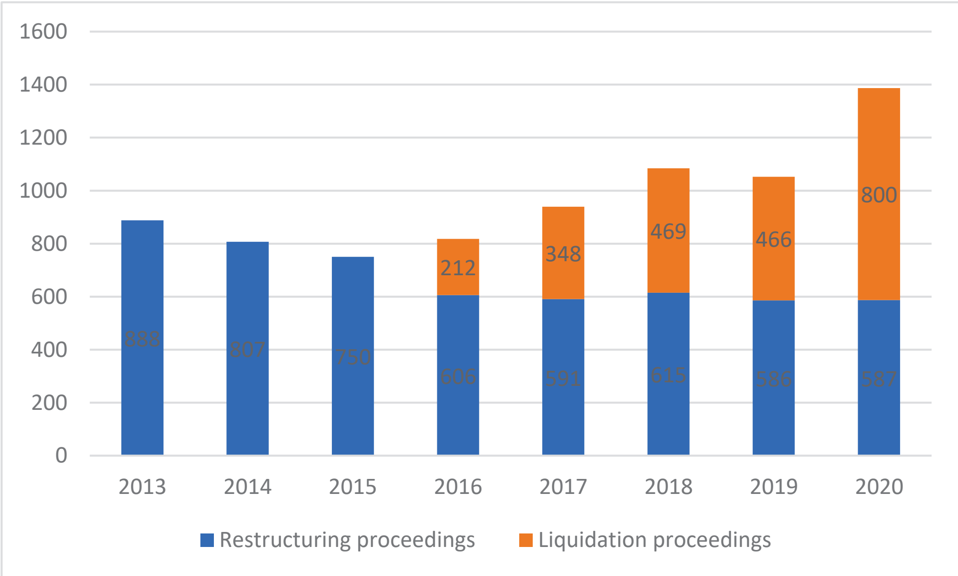


Figure 1. The number of liquidation proceedings vs. restructuring proceedings in Poland (pre-2016 restructuring proceedings not recorded) (Polish Institute of Credit Management).

2. Materials and Methods

2.1. Data

2.1.1. Companies' Financial Data

Our sample is quite unique in that it is large and dominated by non-public companies. It consists of two subsets. The former is the dataset of financial accounts, and the latter is the dataset on default events. Both are provided by Coface Group, the world's leading credit insurance provider and the owner of sensitive data on defaulters. Other data on, for e.g., macro statistics are obtained from publicly available sources such as Statistics Poland (GUS).

As for companies' financial statements, we assembled financial accounts for as many as 15,122 non-financial Polish companies. According to the Statistical Classification of Economic Activities in the European Community (NACE), 39.8% of our companies come from manufacturing, 30.8% represent the wholesale and retail trade, and the repair of motor vehicles and motorcycles, 12.2%—construction, 5.5%—transportation and storage, 2.8%—information and communication, 2.7%—professional, scientific, and technical activities, and 1.7%—other. All our entities are limited companies, with limited liability companies outnumbering joint stock companies by the ratio of 5:1. The dataset consists of 193,420 company periods from 2006 to 2018. We concede the data are of poor quality in that many missing cells are encountered or contradictory records reported (e.g., subtotals in the balance sheets do not add up). We made a substantial effort to validate, cross-check, and, if necessary, correct the dataset. As a result, we identified 143,451 useable annual company-years, i.e., a modest 73% of all company-years possible assuming all companies produced annual numbers for 2006–2018. Given the minuscule size of the 2006 and 2018 sub-samples, we arbitrarily excluded these years from our sample (see Figure 2). Consequently, when the period of analysis is limited to 2007–2017, the completeness of our dataset significantly improves to 86%. We regard this as satisfactory, as 100% is impossible by definition—some entities went down or exited for any other reason during the sample period. As shown in Figure 2, the number of company-years decreases towards the end of the period, which we associate with the fact that many non-public companies publish their accounts with a big lag. Moreover, some financial accounts arrive in multi-year packages, caused by bad release timing (clearly the company's fault) or for poor data-collection reasons, which may well be beyond the company's control. The low quality of the data is by no means disheartening. Quite the opposite. To be able to prove the applicability of the doubly stochastic Poisson process to non-public companies' data characterized by poor quality, frequently susceptible, and incomplete inputs does lie at the heart of this research, and as such, present a challenge rather than a problem.

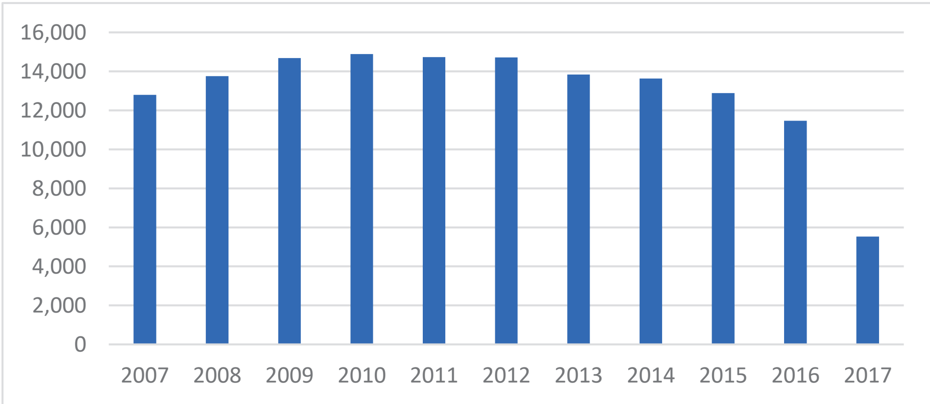


Figure 2. The number of company-years from 2007 to 2017.

Table 1 provides an accurate number of available company-years for each year from 2007 to 2017. The annual frequencies between 2007 and 2016, ranging between 11,470 in 2016 to 14,887 in 2010, are rather stable.

Table 1. The number of company-years in the sample 2007–2017.

Year	No. of Annual Financial Statements	Year	No. of Annual Financial Statements
2007	12,798	2013	13,834
2008	13,756	2014	13,631
2009	14,483	2015	12,883
2010	14,887	2016	11,470
2011	14,729	2017	5525
2012	14,712		

2.1.2. Bankruptcy Events

As for default events, we assembled various data subsets from Coface. Firstly, we obtained a list of 1240 companies ticked as bankrupt by Coface itself. Upon closer scrutiny, we realized the definitions of bankruptcy underpinning the tick were different from one firm to another. For example, in some instances, the company was implicated as bankrupt after the court decision, but in others, after a mere petition to court. Consequently, we also collected a separate dataset (from Coface) of 4095 events related (in any way) to the bankruptcy process. The list covers as many as 41 different, typically legal, categories, such as:

- Filings of a request with a court (by a creditor or a debtor, when applicable) to initiate either bankruptcy proceedings or one of many different restructuring/composition/rehabilitation proceedings;
- Different court decisions made at any different stage, e.g., a court approval (of one of many different arrangement schemes reached by the parties), a dismissal of the bankruptcy petition, a refusal to open restructuring proceedings, a discontinuation of proceedings, etc.

The list includes rather vague events coded by Coface as "bankruptcy" or "other bankruptcy event", too. For some reason (unclear to us and not explained by the data provider), the events recorded come mainly after 2003, with the highest frequencies after

2010. We subsequently verified that the "bankruptcy" events affected as many as 1240 different companies—the number of bankrupt firms originally identified by Coface. Again, instead of being disheartened by the vagueness and proliferation of different default events not untypical of non-public companies, we were rather determined to select the best event definition, most coherent across time, introducing the least heterogeneity into already high noise brought about by the poor quality of the financial data. The price we were prepared to pay was the reduction in the number of defaulters. A structural legal shift in the Polish bankruptcy law from 2016 onward made some "bankruptcy events" disappear. Other events (related to restructuring proceedings in particular) were only defined in 2016. Eventually, we decided to select the event, coded by Coface as 736, to mean "the declaration of bankruptcy to liquidate the debtors' assets". The event was identified for 455 companies from 2007 onwards, with only a few cases of registered liquidation recorded before 2012. Figure 3 shows that the default data are mostly represented by the last few years with no sign of structural change around 2015–2016. In summary, we used the 2007–2017 period for financial statements, and 2008–2018 for the information on default events. The financial statements one, two, and three years prior to the default event are referred to as tau = 1, tau = 2, and tau = 3.

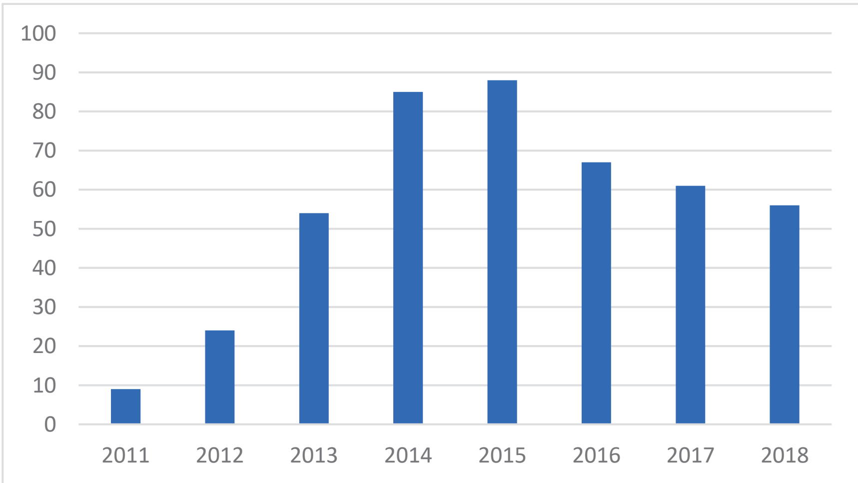


Figure 3. The number of companies declared bankrupt between 2011 and 2018.

2.2. Model

2.2.1. A Poisson Process

A Poisson process is a simple stochastic process widely used in modeling the times at which discrete events occur. It can be thought of as the continuous-time version of the Bernoulli process. In contrast to the known average time between events, the exact timing of events in a Poisson process is random. The process does not have memory, as the arrival of an event is independent of the event before. A Poisson process is a non-explosive counting process M with deterministic intensity λ, such that

$$\int_0^t \lambda_s ds < \infty \tag{1}$$

with the property that, for all t and $s > t$, the random variable M_s-M_t has the Poisson distribution with parameter

$$\int_t^s \lambda_u du \tag{2}$$

If a random variable N with outcomes $n = 0, 1, 2, \ldots$ has the Poisson distribution with parameter β, then the probability of n occurrences equals:

$$P(N = n) = e^{-\beta}\frac{\beta^n}{n!} \tag{3}$$

Consequently, the times between arrivals are independent exponentially distributed with mean $1/\lambda$.

We talk about the doubly stochastic Poisson process when the default intensity is stochastic and its intertemporal variation is allowed to depend on observable or unobservable state variables X_t that are linked to the probability of default. Hence, under a standard doubly stochastic assumption, firms' default times are correlated only as implied by the correlation of factors determining their default intensities. The conditional probability of default within s years is then

$$q(X_t, s) = E_t\Big[\int_t^{t+s} e^{-\int_t^z \lambda(u)du} \lambda z dz | X_t\Big] \tag{4}$$

Thanks to the doubly stochastic assumption, the dynamics of the state variables are not affected by default and the estimation of the model parameters is rather simple. The maximum likelihood estimator of the default probability can be obtained from the separate maximum likelihood estimations of a vector determining the dependence of the default intensity, and a vector determining the time-series behavior of the underlying state vector X_t of covariates.

Duffie et al. (2007) go one step further and estimate the probabilities of default over several future periods. For that, they need to know the stochastic process λ_t, or to understand the time-series dynamics of the explanatory state variables. Of many potential candidates for the specification of the behavior of X_t, they choose a simple Gaussian vector autoregressive model. Subsequently, they propose a maximum likelihood estimator, which, under some conditions, is shown to be consistent, efficient, and asymptotically normal.

The problems associated with the uncertain knowledge of the future values of state covariates and a potential misspecification of the model is overcome by Duan et al. (2012). Instead of modeling λ_t as some functions of state variables X_t, they propose directly a function $f_t(\tau)$ of state variables available at time t and the forward starting time of interest, τ. Following Duan et al. (2012), we also assume $f_{it}(\tau)$, for the i-th firm, to be

$$f_{it}(\tau) = e^{\alpha_0(\tau)+\alpha_1(\tau)x_{it,1}+\alpha_2(\tau)x_{it,2}+\ldots+\alpha_k(\tau)x_{it,k}} \tag{5}$$

If $\tau = 0$, our forward intensity set-up is the same as the spot intensity formulation of Duffie et al. (2007). The forward intensity method allows estimating future default probabilities without explicitly simulating the high-dimensional state variable process, hence only the data known at the time of performing the prediction are used. Thanks to the pseudo-likelihood function, the estimation of the forward default parameters are unrelated to each other. Being a variant of a standard doubly stochastic assumption, firms' survival and default probabilities are assumed to depend upon internal or external factors. Any dependency may only result from sharing of the common macro-economic factors and/or any correlation among the firm-specific attributes.

2.2.2. Covariates

The approach to the selection of a set of covariates differs widely. Duffie et al. (2007) mention several variables (e.g., 10-year Treasury yield, personal income growth rate, GDP growth rate, average Aaa-to-Baa bond yield spread, firm size), yet in the final analysis they use only four: distance to default, the three months Treasury bill rate, the (trailing) one-year return on both the S&P 500, and the stock. Their repeated emphasis on the inadequacy of distance to default, which is nothing but a volatility adjusted leverage measure based on Merton (1974), may suggest the comprehensive selection of covariates is not the main priority for Duffie et al. (2007), as long as they identify a model outperforming one that is solely dependent on the distance-to-default itself. Duan et al. (2012) are more generous in the selection of covariates. On top of the variables used in Duffie et al. (2007), they use the cash-to-assets ratio, net income-to-assets ratio, a firm's market equity value, P/BV, idiosyncratic volatility, etc. Other traditional firm-level risk factors such as leverage, profitability, growth, or liquidity, each in many different formats, are also available. An increasing consensus among researchers exists that adding macro variables to the company-specific covariates improves prediction (Beaver et al. 2005; Shumway 2001; Berent et al. 2017).

Remembering the objective of this paper is to show the applicability of the doubly stochastic Poisson model to a large dataset of low-quality corporate data from emerging markets, we are less preoccupied with the optimal selection of the state variables. Quite the opposite. Knowing the doubly stochastic assumption may, by definition, inadequately represent the default clustering, we opted to enlarge the size of our input beyond Duan et al. (2012). Multicollinearity-type problems that result from the presence of correlated covariates are not our major concern. Consequently, we chose five company-specific areas: liquidity, profitability, leverage, size, and rotation to be represented by two ratios each. The macro variables selected (GDP growth, interest rates, and inflation) are fairly standard, too. Table 2 presents the list and the definitions of the state variables used.

Table 2. The list of firm-specific and macro state variables.

Micro/Macro	Variable	Symbol	Definition
Micro	Liquidity	Cash/TA	Cash and cash equivalents to total assets
		CA/CL	Current assets to current liabilities
	Profitability	NP/TA	Net profit to total assets
		EBIT/Rev	Operating income to revenue
	Leverage	ND/EBIT	Interest bearing liability minus cash and cash equivalents to operating profit
		ND/E	Interest bearing liability minus cash and cash equivalents to equity
	Rotation	Rev/STR	Revenue to short-term receivables
		Rev/TA	Revenue to total assets
	Size	TA	Total assets
		Rev	Revenue
Macro	Gross Domestic Product growth	GDP	Gross domestic product
	Inflation	CPI	Consumer price index
	Interest Rates	WIBOR3	Three-month WIBOR

2.3. Accuracy Ratio

The choice among so many "goodness-of-fit" measures in reporting empirical results is critical in the default literature. The ambiguity surrounding this choice may indeed result in major confusion. It should certainly be driven by the research objective and the costs of misclassification. For example, if the priority is to let no (future) bankrupt firm to be treated as healthy, then the count of false negatives is pivotal. Even then, it is not obvious how to report it in relative terms: as the percentage of all bankrupt firms—a so-called false negative rate, or as the percentage of all firms diagnosed as healthy—a so-called false negative discovery rate. A very low false negative rate value may not correspond to a low false negative discovery rate value. Moreover, if one wanted to maliciously report a close to zero false negative rate, one should classify all (most) firms as bankrupt. A different measure, a so-called accuracy rate, takes into account the correct classification of both healthy and unhealthy firms. One should be warned, however, that in an unbalanced population (sample), with the positives being outnumbered by the negatives (as frequently happens), the simple allocation of a negative rating to all objects would, again, produce a very high success ratio.

The almost unparalleled richness of the vocabulary used in this context, a legacy of the binary classification being popular in so many areas (e.g., medicine, epidemiology, meteorology, information theory, machine learning, or methodology) makes the confusion even bigger. For example, depending on the context:

- A true positive rate (a percentage of all positives being diagnosed as positive) may also be referred to as sensitivity, recall, or power;
- A false negative rate (a percentage of all positives being misclassified) can also be called a miss rate, or beta;
- A proportion of negatives being misclassified as positives is referred to as false positive rate, false alarm probability, fall-out, or alfa;
- A proportion of all negatives being correctly tagged as negative is also known as a true negative rate, or specificity.

Similarly,

- A percentage of true positives in the group of all positive designations is called either a positive predictive value, a post-test positive probability, or simply precision;
- Accordingly, a percentage of diagnosed negatives which are indeed negative is known as a negative predictive value (NPV), or a post-test negative probability;
- The ratio of misclassified negatives, in relation to all those positively diagnosed, is referred to as a false positive discovery rate; and
- The ratio of misclassified positives, in relation to all those negatively diagnosed, is referred to as a false negative discovery rate, or a false omission rate.

The list of different names seems virtually endless, and so is the scope for confusion, even if there was clarity in what one really would like to optimize. A classification success of 95%, for example, may therefore mean many different things depending on how the success is defined. In many contexts, a high success is guaranteed with little or no effort. The topic of how the classification success should be measured certainly deserves separate treatment, not only for its linguistic capacity. The wide confusion surrounding the reported "accuracy" of various COVID-19 diagnostic tests has recently demonstrated the huge potential for misunderstanding (or even manipulation).

As mentioned earlier, we are less determined to maximize the goodness-of-fit in terms of false/true positive/negative designations for any given cut-off point. Instead, following, for e.g., Duffie et al. (2007) and Duan et al. (2012), we report the diagnostic ability of a binary classifier to classify correctly across all cut-off points simultaneously; hence, we want to report the diagnostic ability of the model itself, rather than that of a particular cut-off point. To do this, we apply a standard binary-classification receiver operating characteristic, or ROC analysis, and construct a ROC curve, a plot of the true positive rate (the percentage of bankrupt companies correctly identified as bankrupt) against the false positive rate (the percentage of healthy firms identified as bankrupt) across all cut-off points. We subsequently calculate the AUC—the area under (ROC) curve. For a perfect segregation (all bankrupt companies identified as bankrupt, and all healthy companies identified as healthy—possible only for perfectly separate score distributions), the AUC is 100%. When the binary classification is random, then AUC = 50%. We note that the accuracy rate, defined previously as the percentage of correct (true and false) designations, is only vaguely related to the accuracy ratio concept computed in the context of the ROC analysis. The former is calculated for a given cut-off point, and the latter for a full spectrum of them. The accuracy ratio (AR), closely related to a famous Gini Index, is calculated as

$$AR = 2 \times (AUC - 0.5) \tag{6}$$

The perfect discrimination is 100% again, and the random one is fittingly no longer 0.5 (as it is for AUC), but is zero. Table 3 lists some pairs of AUC and AR.

Table 3. Area under curve (AUC) and a corresponding accuracy ratio (AR) for selected levels.

AUC	AR
50%	0%
60%	20%
70%	40%
80%	60%
90%	80%
95%	90%
100%	100%

2.4. ROC Curves, Statistical Inference and Out-of-Sample Testing

To prove the doubly stochastic Poisson process is successful in modelling default in the context of a large database of low-quality data, we produce both in-sample as well as cross-section out-of-sample ROC curves and compute the corresponding in- and out-of-sample accuracy ratios. Subsequently, two statistical tests, the Hanley–McNeil test (Hanley and McNeil 1982; Beaver et al. 2005) and the DeLong test (DeLong et al. 1988; Bharath and Shumway 2008), are performed to check whether a sample AR does not come from a random variation. The null hypothesis is therefore

$$H_0\text{: AUC} = 0.5 \tag{7}$$

against

$$H_1\text{: AUC} \neq 0.5 \tag{8}$$

To perform out-of-sample tests, the whole sample (both default and surviving firms) is randomly split into two equal sub-samples: one used in estimation, and one used in out-of-sample testing. We additionally perform diagnostic tests to prove the split is indeed random. The estimated vector of parameters from an in-sample estimation, applied to a group of companies that are not used in estimation, produces an out-of-sample ROC curve and the corresponding AR. The Hanley–McNeil and DeLong tests are computed to check whether the out-of-sample results are statistically significant. We expect in-sample accuracy ratios to be much larger than those derived from the out-of-sample data, even though empirical evidence from the developed markets suggests the difference between in- and out-of-sample is usually quite modest. Both are expected to be significantly different from the random rating designation.

3. Results

3.1. Descriptive Statistics

Below, we present descriptive statistics, i.e., the quartile values for the first (25th percentile, or 25), the second (50th percentile, or 50), and the third quartile (75th percentile, or 75) for both companies that went bust during our sample period 2008–2018 and for those that survived. The ratios analyzed are those used in the state variables model, i.e., liquidity, profitability, leverage, rotation, and size, all represented by two indices (see Table 2). For the default firms, we show the numbers one (tau = 1), two (tau = 2), and three (tau = 3) years before the reported default. We also report the numbers for surviving companies.

3.1.1. Liquidity

Tables 4 and 5 contain liquidity measures, as defined in Table 2. For clarity, the default companies are all placed on separate graphs (on the left), with the survivors being compared only to the most survivor-like bankrupt companies, i.e., firms three years before default (on the right). Bankrupt companies tend to have less and less cash in relation to their assets as they approach default. This is true at every quartile. Moreover, as both Table 4 and Figure 4b show, the survivors have more than twice as much cash (in relation to assets) as the bankrupt firms three years before default.

Table 4. Cash/TA one, two, and three years before default.

Cash/TA	tau = 1	tau = 2	tau = 3	Survivors
25	0.0037	0.0066	0.0078	0.0162
50	0.0159	0.0202	0.0222	0.0574
75	0.0462	0.0514	0.0587	0.1646

Table 5. CA/CL one, two, and three years before default.

CA/CL	tau = 1	tau = 2	tau = 3	Survivors
25	0.3736	0.6701	0.7777	1.0663
50	0.7376	0.9138	1.0188	1.4895
75	1.0585	1.1735	1.3080	2.4051

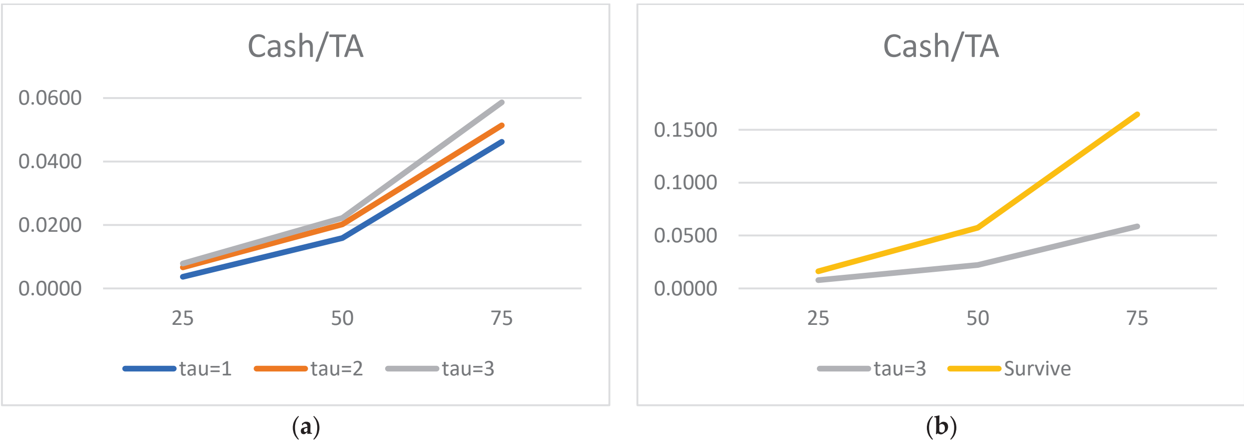


Figure 4. Cash to total asset: (**a**) default companies; (**b**) survivors vs. default companies three years before default.

The conclusions barely change when liquidity is measured in terms of current assets to current liabilities. The closer to default, the lower the ratio, with the survivors well ahead of the bankrupt firms three years before default (Figure 5).

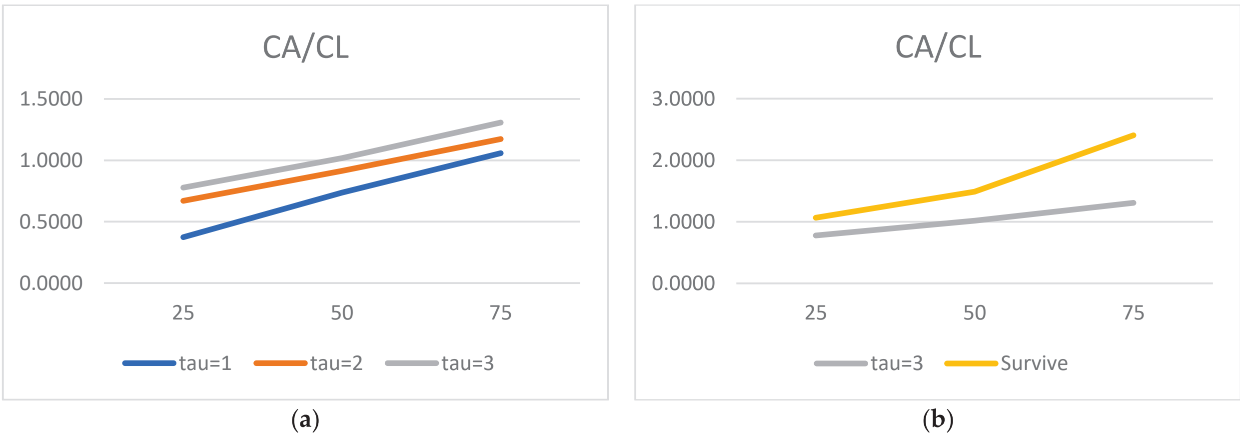


Figure 5. Current assets to current liabilities: (**a**) default companies; (**b**) survivors vs. default companies three years before default.

3.1.2. Profitability

The conclusions on profitability are as uncontentious as on liquidity (see Tables 6 and 7). In contrast to the survivors (always profitable on both net and operating levels), losses on both net and operating levels are reported for bankrupt companies in the lowest quartile already three years before default. As for the medians, they are negative for both NP/TA and EBIT margins one year before default. The closer to default, the smaller the net profit (in relation to assets) at each quartile. Similarly, the operating margins get worse towards default. In short, the closer to default, both the return on assets and the operating margins deteriorate.

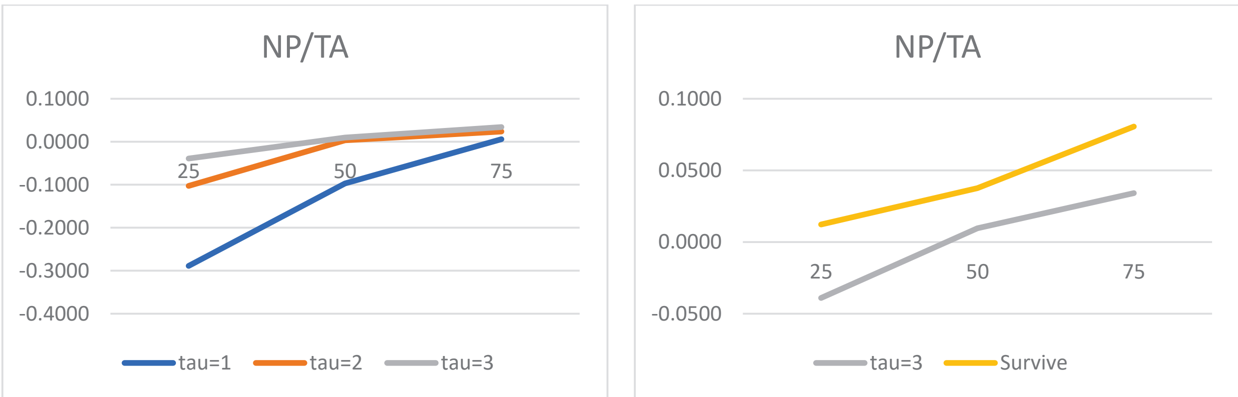
Table 6. NP/TA one, two, and three years before default.

NP/TA	tau = 1	tau = 2	tau = 3	Survivors
25	−0.2888	−0.1027	−0.0390	0.0091
50	−0.0970	0.0034	0.0097	0.0454
75	0.0059	0.0238	0.0341	0.1052

Table 7. EBIT/Rev one, two, and three years before default.

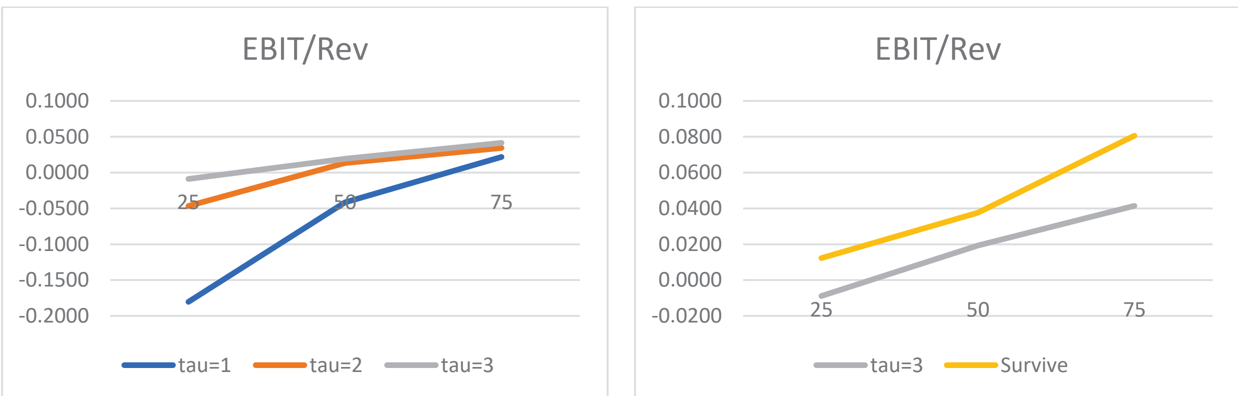
EBIT/Rev	tau = 1	tau = 2	tau = 3	Survivors
25	−0.1804	−0.0466	−0.0089	0.0123
50	−0.0416	0.0133	0.0192	0.0376
75	0.0218	0.0344	0.0414	0.0806

Figures 6a and 7a graphically illustrate the miserable financial condition of the bankrupt companies one year before default in terms of their profitability, particularly in the first quartile. Figures 6b and 7b show that the distance in profitability between the surviving and the bankrupt companies in our sample, even three years before default, could hardly be bigger.



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

Figure 6. Net profit to total assets: (**a**) default companies; (**b**) survivors vs. default companies three years before default.



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

Figure 7. EBIT margin: (**a**) default companies; (**b**) survivors vs. default companies three years before default.

3.1.3. Leverage

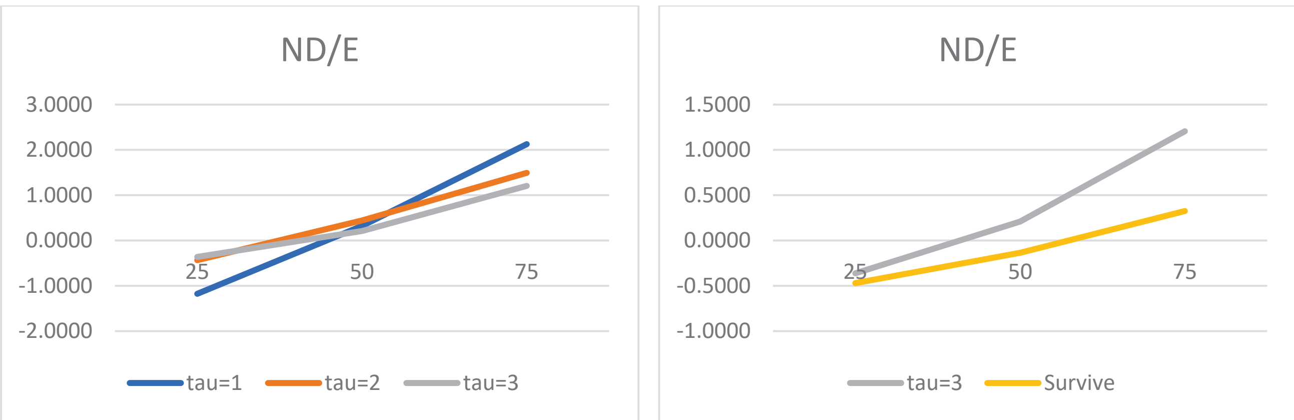
The leverage ratios (ND/E and ND/EBIT) are more difficult to interpret (see Tables 8 and 9, Figures 8 and 9). The companies with financial problems may have both huge (net) debt and low, potentially negative EBIT and equity. In contrast, healthy companies may have low, potentially negative net debt and highly positive EBIT and E. After scanning both the default firms and the survivors in our sample in terms of the net debt position, we find that as much as 55% of healthy companies have more cash and cash equivalents than debt, hence they have a negative net debt position. This drops to 30%, 25%, and 15% for companies facing bankruptcy within three, two, and one year, respectively. Similarly, negative equity is not practically reported for the survivors in our sample. Yet, for the bankrupt companies, as many as 30% of firms show negative equity one year prior to default, dropping to around 10% two and three years before default. With regard to EBIT, 25%, 35%, and 55% of defaulters have negative operating profit three, two, and one year prior to default, respectively. Within surviving firms, less than 5% are in the red on the operating level.

Table 8. ND/E one, two, and three years before default.

ND/E	tau = 1	tau = 2	tau = 3	Survivors
25	−1.1764	−0.4346	−0.3620	−0.4701
50	0.3232	0.4436	0.2082	−0.1363
75	2.1259	1.4931	1.2056	0.3250

Table 9. ND/EBIT one, two, and three years before default.

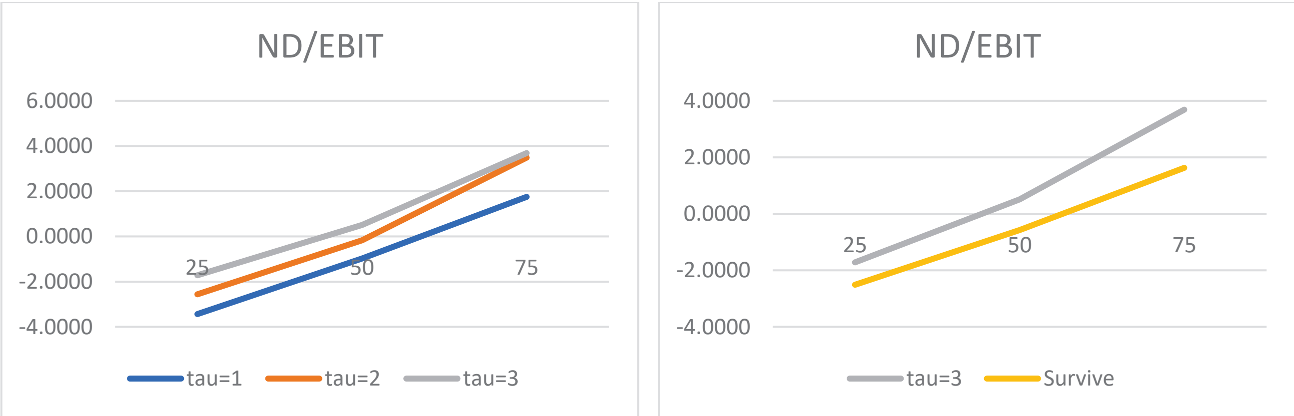
ND/EBIT	tau = 1	tau = 2	tau = 3	Survivors
25	−3.4364	−2.5609	−1.7182	−2.5117
50	−0.9796	−0.1704	0.5084	−0.5766
75	1.7503	3.4877	3.6871	1.6281



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

Figure 8. Net debt to equity: (**a**) default companies; (**b**) survivors vs. default companies three years before default.

As a result, for example, the bankrupt companies one year before default post the most negative value of ND/E in the first quartile, and the highest positive value in the third (Table 8, Figure 8a). There is very little order in ND/E over time (to default). In terms of ND/EBIT (Table 9 and Figure 9b), the survivors frequently exhibit the lowest values of the ratio.



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

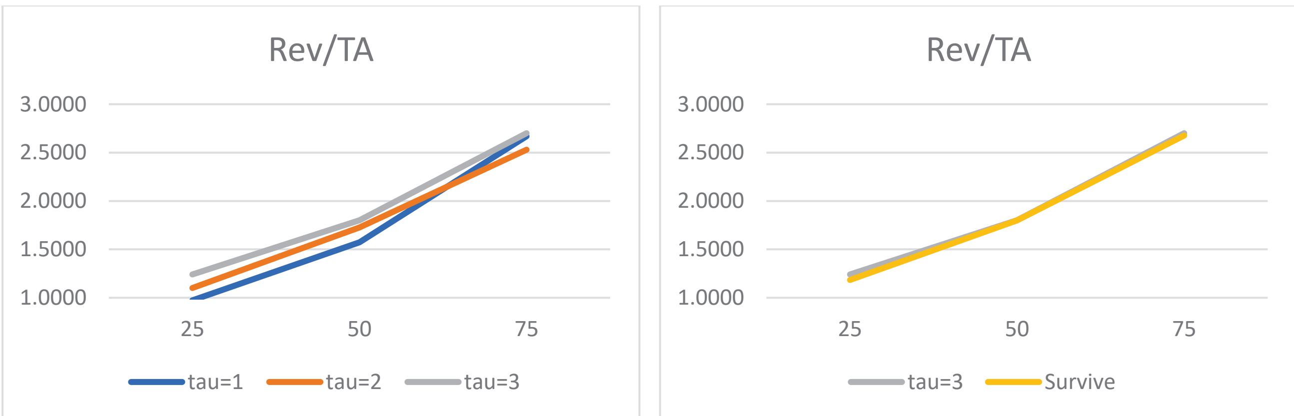
Figure 9. Net debt to EBIT: (**a**) default companies; (**b**) survivors vs. default companies three years before default.

3.1.4. Rotation

As illustrated in Tables 10 and 11 and Figures 10 and 11, the assets rotation is not a strong discriminator, regardless of whether we compute the total asset or the short-term receivables rotation ratios. For most cases, the revenue of (future) bankrupt firms in relation to assets shrinks as the company moves towards default. The sample distributions across time are (surprisingly) close to each other. Revenue tends to be 1.0–1.2, 1.6–1.8, and 2.5–2.7 times higher than the total assets for, respectively, the first, the second, and the third quartile, regardless of whether the rotation is measured one, two, or three years before default. The multiples for the survivors fit well into these ranges, too. As illustrated by Table 11 and Figure 11b, the short-term receivables rotations are even more homogeneous, with little discrimination between the survivors and the defaulters.

Table 10. Rev/TA one, two, and three years before default.

Rev/TA	tau = 1	tau = 2	tau = 3	Survivors
25	0.9733	1.1009	1.2409	1.1839
50	1.5727	1.7264	1.8018	1.7991
75	2.6691	2.5298	2.7008	2.6765

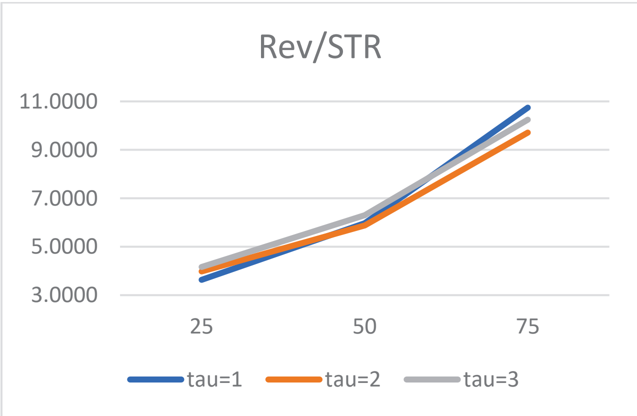


(**a**) (**b**)

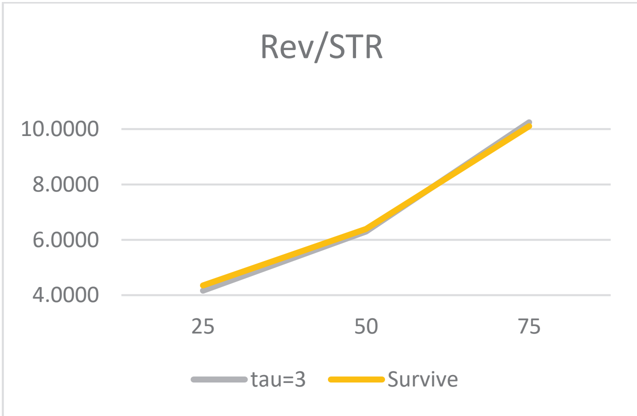
Figure 10. Total assets rotation: (**a**) default companies; (**b**) survivors vs. default companies three years before default.

Table 11. Rev/STR one, two, and three years before default.

Rev/STR	tau = 1	tau = 2	tau = 3	Survivors
25	3.6326	3.9797	4.1633	4.3510
50	5.9690	5.8769	6.2982	6.3908
75	10.7425	9.7118	10.2449	10.1043



(**a**)



(**b**)

Figure 11. Short-term receivables rotation: (**a**) default companies; (**b**) survivors vs. default companies three years before default.

3.1.5. Size

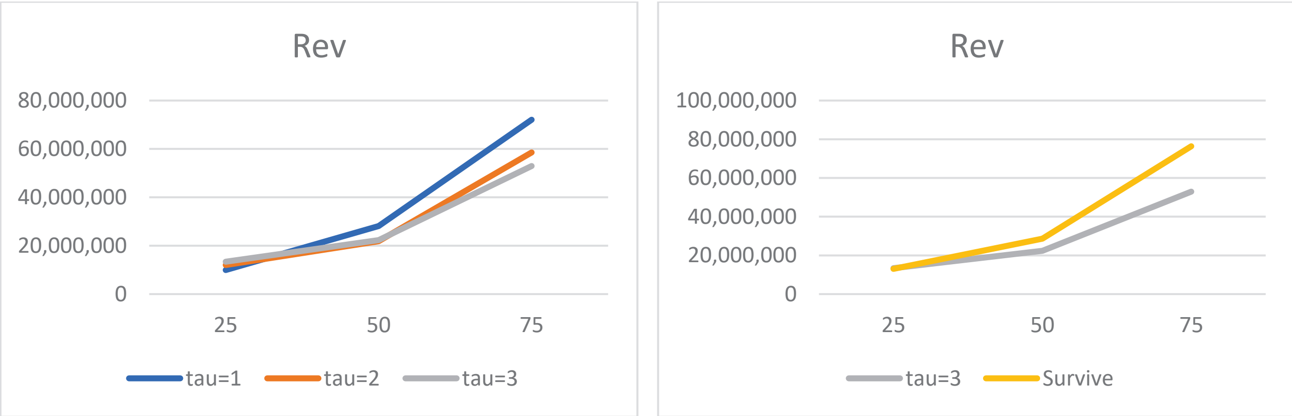
As Tables 12 and 13 and Figures 12a and 13 show, the companies that went bust during our sample period had lower revenue and lower assets one year before default compared to what they had earlier (and what was recorded for the survivors, on average) for only the first quartile. Surprisingly, the medians and the values for the third quartile are higher one year prior to default and do not differ materially from the values for the surviving firms. The survivors are also (marginally) smaller than some future bankrupt firms, according to the first quartile statistics (see Figures 12b and 13b). We are not able to easily explain this finding, but comment later on the ambiguity of the size statistics found in the bankruptcy literature.

Table 12. Rev one, two, and three years before default (PLN).

Rev	tau = 1	tau = 2	tau = 3	Survivors
25	9,943,760	12,024,161	13,391,353	13,041,557
50	28,098,943	21,865,540	22,341,462	28,603,362
75	72,078,970	58,552,894	52,935,266	76,376,322

Table 13. TA one, two, and three years before default (PLN).

TA	tau = 1	tau = 2	tau = 3	Survivors
25	5,757,452	6,628,929	7,228,960	7,049,281
50	15,068 820	13,947,311	15,023 695	17,191,949
75	51,063,916	48,831,692	37,612,128	50,360,517



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

Figure 12. Revenue: (**a**) default companies; (**b**) survivors vs. default companies three years before default.

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

Figure 13. Total assets: (**a**) default companies; (**b**) survivors vs. default companies three years before default.

3.2. Parameter Estimates

In Table 14, we present the maximum pseudo-likelihood estimates for $\alpha(\tau)$. These parameters quantify the impact of various firm-specific factors on its default probability. The set of $\alpha(\tau)$ is estimated separately for tau = 1, tau = 2, and tau = 3. Given the rather generous and overlapping representation of the various firm attributes in our model, which makes the parameters estimates vulnerable to problems related to multicollinearity, we are more than encouraged to see that the signs of the parameters are almost perfectly realigned with what we expected. In particular, as Table 14 shows, the higher the liquidity of the company, be it in the form of cash or working capital, the lower the probability of the company being unable to pay back its debt and interest on it—hence, the lower the probability of default. In other words, the forward default intensities are estimated to increase, as expected, with the decrease of cash to assets and current assets to current liabilities ratios. Interestingly, the estimated negative α-s for both Cash/TA and CA/CL increase (in absolute values) with every year nearer default. This would suggest, unsurprisingly, the recorded drop in liquidity is more and more punishing when approaching financial distress.

Similar results are reported for profitability. All parameters are negative, as expected, for both return on assets and operating margins; i.e., the higher the profitability, the lower the default intensity and the probability of default. The parameters are also negative for all time horizons, i.e., one, two, and three years before default, with the strength of the sensitivity of default to the drop in profitability (proxied by the return on assets) increasing with the time approaching default.

Table 14. The maximum pseudo-likelihood estimates of $\alpha(\tau)$ for company-specific variables.

	Cash/TA	CA/CL	NP/TA	EBIT/Rev	ND/EBIT	ND/E	Rev/TA	Rev/STR	REV	TA
tau = 1	−0.2473	−1.972	−0.4410	−0.0059	−0.0004	0.0041	−0.5661	−0.1792	0.2547	−0.4487
tau = 2	−0.1596	−1.598	−0.2764	−0.0494	0.0027	0.0041	−0.2469	−0.2524	−0.0602	−0.2571
tau = 3	−0.1056	−1.258	−0.1882	−0.0005	0.0029	0.0056	0.0156	−0.2571	−0.1661	−0.1956

The evidence on leverage is also appealing. The higher net debt (to either E or EBIT), the higher the forward default intensity and the probability of failure (in five out of six cases, Table 14). Still, we should remember the ambiguity of the leverage ratios at the lower end of the distribution discussed above, where negative levels of ND/E or ND/EBIT may be recorded for either good companies (ND < 0) or bad ones (E < 0, EBIT < 0). We will return to the issue and the robustness of these estimates in the Discussion section.

The impact of rotation on the probability of default is also coherent throughout time and proxies used. The companies with better utilization of their assets (either total assets or just short-term receivables) tend to have a lower likelihood of default. Again, the strength of this relationship seems to increase as a company approaches default.

Unsurprisingly, the size of the company tends to adversely affect the probability of default. Five out of six size parameters (for total assets and revenue) are negative. The larger the company, the greater the financial flexibility, the higher the diversification, and the lower the idiosyncratic risk. It is no surprise, then, that there is also a lower probability of default.

Leaving more detailed debate to the Discussion section, we are pleased to summarize that 27 out of 30 parameters have the signs as expected. This comes as a surprise to us, given the low quality of the data used as well as the fact that we use two ratios for each company's attribute. Table 15 includes the parameters estimated for the macro factors. The weaker the economy, as measured by the GDP growth, the higher the probability of default. Moreover, as the time to default narrows, the effect becomes stronger. The negative relation between default intensity and interest rates looks less appealing. We would expect the opposite: the higher the interest rates, the heavier the debt cost burden, and the higher the probability of default. We hasten to add at this stage that our result is frequently reported in this type of research. The impact of inflation is less conclusive. We come back to the interpretation of our results including macro factors in the light of the literature later on in the Discussion section.

Table 15. The maximum pseudo-likelihood estimates of $\alpha(\tau)$ for macro variables.

	GDP	CPI	WIBOR3
tau = 1	−2.5770	1.7340	−19.9000
tau = 2	−2.1240	2.8420	−24.2600
tau = 3	−2.0510	−0.9942	−24.8800

3.3. Accuracy Ratios

As explained in the Materials and Methods section, to evaluate the usefulness of the model, we compute a Gini-index like measure, an accuracy ratio AR, which is a standard measure in the binary classification analysis. We reiterate that the optimization of the accurate designation of companies to positive (bankrupt) and negative (non-bankrupt) groups for a given cut-off point is not our priority. Instead, we are interested in predicting default probabilities, which in turn separate the two groups regardless of the cut-off point selected. In other words, we are determined to measure the goodness-of-separation across all cut-off points simultaneously. To achieve this, we compute the accuracy ratios derived from cumulative accuracy profiles using true positive and false positive rates. We report both in-sample as well as out-of-sample estimates. All results are tested for statistical significance using Hanley–McNeil and DeLong tests.

The in-sample accuracy is much higher than we could ever have hoped for working with such a patchy dataset. AR = 0.8823 for tau = 1 implies the segregation of default and healthy firms one year before default, which is at least as good, if not better, than those achieved by Duan et al. (2012) for the US public firms. The same applies to two and three years before default, where accuracy rates are still above 0.8000. The results are statistically significant at the 1% level using either test. Table 16 reports all our accuracy ratios, with tau = 0 results provided for comparison only. The accuracy of the model that separates the actual defaults from the survivors, based on what effectively is the ex-post data (tau = 0), is almost perfect (AR = 0.9599).

Table 16. In-sample and out-of-sample accuracy ratios.

	In-Sample	Out-of-Sample
tau = 0	0.9599 ***	0.8221 ***
tau = 1	0.8823 ***	0.8475 ***
tau = 2	0.8299 ***	0.8114 ***
tau = 3	0.8028 ***	0.7590 ***

*** significant at the 1% level.

We are fully aware that our sample is quite large, certainly larger in terms of number of companies than that of Duan et al. (2012). The chance of over-fitting is a real threat; hence, out-of-sample tests are indispensable. As expected, our out-of-sample accuracies are lower. Nevertheless, they still comfortably beat those reported by Duan et al. (2012), and are only marginally lower than the one-year accuracy reported by Duffie et al. (2007), even if Duffie et al. (2007) do not report two- and three-year statistics. The scale of the drops relative to the in-sample results is relatively small, ranging from 1.85 percentage points for tau = 2 to 3.48 and 4.38 percentage points for tau = 1 and tau = 3, respectively. We also highlight the out-of-sample accuracy for tau = 0 dropping drastically—a clear sign of over-fitting in the context of in-sample estimation. Both Hanley–McNeil and DeLong tests confirm that it is practically impossible this result is an outcome of the random variation of a neutral scoring system.

We also report the results graphically. Figure 14 shows areas under the curve versus the 45° line. The better the segregation of healthy and default firms, the steeper the curve at first. In-sample over-fitting for tau = 0 is easy to see, with all other graphs pointing to a strong discrimination power of the model. The bell-shaped curves represent the difference between the ROC curves and the 45° lines.

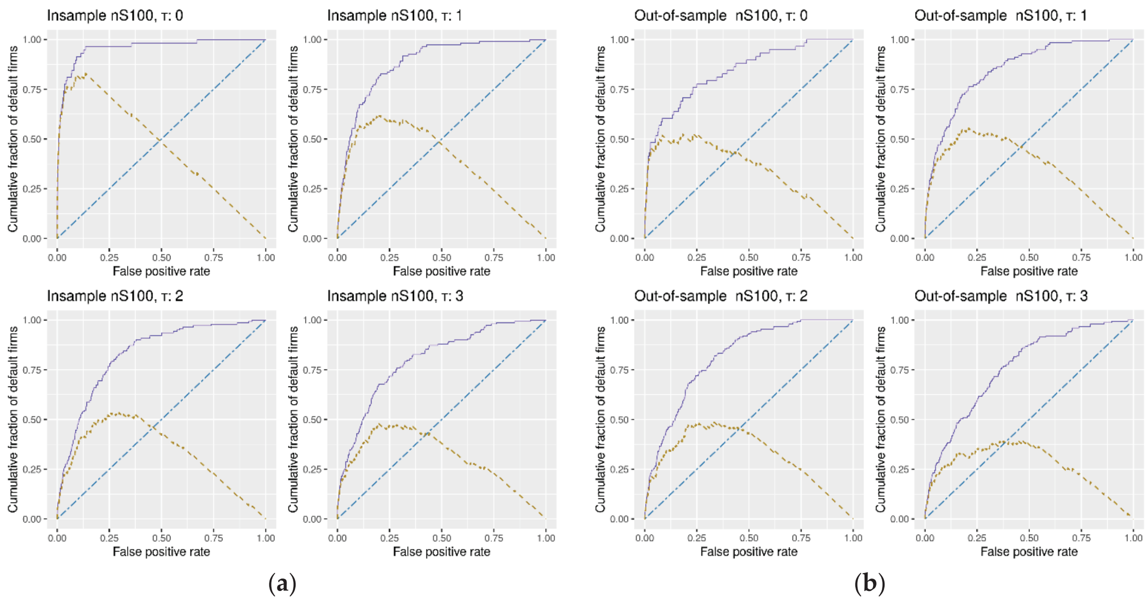


Figure 14. ROC curves for tau = 1, tau = 2, tau = 3 (tau = 0 for comparison only): (**a**) in-sample; (**b**) out-of-sample.

4. Discussion

The structural models (Merton 1974; Fischer et al. 1989; Leland 1994) make use of the distance to default to predict the moment when the firm's liabilities exceed its equity. This is clearly an approach determined endogenously. Other research (Beaver 1966, 1968; Altman 1968; Ohlson 1980; Zmijewski 1984; Kavvathas 2000; Shumway 2001; Chava and Jarrow 2004; Hillegeist et al. 2004) applies various statistical tools to optimize the discrimination between (future) defaulters and healthy companies using reduced-form formulations. Our model, instead, assumes that at each point in time, default occurs at random, with the probability of default, following Duffie et al. (2007) and Duan et al. (2012), depending on some company-specific and/or macroeconomic explanatory variables. Swapping the model of Duffie et al. (2007), which requires the knowledge of the exact level for the future state random variables, for the model of Duan et al. (2012), we can fully rely on the information available to us at the moment of making a prediction. If the model for the dynamics of the covariates proposed by Duffie et al. (2007) is mis-specified, then the predictions are contestable. The model of Duan et al. (2012), with its maximum pseudo-likelihood estimation procedure, is void of these problems. We should therefore agree with Duan et al. (2012) that apart from the computational efficiency, their approach is more robust, especially for long-term predictions. We use the model proposed by Duan et al. (2012), but throughout the paper refer to it as the Duffie–Duan model.

As for the choice of covariates, the approaches vary widely, not only in the context of broadly conceived default debate, but also within the doubly stochastic Poisson literature. For example, Duffie et al. (2007) choose two firm-specific (the share price performance and the distance-to-default) and two macro factors (the three-month Treasury bill rate and the trailing one-year return on the S&P index). Duan et al. (2012) choose the same two macro factors and complement them with as many as six company-specific variables. In contrast to Duffie et al. (2007), Duan et al. (2012) opt for more traditional factors measuring a company's liquidity, profitability, and size. In addition, they add the idiosyncratic volatility and P/BV ratio. Caporale et al. (2017), applying the Duffie–Duan model to general insurance firms in the UK, are even more "lavish" in the use of state variables: they go for no less than twelve micro and six macro factors. Apart from the insurance-specific variables, all other standard company-specific candidates are represented: liquidity, profitability, leverage, size, growth, etc. Macro conditions are proxied by interest rates, inflation, and GDP growth, but also by exchange rate and FDI indices.

With our standard five micro and three macro factors, we are located somewhere in-between. As for the company-specific set (liquidity, profitability, leverage, rotation, and size), we attempt to link our research to the seminal papers of Altman (1968). We include the assets rotation for its Dupont analysis appeal. As for the firm size, large firms are thought to have more financial flexibility than small firms, hence size should be crucial. Yet, as Duffie et al. (2007) demonstrate, the size insignificance results from the wide representation of other variables in the model (see also Shumway 2001). In terms of the macro set (inflation, interest rates, and GDP growth), we could hardly be more conservative. We admit that in choosing more variables, we have also been encouraged by Caporale et al. (2017), who show that all of their chosen factors (18 in total) proved statistically significant.

We would now like to turn to the state variable estimated parameters and acknowledge the truly surprising consistency of the covariate signs. As many as 27 out of 30 alphas are consistent with the theory. We show evidence that the lower the liquidity, profitability, asset rotation, and size, the higher the probability of default. Conversely, the lower the leverage, the higher the chance of the survival. Moreover, this result is surprisingly robust. The covariates' parameters barely move with the change of a state variable vector. We have verified it with as many as 100 different versions of the model. The accuracy ratios do go down when the number of state variables is drastically reduced, but the parameter signs (and frequently their magnitude) are broadly unchanged.

Regarding the time sequence towards default, just like in Duan et al. (2012), we report ever higher (in absolute terms) negative coefficients for both versions of the liquidity ratios.

The closer to the default moment, the more painful the drop in liquidity. We also report a similar time trend for the return on assets: the closer to default, the lower the profitability, resulting in a higher probability of failure. As for rotation, it is also observed for total assets turnover.

In contrast to the micro variables, some of our results on the macro factors are less intuitive. We hasten to add it is a regularly received outcome in the default literature using doubly stochastic formulations. For example, Duffie et al. (2007), just like us, report a negative relationship between short-term interest rates and default intensities; i.e., the higher the interest rates (the higher the costs of debt), the lower the chance of going bust. They argue that this counter-intuitive result may be explained by the fact that "short rates are often increased by the US Federal Reserve in order to 'cool down' business expansions" (p. 650). Similarly, to find a positive relation between the S&P index and default intensity is to say that when the equity market performs well, firms are more likely to default. This clearly counter-intuitive result is produced by both Duan et al. (2012) and Duffie et al. (2007). The correlation between the S&P 500 index return and other firm-specific attributes are quoted to be responsible for this outcome—i.e., in the boom years, financial ratios tend to overstate the true financial health. We must admit that we find these explanations somewhat arbitrary. What we take from this debate, however, is simple: adding the macro dimension is beneficial to the power of the model (see also Beaver et al. 2005; Shumway 2001; or Berent et al. 2017). This is precisely what we aim at in our paper. It is less important for us to measure the strength or to rank the importance of various macro (and micro) constituents. It is the performance of the whole model that is at stake here.

We turn now to the measures of the model's goodness-of-fit—the pivotal part of this research. As already mentioned, our accuracy ratios are very high by all standards. This is further documented by Table 17. Compared to Duffie et al. (2007) and Duan et al. (2012), two seminal papers in this area, our results (in- and out-of-sample) tend to be quite good: better than those of Duan et al. (2012) and similar to the (out-of-sample) results of Duffie et al. (2007). The latter do not publish in-sample statistics, with only one and five year ahead out-of-sample accuracy ratios released.

Table 17. In-sample and out-of-sample accuracy ratios compared to the seminal results of Duffie et al. (2007) and Duan et al. (2012).

	In-Sample Berent & Rejman	**In-Sample Duan et al. (2012) [1]**	**Out-of-Sample Berent & Rejman**	**Out-of-Sample Duan et al. (2012) [1]**	**Out-of-Sample Duffie et al. (2007) [2]**
tau = 1	0.8780	0.8352	0.8452	0.8336	0.8800
tau = 2	0.8270	0.7410	0.8046	0.7337	n/a
tau = 3	0.8045	0.6667	0.7608	0.6547	n/a
tau = 5	n/a	n/a	n/a	n/a	0.6900

[1] The accuracy ratios in Duan et al. (2012), who run the model with monthly data, are reproduced for 12, 24, and 36 months, respectively. [2] Duffie et al. (2007) do not release in-sample statistics. We acknowledge their accuracy ratio is slightly different than that derived from a standard ROC analysis.

To put our results into perspective, we quote Duffie et al. (2007), who in turn place their results in the context of other research. They note that their out-of-sample accuracy ratio of 88% favorably compares to the 65% produced by Moody's credit ratings, 69% based on ratings adjustments on Watchlist and Outlook, and 74% based on bond yield spreads, all reported by Hamilton and Cantor (2004). Duffie et al. (2007) also quote Bharath and Shumway (2008) who, using KMV estimated default frequencies, place approximately 69% of the defaulting firms in the lowest decile. The model of Beaver et al. (2005), based on accounting ratios, places 80% of the year-ahead defaulters in the lowest two deciles, out of sample, for the period 1994–2002. Against all these measures, the accuracy ratios of Duffie et al. (2007) compare more than favorably. We reiterate that our (out-of-sample) accuracy ratios are very close to those of Duffie et al. (2007).

We are particularly (positively) surprised by the level of our out-of-sample fit. Having no prior evidence from any other research on the accuracy of the doubly stochastic Poisson formulation in modelling default within the context of non-public emerging market companies, we could not have expected the levels of accuracy compared to the models using high-quality data from developed markets. What we lost in the quality of the data, we hoped to recoup in the sample size and the power of the tests. Indeed, we believe our results may have benefited from the size of the sample, as our over 15,000 population of firms modelled significantly outnumber the sample size of either Duan et al. (2012), who use 12,268 firms, and of Duffie et al. (2007), with 2770 firms. Others using the Duffie–Duan model apply it to even smaller samples, as evidenced by, for e.g., the 366 (one sector) companies used in Caporale et al. (2017).

It is also possible that our results are so strong because the defaulters representing non-public companies are so weak financially, compared to the public companies from the developed markets used elsewhere, that the discrimination between the positives and the negatives in our sample is simply easier, no matter what model is used. This said, we have not seen many examples of emerging market data outperforming the results originating from the developed markets default literature.

Our accuracy ratio is not the same as the area under curve, or AUC, also used in the literature. The latter amounts to 0.5 for a random classifier, an equivalent of the accuracy ratio of zero (see Table 3). Hence, the accuracy ratios of 0.8475, 0.8114, or 0.7590, that we record out-of-sample for tau = 1, tau = 2, and tau = 3, respectively, imply an AUC of 0.9238, 0.9057, and 0.8795, respectively. The difference between 0.7590 and 0.8795 in reported statistics is worth noting. The awareness of what measure is used to gauge the goodness-of-fit is important in any research. In the default literature, it is critical.

We also note the relatively low drop in the out-of-sample accuracy of 2–4 p.p., compared to the in-sample statistics. The finding is broadly in line with Duan et al. (2012). As illustrated by Table 17, the drop reported in Duan et al. (2012) is even smaller and amounts to merely one percentage point. As for the statistical inference, neither Duffie et al. (2007) nor Duan et al. (2012) deliver any statistical tests for the accuracy ratios. In contrast, our results are fully documented to be statistically significant at the 1% level using two different statistical tests.

In summary, we would like to reiterate the coherence and the robustness of the results obtained. Firstly, we note again that the signs of covariates, including those for critically important variables, are almost perfectly in line with the theoretical expectations. This is particularly rewarding, as we decided to represent every and each micro factor with two closely related ratios. We have verified this result by running an additional 100 different model specifications, e.g., with and without macro input, with micro variables represented by one or two ratios each, with or without the size control variable, etc. The signs stayed broadly unaffected. Secondly, the coefficient magnitudes confirm the time-dependent relationships. The closer to default, the greater the drop in liquidity and profitability, for example, resulting in ever higher default forward intensities. Thirdly, both in- and out-of-sample accuracy ratios are large and comparable in size with those generated for developed markets and public companies. Fourthly, the accuracy ratios monotonically change with time—the further away from default, the lower the power of the model discrimination. This is precisely what we expect. Fifthly, the accuracy of the model changes in line with the addition (or subtraction) of additional variables to the model. For example, scrapping the double representation of company-specific ratios results in the loss of accuracy ratios of around 3–4 percentage points. Aborting the macro factors cuts AR by another 4–5 percentage points. Finally, all of our replications result in both in- and out-of-sample results that are statistically significant at the 1% level, using two different tests. This all proves the model is far more resilient than traditionally used linear regression specifications to different model specification changes.

Last but not least, we reiterate that the doubly stochastic Poisson model used in this paper is still dependent, by definition, on the doubly stochastic assumption under which firms' default times are correlated only in the way implied by the correlation of factors determining their default intensities. Notwithstanding our highly satisfactory and robust results, these results should be viewed with caution, as no doubt the model implies unrealistically low estimates of default correlations compared to the sample correlations. This has been reported by, for e.g., Das et al. (2007) and Duffie et al. (2000). It should also be remembered that the overlapped pseudo-likelihood function proposed by Duan et al. (2012) used in this paper does violate the standard assumptions, and the implications of this are not immediately clear. Still, the potential biases introduced affect the results of both Duffie et al. (2007) and/or Duan et al. (2012) as much they affect our findings.

5. Conclusions

The doubly stochastic Poisson process has proved to produce very robust default probabilities, not only for public firms operating in the high-quality reporting environment. In this paper, having conducted a meticulous data cross-check, we illustrated that the model is also capable of producing strong and reliable results for low-quality data on non-public firms from capital markets characterized by lower quality governance and transparency standards (Aluchna et al. 2019). Being assured of the model's applicability, we must acknowledge that the wide range of unanswered questions, on top of those already mentioned in the Discussion section, still remains. Below, we present just a few of them in the form of an exemplary, less than complete, checklist, starting with the most obvious (based on the literature).

1. We start with the most popular topic of default literature, i.e., the determinants of the default. What does exactly determine (describe) bankruptcy? In this paper, we selected some micro and macro factors without much attention attached to this issue. This is not to say it is unimportant. Quite the opposite. To know which company-specific attributes and macro influences add most to our understanding of forward default intensity is always rewarding.
2. We acknowledge that Duan et al. (2012), via quite intricate, and sometimes quite arbitrary, in-house developed procedures, make all their input monthly. Given the practical importance of the timely reporting of default probabilities, the rationale for it is self-evident. We wonder how much this approach would affect our results in terms of both the goodness-of-fit and the robustness of our results.
3. In order to capture the dynamics of the covariates, on top of the levels, Duan et al. (2012) introduce the first changes (called trends) in the values of state variables, too. Would that materially change our results?
4. In this research, we identified default as a well-defined legal event within the context of Poland's bankruptcy law ("the declaration of bankruptcy to liquidate the debtors' assets"). Our decision was, to some extent, arbitrary, yet we acknowledge we have collected information on as many as 41 different legal events (starting from various court petitions, numerous different court decisions, to actual declarations of many types of bankruptcy proceedings). Would the model work for different types of "default definitions"? Would it discriminate successfully between event-positive and event-negative firms, regardless of the event definition?
5. In defining survivors, we adopted a very strong criterion: no bankruptcy event of any sort during the sample period. No doubt, other, less restrictive rulings are also possible.
6. Finally, we would like to emphasize the arbitrary nature of the match between the default event date and the date of the financial statement identified as the one that precedes the event—a typical issue in all default research. Simple lagging is an option. Yet, there are many other ways to do it, utilizing, for example, the information on actual filing timing. The timing of the filing may also be used as a proper signal itself (companies that file the accounts with a delay tend to be less financially reliable). The use of this information is certain to be particularly valuable in the context of non-public, low-quality players.

Given the nature of the data (which calls for many cross-checks and interpretations) and the novelty of the methodological approach, the list of various aspects worth researching is practically endless, ranging from detailed analysis of input data to searching for better estimators. Whatever the look of a complete list, it is comforting to know that compiling such a list makes sense in the first place, as we successfully demonstrated in this paper the applicability of the doubly stochastic formulation in the context of emerging market non-public firms producing low-quality inputs.

Author Contributions: Conceptualization, T.B.; Data curation, R.R.; Investigation, T.B. and R.R.; Methodology, T.B.; Writing—original draft, T.B.; Writing—review & editing, T.B. and R.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Science Centre of Poland, grant number 2015/17/B/HS4/02081.

Acknowledgments: We would like to kindly acknowledge the contribution of Bogusław Bławat, Marek Dietl, and Przemysław Krzyk, all of whom were involved at earlier stages of our research.

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

References

Acharya, Viral V., Matteo Crosignani, Tim Eisert, and Christian Eufinger. 2020. Zombie Credit and (Dis-)inflation: Evidence from Europe. *National Bureau of Economic Research*. Available online: https://www.nber.org/system/files/working_papers/w27158/w27158.pdf (accessed on 23 July 2021).

Altman, Edward I. 1968. Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance* 23: 589–609. [CrossRef]

Altman, Edward I. 2020. Covid-19 and the credit cycle. *Journal of Credit Risk* 16: 67–94.

Aluchna, Maria, Tomasz Berent, and Bogumił Kamiński. 2019. Dividend payouts and shareholder structure: Evidence from the Warsaw Stock Exchange. *Eastern European Economics* 57: 227–50. [CrossRef]

Banerjee, Ryan, and Boris Hofmann. 2018. The Rise of Zombie Firms: Causes and consequences. *BIS Quarterly Review September*. pp. 67–78. Available online: https://www.bis.org/publ/qtrpdf/r_qt1809g.pdf (accessed on 20 October 2021).

Beaver, William H. 1966. Financial ratios as predictors of failure. *Journal of Accounting Research* 4: 71–111. [CrossRef]

Beaver, William H. 1968. Market prices, financial ratios, and the prediction of failure. *Journal of Accounting Research* 6: 179–92. [CrossRef]

Beaver, William H., Maureen F. McNichols, and Jung-Wu Rhie. 2005. Have financial statements become less informative? Evidence from the ability of financial ratios to predict bankruptcy. *Review of Accounting Studies* 10: 93–122. [CrossRef]

Berent, Tomasz, Bogusław Bławat, Marek Dietl, Przemysław Krzyk, and Radosław Rejman. 2017. Firm's default? new methodological approach and preliminary evidence from Poland. Equilibrium. *Quarterly Journal of Economics and Economic Policy* 12: 753–73.

Bharath, Sreedhar T., and Tyler Shumway. 2008. Forecasting default with the Merton model. *The Review of Financial Studies* 21: 1339–69. [CrossRef]

Caporale, Guglielmo Maria, Mario Cerrato, and Xuan Zhang. 2017. Analysing the determinants of insolvency risk for general insurance firms in the UK. *Journal of Banking & Finance* 84: 107–22.

Chava, Sudheer, and Robert A. Jarrow. 2004. Bankruptcy prediction with industry effects. *Review of Finance* 8: 537–69. [CrossRef]

Collin-Dufresne, Pierre, and Robert S. Goldstein. 2001. Do credit spreads reflect stationary leverage ratios? *The Journal of Finance* 56: 1929–57. [CrossRef]

Das, Sanjiv R., Darrell Duffie, Nikunj Kapadia, and Leandro Saita. 2007. Common failings: How corporate defaults are correlated. *The Journal of Finance* 62: 93–118. [CrossRef]

DeLong, Elizabeth R., David M. DeLong, and Daniel L. Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 54: 837–45. [CrossRef]

Duan, Jin-Chuan, Jie Sun, and Tao Wang. 2012. Multiperiod corporate default prediction—A forward intensity approach. *Journal of Econometric* 170: 191–209. [CrossRef]

Duan, Jin-Chuan, Baeho Kim, Woojin Kim, and Donghwa Shin. 2018. Default probabilities of privately held firms. *Journal of Banking & Finance* 94: 235–50.

Duffie, Darrell, Andreas Eckner, Guillaume Horel, and Leandro Saita. 2000. Frailty Correlated Default. *Journal of Finance* 64: 2089–123. [CrossRef]

Duffie, Darrell, Leandro Saita, and Ke Wang. 2007. Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics* 83: 635–65. [CrossRef]

Fischer, Edwin O., Robert Heinkel, and Josef Zechner. 1989. Dynamic capital structure choice: Theory and tests. *The Journal of Finance* 44: 19–40. [CrossRef]

Giesecke, Kay. 2004. Correlated default with incomplete information. *Journal of Banking & Finance* 28: 1521–45.

Hamilton, David T., and Richard Cantor. 2004. Rating transition and default rates conditioned on outlooks. *The Journal of Fixed Income* 14: 54–70. [CrossRef]
Hanley, James A., and Barbara J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36. [CrossRef] [PubMed]
Hillegeist, Stephen A., Elizabeth K. Keating, Donald P. Cram, and Kyle G. Lundstedt. 2004. Assessing the probability of bankruptcy. *Review of Accounting Studies* 9: 5–34. [CrossRef]
Jarrow, Robert A., and Fan Yu. 2001. Counterparty risk and the pricing of defaultable securities. *The Journal of Finance* 56: 1765–99. [CrossRef]
Kalbfleisch, John D., and Ross Prentice. 2002. *The Statistical Analysis of Failure Time Data*, 2nd ed. Hoboken: Wiley.
Kavvathas, Dimitrios. 2000. Estimating Credit Rating Transition Probabilities for Corporate Bonds. AFA 2001 New Orleans Meetings. Available online: https://ssrn.com/abstract=252517 (accessed on 23 July 2021).
Lancaster, Tom. 1990. *The Econometric Analysis of Transition Data*. Cambridge: Cambridge University Press.
Leland, Hayne E. 1994. Corporate debt value, bond covenants, and optimal capital structure. *The Journal of Finance* 49: 1213–52. [CrossRef]
Merton, Robert C. 1974. On the pricing of corporate debt: The risk structure of interest rates. *The Journal of Finance* 29: 449–70.
Ohlson, James A. 1980. Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research* 18: 109–31. [CrossRef]
Schoenbucher, Philipp. 2003. Information-Driven Default Contagion. Department of Mathematics, ETH Zurich. Available online: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.139.2714&rep=rep1&type=pdf (accessed on 23 July 2021).
Shumway, Tyler. 2001. Forecasting bankruptcy more accurately: A simple hazard model. *The Journal of Business* 74: 101–24. [CrossRef]
Zmijewski, Mark E. 1984. Methodological issues related to the estimation of financial distress prediction models. *Journal of Accounting Research* 25: 59–82. [CrossRef]

risks

MDPI

Article

Risk Management and Financial Stability in the Polish Public Hospitals: The Moderating Effect of the Stakeholders' Engagement in the Decision-Making

Aldona Frączkiewicz-Wronka [1], Tomasz Ingram [2], Karolina Szymaniec-Mlicka [1,*] and Piotr Tworek [1]

[1] Department of Public Management & Social Science, University of Economics in Katowice, 40-287 Katowice, Poland; afw@ue.katowice.pl (A.F.-W.); piotr.tworek@ue.katowice.pl (P.T.)
[2] Department of Entrepreneurship & Management Innovation, University of Economics in Katowice, 40-287 Katowice, Poland; tomasz.ingram@uekat.pl
* Correspondence: karolina.szymaniec-mlicka@uekat.pl

Abstract: Public healthcare organizations usually operate under significant financial strain and frequently strive for survival. Thus, in most cases, financial stability is a "holy grail" of public healthcare organizations in general and hospitals in particular. The financial stability itself is partly dependent upon the ability to manage risk associated with hospital actions. In the paper, we seek to address the question related to the moderating role of stakeholders' engagement in the relationship between risk management practices and a hospital's financial stability. To answer this question, we designed and carried out empirical research on a sample of 103 out of 274 Polish public hospitals operating at the first-level (closest to the patient). Results show that risk management practices are positively related to financial stability. Hospitals with well-developed risk management practices are better prepared and find appropriate answers to threats, helping them attain financial stability. We also found that stakeholder engagement acts as a moderator of the relationship between risk management practices and financial stability. Research results indicate that with more sophisticated risk management practices, stakeholder engagement in decision-making leads to statistically lower financial stability. On the other hand, high levels of stakeholders' engagement help when risk management practices are underdeveloped.

Keywords: public management; risk management; public hospitals; financial stability; stakeholders' engagement; survey research; Poland

Citation: Frączkiewicz-Wronka, Aldona, Tomasz Ingram, Karolina Szymaniec-Mlicka, and Piotr Tworek. 2021. Risk Management and Financial Stability in the Polish Public Hospitals: The Moderating Effect of the Stakeholders' Engagement in the Decision-Making. *Risks* 9: 87. https://doi.org/10.3390/risks9050087

Academic Editors: Józef Dziechciarz and Krzysztof Jajuga

Received: 23 March 2021
Accepted: 28 April 2021
Published: 6 May 2021

1. Introduction

One of the critical issues that need to be solved in contemporary healthcare organizations is analyzing many diverse and complex interdependencies emerging in the decision-making process, both between interest groups, within these units and with respect to hospitals' relationships with the external environment and emerging risks. Such analysis is essential to identify and address problems related to the effective delivery of health services and the pursuit of financial sustainability.

The shape and scope of health services provided in a given country are a consequence of the adopted healthcare model. Health protection is an essential instrument for creating both the individual ability to function in the labor market and society. It contributes to the state's economic functioning, hence the need for public entities to shape the health protection policy introduced by the European Union (EU from here onwards), especially cohesion policy. We define health policy as various forms of intervening—as a consequence of making systemic choices—in the natural course of events causing health effects. The primary goal of healthcare policy is to ensure citizens' access to healthcare services. Achieving this goal requires the state to design and implement solutions consistent with the practiced economic (the principles of collecting and allocating public funds) and political (implementing the principles of social justice) doctrine, taking the form of the healthcare system. The implementation of citizens' access to healthcare is carried out by the state defining the rules of functioning of entities providing health services and, in particular, their functions, tasks, rules for the provision of services, the required qualifications of the staff, and the method of financing their activities.

Hospitals play a unique role in the healthcare system, mainly due to the scope of tasks and the value of funding allocated to implementing entrusted functions. By providing health services financed from public funds, hospitals are the guarantor of obligations incurred with voters in the economic and social sphere. That explains the importance of the stakeholders and their engagement in the decision-making process. The difficulty in managing hospitals results from the need to reconcile economic efficiency with the social and political consequences of the decisions made, reflected in stakeholders' reactions in the closer and more distant environment that creates the management ecosystem. External and internal factors are controlling such ecosystems (Banoun et al. 2016). The boundaries of the ecosystem, in which stakeholders, through their decisions—supporting or blocking, create risks for hospitals to achieve the desired outcomes (Vargo et al. 2015, 2017), are comprised of the designated by the state norms and business activity principles. Research carried out in the years 2007–2016 allowed for the identification of key stakeholders influencing managerial decision-making processes in Polish public hospitals (Table 1) (Frączkiewicz-Wronka 2018).

Managers identify stakeholders and perceive them as important players in the areas where public organizations operate. All that signifies that making decisions in hospitals that provide health services financed from public funds is more than complicated. In particular, managers have to assess the consequences of not satisfying stakeholders in their understanding of values, which carries numerous risks. Hence, the need to study the relationships between risk management practices, stakeholder management, and financial stability by hospitals providing health services financed from public funds.

Table 1. The identification of key stakeholders in public hospitals in Poland and associated risk of failure to meet their expectations.

Stakeholder	Interest	Impact	Relevance	Identified Risks
Founding body	High level of medical security in a given area; secured provision of health care services; stable financial condition; achievement of statutory objectives; improved image of the organization.	Grants subsidies for provision of services, approves a plan for development of the entity reporting to it.	Attempts are made to take into consideration the expectations and suggestions of a social supervisory board.	Lack of acceptance for actions taken by managers in the hospital and, as a consequence, reduction in financial support and/or boardroom changes in the hospital.
Patients	High quality and availability of medical services; well-developed and modern hospital infrastructure; a comprehensive range of medical services; competent and friendly staff; a variety of medical services on offer.	Their positive feedback is an incentive for and an indicator of future development and a way to attract new patients; patients' preferences determine the performance of the contract; claims may affect the entity's financial condition.	Decisions which are made do not always take into account the expectations of patients' families.	Change of a service and, consequently, a risk that the contract with the NHF may not be completed. Negative feedback, once spread, may damage the organization's positive image.

Table 1. *Cont*.

Stakeholder	Interest	Impact	Relevance	Identified Risks
Ministries (e.g., Ministry of Health, Ministry of Labor and Social Policy)	Tasks performed in compliance with legal requirements (acts and ordinances); adherence to legal standards in the area of public obligations; provision of top quality services, in line with valid regulations and standards; ensured and secured medical services in a given area; an increase in one's own political capital.	Indirect impact through legal regulations, decide about some funds allocated to health care units.	It is important to meet their requirements and perform a contract in compliance with accepted documents, without the need to incur additional costs of service provision.	Withdrawal of funds allocated for operations. Refusal to finance activities planned for the future.
The National Health Fund (NHF)	Correct performance of contractual provisions; a wider range of services; maintaining the right cash flow from provision of services; timely accounting for service provision; furnishing of complete and up-to-date information.	Decides about awarding contracts for provision of services. If a contract is not signed the entity is not able to continue its operations.	The adopted strategy must take into account the legal regulations.	Inability to sign a contract for provision of medical services.
Local government	Availability and high quality of services for the local community; fulfilment of statutory obligations; ensuring highly specialized medical care for inhabitants; pursuing the political interest (health care tends to be one of the main points on the political agenda).	Through a decision-making process related to financial support they approve a specific strategy of the health care unit.	Maintenance of good relationships by meeting the contractual provisions.	Making a decision on replacement of managerial staff. Refusal to grant funds.

Source: Frączkiewicz-Wronka (2018).

2. Legal and Economic Determinants of Hospital Functioning

The fundamental problem that each healthcare system has to deal with is providing funds to finance its activities. In Poland, public expenditure on healthcare amounts to approximately 6.3% of GDP. Lithuania (6.8%), Estonia (6.4%), and Latvia (5.9%) have less or the same amount of funds for healthcare than Poland. The so-called "old union" states spend much more for this purpose, taking, for example, France (11.2%), Germany (11.2%) (Organisation for Economic Cooperation and Development 2020). Poland is one of the countries with a low share of healthcare expenditure in current public spending. In 2017, Poland's public expenditure on health and health care in 2017 amounted to PLN 90.4 billion. It accounted for 4.55% of GDP, while current private expenditure was equal to PLN 39.7 billion and accounted for 2% of GDP (Główny Urząd Statystyczny 2019). The primary source of healthcare financing is comprised of compulsory health insurance contributions; hence, the institution managing them, the National Health Fund (in polish: NFZ), is a crucial stakeholder for every entity providing health services.

On the basis of GUS reports, current expenditure on healthcare in 2017 according to functions, points out that most funds are spent on medical services, 58.3% (hospital plus outpatient). Data presented in Table 2.

Table 2. Current expenditure structure on healthcare in 2017 in Poland (Główny Urząd Statystyczny 2019).

Expenditure Structure on Healthcare in 2017	Structure
Government schemes	10.4%
Compulsory contributory health insurance schemes	59.1%
Voluntary health insurance schemes	5.7%
NPISH financing schemes	0.8%
Enterprise financing schemes	1.2%
Household out-of-pocket payment	22.8%

It is followed by spending on medical supplies, including drugs (22.7%) and long-term healthcare (6%). The lowest expenses are incurred on rehabilitation services (4.8%), prevention and public health (2.3%), and healthcare management and administration (1.8%). The share of the public and private sectors in financing individual healthcare functions in Poland depends on the type of service. For example, the "Healthcare" segment was financed by public funds in 78% and from the private sector is 22%. Current expenditure on healthcare in 2017 points out that the most considerable amount (39.4%) of current healthcare expenditure goes to hospitals, mainly the so-called general. It was followed by spending on outpatient healthcare facilities (27.1%) as well as retailers and other suppliers of medical goods (22.4%), mainly for pharmacies for drug reimbursement. The share of the public and private sectors in financing individual healthcare providers varied depending on the provider type. For example, 'hospitals' were financed in 94.1% by public funds and just in 5.9% by the private sector (Główny Urząd Statystyczny 2019).

The hospital is the most important but also the costliest entity in any healthcare system. Due to their function in the healthcare system, these entities often and quickly fall into debt, pictured in Figure 1.

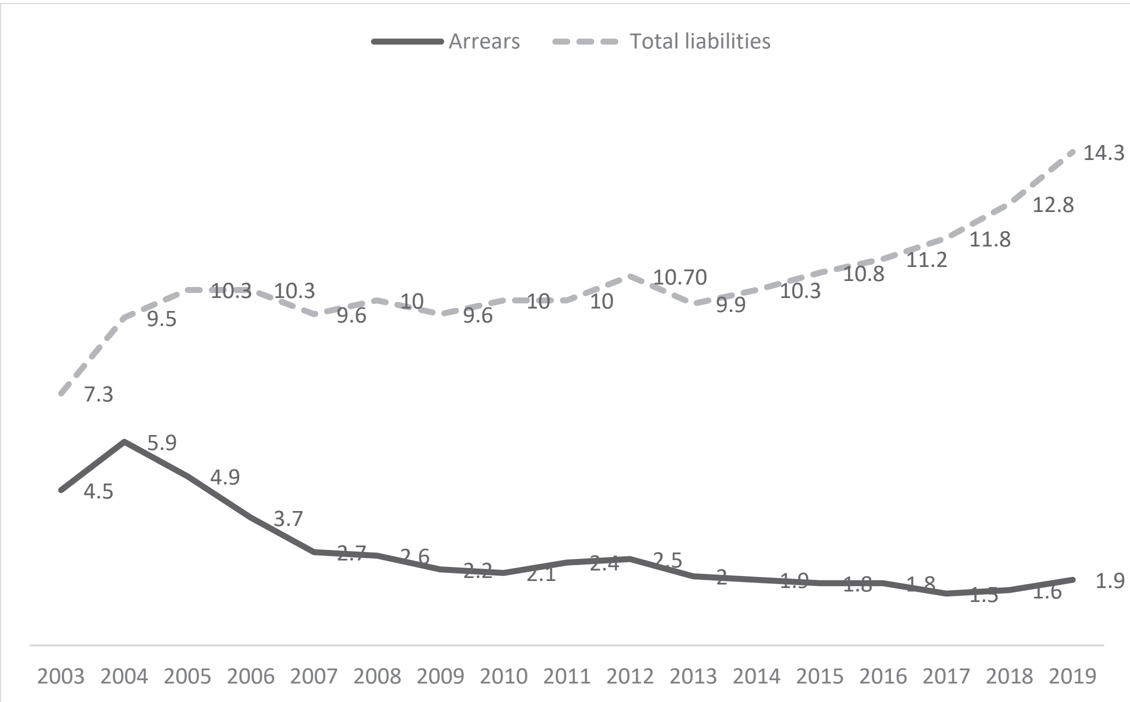


Figure 1. Evolution of hospitals' liabilities in billions of PLN.

Table 3 presents task and macroeconomic environment in which hospitals operate in Poland were shaped mainly due to the legal regulations introduced in 1991–2017.

Table 3. Analysis of the legal changes in the functioning of hospitals.

1991	Healthcare Units Act from 30 August 1991 [Ustawa z dnia 30 sierpnia 1991 r. o zakładach opieki zdrowotnej], Polish Journal of Laws 1991 No. 91, act: 408.
• Changing the source of financing and organization of healthcare entities. • Transferring ownership rights of the independent public healthcare institutions (in polish SPZOZ) to the local government. • Creating conditions for a gradual departure from hospital administration towards active hospital management. • Initiating the separation of a group of managers purposefully trained for management in the healthcare system's entities.	
1997	Common Health Insurance Act from 6 February 1997 [Ustawa z dnia 6 lutego 1997 r. o powszechnym ubezpieczeniu zdrowotnym], Polish Journal of Laws 1997, No. 28, act: 153.
• Departing from the budget system of healthcare financing and transitioning to a hybrid system, which was based on budget and insurance. • Decentralizing the financing as a consequence of the emergence of regional and industry Healthcare Funds (Kasa Chorych) as a paying institution. • Stimulating the growth of competitiveness in the market of public entities as a consequence of limiting the possibility of concluding contracts by the Healthcare Funds only to public entities (SPZOZ). • Creating the additional income sources as a subsidy from the state budget, which was intended solely for the implementation of health policy programs commissioned by the state.	
2003	Common Health Insurance in National Health Fund Act from 23 January 2003 [Ustawa z dnia 23 stycznia 2003 r. o powszechnym ubezpieczeniu zdrowotnym w Narodowym Funduszu Zdrowia], Polish Journal of Laws 2003, No. 45, act: 391.
• Centralizing the financing as a consequence of the liquidation of Healthcare Funds and establishing the National Health Fund (NFZ), including the voivodeship branches (NFZ OW). • Introducing the obligation to prepare healthcare plans for individual voivodeships. • Weakening of the competitive behavior and consolidation of mechanisms promoting low economic efficiency in hospital operations.	
2004	Publicly Funded Healthcare Services Act from 27 August 2004 [Ustawa z dnia 27 sierpnia 2004 r. o świadczeniach opieki zdrowotnej finansowanych ze środków publicznych], Polish Journal of Laws 2004, No. 210, act: 2135.
• Improving the quality of health services as a consequence of conditioning of receiving funds on the quality of an offer (confirmation of desirable infrastructural conditions and human resources) prepared by the entities applying for contract on realizing health services from public funds. These funds were at the disposal of NFZ (Rudawska 2011). • Admitting public and non-public entities to participate in competitions for contracts to provide healthcare services on the same terms.	
2010	Healing Activities Act from 15 April 2011 [Ustawa z dnia 15 kwietnia 2011 r. o działalności leczniczej], Polish Journal of Laws 2011, No. 112, act: 654.
• Creating a legal possibility of commercialization of the healthcare facilities. • The government prepared a system of incentives based on the repayment of a partially paid by the local government debt of independent public healthcare institutions (up to the sum of public-law liabilities) (Węgrzyn 2012).	
2017	The change of the Publicly Funded Healthcare Services Act from 23 March 2017 [Ustawa z dnia 23 marca 2017 r. o zmianie ustawy o świadczeniach opieki zdrowotnej finansowanych ze środków publicznych], Polish Journal of Laws 2017, act: 844.
• Launching the so-called hospital networks as a consequence of changes in the rules of financing health services as part of hospital treatment, and partly also as a part of outpatient specialist care.	

The launch of the primary hospital healthcare system (the so-called hospital network) in 2017 directly impacted the hospitals' functioning (Dubas-Jakóbczyk et al. 2019). As part of the security system, hospitals were divided into: 1st-degree hospitals, 2nd-degree hospitals, 3rd-degree hospitals, and additional into oncology, pulmonary, pediatric, and nationwide hospitals. The changes aimed to optimize the number of specialist departments and enable better coordination of inpatient and outpatient services. The costs of services provided are settled based on contracts concluded within the hospital network in a flat-rate form. The lump-sum amount depends on the number and structure of services provided and reported by the service provider in the period preceding the new contract. The lump-sum means that a given hospital receives a budget for hospital services, but it also receives funds for outpatient specialist care carried out by adequate clinics, services in the field of therapeutic rehabilitation, and even night and holiday healthcare services. According to the legislator's intention, the introduced solution was aiming to: (1) improve the organization of healthcare services provided by hospitals and hospital clinics and improve patients' access to specialist treatment in hospitals, (2) guarantee an appropriate level, as well as continuity and stability of hospital financing, (3) limit the phenomenon of dispersion of public funds allocated to the financing of guaranteed services, (4) ensure a certain stabilization of the continuity of financing the activities of medical entities included in the network, and (5) facilitate the management of hospitals.

The hospital's operation effects, both in the economic terms and the evaluation of access to health services guaranteed by the state, are the subject of constant assessment by various stakeholders. The need to satisfy stakeholders contributes to problems with achieving financial stability, and the lack of it is an impulse to look for solutions that would allow improving the problematic financial condition of Polish hospitals.

3. Risk and Stakeholders in the Decision-Making Processes in Public Hospitals

Risk is inextricably linked with starting a business, and the ability to assess it by managers influences the results achieved by organizations operating in various sectors (Power 2016; Raczkowski and Tworek 2017). In terms of terminology, risk is considered a prediction of the possible likelihood of a loss of resources or not obtaining income than the assumption made (Mennen and Van Tuyll 2015). Since risk is part of the decision-making process, it requires explicit recognition, identification, monitoring, and management. The theoretical framework for defining risk has its sources in research trends related to the organizations' functioning, especially in the aspects of examining the determinants of making strategic choices (Fone and Young 2007; Mennen and Van Tuyll 2015; Power 2016; Raczkowski and Tworek 2017; Young and Tippins 2001) and the functioning of the organization in conditions of uncertainty (Collins 1992). Practically, the risk relates to organizations' management functions, especially the ability to anticipate certain events or achieve expected or undesirable outcomes (Tworek 2016; Young and Tippins 2001). The key to the operationalization of the concept of risk is the assumption that it is quantifiable (i.e., measurable). Keeping in mind the nature of the organizations, we carried out our research, focused on describing hospitals' operation risk.

In the practice of the organization's functioning, it is essential to correctly determine the specific and non-specific risks for a given type of activity and organization. In healthcare organizations, the most tragic consequence of risk is the loss of the patient's life (Carroll 2009; Kwiecińska-Bożek 2018). The American Society for Healthcare Risk Management (ASHRM) assumes that health-specific risks relate to the deterioration of a patient's health and/or patient safety (Carroll 2009). According to the World Health Organization (WHO), the main risk category in healthcare is the risk of in-hospital infections (World Health Organziation 2011).

In the theory of healthcare economics, both the threat and the opportunity mean a different dimension of risk effects, but ultimately the effects of risk on the organization assume an economic dimension (Sohn 2016). In general, risk should be examined through the lenses of the many sub-categories that make up the overall picture of economic risk categories in hospitals. This issue relates uniquely to the methodological perspective of economic risk management in hospitals' operations (Kavaler and Spiegel 2003; Kolluru et al. 1996; Roberts 2002). Correctly performed quantification/measurement of risk may be reflected in the hospital's economic calculus because its incorrect estimation in the decision-making may have economic consequences for the organization. This aspect of considerations relates to the financial risk as a critical issue in achieving economic efficiency by the hospital (McCue and MCluer 2008; Ozcan and McCue 1996).

Hypothesis 1 (H1). *Risk management practices are positively related to financial stability.*

Risk management in healthcare should be considered both from subjective and objective points of view. This issue relates especially to stakeholders interested in hospital risk management (Adil 2008; Bennet et al. 2010; Frączkiewicz-Wronka 2018). On the other hand, risk management, considered from the subject's point of view, is related to the processual approach connected with the universal division of risk into three subprocesses: identification, analysis, assessment, and reaction to risk (Elleuch et al. 2014, The Orange Book 2004). In addition to research of a theoretical or empirical nature, professional literature exhibits the utilitarian aspect of risk management theory (Carroll 2009; Kavaler and Spiegel 2003; Kolluru et al. 1996; Roberts 2002). In theory and practice, emphasizing the need to monitor and control the risk management process (Beck de Silva Etges et al. 2016; The Orange Book 2004; Tworek 2016). The empirical research results in public organizations indicate that it usually takes place as part of the control and internal audit (Bakalikwira et al. 2017; Chambers et al. 2017; Sarens et al. 2010). Hospital risk management is perceived as a highly specialized sub-discipline of knowledge (Hood et al. 2003).

Research on financial risk management in hospitals focuses mainly on operational, investment, and financial activities, resulting from applicable legal regulations. The critical issue in the research on financial risk in hospitals is to define the common denominator, i.e., the value around which the risk should be assessed (the value of reference) (Tworek 2016). Concerning the achievements of the theory of finance (see more Jajuga 2019; Skoczylas and Waśniewski 2014; Wędzki 2012), in this context, reference should be made to the three financial goals of the organization, i.e., financial liquidity, risk, and profitability (Feng 2011; Stroh 2005). Public hospitals in Poland are assessed, among others, by business profitability rations (as of the Decree of the Minister of Health from 12 April 2017, on economic and financial indicators necessary to prepare an analysis and forecast of the economics and financial situation of independent public healthcare institutions). The indicators used to assess public hospitals are as follows: net profitability ration, operating profit ratio, and asset profitability ratio. The profitability analysis is complemented by the efficiency, liquidity, and debt analysis. The legal regulation of hospitals' economic and financial assessment by indicating obligatory indicators was the Ministry's response to the demands of the environment regarding the possibility of comparing the results achieved by independent public health care institutions (Zaleska 2017). Financial liquidity is the inability to meet current liabilities, and the risk is the possibility of not achieving the intended financial results, while profitability means the ability to achieve positive financial results. These three categories, between which there are interrelationships and dependencies, create an overall picture of the risk related to hospitals' lack of financial stability, with the crucial issue being the lack of stable sources of financing in Polish hospitals. This issue is critical for the balancing of the hospital's operations. This problem relates to the systemic risk category (Agnew et al. 2006).

Hypothesis 2 (H2). *Stakeholders' engagement in decision-making is positively related to financial stability.*

Financial risk in healthcare entities' activities, considered a macroeconomic category, relates to the proper valuation of services in all types of services financed by the NFZ. In turn, the risk considered in the microeconomic category means that a public hospital budget is not correctly estimated. Considering the main problem of healthcare in Poland is the insufficient financing level, Polish public hospitals' main challenge is the lack of financial stability, resulting from debt and financial liquidity. In management practice, stakeholders are interested in hospital profitability, i.e., achieving the desired values of financial ratios indicated in the Decree of the Ministry of Health from 12 April 2007. These indicators might be helpful in the evaluation of the healthcare unit by founding bodies, payer, investors, banks, and other stakeholders who are related to a particular healthcare organization (Zaleska 2017). That means hospitals should be interested in minimizing their business risk and maximizing the benefits expressed as a cash equivalent.

On the other hand, stakeholders will be interested in increasing the cash flow as a common denominator for determining entire organizations' risk (Tworek 2016). Therefore, stakeholders' engagement in the financial decision-making process (Burke and Demirag 2017; Lin et al. 2017; Tseng et al. 2020) is a prerequisite for effective hospital financial management (Ozcan and McCue 1996; Zheng et al. 2019). Its absence may lead to the hospital's liquidation by the founding bodies who play the primary stakeholder's role. Moreover, this is the most severe financial effect of the occurrence of economic risk.

According to Nieszporska (2012, p. 151), "(...) the risk categories identified in the activities of Polish public hospitals focus on five aspects/problem areas: (1) significant—equated with cost-effectiveness or the possibility of financial losses, (2) operational indicators—understood as a system of facility's ease of adaptation to changes, with particular emphasis on in-hospital rules, (3) internal control—understood as the evaluation of the control system in each organizational unit of the hospital, (4) the quality of management—represented by all activities related to the organization and modification of hospital structures and the transfer of information, (5) external factors—understood as comprehensive effectiveness in introducing changes to the hospital environment".

Hypothesis 3 (H3). *Stakeholders' engagement in the decision-making moderates the relationship between risk management practices and financial stability.*

The conducted considerations allow for the formulation of a hypothesis that examining the relationship between the issue of stakeholder management and risk management in public hospitals as well as financial stability considered in terms of the financial condition of the organization is an important issue, the results of which may affect the management process of a public hospital (Beck de Silva Etges et al. 2016; Dixit 2017; Mahama et al. 2020). The context of considerations regarding stakeholder management in the public sector, taking into account risk-related issues, is extensive (Borraz 2007; Dansoh et al. 2020; Hunt 2010; Klinke and Renn 2012; Professeure 2004; Rixon 2010). In general, authors in the literature postulate the implementation of the risk management principles to organizations providing health services outlined in the New Public Management (NPM) model (Beck de Silva Etges et al. 2016; Flemig et al. 2016; Hinna et al. 2018; Krewski et al. 2007; Li et al. 2020; Osborne et al. 2020; Oulasvirta and Anttiroiko 2017; Rana et al. 2019).

The scientific literature indicates that in private organizations, the decision-making processes are much smoother and calmer, while public organizations experience more turbulence, breaks, recirculation, and conflicts (Nutt 1999). As Nutt (2005) emphasized, the decision-making process in public organizations requires recognizing rulers' opinions, understanding the mandates and obligations of the organization, and balancing users' needs. The specificity of public organizations' operation also entails a growing demand for opening the external participation process. There is an increasing need to set public expectations about how services are delivered in public organizations and involving more people in the decision-making process. Simultaneously, the legibility of decision-making criteria decreases, and more time is needed to make decisions; there is also a need to consider "soft" criteria and the ones' that will ensure equality in access to services. Frączkiewicz-Wronka (2012, p. 42) argued that "(...) decisions in public organizations are often made in a forced manner and far from rational objectivity, because—as practice shows—reason requires the manager to make political choices instead of looking for economic rationality in solving many tasks that are encountered by the organization". Dillon et al. (2010, p. 236) underlined that "(...) in the end political bargaining appears to be the main determinant of the decision outcome".

One of the distinguishing features of management in public organizations is the existence of many stakeholders who influence the decision-making process by using, among other things, mechanisms of regulation, cooperation, setting directions for changes, legitimation, and control (Gomes et al. 2010). Literature defines a stakeholder as a person, a group of people, or as organizations affected by the organizations' functioning or the effects of its functioning. Bryson (2004) believed that attention to stakeholders is essential in the entire process of strategic management because success in a public organization depends on satisfying key stakeholders. At the same time, satisfaction signifies meeting needs that are perceived as valuable in the stakeholder's individual hierarchy. As Rainey (2003) claimed, public institutions arise and live by satisfying the interests of those influential enough to maintain the political *raison d'être* of the organization and secure the resources that flow in with it. Therefore, public organizations have a special responsibility towards their stakeholders, and this means for managers an obligation to take into account their expectations in the decision-making process. It should also be remembered that stakeholders are interested in both the decision-making process and its results (Osborne et al. 2014). Public value for stakeholders can only be created with the support of the organization's key stakeholders (Langrafe et al. 2020; Williams and Shearer 2011). Due to the unique role of stakeholders in the functioning of public organizations, managers must take into account in the decision-making process the effects that these decisions have on the benefits and losses incurred by individual stakeholders (Johnsen 2015), which is facilitated by the use of the participation process of key stakeholders in the decision-making process (Quick and Bryson 2016). Including stakeholders in the decision-making process allows for obtaining a larger pool of information, increases the legitimacy of decisions made, and improves their quality (Beierle 2002; George et al. 2016). At the same time, the power of veto or resistance to change is reduced (Edelenbos and Klijn 2006). As Elias (2019, p. 313) claimed, stakeholders' engagement "can address problematic situations holistically and give due regard to competing interests". Although the need to involve stakeholders in decision-making processes is becoming more transparent, the process is not free from problems. One of the most important is the necessity to engage resources, mainly financial ones, necessary to implement the participation process (McEvoy et al. 2019). The necessity to incur the high costs of organizing the participation process may negate its benefits. Schalk (2015) also pointed to other problems that arise in stakeholders' engagement in the decision-making process—too much information increases the complexity of the problem, and the time needed for the decision-making process is also longer.

Stakeholders' engagement requires the creation of appropriate conditions for the implementation of an effective participation process, including determining the purpose of stakeholders' engagement, identifying limitations, identifying stakeholders, determining the degree of engagement, ensuring the appropriate quality of the engagement techniques used, informing stakeholders about their role in the process, as well as monitoring the activity of individual stakeholders. It is also vital to present stakeholders to the extent to which they influenced the decision made, the effects of their engagement, which will motivate them to participate in subsequent projects (Tončinić et al. 2020).

Reed et al. (2018, p. 9) indicated that "(...) different modes of engagement are possible, and typically lie along an information or knowledge exchange continuum, from approaches based more on one-way flows of information and knowledge to publics and stakeholders (communication mode) and seeking feedback from publics and stakeholders (consultation mode) to more two-way knowledge exchange and joint formulation of goals and outcomes (more deliberative and coproductive modes)".

In healthcare, patient engagement in decisions regarding the treatment process is relatively well recognized; however, the problem of broader stakeholders' engagement in decision-making processes regarding the management of medical entities or the healthcare system's functioning enjoys less research interest (Malfait et al. 2018; McCarron et al. 2019). Simultaneously, as noted by Petkovic et al. (2020), although healthcare organizations have many stakeholders, the research focus is mainly on the patient and public engagement at the system level. In particular, in the healthcare literature, stakeholders' engagement should help cope with social and economic changes such as increasing healthcare demand and higher patient expectations, considering budget constraints. It builds trust in the healthcare system and engages communities and individuals in healthcare (Cleemput et al. 2015).

One way to engage stakeholders in the decision-making process is to create formal advisory bodies—the so-called stakeholder committees (Malfait et al. 2017, 2018). Analyses conducted by Malfait et al. (2017) with a team on the functioning of the stakeholder committees in Belgian hospitals indicated that the success factors for the actual stakeholder participation in the decision-making process are: (1) close cooperation with the management board, (2) focusing on the operational level of the activity as being more practical and closer to patients than the strategic level, (3) transferring greater autonomy to the stakeholder committee activity, also by enabling the choice of the topics taken—as well as—(4) enabling stakeholders to prepare for decision-making, e.g., by sharing materials.

The issue of stakeholders' engagement was also dealt with by McCarron et al. (2019). They noticed a solid need to build stakeholders' capacity and competence to participate in the decision-making (McCarron et al. 2019). That is particularly important because stakeholders' engagement in the decision-making process raises the issue of knowledge that individual stakeholders have, reflected in a stronger medical professional's position than patients (O'Shea et al. 2019). As a consequence of the synthesis of the literature review, Djellouli et al. (2019) indicated that a recurring conclusion from the research is the belief of stakeholders that although they contributed to the activities undertaken, they did not influence the decision-making process because managers made decisions. The qualitative research carried out by Szymaniec-Mlicka (2017) shows that the directors of hospitals in Poland do not actively engage the hospital's stakeholders in the decision-making process, treating such activities more as an unfortunate necessity. Usually, actions towards stakeholders are limited to informing them about the decisions made. However, there is a trend among directors—if they engage stakeholders, they are more likely to take action concerning internal than external stakeholders. Research results of Cleemput et al. (2015) were aimed at identifying the benefits and risks related to stakeholders' engagement in the decision-making process. They pointed out the benefits of engagement could be "increasing awareness among the general public and patients about the challenges and costs of healthcare and enriched decision processes with expertise from patients' experience. (. . .) Subjectivity, insufficient resources to participate and weigh on the process, difficulties in finding effective ways to express a collective opinion, the risk of manipulation, lobbying or power games of other stakeholders" were identified as potential risks (Cleemput et al. 2015, p. 447).

Jansen et al. (2018) developed a checklist of 29 questions relating to critical stages of stakeholders' engagements in setting health priorities. As key areas, they included what follows: (1) proactively identifying potentially adversely affected stakeholders, (2) comprehensively including them in the decision-making process, (3) ensuring meaningful participation, (4) communication of recommendations or decisions, and (5) the organization of evaluation and appeal mechanisms (Jansen et al. 2018).

On the other hand, Norris et al. (2017) focused on analyzing how the hospital stakeholders define engagement. Research has shown that stakeholders define engagement similarly, as "(. . .) an active and committed decision-making about a meaningful problem through respectful interactions and dialog where everyone's voice is considered" (Norris et al. 2017, p. 1).

Wortley et al. (2016), using a literature review, identified the determinants of the choice of the method of stakeholders' engagement in the decision-making process regarding health technology assessment, which included: perceived complexity of the policy-making issue, perceived impact of the decision, transparency, and opportunities for public engagement in governance, time, and resource constraints. "The influence of these factors vary depending on the context, indicating that a one size fits all approach to public engagement may not be effective" (Wortley et al. 2016, p. 872).

4. Research Methods and Way of Data Collection

We commenced the research by analyzing the literature in the EBSCOhost, Emerald Management, Science Direct, Scopus, Web of Science, and SpringerLink databases. The bibliometric analyses covered the period 1978–2017. We used the following keywords in the search process: decision-making in public hospitals, risk management in public hospitals, identification, and stakeholder management in public hospitals. The literature synthesis made it possible to formulate the research problem, pose research questions and statistical hypotheses emerging from them, design a research model, and prepare a research questionnaire.

Based on the literature review, we formulated three research hypotheses. Figure 2 illustrates relationships between main research constructs.

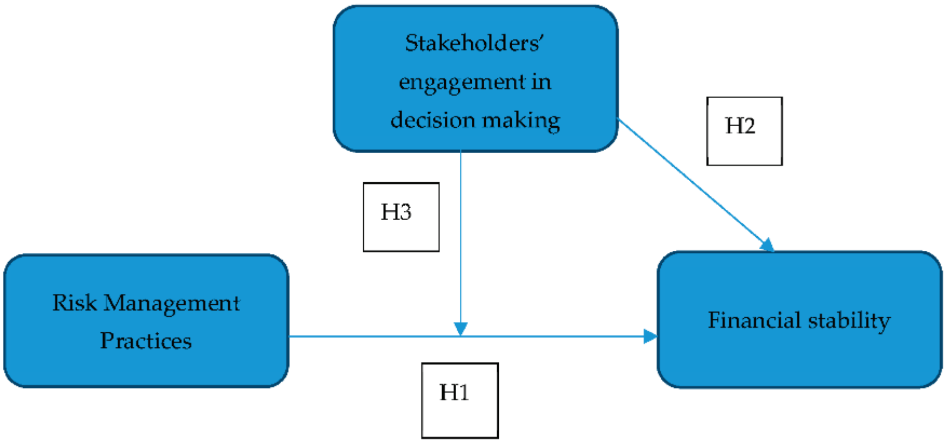


Figure 2. Hypothesized relationships between financial stability, risk management, and stakeholders' engagement in decision-making. Hypothesis 1 (H1). Risk management practices are positively related to financial stability. Hypothesis 2 (H2). Stakeholders' engagement in decision-making is positively related to financial stability. Hypothesis 3 (H3). Stakeholders' engagement in the decision-making moderates the relationship between risk management practices and financial stability.

In the next step, we prepared a measurement tool that consisted of 3 scales. The questionnaire contained questions examining the organizations' financial stability, stakeholders' engagement in decision-making, and risk management in the organization. We gathered the answers on the 7-point Likert scale. Additionally, we asked respondents about the hospital geographical location, the size of the contract with the NFZ, the gender and position of the respondent.

To measure the financial stability (latent variable), we used a 3-item long scale (Cornforth 1978; Snow 1992; Wronka-Pośpiech 2014) (Cronbach's alpha = 0.865). Our respondents were asked to answer subsequent questions: "Our hospital has diversified sources of income, and that guarantees us financial safety,"; "Our hospital can acquire sufficient funds necessary to fulfill its strategy,"; and "Our hospital has sufficient liquid financial resources to handle the short decrease in our incomes." Exploratory factor analysis (KMO = 0.724; Bartlett's test: approx. Chi-square = 146,322 with 3 degrees of freedom $p = 0.000$) carried out using the principal components analysis extraction method with rotation Varimax informed us that all three items load significantly (loading strength from 0.879 to 0.912) to a single metavariable that explains nearly 80% of the variance.

Risk management practices (latent variable) were investigated using the tool employed in a similar research carried out in public organizations from Belgium by Sarens et al. (2010) (we adapted the original scale, and the measurement was done using a 5-item long scale (Cronbach's alpha = 0.857). Our respondents were asked to answer subsequent questions: "In our hospital, we understand that the lack of reaction to threats or risks signifies the loss of resources important for our functioning"; "In the hospital, we have procedures indicating how to react to incoming threats in diverse fields of our activity"; "In the hospital, we have employees trained how to manage risk or threats"; "In the hospital, we have employees responsible for managing threats"; "In the hospital, we have procedures indicating how to react for incoming threats or risks from external stakeholders". Exploratory factor analysis (KMO = 0.750; Bartlett's test: approx. Chi-square = 333,673 with 10 degrees of freedom and $p = 0.000$) carried out using principal components analysis revealed that all the items load to single metavariable (loading strength ranging from 0.696 to 0.939), explaining 66.30% of the variance.

Stakeholders' engagement in decision-making perceived as latent variable, was measured using a 6-item long scale adapted from Amaeshi and Crane (2006) (Cronbach's alpha = 0.860). The decision-making processes refer to strategic aspects of hospitals operation, and the engagement refers to the process of informing and asking for the opinion of stakeholders during scheduled meetings, by email, etc. We left the understanding of the engagement to the respondents. The sample question was as follows: "Key external stakeholders are encouraged to participate in the decision process at every stage of newly undertaken actions" or "We analyze circumstances, methods, and results of engaging external key stakeholders in the decision process". Exploratory factor analysis (KMO = 0.806; Bartlett's test: approx. Chi-square = 357,816 with 15 degrees of freedom and $p = 0.000$) carried out using principal components analysis revealed that all the items load to single metavariable (loading strength ranging from 0.524 to 0.885), explaining 62.37% of the variance.

We started our research in January 2018 by testing the questionnaire on a sample of 31 employees of healthcare entities, students of postgraduate studies in Management in Healthcare at the University of Economics in Katowice and the Medical University of Silesia. Apart from filling in the questionnaire, the respondents were also asked to submit any comments on the content or formulating the questionnaire's questions.

The primary data collection stage was carried out from May 2018 to December 2019 by the Research and Development Centre (from here onwards: RDC) at the UE Katowice (cbir@ue.katowice.pl). The RDC started gathering data by sending traditional mail to all first-level hospitals qualified to the "hospital's network" in Poland. Then, an employee of the center called the person indicated as the respondent (director or deputy director or chief accountant), asking to complete and return the questionnaire to the university's address. It should be emphasized that collecting data in hospitals is difficult as hospital managers are reluctant to spend their time-sharing information.

In our study, we used purposive sampling. As of October 2017, in Poland, 594 medical facilities are qualified for the so-called "network of hospitals" because they meet the statutory conditions, and NFZ guarantees lump-sum financing for these facilities. The hospitals' network was expanded to include medical facilities that, for at least the last two years (2015–2017), provided services as part of an admission room or hospital emergency department under a contract with the NFZ, and have specific hospital departments listed in the Act. As a result, 355 Polish hospitals remain outside the network, 16 of which are public institutions, and the rest are private hospitals.

We aimed at studying only first-level hospitals (274 in total). Such a decision was recognized as true that, due to the territorial range of their activities, they secure the basic needs of residents in the field of residential treatment. First-level hospitals operate in each of the 16 voivodships in Poland and are arranged in such a way as to provide citizens with the best possible access to hospital health services. By definition, these are hospitals operating in the districts, less often in the commune. They constitute the most homogeneous group of entities in terms of the ownership structure (founding entity—district/commune), the hospital wards' medical profile, the scope of services provided, and financing sources. According to the research conducted in 2013–2018, such hospitals operate in a task environment dominated by similar stakeholders (Austen and Frączkiewicz-Wronka 2018). However, in various conditions, depending on the local government units, the characteristics of the interest, influence, and importance of individual stakeholders may vary.

Due to the population's size (hospital register with contact details downloaded from the Ministry of Health website), we decided to send the questionnaire to all first-level hospitals meeting the criterion of a medical facility qualified for the network. Detailed sample selection and composition are presented in the Table 4.

Table 4. Sample selection and composition.

	Voivodeship	Sampling Frame—Number of the First-Level Hospitals in Voivodeships	Number of Received Questionnaires	Number of Discarded Questionnaires	Number of Questionnaires Included in Analyses
1	Lower Silesia Province	20	15	7	8
2	Kuyavian-Pomeranian Province	16	8	5	3
3	Lublin Province	18	7	5	2
4	Lubuskie Province	10	5	4	1
5	Łódź Province	15	15	0	15
6	Lesser Poland Province	11	9	5	5
7	Masovian Province	37	17	4	13
8	Holy Cross Province	8	6	2	4
9	Pomeranian Province	11	8	3	5
10	Podkarpackie Provnice	12	9	2	7
11	Podlasie Province	14	8	4	5
12	Opole Province	12	4	2	2
13	West Pomeranian Province	15	8	3	5
14	Greater Poland Province	24	12	3	9
15	Warmia-Masuria Province	19	8	1	7
16	Silesia Province	32	16	2	12
	Together	274	155	52	103

In the collecting information phase, we obtained 155 responses, of which 103 fully completed ones (containing all the answers required in the form) were qualified for further statistical analyses. Therefore, the effective sample amounts to 37.59% of the sampling frame.

5. Research Results

To test our research hypotheses, in the first step, we employed descriptives of financial stability, risk management practices, and stakeholders' engagement (see Table 5).

Table 5. Descriptives.

Constructs	Financial Stability	Risk Management Practices	Stakeholders' Engagement
Mean	3.6893	5.6951	5.2994
Std. Deviation	1.47474	0.97612	0.94695

N = 103.

Analysis of Table 4 reveals that, on average, hospitals in the sample tend to assess their financial stability below the mid-point of the scale (mean = 3.69) with a relatively significant standard deviation equal to 1.47. On the other hand, hospitals are assessing much higher risk management practices in the unit (mean = 5.69) and stakeholders' engagement (mean

= 5.30). Respondents also report many coherent levels of risk management practices and stakeholders' engagement (standard deviation in both cases is lower than 1).

To fully understand the relationships between studied constructs, we further used structural equation modeling in Mplus 8.2 for Mac. For this purpose, we estimated three models, the first (model 1) with relationships between two primary constructs assuming the influence of risk management practices on financial stability; the second, with added the influence of stakeholders' engagement on financial stability (model 2); and third, in which we additionally account for the interaction between risk management practices and stakeholders' engagement. In all three estimations, we treated constructs as latent and first-order reflective; however, we were forced to use random type analysis due to latent constructs interaction in the model in the third case. Thus, in the third model, we cannot supply model fit statistics other than the Akaike Information Criterion (Hooper et al. 2008). The results are presented in Table 6.

Table 6. Model estimation results.

Model	Model 1	Model 2	Model 3
	The Effect of Risk Management Practices on Financial Stability	The Effect of Risk Management Practices and Stakeholders' Engagement on Financial Stability	The Effect of Risk Management Practices on Financial Stability Moderated by Stakeholders' Engagement
MODEL FIT STATISTICS			
RMSEA	0.074	0.092	-
CFI (Compound Fit Index)	0.980	0.913	-
TLI (Tucker–Lewis Index)	0.968	0.889	-
Akaike Information Criterion (AIC)	2304.632	3838.519	3834.898
r^2	0.060	0.110	0.121
MODEL ESTIMATION RESULTS			
Independent variables (IV)	Estimate β (S.E. σ)	Estimate β (S.E. σ)	Estimate β (S.E. σ)
Risk management practices	0.572 (0.269) *	1.003 (0.465)	0.265 (0.253)
Stakeholders' engagement	-	−0.623 (0.619)	−0.221 (0.224)
Risk management practices × stakeholders' engagement (the interaction)	-	-	−0.207 (0.088)
Constant	2.230 (0.440)	2.106 (0.431)	0.904 (0.061)

Note: * means that statistically significant parameters at $p < 0.01$ are highlighted.

Analysis of the estimated leads to several observations. According to Hooper et al. (2008), the first model is moderately fitted, with RMSEA lower than 0.074 and CFI and TLI both significantly above the 0.9 cut-off line; however, it explains only 6% of the variability financial stability ($r^2 = 0.06$). The second model is significantly worse fitted with RMSEA (root mean square error of approximation) equal to 0.092 (0.08 and below as the acceptable although moderately fit indicator), and CFI slightly above the 0.9 cut-off line (equal to 0.913) and TLI slightly below it (equal to 0.889). In the third model, we were unable to provide fit statistics; however, AIC is lower than in model 2 (AIC of model 3 = 3834.898 vs. AIC of model 2 = 3838.519), which signifies a small improvement in the model fit (the decrease of AIC coefficient signifies improved fit (Hooper et al. 2008). Both models 2 and 3 explain a higher percentage of the variability of financial stability than model 1 (r^2 coefficient equal to 0.11 and 0.121, respectively), although it still accounts for a small part of changes in the level of financial stability.

Referring to Hypothesis H1, in the first model, risk management practices are a significant predictor of financial stability, according to the structural equation modeling estimation. The coefficient signifies that to higher perceived financial stability (coefficients in the first model: $\beta = 0.572$; $\sigma = 0.0269$). In the second model, after accounting for stakeholders' engagement, the relationship is still significant (coefficients in the second model: $\beta = 1.003$; $\sigma = 0.465$). In the third model, when considering the influence of interaction effect between risk management practices and stakeholders' engagement on financial stability, risk management practices themselves stop being important ($\beta = 0.265$; $\sigma = 0.253$). That brings partial support for our Hypothesis H1, while risk management practices, if accounted alone, comprise an essential factor determining financial stability, however, they stop playing this role when considering for stakeholders' engagement interaction.

Referring to our second hypothesis, stakeholders' engagement by itself is not a significant predictor of financial stability of the hospital neither in model 2 ($\beta = -0.623$; $\sigma = 0.619$) nor the model 3 ($\beta = -0.221$; $\sigma = 0.224$). Thus, we are forced to reject our second hypothesis. Finally, the interaction between risk management practices and financial stability plays a vital role in explaining the relationship between risk management practices and the hospitals' financial stability (model 3: $\beta = -0.207$; $\sigma = 0.088$). To better understand the moderated relationships' nature, we plotted two-way interaction using an excel tool adapted from www.jeremydawson.co.uk/slopes.htm (accessed on 30 April 2021). After introducing the data to the excel sheet, we were able to draw a plot, illustrated in Figure 3, representing the influence of interactions between risk management practices and stakeholders engagement on financial stability.

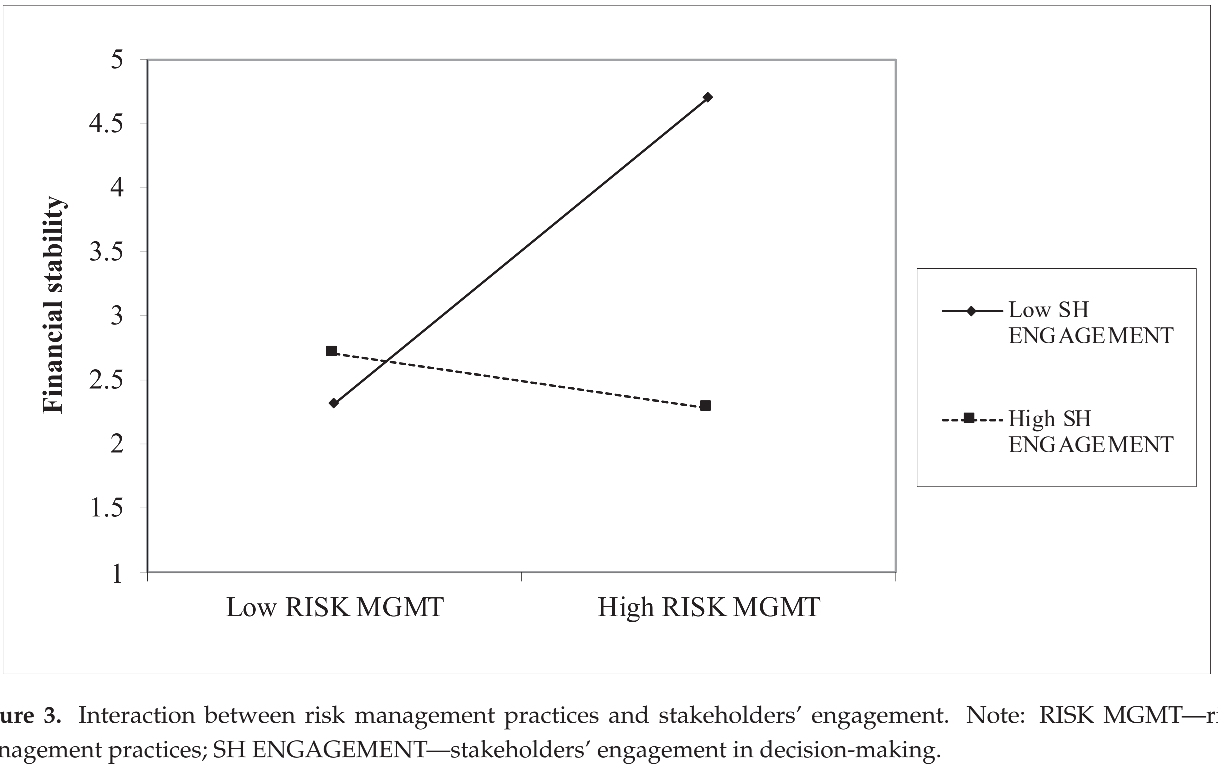


Figure 3. Interaction between risk management practices and stakeholders' engagement. Note: RISK MGMT—risk management practices; SH ENGAGEMENT—stakeholders' engagement in decision-making.

Interpreting the drawing, with high levels of stakeholders' engagement, the increase of risk management practices negatively affects financial stability. However, when stakeholders' engagement is low, the increase in attention paid to risk management practices pays back in increased financial stability. That brings support to our Hypothesis H3.

6. Discussion

In healthcare organizations in general and the hospital environment, both risk and risk management attract significant attention. It mostly triggers interest in the risk related to patient safety (Cagliano et al. 2011; Sheppard et al. 2013); however, there are also works related to cybersecurity risk management (Coronado and Wong 2014); managing healthcare waste (Akpieyi et al. 2015). However, as we outlined earlier in the text, it might be argued that the sources of risk for Polish hospitals should also be studied from a purely managerial and financial perspective. In such a case, organizational procedures—risk management practices—allowing hospitals to become prepared and adequately respond to the unforeseen obstacles and adversities may become essential factors guaranteeing its continuous operation under different financial strains (Beck de Silva Etges et al. 2016; Mahama et al. 2020; Zheng et al. 2019). Thus, well-crafted and executed risk management practices should lead to better financial stability. Especially, when they are coupled with shareholder engagement in the process (Bromiley et al. 2015). Thus, in our study, we tested how risk management practices in healthcare organizations influence a hospital's perceived financial stability considering principal hospital shareholders' engagement.

According to our research results, risk management practices affect financial stability in researched units only when examined without the interaction with the shareholder engagement. When we considered the interaction with the shareholder engagement, the influence of risk management practices on financial stability is nonsignificant. That demonstrates that risk management practices per se can influence financial stability, and hospitals shall seek to develop both appropriate tools to identify and manage risks and encourage their employees to act in a critical or difficult situation. However, in the meantime, other factors are affecting the financial stability of a hospital, which were not included in the model (Sowada et al. 2020). We assume that improvements in the decision-making processes, better financial planning, more accessible and better-suited information systems, and apt human resource practices contribute to financial stability to a more considerable extent (Griffith 2000).

Next, we also tested if the shareholders' engagement in Polish hospitals' decision-making processes led to improved financial stability. However, our model estimations brought no support for the second hypothesis; thus, we found no evidence of the influence of shareholders' engagement in the decision-making on financial stability. That result might be perceived as controversial, considering the hospital's obligation to account for stakeholders' expectations in the decision-making processes (Norris et al. 2017; Osborne et al. 2015; Rixon 2010; Wortley et al. 2016). It also contrasts research results that emphasize potential profits of stakeholders' engagement, namely: acquiring more thorough information, the better legitimization of decisions, or the quality of the decision that is made (Beierle 2002; George et al. 2016). On the other hand, it shows that stakeholder participation is not always desirable and does not lead to expected results, although we did not include nonlinear effects in our hypothesis, producing a different result (Schalk 2015).

Finally, with respect to our third hypotheses, we tested how risk management practices with shareholder engagement affect financial stability. Research results show that these two factors, when taken together, affect financial stability significantly. In particular, when risk management practices are becoming more and more sophisticated, higher stakeholder engagement in the decision-making processes leads to hospitals' statistically lower financial stability. This means that when awareness of the risks increases within the hospital, stakeholders that are actively engaged in decisions limits hospitals' ability to respond appropriately to identified risks. We assume that the adverse impact of shareholders' engagement in the decision-making processes with high-risk management standards may remain a result of extending the time needed to make a decision considering too broad or irrelevant information (Schalk 2015), and critical situation, in which risk management processes are employed, require prompt answers. It might also be caused by the lack of appropriate conditions for stakeholders to actively participate in the process (Reed et al. 2018), or the lack of understanding of the problem by stakeholders and their irrelevant or even harmful pressures on inappropriate solutions

(Wheeler and Sillanpää 1997), especially when their power is too large (Malfait et al. 2017; Malfait et al. 2018; McCarron et al. 2019; Petkovic et al. 2020).

Concerning limitations, we identified four main issues that hinder our study. Firstly, due to the relatively small sample, we could not employ more sophisticated data analysis methods. Our research also suffers from a single response bias (Burchett and Ben-Porath 2019). We tried to mitigate it by directly contacting every single respondent, but we still gathered opinions of single respondents within each hospital. In future studies, we suggest gathering data from multiple sources of information, for example, manager and stakeholder or manager and chosen employee (Turner et al. 2017). Secondly, we gathered opinions on financial stability—in our case, it would make much sense to combine these declarative statements with data flowing from financial reports (Min et al. 2020). However, in the process of data gathering, it exceeded our possibilities of reaching such information, and because of the difficult situation of numerous Polish hospitals, we decided to exclude this type of questions from the survey. We are certain that the future in-depth qualitative study would help to better address this issue. Additionally, we are convinced that testing financial stability both on the basis of declarative statement and "hard data" would bring a more comprehensive, bias free, and valuable standpoint to develop discussion. Thirdly, our study focused on detecting relationships between variables, but it fails to explain why these relationships exist. We suppose that future studies should understand the processes behind relationships between risk management and financial stability (Feng 2011; Karam et al. 2018; McCue and MCluer 2008; Ozcan and McCue 1996). It would also help to contextualize essential, outlined relationships better.

The contextualization of our model offers, in our opinion, a fascinating field of future research. In particular, we assume that it would be profitable for better understanding relationships between studied constructs to include at least three variables helping to explain the main relationship: leadership, while appropriate style might foster and encourage employees respond to a critical situation (Crosby and Bryson 2018); professionalization of management, which might translate existing procedures to life—competent managers would enable and empower employees to identify and better respond to adversities (Gerard 2019; Ingram and Glod 2016), and finally, risk management maturity, which is strongly related to professionalization. Maturity is understood as skills that demonstrate hospitals' readiness to perform specific tasks and the state of being complete, perfect, or ready (Antonucci 2016). In this respect, it seems reasonable to focus on both process maturity (Fraser and Vaishnavi 1997), object maturity (Mahama et al. 2020) as well as people's capability (Nonaka 1994).

Moreover, future studies would also embrace the topic of mutual relationships between risk management practices and financial stability. Clear identification of cause and effect in this regard is difficult while both of these phenomena reinforce themselves mutually. These relationships should be studied more deeply in the future in a richer context and on the larger sample, enabling inclusion of more contextual variables to the model. Most importantly, contextualization should also lead to including the hospital ecosystem as a dynamic moderator of the primary relationship. In such a case, we encourage scholars to perceive ecosystems as a community that consists of the living organisms and the nonliving components of particular natural environment space, interacting as a system. There are many relationships between these organisms and components that allow them to function in harmony and balance. Ecosystems are controlled by external and internal factors (Chapin et al. 2002; Banoun et al. 2016). On the ground of social sciences, the term ecosystem has been mainly applied in recent years to social innovation to describe the enabling environment that needs to be put in place if social innovations are to achieve their ultimate ambition of systemic change (Biggs et al. 2010; Pel et al. 2020; Vargo et al. 2015, 2017). In this context, an ecosystem approach provides a framework for both understanding all the interactions and resources relating to actors involved in social innovation work at a given time and for identifying what changes need to happen in order to build a field that is 'more than the sum of its parts'. In future research, including the ecosystem into the model should mainly focus on testing the interactions of its numerous components (legal regulations, level of financing, political changes, the competition level, etc.) with risk management and shareholders engagement as well as its impact on the financial stability of a hospital. Thus, we argue that the environmental factors that may have the most significant impact on the process of implementing risk management policy in Polish hospitals include social expectations that public services will be of higher quality and better access, which "force" managers to seek opportunities to streamline processes within the organization (Noronha and Mekoth 2013). A strong factor facilitating change is also the ageing society (Buliński and Błachnio 2017), which creates the need to develop new areas of healthcare services that would meet emerging health challenges, including the development of healthcare services based on telemedicine, which will improve the availability of healthcare services for people with limited mobility.

7. Conclusions

On the practical side, our research leads to several recommendations. We divided them into three groups: implications for organizational stakeholders (mainly managers), implications for shareholders (owner entity), and national health system stakeholders. In the first instance, since risk management practices are proven, at least in isolation, to lead to improved financial stability, hospital managers are encouraged to motivate employees to actively monitor the situation with respect to potential risks or crises (Hunt 2010; Kloutsiniotis and Mihail 2017; Li et al. 2020; Oulasvirta and Anttiroiko 2017; Rana et al. 2019). Hospitals should also focus on developing procedures to mitigate risks and train employees to facilitate adequate reaction to a critical situation (Agnew et al. 2006; Ferdosi et al. 2020; Roberts 2002). It seems similarly reasonable and justifiable to create emergency response teams of employees trained specially to counter unforeseen crises (Ab Aziz et al. 2019; Hunt 2010).

Further, we prove that under certain conditions (with well-developed risk management practices that are most likely the result and reflection of managers' high professional competencies in a hospital), shareholders' engagement might be detrimental to financial stability. Professional managers might not need additional help from stakeholders when making tough times (Hinna et al. 2018; Noordegraaf and Van der Meulen 2008). Although, when risk management practices are nonexistent, shareholders should actively engage in decision-making processes to enhance quality. Thus, shareholders should balance their engagement based on constant evaluation of actual needs to shape financial stability actively (Li et al. 2020; Rixon 2010; Wu 2012).

Finally, we suggest including the assessment of risk management procedures and preparedness into the evaluation criteria for hospitals for the system level stakeholders. Additionally, the care for healthcare managers' qualifications seems to be on point (Dwyer et al. 2006) since these might affect hospital decision-makers' attitude towards risk management. At a systemic level, we believe that stakeholders should support the creation of capacity building, perceived as "(. . .) the process by which individuals and organizations develop or strengthen abilities related to understanding, providing input to, conducting, or utilizing risk management as a tool for better health policy and decision making, as well as developing awareness and support in the environment within hospitals acting for implementation risk management as a tool for an effective hospital management" (Pichler et al. 2019, p. 363). Finally, results suggest that system-level stakeholders should focus on controlling shareholders' level of engagement in decision-making, especially when managers are qualified professionals (Kwiecińska-Bożek 2018; Linnander et al. 2017).

Author Contributions: Conceptualization, A.F.-W., T.I., K.S.-M., P.T.; methodology, A.F.-W., T.I., K.S.-M., P.T.; software, A.F.-W., T.I., K.S.-M., P.T.; validation, A.F.-W., T.I., K.S.-M., P.T.; formal analysis, A.F.-W., T.I., K.S.-M., P.T.; investigation, A.F.-W., T.I., K.S.-M., P.T.; resources, A.F.-W., T.I., K.S.-M., P.T.; data curation, A.F.-W., T.I., K.S.-M., P.T.; writing—original draft preparation, A.F.-W., T.I., K.S.-M., P.T.; writing—review and editing, A.F.-W., T.I., K.S.-M., P.T.; visualization, A.F.-W., T.I., K.S.-M., P.T.; supervision, A.F.-W., T.I., K.S.-M., P.T.; project administration, A.F.-W., T.I., K.S.-M., P.T.; funding acquisition, A.F.-W., T.I., K.S.-M., P.T. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Polish Ministry of Science and Higher Education under University of Economics in Katowice the research potential fund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data are available on request from the corresponding author.

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

References

Ab Aziz, Nur Farhani, Farid Wajdi Akashah, and Aniza Abdul Aziz. 2019. Conceptual framework for risk communication between emergency response team and management team at healthcare facilities: A Malaysian perspective. *International Journal of Disaster Risk Reduction* 41: 101282. [CrossRef]

Adil, Mahmood. 2008. Risk communication in healthcare: An overview. *Journal of Communication in Healthcare* 1: 363–72. [CrossRef]

Agnew, John E., Nicholas Komaromy, and Richard E. Smith. 2006. Healthcare Institution Risk Assessments: Concentration on "Process" or "Outcome"? *Journal of Risk Research* 9: 503–23. [CrossRef]

Akpieyi, Alex, Terry L. Tudor, and Cleber Dutra. 2015. The utilisation of risk-based frameworks for managing healthcare waste: A case study of the National Health Service in London. *Safety Science* 72: 127–32. [CrossRef]

Amaeshi, Kenneth M., and Andrew Crane. 2006. Stakeholder engagement: A mechanism for sustainable aviation. *Corporate Social Responsibility and Environmental Management* 13: 245–60. [CrossRef]

Antonucci, Domenic. 2016. *Risk Maturity Models*. London: Kogan Page Limited.

Austen, Agata, and Aldona Frączkiewicz-Wronka. 2018. Stakeholders and resources in public hospitals: Towards an integrated view. *Argumenta Oeconomica* 1: 253–74. [CrossRef]

Bakalikwira, Lasuli, Juma Bananuka, Twaha K. Kigongo, Doreen Musimenta, and Veronica Mukyala. 2017. Accountability in the public health care systems: A developing economy perspective. *Cogent Business & Management* 4: 1–14. [CrossRef]

Banoun, Arnaud, Lucas Dufour, and Meena Andiappan. 2016. Evolution of a service ecosystem: Longitudinal evidence from multiple shared services centers based on the economies of worth framework. *Journal of Business Research* 69: 2990–98. [CrossRef]

Beck de Silva Etges, Ana Paula, Joana Siqueira de Souza, Francisco José Kliemann Neto, and Elaine Aparecida Felix. 2016. A proposed enterprise risk management model for health organizations. *Journal of Risk Research* 22: 513–31. [CrossRef]

Beierle, Thomas C. 2002. The quality of stakeholder-based decisions. *Risk Analysis* 22: 739–49. [CrossRef] [PubMed]

Bennet, Peter, Kenneth Calman, Sarah Curtis, and Denis Fischbacher-Smith. 2010. *Risk Communication and Public Health*. Oxford: Oxford University Press.

Biggs, Reinette, Frances R. Westley, and Stephen R. Carpenter. 2010. Navigating the back loop: Fostering social innovation and transformation in ecosystem management. *Ecology and Society* 15: 28. [CrossRef]

Borraz, Olivier. 2007. Risk and Public Problems. *Journal of Risk Research* 10: 941–57. [CrossRef]

Bromiley, Philip, Michael McShane, Anil Nair, and Elzotbek Rustambekov. 2015. Enterprise risk management: Review, critique, and research directions. *Long Range Planning* 48: 265–76. [CrossRef]

Bryson, John M. 2004. What to do when stakeholders matter: Stakeholder identification and analysis techniques. *Public Management Review* 6: 21–53. [CrossRef]

Buliński, Leszek, and Aleksandra Błachnio. 2017. Health in old age, and patients' approaches to telemedicine in Poland. *Annals of Agricultural and Environmental Medicine* 24: 322–28. [CrossRef]

Burchett, Danielle, and Yossef S. Ben-Porath. 2019. Methodological considerations for developing and evaluating response bias indicators. *Psychological Assessment* 31: 1497–511. [CrossRef] [PubMed]

Burke, Richard, and Istemi Demirag. 2017. Risk transfer and stakeholder relationships in Public Private Partnerships. *Accounting Forum* 41: 28–43. [CrossRef]

Cagliano, Anna Corina, Sabrina Grimaldi, and Carlo Rafele. 2011. A systemic methodology for risk management in healthcare sector. *Safety Science* 49: 695–708. [CrossRef]

Carroll, Robert, ed. 2009. *Risk Management Handbook for Health Care Organizations*. Hoboken: John Wiley & Sons Inc.

Chambers, Naomi, Gill Harvey, and Russell Mannion. 2017. Who should serve on health care boards? What should they do and how should they behave? A fresh look at the literature and the evidence. *Cogent Business & Management* 4: 1–14. [CrossRef]

Chapin, F. Stuart, Pamela A. Matson, and Harold A. Mooney. 2002. *Principles of Terrestrial Ecosystem Ecology*. New York: Springer.

Cleemput, Irina, Wendy Christiaens, Laurence Kohn, Christian Léonard, François Daue, and Alain Denis. 2015. Acceptability and Perceived Benefits and Risks of Public and Patient Involvement in Health Care Policy: A Delphi Survey in Belgian Stakeholders. *Value Health* 18: 477–83. [CrossRef] [PubMed]

Collins, David. 1992. The Strategic Management of Uncertainty. *European Management Journal* 10: 125–35. [CrossRef]

Cornforth, Chris. 1978. What makes boards effective? An examination of the relationships between board inputs, structures, processes and effectiveness in non-profit organizations. *Corporate Governance: An International Review* 9: 217–27. [CrossRef]
Coronado, Anthony J., and Timothy L. Wong. 2014. Healthcare cybersecurity risk management: Keys to an effective plan. *Biomedical Instrumentation & Technology* 48: 26–30. [CrossRef]
Crosby, Barbara C., and John M. Bryson. 2018. Why leadership of public leadership research matters: And what to do about it. *Public Management Review* 20: 1265–86. [CrossRef]
Dansoh, Ayirebi, Samuel Frimpong, Godslove Amppatwaum, Goodenough Dennis Oppong, and Robert Kyei-Osei. 2020. Exploring the role of traditional authorities in managing the public as stakeholders on PPP projects: A case study. *International Journal of Construction Management* 20: 628–41. [CrossRef]
Dillon, Stuart, John Buchanan, and Jim Corner. 2010. Comparing Public and Private Sector Decision Making: Problem Structuring and Information Quality Issues. Paper presented at 45th Annual Conference of the ORSNZ, Auckland, New Zealand, November 29–30.
Dixit, Sunil K. 2017. A new multiperspective emphasis on the public hospital governance. *International Journal of Healthcare Management* 13: 267–75. [CrossRef]
Djellouli, Nehla, Lorelei Jones, Helen Barratt, Angus I. G. Ramsay, Steven Towndrow, and Sandy Oliver. 2019. Involving the public in decision-making about large-scale changes to health services: A scoping review. *Health Policy* 123: 635–45. [CrossRef]
Dubas-Jakóbczyk, Katarzyna, Iwona Kowalska-Bobko, and Christoph Sowada. 2019. The 2017 reform of the hospital sector in Poland—The challenge of consistent design. *Health Policy* 123: 538–43. [CrossRef]
Dwyer, Joseph, Michael Paskavitz, Sylvia Vriesendorp, and Sarah Johnson. 2006. An urgent call to professionalize leadership and management in health care worldwide. *Management Sciences for Health Occasional Papers* 4: 1–14.
Edelenbos, Jurian, and Erik-Hans Klijn. 2006. Managing Stakeholder Involvement in Decision Making: A Comparative Analysis of Six Interactive Processes in the Netherlands. *Journal of Public Administration Research and Theory* 16: 417–46. [CrossRef]
Elias, Arun A. 2019. Strategy Development Through Stakeholder Involvement: A New Zealand Study. *Global Journal of Flexible Systems Management* 20: 313–22. [CrossRef]
Elleuch, Hatem, Wafik Hachicha, and Habib Chabchoub. 2014. A combined approach for supply chain risk management: Description and application to a real hospital pharmaceutical case study. *Journal of Risk Research* 17: 641–63. [CrossRef]
Feng, Ziqin. 2011. Financial risks from three dimensions and risk identification model of enterprise. *Strategic Finance* 6: 71–80. [CrossRef]
Ferdosi, Masoud, Reza Rezayatmand, and Yasamin Molavi Taleghani. 2020. Risk Management in Executive Levels of Healthcare Organizations: Insights from a Scoping Review 2018. *Risk Management and Healthcare Policy* 13: 215–43. [CrossRef] [PubMed]
Flemig, Sophie, Stephen Osborne, and Tony Kinder. 2016. Risky Business—Reconceptualizing Risk and Innovation in Public Services. *Public Money & Management* 36: 425–32.
Fone, Martin, and Peter C. Young. 2007. *Managing Risks in Public Organizations*. London: Palgrave Macmillan.
Frączkiewicz-Wronka, Aldona. 2012. Rola i znaczenie interesariuszy w procesie podejmowania decyzji w organizacjach publicznych [Role and importance of stakeholders in decision making in public organizations]. In *Wykorzystanie Analizy Interesariuszy w Zarządzaniu Organizacją Zdrowotną [Using Stakeholder Analysis in Managing a Health Organization].* Edited by Aldona Frączkiewicz-Wronka. Katowice: Śląsk, pp. 15–46.
Frączkiewicz-Wronka, Aldona. 2018. Stakeholders as a source of risk in public hospitals in Poland—selected problems. In *Public Relations w Przestrzeni Publicznej [Public Relations in Public Space]*. Edited by Anna Adamus-Matuszyńska. Prace Naukowe Uniwersytetu Ekonomicznego w Katowicach. Katowicach: Uniwersytetu Ekonomicznego w Katowicach, pp. 150–57.
Fraser, Martin D., and Vijay Vaishnavi. 1997. A formal specifications maturity model. *Communications of the ACM* 40: 95–103. [CrossRef]
George, Bert, Sebastian Desmidt, and Julie De Moyer. 2016. Strategic decision quality in Flemish municipalities. *Public Money & Management* 36: 317–24. [CrossRef]
Gerard, Nathan. 2019. Perils of professionalization: Chronicling a crisis and renewing the potential of healthcare management. *Health Care Analysis* 27: 269–88. [CrossRef]
Główny Urząd Statystyczny (GUS) [Main Statistical Office]. 2019. *Ochrona Zdrowia w Gospodarstwach Domowych w 2017 r. [Healthcare in Households in 2017].* Warsaw: GUS.
Gomes, Ricardo Corrêa, Joyce Liddle, and Luciana O.M. Gomes. 2010. A five-sided model of stakeholder influence. A cross-national analysis of decision making in local government. *Public Management Review* 12: 701–24. [CrossRef]
Griffith, John R. 2000. Championship management for healthcare organizations. *Journal of Healthcare Management* 45: 17–30. [CrossRef] [PubMed]
Hinna, Alessandro, Danila Scarozza, and Fabrizio Rotundi. 2018. Implementing Risk Management in the Italian Public Sector: Hybridization between Old and New Practices. *International Journal of Public Administration* 41: 110–28. [CrossRef]
Hood, John, Andrew Mills, and William M. Stein. 2003. Development in UK public sector risk management: The implications for PPP/PFI projects. In *Public-Private Partnership: Managing Risks and Opportunities*. Edited by Akintola Akintoye, Cliff Hardcastle and Matthias Beck. Hoboken: Blackwell Science Ltd.
Hooper, Daire, Joseph Coughlan, and Michael R. Mullen. 2008. Evaluating Model Fit: A Synthesis of the Structural Equation Modelling Literature. *Electronic Journal of Business Research Methods* 6: 53–60. [CrossRef]
Hunt, Brian. 2010. Risk and Crisis Management in the Public Sector. *Public Management Review* 12: 747–51. [CrossRef]

Ingram, Tomasz, and Wojciech Glod. 2016. Talent management in healthcare organizations-qualitative research results. *Procedia Economics and Finance* 39: 339–46. [CrossRef]

Jajuga, Krzysztof, ed. 2019. *Zarządzanie Ryzykiem [Risk Management]*. Warszawa: PWN.

Jansen, Maarten P. M., Rob Baltussen, and Kristine Bærøe. 2018. Stakeholder Participation for Legitimate Priority Setting: A Checklist. *International Journal of Health Policy and Management* 7: 973–76. [CrossRef]

Johnsen, Åge. 2015. Strategic Management Thinking and Practice in the Public Sector: A Strategic Planning for All Seasons? *Financial Accountability & Management* 31: 243–268. [CrossRef]

Karam, Marlène, Isabelle Brault, Thérèse Van Durme, and Jean Macq. 2018. Comparing interprofessional and interorganizational collaboration in healthcare: A systematic review of the qualitative research. *International Journal of Nursing Studies* 79: 70–83. [CrossRef] [PubMed]

Kavaler, Florence, and Allen D. Spiegel. 2003. *Risk Management in Health Care Institutions: A Strategic Approach*. Burlington: Jones & Bartlett Publishers.

Klinke, Andreas, and Ortwin Renn. 2012. Adaptive and Integrative Governance on Risk and Uncertainty. *Journal of Risk Research* 15: 273–92. [CrossRef]

Kloutsiniotis, Panagiotis V., and Dimitrios M. Mihail. 2017. Linking innovative human resource practices, employee attitudes and intention to leave in healthcare services. *Employee Relations* 39: 34–53. [CrossRef]

Kolluru, Rao V., Steven M. Bartell, Robin M. Pitblado, and Scott R. Stricoff. 1996. *Risk Assessment and Management Handbook. For Environmental, Health and Safety Professional*. New York: McGraw-Hill.

Krewski, Daniel, Victoria Hogan, Michelle C. Turner, Patricia L. Zeman, Ian McDowell, Nancy Edwards, and Joseph Loses. 2007. An Integrated Framework for Risk Management and Population Health. *Human and Ecological Risk Assessment: An International Journal* 13: 1288–312. [CrossRef]

Kwiecińska-Bożek, Lidia. 2018. Risk Management in Public Hospitals in Poland—Selected Issues. In *Public Relations w Przestrzeni Publicznej [Public Relations in Public Space]*. Edited by Anna Adamus-Matuszyńska. Prace Naukowe Uniwersytetu Ekonomicznego w Katowicach. Katowicach: Uniwersytetu Ekonomicznego w Katowicach, pp. 140–49.

Langrafe, Taiguara de Freitas, Simone Ruchdi Barakat, Fabricio Stocker, and Joao Maurício Gama Boaventura. 2020. A stakeholder theory approach to creating value in higher education institutions. *The Bottom Line* 33. [CrossRef]

Li, Yanwei, Araz Taeihagh, Martin de Jong, and Andreas Klinke. 2020. Toward a Commonly Shared Public Policy Perspective for Analyzing Risk Coping Strategies. *Risk Analysis*, 1–14. [CrossRef]

Lin, Lexin, Claudia Rivera, Marcus Abrahamsson, and Henrik Tehler. 2017. Communicating risk in disaster risk management systems—experimental evidence of the perceived usefulness of risk descriptions. *Journal of Risk Research* 20: 1534–53. [CrossRef]

Linnander, Erika L., Jeannie M. Mantopoulos, Nikole Allen, Ingrid M. Nembhard, and Elizabeth H. Bradley. 2017. Professionalizing healthcare management: A descriptive case study. *International Journal of Health Policy and Management* 6: 555–60. [CrossRef] [PubMed]

Mahama, Habib, Mohamed Elbashirs, Stave Sutton, and Vicky Arnold. 2020. New development: Enabling enterprise risk management maturity in public sector organizations. *Public Money & Management*, 1–5. [CrossRef]

Malfait, Simon, Ann Van Hecke, Johan Hellings, Griet De Bodt, and Kristof Eeckloo. 2017. The impact of stakeholder involvement in hospital policy decision making: A study of the hospital's business processes. *International Journal of Clinical and Laboratory Medicine* 72: 63–71. [CrossRef]

Malfait, Simon, Ann Van Hecke, Griet De Bodt, Nele Palsterman, and Kristof Eeckloo. 2018. Patient and public involvement in hospital policy-making: Identifying key elements for effective participation. *Health Policy* 122: 380–88. [CrossRef]

McCarron, Tamara L., Karen Moffat, Gloria Wilkinson, Sandra Zelinsky, Jamie M. Boyd, Deborah White, Derek Hassay, Diane L. Lorenzetti, Nancy J. Marlett, and Thomas Noseworthy. 2019. Understanding patient engagement in health system decision-making: A co-designed scoping review. *Systematic Reviews* 8: 97. [CrossRef]

McCue, Michael J., and Forrest R. MCluer. 2008. Financial and Operational Ratios for Bond-Insured Hospitals. *Hospital Topics* 86: 3–8. [CrossRef] [PubMed]

McEvoy, Rachel, Edel Tierney, and Anne MacFarlane. 2019. 'Participation is integral': Understanding the levers and barriers to the implementation of community participation in primary healthcare: A qualitative study using normalisation process theory. *BMC Health Services Research* 19: 515. [CrossRef] [PubMed]

Mennen, Marcel G., and Maaike Van Tuyll. 2015. Dealing with Future Risks in the Netherlands: The National Security Strategy and the National Risk Assessment. *Journal of Risk Research* 18: 860–76. [CrossRef]

Min, Naon, Ruowen Shen, David Berlan, and Keon-Hyung Lee. 2020. How Organizational Identity Affects Hospital Performance: Comparing Predictive Power of Mission Statements and Sector Affiliation. *Public Performance & Management Review* 43: 845–70. [CrossRef]

Nieszporska, Sylwia. 2012. Ryzyko w ochronie zdrowia [The Risk in Heath Care]. *Logistyka* 5: 149–52.

Nonaka, Ikujiro. 1994. A dynamic theory of organizational knowledge creation. *Organization Science* 5: 14–37. [CrossRef]

Noordegraaf, Mirko, and Martijn Van der Meulen. 2008. Professional power play: Organizing management in health care. *Public Administration* 86: 1055–69. [CrossRef]

Noronha, Aruna Mesquita, and Nandakumar Mekoth. 2013. Social support expectations from healthcare systems: Antecedents and emotions. *International Journal of Healthcare Management* 6: 269–75. [CrossRef]

Norris, Jill M., Deborah E. White, Lorelli Nowell, Kelly Mrklas, and Henry T. Stelfox. 2017. How do stakeholders from multiple hierarchical levels of a large provincial health system define engagement? A qualitative study. *Implementation Science* 12: 1–13. [CrossRef] [PubMed]

Nutt, Paul. 1999. Public-private differences and the assessment of alternatives for decision making. *Journal of Public Administration Research and Theory* 9: 305–49. [CrossRef]

Nutt, Paul. 2005. Comparing Public and Private Sector Decision-Making Practices. *Journal of Public Administration Research and Theory* 16: 289–318. [CrossRef]

O'Shea, Alison, Annette L. Boaz, and Mary Chambers. 2019. A Hierarchy of Power: The Place of Patient and Public Involvement in Healthcare Service Development. *Frontiers in Sociology* 4: 38. [CrossRef] [PubMed]

Organisation for Economic Cooperation and Development (OECD). 2020. *Health at a Glance 2019: OECD Indicators*. Paris: OECD Publishing. [CrossRef]

Osborne, Stephen. P., Zoe Radnor, Tony Kinder, and Isabel Vidal. 2014. Sustainable public service organisations: A public service-dominant approach. *Society and Economy* 36: 313–38. [CrossRef]

Osborne, Stephen P., Zoe Radnor, Tony Kinder, and Isabel Vidal. 2015. The SERVICE Framework: A Public-service-dominant Approach to Sustainable Public Services. *British Journal of Management* 26: 424–38. [CrossRef]

Osborne, Stephen P., Taco Brandsen, Valentina Mele, Juraj Nemec, Marieke van Genugten, and S. Sophie Fleming. 2020. Risking innovation. Understanding risk and public service innovation—Evidence from a four nation study. *Public Money & Management* 40: 52–56. [CrossRef]

Oulasvirta, Lasse, and Ari-Veikko Anttiroiko. 2017. Adoption of comprehensive risk management in local government. *Local Government Studies* 43: 451–74. [CrossRef]

Ozcan, Yasar A., and Michael J. McCue. 1996. Development of a Financial Performance Index for Hospitals: DEA Approach. *Journal of the Operational Research Society* 47: 18–26. [CrossRef]

Pel, Bonno, Julia Wittmayer, Jens Dorland, and Michael S. Jorgensen. 2020. Unpacking the Social Innovation Ecosystem: An empirically grounded typology of empowering network constellations. *Innovation: The European Journal of Social Science Research* 33: 311–36. [CrossRef]

Petkovic, Jennifer, Alison Riddle, Elie A. Akl, Joanne Khabsa, Lyubove Lytvyn, Pearl Atwere, Pauline Cambell, Kalipso Chalkidou, Stephanie M. Chang, Sally Crowe, and et al. 2020. Protocol for the development of guidance for stakeholder engagement in health and healthcare guideline development and implementation. *Systematic Reviews* 9: 1–11. [CrossRef] [PubMed]

Pichler, Franz, Wija Oortwijn, Alric Ruether, and Rebecca Trowman. 2019. Defining capacity building in the context of HTA: A proposal by the HTAi Scientific Development and Capacity Building Committee. *International Journal of Technology Assessment in Health Care* 35: 362–66. [CrossRef] [PubMed]

Power, Michael. 2016. *Riskwork: Essays on the Organizational Life of Risk Management*. Oxford: Oxford University Press.

Professeure, Lucie Rouillard. 2004. Managing Risk: A New Approach to Government Intervention. *Public Management Review* 6: 95–111. [CrossRef]

Quick, Kathryn S., and John M. Bryson. 2016. Public Participation. In *Handbook on Theories of Governance*. Edited by Chris Ansell and Jacob Torfing. Cheltenham: Edward Elgar Press.

Raczkowski, Konrad, and Piotr Tworek. 2017. What Does Risk Management in an Economy Really Mean? In *Risk Management in the Public Administration*. Edited by Konrad Raczkowski. London: Palgrave Macmillan.

Rainey, Hal G. 2003. *Understanding and Managing Public Organizations*. San Francisco: Jossey-Bass.

Rana, Tarek, Danture Wickramasinghe, and Enrico Bracci. 2019. New development: Integrating risk management in management control systems—lessons for public sector managers. *Public Money & Management* 39: 148–51. [CrossRef]

Reed, Mark, Steven Vella, Edward Challies, Joris de Vente, Lynne Frewer, Daniela Hohenwallner-Ries, Tobias Huber, Rosmarie Neumann, Elizabeth Oughton, Julian Sidoli del Ceno, and et al. 2018. A theory of participation: What makes stakeholder and public engagement in environmental management work? *Restoration Ecology* 26: 7–17. [CrossRef]

Rixon, Daphne. 2010. Stakeholder Engagement in Public Sector Agencies: Ascending the Rungs of the Accountability Ladder. *International Journal of Public Administration* 33: 374–556. [CrossRef]

Roberts, Geoff. 2002. *Risk Management in Healthcar*. London: Witherby & Co.

Rudawska, Iga. 2011. Ku rynkowi w opiece zdrowotnej [Towards a market in healthcare]. *Polskie Stowarzyszenie Zarządzania Wiedzą* 55: 131–38.

Sarens, Gerrit, Christian de Visscher, and Diane van Gils. 2010. Risk management and internal control in the public sector: An in-depth analysis of Belgian Social Security Public Institutions. *Federale Overheidsdienst Financiën—België* 70: 65–90.

Schalk, Jelmer. 2015. Linking Stakeholder Involvement to Policy Performance: Nonlinear Effects in Dutch Local Government Policy Making. *American Review of Public Administration* 47: 479–59. [CrossRef]

Sheppard, Faye, Marcie Williams, and Victor R. Klein. 2013. TeamSTEPPS and patient safety in healthcare. *Journal of Healthcare Risk Management* 32: 5–10. [CrossRef] [PubMed]

Skoczylas, Wanda, and Piotr Waśniewski. 2014. Wskaźniki finansowe w raporcie z badania sprawozdania finansowego przez biegłego rewidenta [Financial ratios in the report supplementing the opinion on the audit of financial statement by the statutory auditor]. *Finanse, Rynki Finansowe, Ubezpieczenia* 69: 243–52.

Snow, Donald. 1992. *Inside the Environmental Movement: Meeting the Leadership Challenge*. Washington: Island Press.

Sohn, Kitae. 2016. Risk Incomprehension and Its Economic Consequences. *The Journal of Development Studies* 52: 1545–60. [CrossRef]

Sowada, Christoph, Iwona Kowalska-Bobko, and Anna Sagan. 2020. What next after the 'commercialization' of public hospitals? Searching for effective solutions to achieve financial stability of the hospital sector in Poland. *Health Policy* 124: 1050–55. [CrossRef]

Stroh, Patrick J. 2005. Enterprise Risk Management at UnitedHealth Group. *Strategic Finance* 87: 26–27.

Szymaniec-Mlicka, Karolina. 2017. The decision-making process in public healthcare entities. identification of the decision-making process type. *Management* 21: 191–204. [CrossRef]

The Orange Book. Management of Risk—Principles and Concepts. 2004. London: HM Treasury UK.

Tončinić, Sara, Renée Wildt-Liesveld, and Hubertus Vrijhoef. 2020. Evaluation of a digital platform that engages stakeholders in the co-creation of healthcare innovations: A mixed-methods study. *International Journal of Care Coordination* 23: 33–42. [CrossRef]

Tseng, Ming-Lang, Hien Minh Ha, Ming K. Lim, Kuo-Jui Wu, and Mohammad Iranmanesh. 2020. Sustainable supply chain management in stakeholders: Supporting from sustainable supply and process management in the healthcare industry in Wietnam. *International Journal of Logistics Research and Applications*, 1–20. [CrossRef]

Turner, Scott F., Laura B. Cardinal, and Richard M. Burton. 2017. Research design for mixed methods: A triangulation-based framework and roadmap. *Organizational Research Methods* 20: 243–67. [CrossRef]

Tworek, Piotr. 2016. Understanding Public Risk. In *Public Risk Management. Tome I: Perspective of Theory and Practice*. Edited by Piotr Tworek and Józef Myrczek. Katowice: Publishing House of the University of Economics.

Vargo, Stephen. L., Heiko Wieland, and Melissa Archpru Akaka. 2015. Innovation through institutionalization: A service ecosystems perspective. *Industrial Marketing Management* 44: 63–72. [CrossRef]

Vargo, Stephen L., Melissa Archpru Akaka, and Claudia M. Vaughan. 2017. Conceptualizing Value: A Service-ecosystem View. *Journal of Creating Value* 3: 1–8. [CrossRef]

Wędzki, Dariusz. 2012. The sequence of cash flow in bankruptcy prediction: Evidence from Poland. *Zeszyty Teoretyczne Rachunkowosci* 68: 161–79.

Węgrzyn, Maria. 2012. Healthcare as a link in the public finance system. Existing and desired directions of change. *Zeszyty Naukowe Wyższej Szkoły Bankowej w Poznaniu* 41: 187–99.

Wheeler, David, and Maria Sillanpää. 1997. *The Stakeholder Corporation: A Blueprint for Maximising Stakeholder Value*. London: Pitman.

Williams, Iestyn, and Heather Shearer. 2011. Appraising public value: Past, present and futures. *Public Administration* 89: 1367–84. [CrossRef]

World Health Organziation (WHO). 2011. *Report on the Burden of Endemic Health Care-Associated Infection Worldwide, Clean Care is Safer Care*. Geneva: WHO.

Wortley, Sally, Jackie Street, Wendy Lipworth, and Kristen Howard. 2016. What factors determine the choice of public engagement undertaken by health technology assessment decision-making organizations? *Journal of Health Organization and Management* 30: 872–90. [CrossRef] [PubMed]

Wronka-Pośpiech, Martyna. 2014. Sukces i jego pomiar w przedsiębiorstwach społecznych—próba operacjonalizacji [Success and its measurement in social enterprises—An attempt of operationalisation]. *Prace Naukowe Wałbrzyskiej Wyższej Szkoły Zarządzania i Przedsiębiorczości* 27: 363–87.

Wu, Minyu. 2012. Managing stakeholders: An integrative perspective on the source of competitive advantage. *Asian Social Science* 8: 160. [CrossRef]

Young, Peter C., and Steven C. Tippins. 2001. *Managing Business Risk: An Organization—Wide Approach to Risk Management*. New York: American Management Association.

Zaleska, Beata. 2017. Rola wskaźników finansowych w analizie i prognozie sytuacji ekonomiczno-finansowej samodzielnych publicznych zakładów opieki zdrowotnej. [The role of financial ratios in analysis and forecast of economic and financial situation of the independent public healthcare centers]. *Zeszyty Naukowe Wydziału Nauk Ekonomicznych* 21: 373–86.

Zheng, Yi, Wei Wang, Wenbin Liu, and John Mingers. 2019. A performance management framework for the public sector: The balanced stakeholder model. *Journal of the Operational Research Society* 70: 568–39. [CrossRef]

risks

MDPI

Article

An Efficient Method for Pricing Analysis Based on Neural Networks

Yaser Ahmad Arabyat [1,*], Ahmad Ali AlZubi [2], Dyala M. Aldebei [3] and Samerra'a Ziad Al-oqaily [4]

1 Faculty of Business, Department of Economic and Finance, Al-Blaqa Applied University, Al Salt 19110, Jordan
2 Computer Science Department, Community College, King Saud University, P.O. Box 28095, Riyadh 11437, Saudi Arabia; aalzubi@ksu.edu.sa
3 Faculty of Business, Department of Accounting, Al-Blaqa Applied University, Al Salt 19110, Jordan; dyaladebei@bea.edu.jo
4 Business School, Al-Blaqa Applied University, Al Salt 19110, Jordan; samoqaily@bea.edu.jo
* Correspondence: yaser_arabyat@bau.edu.jo

Abstract: The revolution in neural networks is a significant technological shift. It has an impact on not only all aspects of production and life, but also economic research. Neural networks have not only been a significant tool for economic study in recent years, but have also become an important topic of economics research, resulting in a large body of literature. The stock market is an important part of the country's economic development, as well as our daily lives. Large dimensions and multiple collinearity characterize the stock index data. To minimize the number of dimensions in the data, multiple collinearity should be removed, and the stock price can then be forecast. To begin, a deep autoencoder based on the Restricted Boltzmann machine is built to encode high-dimensional input into low-dimensional space. Then, using a BP neural network, a regression model is created between low-dimensional coding sequence and stock price. The deep autoencoder's capacity to extract this feature is superior to that of principal component analysis and factor analysis, according to the findings of the experiments. Utilizing the coded data, the proposed model can lower the computational cost and achieve higher prediction accuracy than using the original high-dimensional data.

Keywords: neural network; stock exchange; accounting systems; finance

Citation: Arabyat, Yaser Ahmad, Ahmad Ali AlZubi, Dyala M. Aldebei, and Samerra'a Ziad Al-oqaily. 2022. An Efficient Method for Pricing Analysis Based on Neural Networks. *Risks* 10: 151. https://doi.org/10.3390/risks10080151

Academic Editors: Krzysztof Jajuga and Józef Dziechciarz

Received: 22 June 2022
Accepted: 20 July 2022
Published: 28 July 2022

1. Introduction

The stock market is a market closely related to our daily lives, and it is very important to the economic development of our country and even the world. As important participants in the stock market, changes in people's sentiments will be quickly reflected in the market. In recent years, with the Internet and the rapid development of technology, the speed and channels of information dissemination are also increasing, and the ways for investors to obtain information are becoming more and more diversified (Hecht et al. 2020). Financial news, social media, etc. have gradually begun to affect investors' investment decisions. Therefore, stock prices will not only be affected by politics, the economy, and the military, but also by "emotional" factors (Lu et al. 2012; Craighead et al. 2009). With the influence of news, Weibo, blogs, post bars, forums and other social networks, financial news is more rapid and intuitive. This induction magnifies investors' attitudes towards the stock market, leading to great uncertainty and volatility in stock prices, making stock price prediction a major problem for research (Brandt and Brandt 2004).

In recent years, neural networks (Hecht et al. 2020) has become a learning boom. After decades of development, neural networks have become a popular technology (Garnet et al. 2002), and have continued to make breakthroughs in various fields, such as stock price forecasting (Brandt and Brandt 2013), time series forecasting, text analysis, computer vision (Van Velthoven et al. 2005; Zohar et al. 2002), etc.

Financial data modeling is a method of abstract representation of financial time series. The most important purpose of this task is to predict the future trend of financial series based on current data and to help decision-makers formulate strategies. It is currently widely used in accounting, business investment, stock valuation and other fields. Among them, the stock market has a higher rate of return on investment compared with other industries and has attracted the attention of a large number of investors. However, the strong volatility of stocks also causes many potential risks. Therefore, being able to accurately predict stock prices can not only obtain considerable profits, but also avoid potential risks.

The stock market has the characteristics of non-linearity, instability, complexity and so on. Regarding stock forecasting methods, scholars at home and abroad have conducted much research. At present, the methods of stock forecasting are roughly divided into two aspects: forecasting methods based on news and forecasting methods based on historical data. In the forecasting method based on news, Hai Nguyen et al. (2015) predicted the trend of stocks by mining people's sentiments about stocks from social media. Li et al. (2015) proposed a method of using news digests of stock-related reports to predict stock prices, and its effect is better than the method of using complete articles for prediction. Prachyachuwong and Vateekul (2021) designed a stock prediction model that combines word embedding and deep learning methods. This model can extract effective information from news and has higher accuracy. The research methods are based on historical data originated earlier, and are more diverse. Rundo et al. (2019) used an improved ARMA model to predict stock prices. The authors of (Chopra and Sharma 2021) used a model combining gray neural networks and Markov methods to study stock prices. Li et al. (2017) used a deep-belief-networks-based stock forecasting method, and introduced the concept of "intrinsic plasticity", which significantly improves the forecasting effect of the network.

Because the domestic stock market is not yet sound, there are many misleading factors, and retail investors participate more; thus the method is complicated. The research methods based on the existing literature are not suitable for the domestic stock market. Compared with the information based on the news, the historical trading data of stocks is more true and objective, and has reference value. Therefore, this article uses the method based on historical data to study stock prices. First of all, this paper constructs a deep autoencoder to reduce the dimensionality and feature-extraction of stock index data, then uses a BP neural network to build a regression model, predicts the stock price based on the DAE dimensionality reduction coding sequence, and finally verifies its effectiveness through experiments.

2. Deep Autoencoder

An autoencoder (AE) is a type of neural network. Unlike other neural network models, AE is an unsupervised learning algorithm that does not require additional label information during training. AE has an input layer, a hidden layer h and an output layer, and its structure is shown in Figure 1.

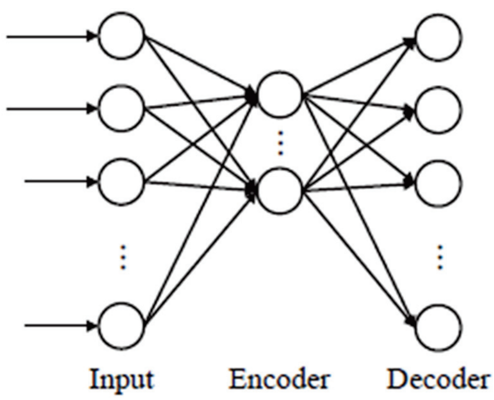


Figure 1. Autoencoder structure.

The AE uses the same structure of input and output, and through repeated training, the function $h\_(\boldsymbol{W},b)\ (x) \approx x$ is trained to make the input and output as similar as possible. The hidden layer added after training is also called an encoder, and the output of the encoder is the result of encoding the input data. The output layer is called the decoder, which can decode the encoded data, but usually does not pay attention to the output of the decoder. The working principle of an AE is expressed as follows:

Encoder:

$$h = f(\boldsymbol{W}x + b) \tag{1}$$

Decoder:

$$\hat{x} = f\left(\boldsymbol{W}'h + b'\right) \tag{2}$$

Among them, x is the input data, f is the activation function, $\boldsymbol{W}$ and b are the weights and bias values and $\hat{x}$ is the output of the decoder.

The use of autoencoders can achieve the purpose of feature extraction and feature dimensionality reduction. By setting the number of hidden layer nodes, dimensionality reduction can be achieved for any dimension. At the same time, the output of the encoder also expresses the higher-level features of the input data.

A deep autoencoder (DAE) is a deep neural network formed by stacking shallow autoencoders. It contains multiple hidden layers. Compared with shallow autoencoders, its advantage is that it can learn features in the data layer by layer. For example, we can input raw data into the DAE network, and generate a coding sequence in the first layer of the network. By analogy, the k-th in the DAE extracts features layer by layer based on the code of the $k-1$ layer output. As the number of network layers increases, the extracted features become more abstract (Wang et al. 2016). Deep autoencoders usually have better feature extraction capabilities than shallow autoencoders, and can extract non-linear features in the high levels of the data. The structure of the deep autoencoder is shown in Figure 2.

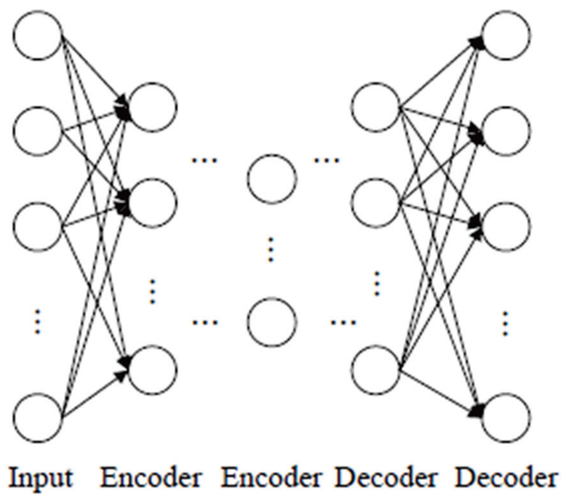


Figure 2. The structure of a DAE.

Although a DAE has the same structure as the multilayer feedforward neural network, its training strategy is different from that of the multilayer feedforward neural network. DAE uses a layer-by-layer greedy training method to pre-train the weights of each layer. Then the pre-trained weights are substituted into the corresponding layer. The global optimization algorithm fine-tunes the weight to obtain the final weight of the DAE.

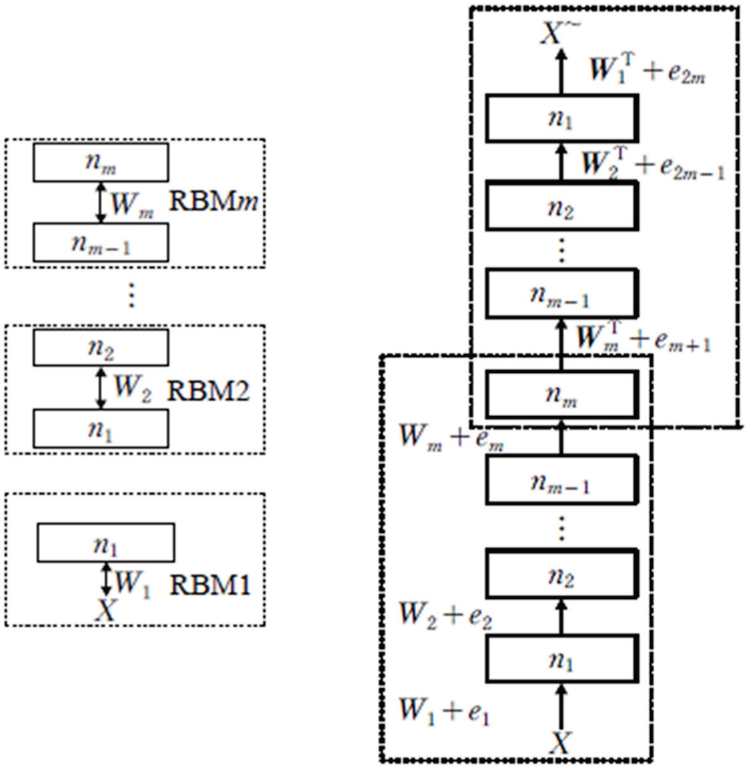
Deep autoencoders are currently widely used and have been successful in many fields; for example, in the field of image recognition, deep autoencoders are used for feature extraction (Xu et al. 2016; Gao et al. 2015) and image compression (Hinton and Salakhutdinov 2006). In the field of natural language processing, deep autoencoders are used for semantic hashing (Salakhutdinov and Hinton 2009), text generation (Babić et al.

2020) and speech processing (Araki et al. 2015). By using the deep autoencoder to reduce the original data, the algorithm becomes more efficient in the low-dimensional data space. Therefore, the deep autoencoder is an important tool and method in deep learning.

3. Proposed Model

3.1. Dimensionality Reduction of Stock Data Based on DAE

There is often multiple collinearity between stock indicators and indicators. For example, the J in KDJ (a stochastic indicator) is obtained by linear operation of K and D. Although these indicators with multi-collinearity provide convenience for technical analysis, they do not have much practical significance in the calculation of neural networks. Instead, they increase the computational cost of the model and the risk of overfitting. Therefore, it is necessary to reduce the dimensionality of the stock indicators to eliminate the multi-collinearity, reduce data redundancy, and greatly reduce the calculation and complexity of the model. This paper constructs a DAE model based on a Restricted Boltzmann Machine (RBM) and uses it for dimensionality reduction and feature extraction of stock index data. The structure of the model is shown in Figure 3.



(**a**) Trained layer by layer using RBM (**b**) After fine-tuning

Figure 3. Proposed model based on RBM.

A DAE is a deep neural network. Although the error back propagation algorithm can be used to train a DAE, the training effect is not ideal. As the number of neural network layers deepens, the gradient disappears when training the neural network using the error back propagation algorithm (Glorot and Bengio 2010). Aiming at the problem of the difficulty of deep neural network training, the literature (Le Roux and Bengio 2008) gives a method based on the RBM training method.

The training process is mainly divided into three steps:

(1) First construct multiple RBMs, as shown in Figure 3a, and pre-train the models separately using a layer-by-layer greedy training strategy. Among them, $\boldsymbol{X}$ is the

original data; each RBM uses the output of the previous RBM as input, and $\boldsymbol{W}$ is the pre-trained weight.

(2) Stack the pre-trained RBM layer by layer to build a symmetric model as shown in Figure 3b. The first to m-th layers of the model are called encoders, and each layer of the encoder uses the corresponding $\boldsymbol{W}$ as weight. The $m + 1$ layer to the 2 m layer of the model are called decoders, and each layer of the decoder uses the corresponding $\boldsymbol{W}^{\mathrm{T}}$ as the weight.

(3) Use the BP algorithm to fine-tune the model and update the weight to $\boldsymbol{W} + e$ to make the final output $\widetilde{\boldsymbol{X}}$ of the model as similar as possible to the input $\boldsymbol{X}$; the output of the m-th layer is the coding result. The following describes the RBM-based DAE training process in detail.

The RBM is an undirected neural network model based on statistical mechanics, and its structure is shown in Figure 4. The RBM has two layers that are the visible and hidden layer.

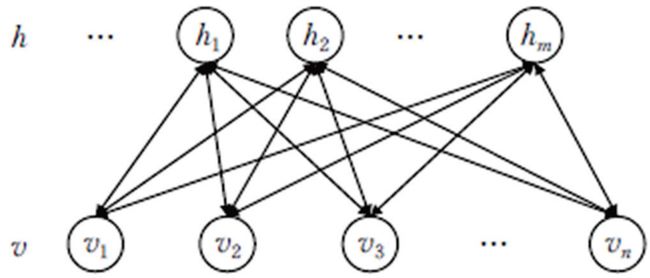


Figure 4. RBM model structure diagram.

For a set of states (v, h) given by the visible layer and the hidden layer, the energy function of the RBM in this state can be defined as:

$$E_\theta(v,h) = -\boldsymbol{a}^{\mathrm{T}}v - \boldsymbol{b}^{\mathrm{T}}h - h^{\mathrm{T}}\boldsymbol{W}v \tag{3}$$

Among them, $\boldsymbol{a},\ \boldsymbol{b},\ \boldsymbol{W}$ are the parameters of the model, a represents the bias value of the visible layer, $\boldsymbol{b}$ represents the bias value of the hidden layer, and $\boldsymbol{W}$ represents the weight matrix between the visible and the hidden layer.

Based on the energy function defined by Equation (3), the joint probability distribution of (h, v) can be obtained:

$$P(v,h|\theta) = \frac{e^{-E(v,h|\theta)}}{\sum_{v,h} e^{-E(v,h|\theta)}} \tag{4}$$

where

$$\theta = \{\boldsymbol{W},\ \boldsymbol{a},\boldsymbol{b}\} \tag{5}$$

With the joint probability distribution of (h, v), the marginal distributions of h and v can be calculated separately. In practical applications, the probability distribution of the visible layer is more important. The distribution function is expressed as:

$$P(v|\theta) = \frac{\sum_h e^{-E(v,h|\theta)}}{\sum_{v,h} e^{-E(v,h|\theta)}} \tag{6}$$

The process of training the RBM is essentially to find a set of θ so that the output of the visible layer is as similar as possible to the distribution of the training sample S. In order to achieve this goal, the likelihood function is defined:

$$\ln L_{\theta,S} = \sum_{i=1}^{n_s} \ln P(v_i) \tag{7}$$

A set of optimal parameters θ can be obtained by seeking the maximum value of the likelihood function, where v_i is the i-th neuron in the visible layer, and n_s is the number of training samples. In order to calculate the maximum value of Equation (7), it is necessary to derive it, and the joint probability distribution $P(h|v)$ appears in the derivation result, because $P(h|v)$ is difficult to calculate directly. Therefore, Gibbs sampling is used to approximate it.

The RBM has good properties. For example, given the state of neurons in the visible layer, the activation conditions of neurons in the hidden layer are independent. Similarly, given the state of neurons in the hidden layer, the activation conditions of neurons in the visible layer are also independent. Therefore, as long as the RBM model is designed reasonably, any discrete probability distribution can be fitted (Hinton 2002). Based on the above properties of the RBM, when the state of the neurons in the visible layer is known, the activation probability of any j-th hidden-layer neuron can be obtained:

$$P\left(h_j = 1 \middle| v, \theta\right) = \sigma\left(\sum_i v_i \boldsymbol{W}_{ij} + b_j\right) \tag{8}$$

where σ is the sigmoid activation function. Since the RBM has only two layers, after obtaining the state of the neurons in the hidden layer, the state of the neurons in the visible layer can be calculated in the same way. The activation probability of the i-th neuron in the visible layer is:

$$P(v_i = 1|h, \theta) = \sigma\left(\sum_j h_j \boldsymbol{W}_{ij} + a_i\right) \tag{9}$$

Repeating k times using Equations (8) and (9) can express the k-th sampling result, and obtain an approximate joint probability distribution $\widetilde{P}(h|v)$, and the RBM training can be completed after m cycles.

Based on the above-mentioned RBM training method, the problem of training a DAE can be simplified to the problem of training multiple RBMs. In the actual training process, the contrast divergence (CD) algorithm is used to train an RBM, and the sampling results of a few Gibbs samples can be used to approximate the joint probability distribution of (h, v). It greatly reduces the training cost for an RBM.

The trained RBMs are stacked into a DAE network as shown in Figure 3b. At this time, the DAE network weights have not reached the optimal state, and the weights need to be fine-tuned. The specific method is that the original stock data $\boldsymbol{X}$ is input to the DAE, and the error back propagation algorithm is used to fine-tune the network to minimize the error of the output decoded vectors $\widetilde{X}$ and $\boldsymbol{X}$. After fine-tuning, the weight in the DAE network reaches the optimal value, and the output of the m-th layer encoder in the middle of the network is a share-coded sequence $\boldsymbol{X}'$ after dimensionality reduction of ticket data.

3.2. Stock Prediction Based on BP Neural Network

A BP neural network is an algorithm based on delta learning rules. Its principle is to use the method of gradient descent to transmit the error generated by each training forward, update the weight and bias value of each layer in turn, repeat iteratively, and finally achieve the iteration ends after the convergence condition. The BP neural network uses a non-linear activation function, so it has good non-linear fitting ability. The stock system is a complex non-linear dynamic system. Its price trend has strong volatility, and there are many factors that affect the price. The linear model cannot solve the stock prediction problem well. Therefore, the neural-network-based model is very suitable for the stock system.

This paper uses a BP neural network to build a regression model to predict stock prices. The model is divided into three layers: input layer, hidden layer and output layer. The neural network structure is shown in Figure 5.

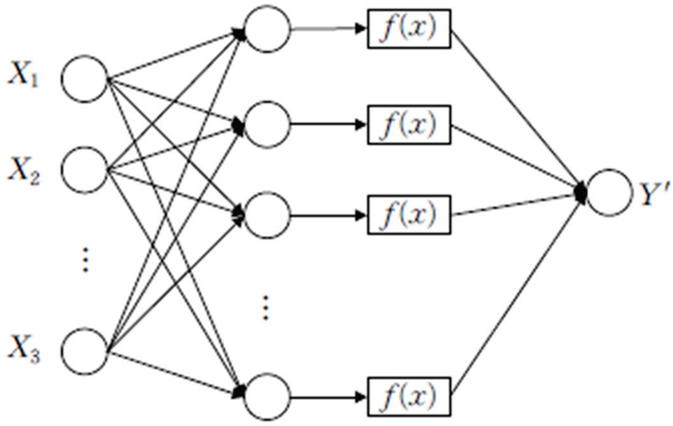


Figure 5. Illustration of BP neural network architecture.

The input of the model is the result $\boldsymbol{X}'$ of the stock data dimensionality reduction of the previous day, and the output is the predicted price $\boldsymbol{Y}'$ of the stock from the next day, where $f(x)$ is the activation function.

3.3. Algorithm Description

The proposed algorithm flow to predict the stock price is as follows in Algorithm 1.

Algorithm 1 Proposed algorithm for stock price prediction

1: For num in layer
2: Initialize RBM (num) weight $\boldsymbol{W}$ and $\boldsymbol{b}$ offset value
3: Enter stock data $\boldsymbol{X}$
4: for I in epoch1
5: Use the method in Section 3.1 to train RBM(num)
6: $\boldsymbol{W}$.append($\boldsymbol{W}$), $\boldsymbol{b}$.append($\boldsymbol{b}$)
7: Return $\boldsymbol{W}$, $\boldsymbol{b}$
8: Use the trained RBM to construct a DAE with a symmetrical structure
9: Load $\boldsymbol{W}$ and $\boldsymbol{b}$ into the corresponding DAE network
10: for i in epoch2
11: Input stock data $\boldsymbol{X}$ to DAE network to fine-tune network parameters
12: End
13: Input stock data $\boldsymbol{X}$ to the trained DAE network
14: Obtain the encoding result $\boldsymbol{X}'$ of the intermediate layer encoder
15: Divide $\boldsymbol{X}'$ into training and test set
16: Build BP neural network
17: For i in epoch3
18: Input the training set to train the BPNN
19: End
20: The BP neural network uses the test set to predict the stock price $\boldsymbol{Y}'$.

Steps (1)~(7) of the proposed algorithm are the process of training the RBM layer by layer. Steps (8)~(12) are the fine-tuning process of the DAE network, and steps (13)~(15) are the code reduction of the original stock data. In the process of dimensioning, steps (16)~(19) are the training process of BP neural network, and step (20) is the prediction process.

4. Experimental Results

4.1. Data Set

This article uses the real transaction data of Kweichow Moutai (600519) as the data set. The data set contains 48 dimensions of data, including commonly used indicators in technical analysis. They are OBV (OBV.OBV, OBV.MAOBV), HSL (HSL.HSL, HSL.MAHSL, VRSI.RSI1), VRSI (VRSI.RSI2, VRSI.RSI3), PSY (PSY.PSY, PSY.PSYMA), WAD (WAD.WAD,

WAD.MAWAD), VR (VR.VR, VR.MAVR), CYR (CYR.CYR, CYR.MACYR), CCI (CCI.CCI), RSI (RSI.RSI1, RSI.RSI2, RSI.RSI3), BIAS (BIAS.BIAS1, BIAS.BIAS2, BIAS.BIAS3), CYF (CYF.CYF), DKX (DKX.DKX, DKX.MADKX), MARSI (MARSI.RSI10, MARSI.RSI6), DMA (DMA.DIF, DMA.DIFMA), ATR (ATR.MTR, ATR.ATR), DPO (DPO.DPO, DPO.MADPO), PAVE (PAVE.CV, PAVE.MCV, PAVE.DIFF), CYE (CYE.CYEL, CYE.CYES), EMV (EMV.EMV, EMV.MAEMV), MACD (MACD.DIF, MACD.DEA, MACD.MACD), MTM (MTM.MTM, MTM.MTMMA) and ROC (ROC.ROC, ROC.MAROC). There are a total of 1156 trading days in the data set; 900 pieces of data are divided into the training set, and 256 pieces of data are used as the test set.

Due to the different dimensions of the data of different stock indicators, the data with larger dimensions has a greater impact on the final result, whereas the data with a smaller dimension has almost no effect on the original result. Data for different dimensions will cause great interference to both the dimensionality reduction stage and the prediction stage, so it is necessary to unify the data to the same order of magnitude. This article uses the linear function normalization ($\mathrm{Min} - \mathrm{Max}$ scaling) method; the normalization method is:

$$x' = \frac{x - \min}{\max - \min} \tag{10}$$

Through the normalization operation, the data is compressed to between (0, 1), which eliminates the influence of a large amount of dimensional data on data dimensionality reduction, and prevents the phenomenon of gradient saturation of neurons when the logistic and tanh activation functions are used.

4.2. Comparison of DAE Dimensionality Reduction Effects of Different Depths

The autoencoder can be trained as a deep autoencoder with multiple hidden layers. Generally, the more layers of the deep autoencoder, the more abstract the extracted features, and the encoded result represents a higher layer in the data. In order to compare the impact of encoding results of different depths on the prediction results, this paper designs five DAEs, and the number of layers is gradually increased from two to six. The results of the five DAE encodings are respectively used for stock price prediction, and the prediction results are shown in Table 1.

Table 1. The prediction effect of using different depths of DAE encoding.

DAE Network Structure	Number of Network Layers	Training Error (MSE)	Test Error (MSE)
48-5	2	0.00051 ± 0.00021	0.0054 ± 0.0009
48-24-5	3	0.00041 ± 0.00016	0.0030 ± 0.0004
48-24-12-5	4	0.00036 ± 0.00015	0.0025 ± 0.0007
48-30-20-10-5	5	0.00047 ± 0.00012	0.0034 ± 0.0005
48-40-30-20-10-5	6	0.00670 ± 0.00170	0.0220 ± 0.0040

From the prediction results in Table 1, it can be seen that the results of using different depths of DAE encoding sequences to predict stock prices are different. Among them, the data after the use of four-layer-structure DAE dimensionality reduction is more effective in stock prediction, and its training error and test errors are all the smallest.

In order to further measure the encoding effect of encoders with different depths, this paper obtains the results $\widetilde{X}$ after DAE decoding of different depths, calculates the Euclidean distance between X and $\widetilde{X}$ to find the reconstruction error, and uses the reconstruction error to measure the degree of information loss in the DAE encoding–decoding process. The calculation results are shown in Table 2.

Table 2. Reconstruction error of DAE at different depths.

Network Structure	Reconstruction Error
48-5	20.60
48-24-5	17.10
48-24-12-5	15.60
48-30-20-10-5	16.70
48-40-30-20-10-5	19.79

The formula for calculating the Euclidean distance of a vector in a multidimensional space is:

$$d_{12} = \sqrt{\sum_{k=1}^{n} (x_{1k} - x_{2k})^2} \tag{11}$$

Among them, n represents the dimension of matrix $\boldsymbol{X}$.

Under normal circumstances, as the number of layers of the DAE deepens, its ability to abstract and generalize becomes stronger. A shallow DAE can complete the dimensionality reduction of the data through only one encoding, and some important information will be lost during the encoding process, so the reconstruction error is relatively large. The deep DAE extracts features layer by layer, which can more completely retain the important features in the data, with less information loss, so the reconstruction error is small. However, the depth of the DAE is not at its maximum. As the number of encodings increases, the extracted features will become more and more abstract, and the detailed information in the data will be lost, so the reconstruction error will be larger.

4.3. Comparison with Other Dimensionality Reduction Methods

Principal component analysis (PCA) and factor analysis (FA) are two commonly used methods in multivariate analysis, which are usually used for dimensionality reduction analysis of observation samples.

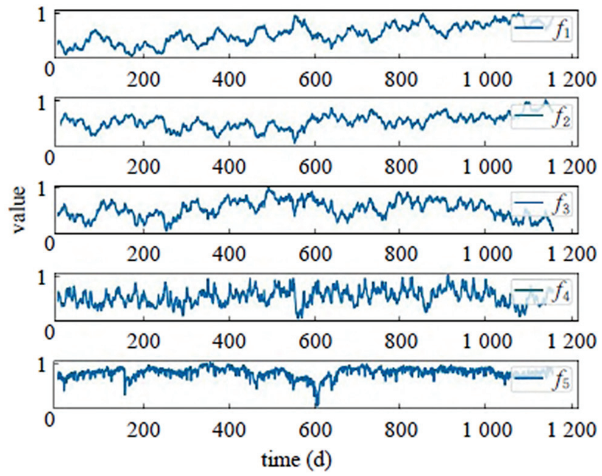
PCA is a linear dimensionality reduction method. The principle is to remove the data in the original high-dimensional space from the data with smaller variance dimensions and simplify it into low-dimensional data, and then rotate the low-dimensional data into the low-dimensional coordinate system. In the coordinate system, the data of each dimension is a linear combination of the original data (Bakshi 2010), the purpose of which is to make the reconstructed data closer to the original data. Therefore, PCA is also called linear autoencoding.

FA is also a method of linear dimensionality reduction. Its principle is to synthesize variables into a few common factors according to their correlation, and replace a group with a strong correlation with one factor. The correlation of variables between different groups is weak, and the correlation of variables in the same group is strong. Similar to the idea of clustering, the original variables can be expressed as a linear combination of common factors (Kaiser 2016).

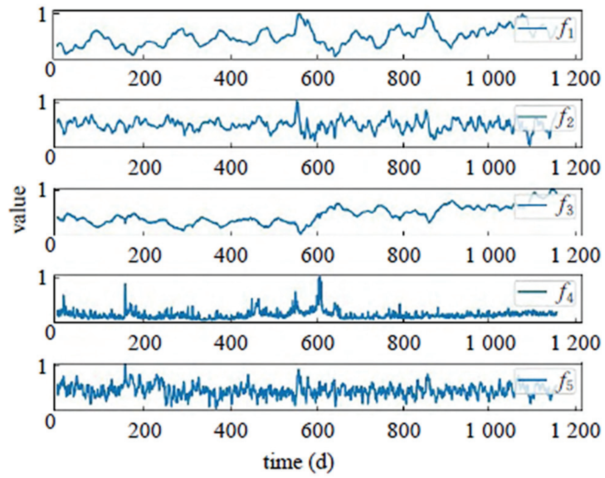
Different from the above two methods, the DAE used in this article is a non-linear dimensionality reduction method. The reason it is called non-linear is that a non-linear activation function can be used between the layers of the DAE, such as sigmoid, tanh, relu, etc., to achieve non-linear mapping of input data and output data. As the number of DAE layers increases, the more times it performs non-linear mapping, and the more abstract the extracted features.

Using the PCA method to reduce dimensionality, when the number of principal components is five, the cumulative variance contribution is 88.27%, so the five-dimensional dimension is selected as the final dimension after dimensionality reduction. FA and DAE are based on the dimension after PCA dimensionality reduction. It is also reduced to five dimensions, which is convenient for comparing the dimensionality reduction effects of different methods. Among them, DAE uses the best four-layer structure model in Section 4.2.

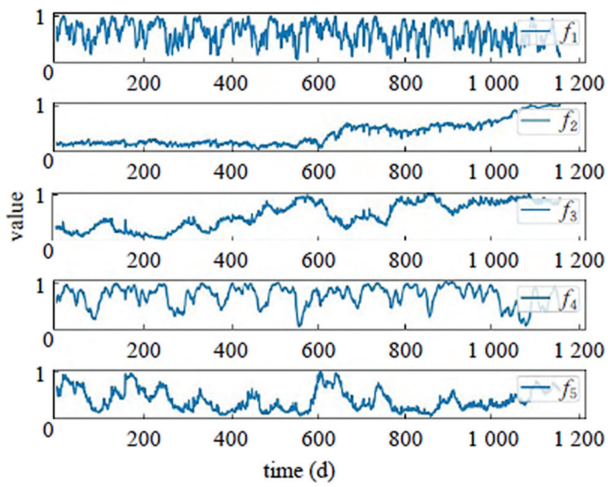
Using the above-mentioned dimensionality reduction methods to reduce the stock data to five dimensions, the coding sequence of each dimension is shown in Figure 6.



(**a**) The result after using PCA for dimensionality reduction



(**b**) The result after dimensionality reduction using FA



(**c**) The result after dimensionality reduction using DAE

Figure 6. Coding sequence of stock data after dimensionality reduction.

Intuitively analyzing the sequence encoded by the three methods in Figure 6, it can be found that the characteristics of the encoded sequence after using PCA dimensionality reduction are not obvious, and it does not reflect the basic situation of the stock trend. The coding sequence after dimensionality reduction using FA and DAE contains the characteristics of stock price trends, among which DAE's stock price trend characteristics are the most obvious (as shown by f_2 in Figure 6c), which can be inferred because the effect of dimensionality reduction using FA and DAE data forecast is better.

In order to verify the above statement, the sequences obtained using the above three methods are used as the input of the BP neural network, and the prediction results are generated using the BP neural network calculation. The prediction results are shown in Figure 7.

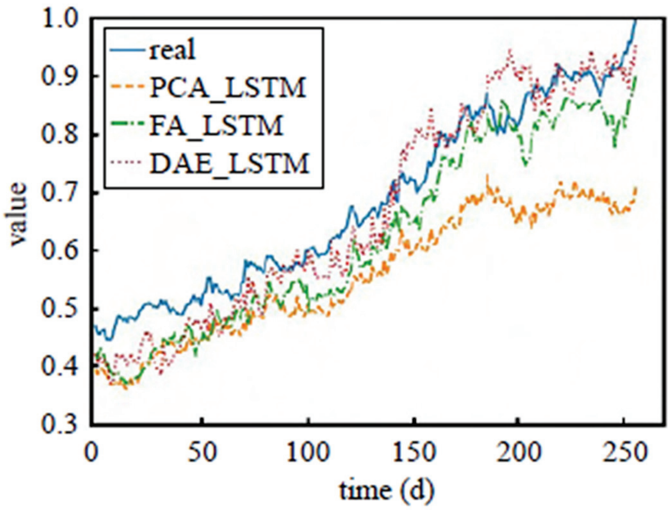


Figure 7. The results of predicting stock prices with different coding sequences.

In order to measure the error between the predicted result and the real result, this article uses three error evaluation indicators, namely MSE (Mean Square Error), MRE (Mean Relative Error) and MAE (Mean Absolute Error). The calculation results of the three errors are shown in Table 3. Analyzing the data in Table 3, we can see that the proposed algorithm has a better prediction effect, whereas PCA-BP and FA-BP have poor prediction results. The single-day rise and fall of the Chinese stock market is only 10%, and the relative error of using the PCA dimensionality reduction code to predict the stock price has reached 17.9%, far exceeding the upper limit of the stock's rise and fall. Therefore, the PCA dimensionality reduction code is used. Predicting stock prices has no practical significance. The data after dimensionality reduction using FA barely meets the practical requirements, but still has a large relative error. The proposed model is used to reduce the dimensionality of the stock data. The reduced data has a higher prediction accuracy and meets practical requirements.

Table 3. Errors in predicting stock prices with different coding sequences.

Algorithm	Error		
	MAE	**MSE**	**MRE**
FA_BP	0.062	0.0045	9.72%
PCA_BP	0.127	0.02	17.9%
Proposed DAE_BP	0.043	0.0025	6.94%

In order to explore the reasons why different encoding methods have different effects on the prediction results, this article further analyzes the encoded sequence. Stock sequence is a time series with strong volatility. From a theoretical point of view, the variables used to predict stock prices also need to have similar strong volatility in order to predict stock prices well. Variance is used in statistics to measure the degree of dispersion of a sequence. The larger the variance, the greater the volatility of the sequence. Similarly, the smaller the variance, the smaller the volatility of the sequence. Therefore, this paper calculates the variance of the coded sequence after dimensionality reduction by various dimensionality reduction methods, which is used to measure the degree of fluctuation of the coded sequence after dimensionality reduction. Among them, Di represents the i-th dimension of the coding sequence.

Through calculation, in Table 4, the variance of the stock price sequence is 0.061. Analysis of the data in Table 4 shows that the variance of the sequence after DAE dimensionality reduction is the closest to the variance of the stock price, so the following conclusions can be drawn. The DAE should be used to compare stock data to perform dimensionality reduction, and the FA and PCA methods should be used for comparison. When the data is reduced to the same dimension, the DAE can retain more useful information in the original data, and the extracted features are more in line with the main features in the original data.

Table 4. The variance of the coding sequence of stock data after dimensionality reduction.

Algorithm	Variance				
	***D*1**	***D*2**	***D*3**	***D*4**	***D*5**
FA	0.049	0.027	0.039	0.030	0.015
PCA	0.039	0.015	0.038	0.009	0.018
DAE	0.053	0.073	0.086	0.051	0.056

4.4. Comparison with Different Forecasting Methods

This article compares the prediction results of the DAE-BP model with the three commonly used models of Multilayer Perceptron (MLP), Support Vector Regression (SVR) and Multiple Linear Regression (MLR). The MLP, SVR and MLR use stocks without dimensionality reduction. The index data predicts the price, and the prediction results are shown in Table 5. From the data in Table 5, it can be seen that although the MLR has the shortest running time, its prediction effect is the worst. Compared with other prediction models, the proposed algorithm has better performance in terms of prediction accuracy and running time.

Table 5. Comparison of prediction results of different methods.

Algorithm	Running Time	Error		
		MAE	**MSE**	**MRE**
SVR	9.16 ms	0.056	0.0039	7.71%
MLR	0.13 ms	0.160	0.029	22.9%
MLP	0.44 ms	0.063	0.0047	9.98%
Proposed	0.29 ms	0.043	0.0025	6.94%

5. Conclusions

This paper designs a stock prediction model based on the DAE-BP neural network. Since the original indicator data of stocks has a high dimensionality and there is multiple collinearity between the indicators, directly using the original indicators to predict stock prices will greatly increase the complexity of the model and computational overhead. Thus, this article uses a deep autoencoder to reduce the dimensionality of stock data. The effect of DAE dimensionality reduction is related to the number of layers. If the number of layers is too deep or too shallow, the reconstruction error will increase. The DAE with a four-layer structure has the best dimensionality reduction effect. This article also compares the dimensionality reduction effects of the three dimensionality reduction methods of DAE, PCA and FA. When the data is reduced to the same dimension, the data after DAE dimensionality reduction can retain the more essential characteristics of the data and use it for stocks. It has high accuracy when forecasting. In comparison with other prediction models, DAE-BP has more advantages in terms of running time and prediction accuracy.

Using the model designed in this paper to predict stock prices still produces a large error; the relative error reaches 6.94%, and although this is less than the single-day limit of stocks, it is still misleading in practical applications. Therefore, the next step is to adjust and optimize the prediction part of the model to further reduce the prediction error and make the model more practical.

Author Contributions: Formal analysis, Y.A.A.; Investigation, D.M.A.; Methodology, S.Z.A.-o.; Project administration, Y.A.A.; Resources, D.M.A.; Software, A.A.A.; Validation, A.A.A.; Visualization, A.A.A.; Writing—review & editing, S.Z.A.-o. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Researchers Supporting Project (No. RSP-2021/395), King Saud University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

Araki, Shoko, Tomoki Hayashi, Marc Delcroix, Masakiyo Fujimoto, Kazuya Takeda, and Tomohiro Nakatani. 2015. Exploring multichannel features for denoising-autoencoder-based speech enhancement. Paper presented at IEEE International Conference on Acoustics, Speech and Signal Processing, South Brisbane, QLD, Australia, April 19–24, pp. 116–20.

Babić, Karlo, Sanda Martinčić-Ipšić, and Ana Meštrović. 2020. Survey of neural text representation models. *Information* 11: 511. [CrossRef]

Bakshi, Bhavik. 2010. Multiscale PCA with application to multivariate statistical process monitoring. *Aiche Journal* 44: 1596–610. [CrossRef]

Brandt, Andreas, and Manfred Brandt. 2004. On the two-class M/M/1 system under preemptive resume and impatience of the prioritized customers. *Queuing Systems* 47: 147–68. [CrossRef]

Brandt, Andreas, and Manfred Brandt. 2013. Workload and busy period for M/GI/1 with a general impatience mechanism. *Queuing Systems* 75: 189–209. [CrossRef]

Chopra, Ritika, and Gagan Deep Sharma. 2021. Application of artificial intelligence in stock market forecasting: A critique, review, and research agenda. *Journal of Risk and Financial Management* 14: 526. [CrossRef]

Craighead, Christopher W., Kirk R. Karwan, and Janis L. Miller. 2009. The effects of severity of failure and customer loyalty on service recovery strategies. *Production and Operations Management* 13: 307–21. [CrossRef]

Gao, Shenghua, Yuting Zhang, Kui Jia, Jiwen Lu, and Yingying Zhang. 2015. Single sample face recognition via learning deep supervised autoencoders. *IEEE Transactions on Information Forensics and Security* 10: 2108–18. [CrossRef]

Garnet, Ofer, Avishai Mandelbaum, and Martin. I. Reiman. 2002. Designing a call center with impatient customers. *Manufacturing & Service Operations Management* 4: 208–27.

Glorot, Xavier, and Yoshua Bengio. 2010. Understanding the difficulty of training deep feedforward neural networks. *Journal of Machine Learning Research* 9: 249–56.

Hai Nguyen, Thien, Kiyoaki Shirai, and Julien Velcin. 2015. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications* 42: 9603–11. [CrossRef]

Hecht, Amelie A., Crystal L. Perez, Michele Polascek, Anne N. Thorndike, Rebecca L. Franckle, and Alyssa J. Moran. 2020. Influence of food and beverage companies on retailer marketing strategies and consumer behavior. *International Journal of Environmental Research and Public Health* 17: 7381. [CrossRef]

Hinton, Geoffrey. 2002. Training products of experts by minimizing contrastive divergence. *Neural Computation* 14: 1771–800. [CrossRef] [PubMed]

Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. 2006. Reducing the dimensionality of data with neural networks. *Science Journal* 313: 504–7. [CrossRef]

Kaiser, Henry F. 2016. The application of electronic computers to factor analysis. *Educational & Psychological Measurement* 20: 141–51.

Le Roux, Nicolas, and Yoshua Bengio. 2008. Representational power of restricted Boltzmann machines and deep belief networks. *Neural Computation* 20: 1631–42. [CrossRef] [PubMed]

Li, Xiaodong, Haoran Xie, Yangqiu Song, Shanfeng Zhu, Qing Li, and Fu Lee Wang. 2015. Does summarization help stock prediction? News impact analysis via summarization. *IEEE Intelligent Systems* 30: 26–34. [CrossRef]

Li, Xiumin, Lin Yang, Fangzheng Xue, and Hongjun Zhou. 2017. Time series prediction of stock price using deep belief networks with intrinsic plasticity. Paper presented at Control and Decision Conference, Chongqing, China, May 28–29, pp. 1237–42.

Lu, Yina, Andrés Musalem, Marcelo Olivares, and Ariel Schilkrut. 2012. Measuring the effect of queues on customer purchases. *Measurement Science* 59: 1743–63. [CrossRef]

Prachyachuwong, Kittisak, and Peerapon Vateekul. 2021. Stock trend prediction using deep learning approach on technical indicator and industrial specific information. *Information* 12: 250. [CrossRef]

Rundo, Francesco, Francesca Trenta, Agatino Luigi di Stallo, and Sebastiano Battiato. 2019. Machine learning for quantitative finance applications: A survey. *Applied Sciences* 9: 5574. [CrossRef]

Salakhutdinov, Ruslan, and Geoffrey Hinton. 2009. Semantic Hashing. *International Journal of Approximate Reasoning* 50: 969–78. [CrossRef]

Van Velthoven, J., Benny Van Houdt, and Chris Blondia. 2005. Response time distribution in a D-MAP/PH/1 queue with general customer impatience. *Stochastic Models* 21: 745–65. [CrossRef]

Wang, Yao, Wan-Dong Cai, and Peng-Cheng Wei. 2016. A deep learning approach for detecting malicious javascript code. *Security & Communication Networks* 9: 1520–34.

Xu, Jun, Lei Xiang, Qingshan Liu, Hannah Gilmore, Jianzhong Wu, Jinghai Tang, and Anant Madabhushi. 2016. Stacked sparse autoencoder for nuclei detection of breast cancer histopathology images. *IEEE Transactions on Medical Imaging* 35: 1193–204. [CrossRef] [PubMed]

Zohar, Ety, Avishai Mandelbaum, and Nahum Shimkin. 2002. Adaptive behavior of impatient customers in tele-queues: Theory and empirical support. *Management Science* 48: 566–83. [CrossRef]

risks

MDPI

Article

An Overview on the Landscape of R Packages for Open Source Scorecard Modelling

Gero Szepannek

Institute of Applied Computer Science, Stralsund University of Applied Sciences, Zur Schwedenschanze 15, 18435 Stralsund, Germany; gero.szepannek@hochschule-stralsund.de

Abstract: The credit scoring industry has a long tradition of using statistical models for loan default probability prediction. Since this time methodology has strongly evolved, and most of the current research is dedicated to modern machine learning algorithms which contrasts with common practice in the finance industry where traditional regression models still denote the gold standard. In addition, strong emphasis is put on a preliminary binning of variables. Reasons for this may be not only the regulatory requirement of model comprehensiveness but also the possibility to integrate analysts' expert knowledge in the modelling process. Although several commercial software companies offer specific solutions for modelling credit scorecards, open-source frameworks for this purpose have been missing for a long time. In recent years, this has changed, and today several R packages for credit scorecard modelling are available. This brings the potential to bridge the gap between academic research and industrial practice. The aim of this paper is to give a structured overview of these packages. It may guide users to select the appropriate functions for the desired purpose. Furthermore, this paper will hopefully contribute to future development activities.

Keywords: credit scorecard development; open source; R

Citation: Szepannek, Gero. 2022. An Overview on the Landscape of R Packages for Open Source Scorecard Modelling. *Risks* 10: 67. https://doi.org/10.3390/risks10030067

Academic Editors: Krzysztof Jajuga and Józef Dziechciarz

Received: 16 February 2022
Accepted: 14 March 2022
Published: 18 March 2022

1. Introduction

In the credit scoring industry, there is a long tradition of using statistical models for loan default probability prediction, and domain specific standards were established long before the hype of machine learning. An overview of the historical evolution of credit risk scoring can be found in Kaszynski (2020) and Anderson (2019). A comprehensive description of the corresponding methodology is given in Thomas et al. (2019) and Kaszynski et al. (2020). The different subsequent steps during the scorecard modelling process are worked out in Anderson (2007), Finlay (2012) and Siddiqi (2006) where the latter is closely related to the credit scoring solution as implemented by the SAS Enterprise Miner software[1]. The typical steps in credit risk scorecard modelling refer to the general process definition for data mining as given by KDD, CRISP-DM or SAS's SEMMA (cf. Azevedo and Santos 2008). It turns out that strong emphasis is laid on possibilities for manual intervention after each modelling step. Therefore, functions to summarize and visualize the intermediate results of each single step are of great importance. The typical development steps are denoted by:

1. Binning and Weights of Evidence (Section 3)
2. Preselection of Variables (Section 4)
3. Multivariate Modelling (Section 5)
4. Performance Evaluation (Section 6)
5. Reject Inference (Section 7)

In contrast, the typical scorecard modelling process is rarely taken into account in current academic benchmark studies (for an overview cf. Louzada et al. 2016). An exception is given in Bischl et al. (2016), where both approaches are covered. A reason for this gap between academic research and business practice may be due to the lack of open source frameworks for scorecard modelling.

Although several commercial software companies, such as SAS, offer specific solutions for credit scorecard modelling (cf. footnote 1), explicit packages for this purpose in R have been missing for a long time and in the CRAN task view on Empirical Finance[2] the explicit topic of scorecard modelling is not covered. A "Guide to Credit Scoring in R" can be found among the CRAN contributed documentations (Sharma 2009) being dedicated to describing the application of different (binary) classification algorithms to credit scoring data rather than to emphasizing the common subsequent modelling stages that are typical for scorecard modelling processes. This can be a result of the circumstances: at that time, no explicit packages were available in R for undertaking this kind of task.

In recent years this has changed, and several packages have been submitted to CRAN with the explicit scope of credit risk scorecard modelling, such as `creditmodel` (Fan 2022), `scorecard` (Xie 2021), `scorecardModelUtils` (Poddar 2019), `smbinning` (Jopia 2019) `woeBinning` (Eichenberg 2018), `woe` (Thoppay 2015), `Information` (Larsen 2016), `InformationValue` (Prabhakaran 2016), `glmdisc` (Ehrhardt and Vandewalle 2020), `glmtree` (Ehrhardt 2020), `Rprofet` (Stratman et al. 2020) and `boottol` (Schiltgen 2015).

Figure 1 gives an overview of the packages and their popularity in terms of the number of their CRAN downloads as well as their activity and existence as observable by their CRAN submission dates. It can be seen that the packages `smbinning`, `InformationValue` and `Information` are among the most popular, and they have been available for quite some time. Another popular toolbox is provided by the package `scorecard`, which has been frequently updated in the recent past as has also happened with the package `creditmodel`.

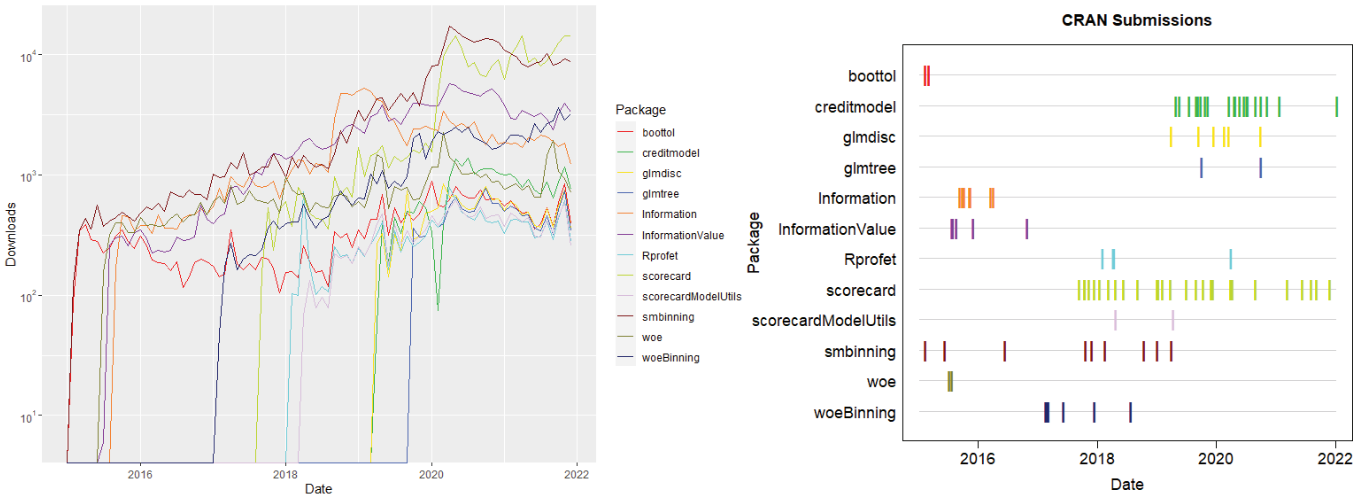


Figure 1. CRAN release activity and download statistics (as returned by `cranlogs`, Csárdi 2019) of packages available on CRAN.

In addition, some packages are available on Github but not on CRAN, such as `creditR` (Dis 2020), `riskr` (Kunst 2020), and `scoringTools` (Ehrhardt 2018).

As all of these packages have become available during the last few years, this paper is dedicated to the question of whether recent developments have made it possible to perform all steps of the entire scorecard development process within R. For this reason, the presentation of the package landscape will be guided by these steps. One section will be dedicated to each stage. In each section, the available packages will be presented together with their advantages and disadvantages. The aim of the paper is to give a structured overview of existing packages. It may guide users in selecting the appropriate functions for the desired purpose by working out pros and cons of existing functions.

As an open source programming language, the R universe is extended by a large community with currently more than 19,000 contributed packages. It is impossible for a single user to know all of them which in turn leads to some redundant development activities by programming multiple solutions for the same task. Moreover, sometimes contributed packages for a similar purpose provide different desirable functionalities but are not compatible with each other because they rely on different kind of input objects. By working out the pros and cons of the functions provided by the aforementioned packages, this paper aims to analyze existing gaps and provide several remedies in the supplementary code (cf. corresponding footnotes).

Note that this paper focuses on the open source statistical programming language R. Within the data science industry other open source frameworks such as python have increased in popularity during the last few years, which is beyond the scope of this paper. For interested readers, some useful python functions for scorecard development are mentioned in Kaszynski et al. (2020), and some websites are dedicated to this purpose[3,4]. In particular, a python implementation of the `scorecard` R package (Xie 2021) is available[5] which means that some of the results as worked out in this paper are directly transferable into the python world. Nonetheless as R denotes the lingua franca of statistics Ligges (2009), it provides access to a huge number of contributed packages and functionalities from the field of statistics outside the aforementioned ones. For this reason, the paper concentrates on R and investigates whether scorecard development can be improved by access to other already existing packages that have initially been designed for other purposes but can improve the analyst's life. If available, such functionalities will also be mentioned in the corresponding sections.

Note that, traditionally, logistic regression is used for credit risk scorecard modelling despite the current hype around modern machine learning methods as they are provided by frameworks such as e.g., `mlr3` (Lang et al. 2019, 2021) or `caret` (Kuhn 2008, 2021). Studies have investigated potential benefits from using modern machine learning algorithms (Baesens et al. 2002; Bischl et al. 2016; Lessmann et al. 2015; Louzada et al. 2016; Szepannek 2017), but regulators and the General Data Protection Regulation (GDPR) require models to be understandable (cf. Financial Stability Board 2017; Goodman and Flaxman 2017). The latter issue can be addressed by methodologies of explainable machine learning (for an overview Bücker et al. 2021), e.g., using frameworks as provided by the packages `DALEX` (Biecek 2018) or `iml` (Molnar et al. 2018) while taking into account to what extent a model actually is explainable (Szepannek 2019). It further turned out that the use of current state-of-the-art ML algorithms is not necessarily always beneficial in the credit scoring context (Chen et al. 2018; Szepannek 2017), and they should be rather carefully analyzed in each specific situation, rather than relying on preferred preferred models (Rudin 2019). For this reason this paper focuses on the traditional way of scorecard modelling as briefly described above.

2. Data

Probably the most common credit scoring data are the German Credit Data provided by Hoffmann (1994) that are contained in the UCI Machine Learning Repository (Dua and Graff 2019). The data consist of 21 variables: a binary target (`creditability`) and 13 categorical as well as seven numeric predictors, and 1000 observations in total with 300 defaults (`level == ''bad''`) and 700 nondefaults (`level == ''good''`). The data are provided by several R packages such as `klaR` (Roever et al. 2020), `woeBinning`, `caret` or `scorecard`. For the examples in this paper, the data from the `scorecard` package are used where in addition the levels of the categorical variables such as `present.employment.since`, `other.debtors.or.guarantors`, `job` or `housing` are sorted according to their expected order w.r.t. credit risk. Note that Groemping (2019) compared the data from the UCI repository to the original papers and made a corrected version of itavailable[6] (cf. also Szepannek and Lübke 2021). Other (partly simulated) example data sets (amongst others loan data of the peer-to-peer lending company Lending Club[7]) are contained within the packages `creditmodel`, `scoringTools` and `smbinning` and `riskr`.

It is common practise to use separate validation data which are not used for model training but only for validation purposes. The manual interventions between the different modelling steps do not allow for repetitive resampling strategies such as k-fold cross validation or bootstrapping for model validation as they are, e.g., provided by the package mlr3 (see Section 6). Instead, usually one single holdout set is used. The package scorecard has a function split_df() that splits data according to a prespecified percentage into training and validation sets. For the examples in the remainder of the paper, the following data are used:

```
### example 1: load data
library(scorecard)
data(germancredit)
# transform character variable purpose into factor
germancredit$purpose <- as.factor(germancredit$purpose)

tv <- split_df(germancredit, y = ''creditability'', ratio = c(0.7, 0.3),
seed = 42, no_dfs = 2, name_dfs = c(''train'', ''valid''))

train <- tv$train
valid <- tv$valid

# several packages require the target variables to take values 0/1
train2 <- train; valid2 <- valid
train2$creditability <- as.integer(train2$creditability == ''good'')
valid2$creditability <- as.integer(valid2$creditability == ''good'')

# the package creditmodel does not support variables of type Factor
train3 <- as.data.frame(train2)
valid3 <- as.data.frame(valid2)
for (j in which(sapply(train3[,-21], is.factor))) {
  train3[,j] <- as.character(train3[,j])
  valid3[,j] <- as.character(valid3[,j])
}
```

Note that some of the packages (smbinning, woe, creditR, riskr, glmdisc, scoringTools, scorecardModelUtils and creditmodel[8]) do require the target variable to take only values 0 and 1 as in the example's data sets train2 and valid2. Although this is of course easily obtained, the package scoringModelUtils contains a function fn_target() that does this job and replaces the original target variable with a new one of name Target.

3. Binning and Weights of Evidence

3.1. Overview

Binning of numeric variables is often considered the most relevant step in scorecard development. An initial automatic algorithm-based binning is manually checked and—if necessary—modified by the analyst variable by variable. On the one hand, this is a very time-consuming task, but, on the other hand, this ensures the dependencies between the explanatory variables and the target in the final model to be plausible and helps detect sampling bias (Verstraeten and den Poel 2005). Furthermore, it allows modelling of nonlinear dependencies by linear logistic regression in the subsequent Multivariate Modelling step. The loss of information by aggregation turned out to be comparatively small while this kind of procedure does not take into account for interactions between several variables and the target variable (Szepannek 2017). The identification of relevant interactions typically needs a lot of business experience, and Sharma (2009) suggests using random forests to identify potential interaction candidates.

3.2. Requirements

It is important to note that binning corresponds not just to exploratory data analysis, but its results have to be considered an integral part of the final model, i.e., the resulting preprocessing has to be applied to new data to be able use the resulting scorecard for business purposes. For this reason, important requirements on an implementation of the binning step are the possibility to: (i) store the binning results for all variables, and (ii) apply the binning to new data with some kind of `predict()` function.

The importance of an option to: (iii) manually modify an initial automatic binning has already been emphasized. This leads to the requirement for a separate function to manipulate an object that stores the binning results. In order to support this: (iv) summary tables and (v) visualizations of the intermediate binning results are helpful. In addition, application of binning in practice has to: (vi) deal with missing data or new levels of categorical variables that did not occur in the training data as, e.g., by regulation it may be required that holding back information (and the resulting missing values) must not lead to an improvement of the final score. Both missing data and new levels should be taken into account by the implemented binning function.

Often, binning is followed by subsequent assignment of numeric weights of evidence to the factor levels x of the binned variable which are given by:

$$WoE(x) = \log\left(\frac{f(x|y=1)}{f(x|y=0)}\right). \tag{1}$$

Note that just like the bins, the WoEs, as computed on the training data are part of the model. Furthermore, an implementation of WoE computation has to account for potentially occurring bins that are empty w.r.t. the target level $y = 0$ (typically by adding a small constant when computing the relative frequencies $f()$). By construction, WoEs are linear in the logit of the target variable and thus well suited for subsequent use of logistic regression. The use of WoEs is rather advantageous for small data sets, and directly using the bins may increase performance if enough data are available (Szepannek 2017). On the other hand, using WoEs fixes mononty between the resulting scorecard points and the default rates of the bins, such that only the sign of the monotonicity has to be checked. It is also usual to associate binned variables with an information value (IV)

$$IV = \sum_x (f(x|y=1) - f(x|y=0))\, WoE(x) \tag{2}$$

based on the WoEs which describe the strength of a single variable to discriminate between both classes.

3.3. Available Methodology for Automatic Binning

Several packages provide functions for automatic binning based on conditional inference trees (Hothorn et al. 2006) from the package `partykit` (Hothorn and Zeileis 2015): `scorecard::woebin()`, `smbinning::smbinning()`, `scorecardModelUtils::iv_table()` and `riskr::superv_bin()`. The implementation in the `scorecardModelUtils` package merges the resulting bins to ensure monotonicity in default rates w.r.t. with the original variable which might or might not be desired. For the same purpose, the package `smbinning` offers a separate function (`smbinning.monotonic()`). In contrast to all previously mentioned packages, the package `woeBinning` implements its own tree algorithm where either initial bins of similar WoE are merged (`woe.binning()`), or the set of bins is binary split (`woe.tree.binning()`) as long as the IV of the resulting variables decreases (increases) by a percentage less (more) than a prespecified percentage (argument `stop.limit`) while the initial bins are created to be of minimum size (`min.perc.total`). The function `creditmodel::get_breaks_all()` uses classification and regression trees (Breiman et al. 1984) of the package `rpart` (Therneau and Atkinson 2019)[9] to create initial bins. An additional argument, `best = TRUE`, merges these bins subsequently according to different criteria such as the maximum number of bins, the minimum percentage of observations per bin, a threshold for the χ^2 test or odds, a minimum population stability (cf. Section 4) or monotonicity of the default rates across the bins (all of these can be specified by the argument bins_control).

In addition to tree-based binning, the scorecard package offers alternative algorithms (argument method) for automatic binning based on either the χ^2 statistic or equal width or size of numeric variables.

An alternative concept for automatic binning is provided by the package glmdisc, which is explicitly designed to be used in combination with logistic regression modelling for credit scoring (Ehrhardt et al. 2019). The bins are optimized to maximize either AIC, BIC or the Gini coefficient (cf. Section 6) of a subsequent logistic regression model (using binned variables, not WoEs) on validation data (argument criterion=). Second order interactions can also be considered (argument interact = TRUE). Note that this approach is comparatively intense in terms of computation time and does not take variable selection into acount (cf. Section 5).

Some packages do not provide their own implementations of an automatic binning but just interface to discretization functions within other packages. Rprofet::BinProfet() uses the function greedy.bin() of the package binr (Izrailev 2015). The package scoringTools contains a variety of functions (chiM_iter(), mdlp_iter(), chi2_iter(), echi2_iter(), modchi2_iter() and topdown_iter()) which provide interfaces to binning algorithms from the package discretization (Kim 2012). The dlookr package (Ryu 2021), which is primarily designed for exploratory data analysis, has an implemented interface (binning_by()) to smbinning::smbinning().

3.4. Manipulation of the Bins

As outlined before, manual inspection and manipulation of the bins is considered a substantial part of the scorecard development process. Two of the aforementioned packages provide functions to support this. Scorecard::woebin() allows passing an argument breaks_list. Each element corresponds to a variable with manual binning and must be named like the corresponding variable. For numeric variables, it must be a vector of break points, and for factor variables, it must be a character vector of the desired bins given by the merged factor levels, separated by "%,%" (cf. output from Example 3 for variable purpose). In addition, a function scorecard::woebin_adj() allows for an interactive adjustment of bins. The package smbinning provides two functions, smbinning.custom() and smbinning.factor.custom().

Manipulation of the bins should be based on an analysis of the binning results. For this purpose, most of the packages provide result tables on a variable level. The subsequent code example illustrates the step of an initial automatic binning as created by the package scorecard:

```
### Example 2: automatic binning
library(scorecard)
bins <- woebin(train, y = ‘‘creditability’’, method = ‘‘tree’’)

# binning results table for variable purpose
options(digits = 3)
bins$purpose[,c(2,4,5,6,7,8)]

# visualize bins for variable purpose
woebin_plot(bins, x = ‘‘purpose’’, line_value = ‘‘woe’’)
```

```
##                                        bin count_distr neg pos posprob    woe
## 1:                      business%,%car (new)       0.3211 148  79   0.348  0.213
## 2:                                car (used)       0.1089  67  10   0.130 -1.061
## 3:            domestic appliances%,%education       0.0622  24  20   0.455  0.659
## 4:               furniture/equipment%,%others       0.1938  90  47   0.343  0.192
## 5: radio/television%,%repairs%,%retraining       0.3140 165  57   0.257 -0.222
```

The resulting table contains several key figures for each bin such as the distribution (absolute and relative frequency of the samples given the level of the target variable), default rate and the bin's WoE. The information value of the binned variable (cf. Section 4) is given in a column `total_iv` (not shown here).

In addition to summary tables, many packages (`glmdisc`, `riskr`, `Rprofet`, `scorecard`, `smbinning`, `woeBinning`) provide a visualization of the bins on a variable level. Figure 2 (left) shows the binning resulting from code in Example 2 which is similar for most packages. A mosaic plot of the bins, which simultaneously visualizes default rates and the size of the bins, is offered by the package `glmdisc` (Figure 2, right) while the names of the bins after automatic binning are not self-explanatory.

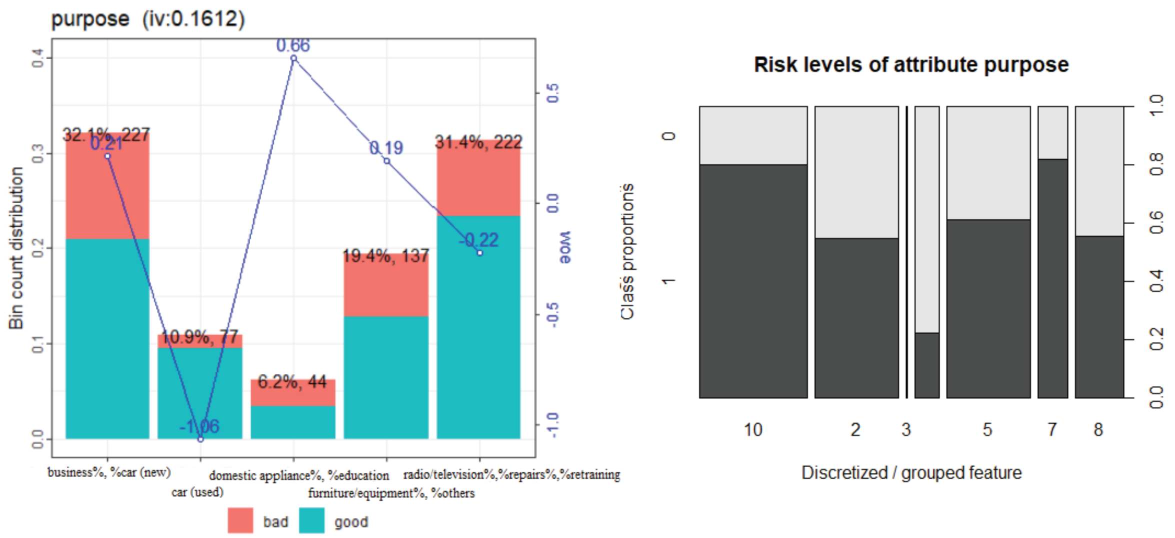


Figure 2. Visualization of the bins for the variable `purpose` as created by the package `scorecard` (**left**) and mosaicplot of the binning result by the package `glmdisc` (**right**).

3.5. Applying Bins to New Data

It has been emphasized that the bins as they are built on training data constitute the first part of a scorecard model. For this reason, it is necessary to store the results of the binning and to have functions to apply it to a data set.

Most of the packages such as `scorecard` (`woebin_ply()`), `smbinning` (`smbinning.gen()` and `smbinning.factor.gen()`), `woeBinning` (`woe.binning.deploy()`), `creditmodel` (`split_bins_all()`), `glmdisc` (`discretize()`) and `scorecardModelUtils` (`num_to_cat()`) provide this functionality. Example 3 illustrates the application of binning results to a data set. Via the `to = ''bin''` argument, either bins or WoEs can be assigned:

```
### Example 3: apply binning to data
train_bins <- scorecard::woebin_ply(train, bins, to = ''bin'')
valid_bins <- scorecard::woebin_ply(valid, bins, to = ''bin'')
```

For `ctree`-based binning (cf. above) a workaround using the `partykit::predict.party()` method for bin assignment can be obtained if the tree model is stored within the results object[10].

More generally, binned variables can be created via the function `cut()` for numeric variables or by using lookup tables for factor variables (cf. Zumel and Mount 2014, p. 23)[11]. It is worth mentioning that several packages (`smbinning` and `riskr`) implement binning only on a single variable level but not simultaneously for several selected variables or all variables of a data frame[12].

3.6. Binning of Categorical Variables

For categorical variables, initially, each level can be considered as a separate bin, but levels of similar default rate and/or meaning could be grouped together. As an additional challenge, there is no natural order of the levels. For these reasons, only some of the packages offer an automatic binning of categorical variables. For example, the package `smbinning` does not offer an automatic merging of levels for factor variables, and its function `smbinning.factor()` only returns the figures similar to the table resulting from Example 2. However, each original level corresponds to only one bin. The bins can be manipulated afterwards via `smbinning.factor.custom()` and further be applied to new data via `smbinning.factor.gen()`. An automatic binning of categorical variables based on conditional inference trees is supported by the packages `riskr` and `scorecard` (`method = "tree"`). Additional merging strategies are provided by the packages `glmdisc` and `creditmodel` (as described above), `scorecard` (`method = "chimerge"`) and `woeBinning` (according to similar WoEs).

Generally, merging levels with a similar default rate should only be done if the level's frequency is large enough to result in a reliable default rate estimate on the sample. By using `woeBinning`'s `woe.binning()` function this can be ensured: Initial bins of a minimum size (`min.perc.total`) are created and smaller factor levels are initially bundled into a positive or negative 'miscellaneous' category according to the sign of the corresponding WoE which is desirable to prevent overfitting. The package `scorecardModelUtils` offers a separate function `cat_new_class()` for this. All levels less frequent than specified by the argument `threshold` are merged together, and a data frame with the resulting mapping table is stored in the output element `$cat_class_new`[13]. The package `creditmodel` provides a function `merge_category` which keeps the `m` most frequent categories and merges all other levels in a new category of name "`other`" but no function is available to apply the same mapping to new data.

Similar to `woeBinning`'s `woe.binning()`, the functions `scorecard::woebin()`[14] and `creditmodel::get_breaks_all()`[15] also merge adjacent levels of similar default rates for categorical variables. An important difference between both implementations consists in how they deal with the missing natural order of the levels and thus the notion of what 'adjacent' means: In `woe.binning()` the levels are sorted according to their WoE before merging. This is not the case for the other two functions where levels are merged along their natural order which is often alphabetical [16]. This might lead to an undesired binning, and as an important conclusion an analyst should think about manually changing the level order for factor variables when working with the package `scorecard`[17].

3.7. Weights of Evidence

Most of the abovementioned packages provide WoEs of the bins within their binning summary tables. To use WoEs within the further modelling steps, it needs a functionality to assign the corresponding WoE value for each bin to the original (/or binned) variables as given by `scorecard::woebin_ply()` (with argument `to = "woe"`), `woeBinning::woe.binning.deploy()` (with argument `add.woe.or.dum.var = "woe"`) and `creditmodel::woe_trans_all()`.

A general way of training, storing and assigning WoEs independently of the package used for binning is given by the function `woe()` in the `klaR` package, probably the first and most comprehensive implementation of WoE computation in R. WoEs for binned variables are computed on the training data and stored in an `S3` object of class `woe` with a corresponding `predict.woe()` method that allows application to new data. Furthermore, via an argument `ids`, a subset of the variables can be selected for which WoEs are to be computed (default: all `factor` variables) and a real value `zeroadj` specified and added to the frequency of bins with empty target levels for computation of $f()$ in Equation (1) to prevent WoEs from resulting in $\pm\infty$. In contrast to other implementations, it allows observation `weights` which can be necessary for reject inference

Reject Inference to be assigned. The subsequent code shows its usage:

```
### Example 4: computing and applying WoEs (based on Example 3)
library(klaR)
# woe() requires variable type factor
train_bins <- dplyr::mutate_if(train_bins, sapply(train_bins, is.character),
                               as.factor)
valid_bins <- dplyr::mutate_if(valid_bins, sapply(valid_bins, is.character),
                               as.factor)

# Compute WoEs on training data
woe_model <- woe(creditability ~ ., data = train_bins)
# ...woes for variable purpose
woe_model$woe$purpose_bin
```

```
# apply WoEs
train_woes <- data.frame(creditability = train_bins$creditability,
woe_model$xnew)
valid_woes <- predict(woe_model, valid_bins)
```

3.8. Short Benchmark Experiment

The example data has been used to compare the performance of the different available packages for automatic binning. For reasons explained above, binning of categorical variables requires expert knowledge on the meaning of the levels. Thus the benchmark is restricted to a comparison for the seven numeric variables in the data set. Note that four of these variables contain small numbers of distinct numeric values such as the number of credits (cf. 2nd column of Table 1). Therefore, the remaining three variables `age`, `amount` and `duration` are the most interesting ones. Further note that (although it is by far the most popular data set used in literature) for reasons of its size and the balance of the target levels, the German credit data might not be representative of typical credit scorecard developments (Szepannek 2017). For this reason, the results should not be overemphasized but rather used to give an idea on differences in performance of the various implementations.

Table 1. Number of bins after automatic binning. Abbreviations of package names: sc = `scorecard`; woeB = `woeBinning` using `woe.binning()`; woeB.T = `woeBinning` using `woe.tree.binning()`; sMU = `scorecardModelUtils`; Rprof = `Rprofet`; smb = `smbinning` and cremo = `creditmodel`.

	Unique	sc	woeB	woeB.T	Glmdisc	sMU	Rprof	smb	Cremo	Riskr
Avg. # bins		6.33	4.33	6	2.67	3.67	11	2.67	2	2.67
duration	32	5	5	5	3	5	13	3	2	3
amount	663	7	4	6	1	3	11	3	2	3
instRate	4	4	4	4	1	4	4	4	2	1
residence	4	4	4	4	3	2	4	4	3	1
age	52	7	4	7	4	3	9	2	2	2
numCredits	4	2	3	3	2	2	3	4	2	1
numLiable	2	2	3	3	2	1	2	2	2	1

Table 1 shows the number of bins resulting from automatic binning as implemented by the different packages. The first row summarizes the average number of bins for the three variables `age`, `amount` and `duration`. The package `Rprofet` (which interfaces to `binr::bins.greedy()`, cf. above) returns the largest numbers of bins. The number of bins as returned by the tree-based binning via `smbinning` and `riskr` as well as `glmdisc` and `creditmodel` are comparatively small.

Table 2 lists the performance of the different binning algorithms. To prevent analyzing the overfitting of the training data (as it would be obtained by increasing the number of bins), the validation data is used for comparison (cf. Example 1). To ensure a fair comparison of all packages, the performance is computed using the same methodology. First, WoEs are assigned to the binned validation data using the package `klaR`. Afterward, univariate Gini coefficients (as one of the most commonly used performance measures for performance evaluation of credit scoring models, cf. Section 6) of the WoE variables are computed using the package `pROC` (Robin et al. 2021). Note that some of the introduced functions for automatic binning allow for a certain degree ofhyperparameterization which could be used to improve the binning results. However, as the scope of automatic binning does not provide a highly tuned perfect model but rather a solid basis for a subsequent manual bin adjustment, all results in the experiment are computed using default parameterization. Further note that, for the package `Rprofet`, no validation performance is available as there exists no `predict()` method. For the packages `riskr`, the workaround has been used as described above to assign bins to validation data[18]. Concerning the results, it also has to be mentioned that the package `glmdisc` optimizes bins w.r.t. subsequent logistic regression based on dummy variables on the bins which further takes into account the multivariate dependencies between the variables and not just discriminative power of the single variables[19].

Table 2. Gini coefficient of WoE transformed variables on validation data.

	LCL	sc	woeB	woeB.T	Glmdisc	sMU	smb	Cremo	Riskr
duration	0.170	0.297	0.259	0.264	0.265	0.299	0.248	0.162	0.248
amount	0.116	0.251	0.179	0.227	0.000	0.196	0.219	0.069	0.219
age	0.078	0.179	0.169	0.222	0.189	0.200	0.187	−0.003	0.187
numLiable	0.000	0.006	0.006	0.006	0.006	0.000	0.006	0.006	0.000
numCredits	0.000	0.068	0.068	0.068	0.068	0.068	0.061	0.068	0.000
residence	0.000	0.006	0.017	0.017	0.017	0.029	0.006	0.017	0.000
instRate	0.000	0.108	0.103	0.103	0.000	0.108	0.108	0.104	0.000

The first column (LCL) of the results contains a 95% lower confidence level of the best binning for each variable using bootstrapping (Robin et al. 2011). Only for the package `creditmodel` results of the automatic binning for the variables `age`, `amount` and `duration` were significantly worse (below LCL) than the best method. In summary, none of the packages clearly dominates the others, and at first glance the choice of the algorithm does not seem to be crucial. In practice, it might be worth trying different algorithms and comparing their results to support the subsequent modelling step of their manual modification (cf. above).

3.9. Summary of Available Packages for Binning

Table 3 summarizes the functionalities for variable binning and WoE assignment that are provided by the different packages as they have been worked above.

Table 3. Summary of the functionalities for binning and WoEs provided by the different packages where ✓ denotes available and ✗ not available. An empty field means that this is not relevant w.r.t. the scope of the package. (1): workaround available (cf. above); (2) separate bin (`00.NA`) is created—binning of new data (`split_bins_all()`) possible but no WoE assignment (`(woe_trans_all)`); (3) always bin 1 assigned; (4) separate function `missing_val()` for imputation; (5) additional function `cat_to_new()` merges levels smaller than threshold (cf. above).

	sc	smb	woeB	Cremo	Riskr	Glmdisc	sMU	Rprof	klaR
automatic binning of numerics	✓	✓	✓	✓	✓	✓	✓	✓	✗
automatic binning of factors	✓	✗	✓	✗	✓	✓	✗	✗	✗
store and predict numerics	✓	✓	✓	✓	(1)	✓	✓	✗	✗
store and predict factors	✓	✓	✓	✗	(1)	✓	✗	✗	✗
supports bin prediction	✓	✓	✓	✗	(1)	✓	✓	✗	✗
supports WoE prediction	✓	✗	✓	✓	✗	✗	✗	✗	✓
summary table	✓	✓	✓	✓	✓	✗	✓	✗	✓
plot	✓	✓	✓	✗	✓	✓	✗	✓	✓
manual modification	✓	✓	✗	✗	✗	✗	✗	✓	✗
multiple variables	✓	✗	✓	✗	✗	✓	✓	✓	✓
supported target levels	✓	✗	✓	✓	✗	✗	✗	✗	✓
adjust WoEs	✓	✗	✓	✗	✗		✓		✓
NAs	✓	✓	✓	(2)	✗	(3)	(4)	✓	
new levels	✗	✗	✓		✗	(3)			
level order irrelevant	✗		✓		✓	✓			
min. level size	✗		✓	✗	✗	✗	(5)	✗	

For an initial automatic binning of variables, most of the packages have implemented strategies based on decisions trees. A short benchmark experiment on the German credit data shows only small differences in performance depending on the package used. Only for the package `creditmodel` using default parameters was a significantly worse performance used. However, because the resulting automatically generated bins should be analyzed and modified if necesseray, the choice of an explicit algorithm for the initial automatic binning becomes less important. In summary, the package `woeBinning` offers quite a comprehensive toolbox with many desirable implemented functionalities, but unfortunately no manual modification of the results from automatic binning is supported. For the latter the `scorecard` package can be used, but it must be used with care for factor variables because its automatic binning of categorical variables suffers from dependence on the natural order of the factor levels. As a remedy, a function has been suggested in the supplementary code (cf. footnote 17) to import the results of `woeBinning`'s automatic binning into the result objects from the `scorecard` package for further processing.

4. Preselection of Variables

4.1. Overview

As outlined above, a major aspect of credit risk scorecard development is to allow for the integration of expert domain knowledge at different stages of the modelling process. In statistics, traditionally criteria such as AIC or BIC are used for variable selection to find a compromise between a model's ability to fit the training data's parsimony in terms of the number of trainable model parameters (cf. Section 5). For scorecard modelling, typically a variable preselection is made, which allows for a plausibility check by analysts and experts. Apart from plausibility checks, several analyses are carried out at this stage, typically consisting of:

- Information values of single variables;
- Population stability analyses of single variables on recent out-of-time data
- Correlation analyses between variables.

4.2. Information Value

Variables with small discriminatory power in terms of their IV (cf. Section 3) are candidates for removal from the development process. While the interpretation ''small" in the context of IV slightly varies depending on who is asked, an example is given in Siddiqi (2006) by $IV < 0.02$. As an important remark and in contrast to a common practice in credit scorecard modelling, in business not just the IV of a variable should be taken into account but rather how much different information a variable will contribute to a scorecard model that is not already included in other variables. For this reason, IVs should be analyzed together with correlations (cf. this Section below). If not just validation data but also an independent test data set is available, a comparison of the IV on training and validation data can be used to check for overfitting of the binning.

Table 4 lists packages that provide functions to compute information values of binned variables. As usual, these packages differ by the type of the target variable that is required. Some allow for factors; others require binary numerics that take the values 0 and 1. An important difference consists in whether (and how) they do WoE adjustment in case of bins where one of the classes is empty. In `creditR` no adjustment is completed, and the resulting IV becomes ∞. Some packages (`creditmodel`, `Information`, `InformationValue` and `smbinning`) return a value different from ∞, but from the documentation it is not clear how it is computed. For the packages `scorecard` and `scorecardModelUtils`, the adjustment is known, and for the package `klaR` the adjustment can be specified in an argument. Note that, depending on the adjustment, the resulting IVs of the affected variables may differ strongly.

Table 4. Packages and functions for computation of IVs.

Package	Function	Target Type	Multiple Variables	WoE Adjustment
`creditR`	`IV.calc.data()`	both, levels 0/1	yes	no
`creditmodel`	`get_iv_all()`	both, levels 0/1	yes	yes
`Information`	`create_infotables()`	numeric 0/1	yes	yes
`InformationValue`	`IV()`	numeric 0/1	no	yes
`klaR`	`woe()`	factor	yes	argument
`riskr`	`pred_ranking()`	numeric 0/1	yes	no
`scorecard`	`iv()`	both	yes	0.99
`scorecardModelUtils`	`iv_table()`	numeric 0/1	yes	0.5
`smbinning`	`smbinning.sumiv()`	numeric 0/1	yes	yes

Example 5 shows how IVs can be computed using the package `klaR` with zero adjustment (which in fact is not necessary here.) The function `woe()` (cf. Example 4) automatically returns IVs for all factor variables.

```
### Example 5: computing IVs  (based on Example 4)
library(klaR)
woe_model <- woe(creditability ~ ., data = train_bins, zeroadj = 0.5)
# ...the IVs are automatically computed and can be assessed via:
woe_model$IV
```

The package `creditR` also offers a function `IV_elimination()` that allows an `iv_threshold` and returns a data set with a subset of variables with IV above threshold for the training data. Similarly, the package `scorecardModelUtils` offers a function `iv_filter()` that returns a list of variable names that pass (/fail) a prespecified threshold.

Beyond computation of IVs, the package `creditR` can be used to compute Gini coefficients for simple logistic regression models on each single variable via the function `Gini.univariate.data()`, and just as for IVs, this can be used for variable subset preselection (`Gini_elimination()`). The function `pred_ranking()` from the package `riskr` returns a summary table containing IV as well as the values of the univariate AUC and KS statistic and an interpretation.

4.3. Population Stability Analysis

To take into account the sample selection bias that results from a customer portfolio shift (e.g., due to new products or marketing activities), the stability of the distribution of the variable's bins over time is considered. For this purpose, typically, the population stability index (PSI) is computed between the (historical) development sample data and a more recent out-of-time (OOT) sample (where typically performance information is not yet available). Basically, the PSI is just the IV (cf. eqn. (2)). While the IV compares two data sets given by the development sample which are split according to the levels of the target variable ($y = 1$ vs. $y = 0$), the PSI compares the entire development sample ($y \in \{0, 1\}$) with an entire out-of-time sample. A large PSI indicates a change in the population w.r.t. the bins. A small PSI close to 0 indicates a stable population and (again referring to Siddiqi (2006)) $PSI < 0.1$ can be interpreted as stable while a $PSI > 0.25$ is an indicator of a population shift. Of course, a decision of inclusion or removal of variables from the development sample should take into account both population stability and the discriminatory power (i.e., IV) of a variable. With reference to the analogy for PSI and IV, the formerly presented functions of IV calculation can also be used for population stability analysis. The function `SSI.calc.data()` from the package `creditR` returns a data frame of PSIs for all variables. The corresponding code (here, for a computation of PSIs between training and validation—not OOT—set) is given in Example 6.

```
### Example 6: population stability analysis for all variables
library(creditR)
SSI.calc.data(train_bins, valid_bins, ‘‘creditability’’)
```

The function `riskr::psi()` calculates the PSI for single variables and also provides a more detailed table on the bin-specific differences (cf. Example 7 for the variable `purpose`). It does contain the absolute and relative distribution of the bins (for reasons of space two columns with the absolute frequencies have been discarded from the output). The PSI of the variable as given by the `value` element of the output corresponds to the sum of the column `index`:

```
### Example 7: PSI for single binned variable purpose (based on Example 3)
library(riskr)
psi(train_bins$purpose_bin, valid_bins$purpose_bin)
```

```
## $value
## [1] 0.00792
##
## $label
## [1] "Insignificant change"
##
## $table
## # A tibble: 5 x 7
##   class          act_percent new_percent diff_percent coefficient     woe   index
##   <fct>                <dbl>       <dbl>        <dbl>       <dbl>   <dbl>   <dbl>
## 1 business%,%c~        0.321       0.355      0.0339        1.11   0.100  3.40e-3
## 2 car (used)           0.109       0.0887    -0.0202        0.815 -0.205  4.13e-3
## 3 domestic app~        0.0622      0.0614    -0.000801      0.987 -0.0130 1.04e-5
## 4 furniture/eq~        0.194       0.191     -0.00265       0.986 -0.0138 3.65e-5
## 5 radio/televi~        0.314       0.304     -0.0102        0.967 -0.0332 3.40e-4
```

Alternatively, the package `smbinning` comes along with a function `smbinning.psi(df, y, x)` which requires both development and OOT sample to be in one data set (`df`) and a variable y that indicates the data set where an observations originates. In addition to a function `get_psi_all()` for PSI calculation, the package `creditmodel` provides a function `get_psi_plots()` to visualize stability of the bins for two data sets using bar plots with juxtaposed bars. The packages `creditR` and `scorecard` further offer functions that can be used for an OOT stability analysis of the final score (cf. Section 5).

4.4. Correlation Analysis

To avoid variability of the estimates of a regression model, its regressors should be of low correlation (cf. e.g., Hastie et al. 2009, chps. 3, 4). As per construction, WoE transformed variables are linear in the logit of the target variable, providing a natural approach in analyzing correlations between these variables. For this purpose, the `caret` package (Kuhn 2008, 2021) offers a function `findCorrelation()` that automatically identifies among any two variables of strong correlation the one that has the larger average (absolute) correlation to all other variables. A major advantage of performing correlation analysis in advance for variable preselection is that it can be used as another way to integrate expert's experience into the modelling. Among variable clusters of high correlations, experts can choose which of these variables should be used or discarded for further modelling. There are some packages that are not originally intended to be used for credit scorecard modelling but that offer functions that can be used for this purpose. The package `corrplot corrplot` offers a function to visualize the correlation matrix and resort it such that groups of correlated variables are next to each other (cf. Figure 3, left). An alternative visualization is given by a phylogenetic tree of the clustered variables using the package `ape` (Paradis and Schliep 2018; Paradis et al. 2021), where the variable clustering is obtained using the package `ClustOfVar` ((Chavent et al. 2012, 2017), cf. Figure 3, right). The code for creation of both plots is given in the following example (note that the choice of the `hclust.method = ‘‘complete’’` in the left plot guarantees a minimum correlation among all variables in a cluster, but all correlations on the training data are below 0.35 in this example).

```
### Example 8: visualizing correlations (based on Example 4)
# reordered correlation matrix
library(corrplot)
# crop redundant prefixes from variable names for plot
X <- train_woes
names(X) <- substr(names(X), 5, 12)
cmat <- cor(X[,-(1:2)])
corrplot(cmat, order = ‘‘hclust’’, method = ‘‘ellipse’’,
hclust.method = ‘‘complete’’)
```

```
# phylogenetic tree
library(ClustOfVar)
library(ape)
vctree  <- hclustvar(X.quanti = X[,-(1:2)])
plot(as.phylo(vctree), type = ‘‘fan’’)
```

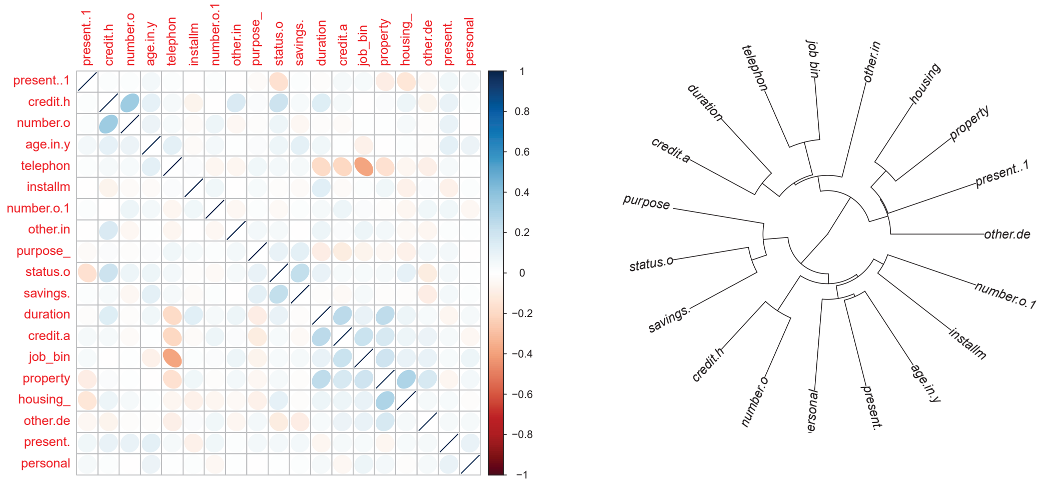


Figure 3. Reordered correlation matrix (**left**) and phylogenetic tree of the clustered variables (**right**).

The package `clustVarLV` (Vigneau et al. 2015, 2020) offers variable clustering such that the correlation between each variable and the first latent principal component of its variable cluster is maximized. The number of clusters `K` has to be prespecified. As it can be seen in the output from Example 9 (only cluster 1 is shown), for each variable the correlation to the cluster's latent component as well as the correlation to the 'closest' next cluster are shown.

```
### Example 9: variable clustering using ClustVarLV (based on Example 4)
library(ClustVarLV)
clverg <- CLV(train_woes[,-(1:2)], method = 1)
plot(clverg)
```

```
summary(clverg, K = 3)
```

```
##                                            cor in group  |cor|next group
## woe_savings.account.and.bonds_bin                  0.73             0.05
## woe_status.of.existing.checking.account_bin        0.72             0.17
## woe_purpose_bin                                    0.51             0.11
```

Among the aforementioned packages dedicated to credit scoring, `creditR` contains a function `variable.clustering()` that performs `cluster`'s `pam` (Maechler et al. 2021) on the transposed data for variable clustering. The (sparsely documented) function `correlation.cluster()` %data, output, "variable", "Group")} can be used to compute average correlations between the variables of each cluster. [20]

The package `Rprofet` provides two functions `WOEClust_hclust()` and `WOEClust_kmeans()` that perform `stats::hclust()` on the transformed data or `ClustOfVar::kmeansvar()` and return a data frame with variable names and cluster index together with the IV of the variable, which may help to select variables from the clusters. Unfortunately, they are only designed to work with output from the package's function `WOEProfet()` and require a list of a specific structure as input argument. In addition to functions `cor_plot()` for visualization of the correlation matrix, `char_cor()` computes a matrix of Cramer's V between or a set of categorical variables and `get_correlation_group()` for detection of groups of correlated (numeric) variables. The package `creditmodel` also contains a function `fast_high_cor_filter()` for an automatic correlation- based variable selection. In a group of highly correlated variables, the one with the highest IV is selected as shown in Example 10.

```
### Example 10: Automatic correlation-based variable selection
(based on Example 4)
library(creditmodel)
# create list of variables sorted according to IV
iv_list = feature_selector(dat_train = train_woes, dat_test = NULL,
target = ‘‘creditability’’,
filter = ‘‘IV’’, iv_cp = 0.02, vars_name = FALSE)
iv_list

# select variables
fast_high_cor_filter(dat = train_woes, com_list = iv_list, p = 0.15,
cor_class = TRUE ,vars_name = TRUE)
```

Similarly, the package `scorecardModelUtils` offers an alternative for an automatic variable preselection based on Cramer's V using the function `cv_filter()`. Among two (categorical) variables of $V >$ `threshold`, the one with lower IV is automatically removed (cf. Example 11). Finally, two functions, `iv_filter()` and `vif_filter()` can be used for variable preselection based on IVs only (w/o taking into account for correlations between the explanatory variables) and based on variance inflation (cf. also Section 5).

```
### Example 11: Cramer's V based variable selection (based on Example 3)
library(scorecardModelUtils)
# package requires 0/1 target:
train_bins2 <- train_bins
train_bins2$creditability <- as.integer(train_bins2$creditability==‘‘good’’)

# first data frames of IVs and Cramer's V have to be computed
ivtable   <- iv_table(train_bins2, ‘‘creditability’’,
cat_var_name = names(train_bins2)[-1])
cvtable   <- cv_table(train_bins2, names(train_bins2)[-1])
selection <- cv_filter(cvtable$cv_val_tab, ivtable$iv_table,threshold = 0.3)
selection
```

4.5. Further Useful Functions to Support Variable Preselection

The package `scorecard` contains a function `var_filter()` that performs an automatic variable selection based on IV and further allows for specifying a maximum percentage of missing or identical values within a variable, but it does not account for correlations among the predictor variables. Alternatively, the package `creditmodel` has a function `feature_selector()` for automatic variable preselection based on IV, PSI, correlation and xgboost variable importance (Chen and Guestrin 2016).

The package `creditR` has two functions to identify variables with missing values (`na_checker()`) and compute the percentage of variables with missing values (`missing_ratio()`). For imputation of numeric variables in a data set with mean or median values, a function `na_filler_contvar()` is available. Of course, this has to be handled with care as the mean or median value will typically not be the same on training and validation data. The package `mlr` (Bischl et al. 2016, 2020) offers imputation that can be applied to new data.

For an assignment of explicit values to missing the package `scorecardModelUtils` also provides a function `missing_val()`. This can be either a function such as ‘‘`mean`’’, ‘‘`median`’’ or ‘‘`mode`’’ or an explicit value such as -99999 which can be meaningful before binning to assign missing values to a separate bin. Similarly, for categorical variables the assignment of a specific level such as ‘‘`missing_value`’’ can be meaningful. A function `missing_elimination()` removes all variables with a percentage above `missing_ratio_threshold` from training (but not from validation) data. The package `creditmodel` offers a convenient function `data_cleansing()` that can be used for automatic deletion of variables with low variance and a high percentage of missing values, to remove duplicated observations and reduce the number of levels of categorical variables. The package `riskr` provides two functions `select_categorical()` and `select_numeric()` to select all (non-/) numeric variables of a data frame.

A univariate summary of all variables is given by the function `univariate()` of the `scorecardModelUtils` package. A summary for numeric variables can be computed using the function `ez_summ_num()` from the package `riskr`. A general overview of packages explicitly designed for exploratory data analysis that provide further functionalities are given in Staniak and Biecek (2019). The packages `scorecard` (`one_hot()` and `var_scale()`) and `creditmodel` (\texttt{one_hot_encoding(), `de_one_hot_encoding()`, `min_max_norm()`) provide functions for one-hot-encoding of categorical and standardization of numeric variables.

5. Multivariate Modelling

5.1. Variable Selection

Traditionally, credit risk scorecards are modelled using logistic regression (cf. e.g., Anderson 2019; Siddiqi 2006; Thomas et al. 2019; Wrzosek et al. 2020), which is in R performed via `glm()` (with `family = binomial`). In addition to the manual variable preselection as described in the former section, typically, a subsequent variable selection is performed which can be completed by the `step()` function. Common criteria for variable selection are AIC (`k = 2`) or BIC (`k = log(nrow(data))`). Example 12 gives an example for BIC based variable selection.

```
### Example 12: BIC variable selection (based on Example 4)
# column 2 (variable foreign.worker_bin excluded as it has only one level)
null   <- glm(creditability ~ 1, data = train_woes[,-2], family = binomial)
full   <- glm(creditability ~ ., data = train_woes[,-2], family = binomial)
bicglm <- step(null, scope=formula(full), direction=‘‘both’’,
k=log(nrow(train_woes)))
```

Note that an initial model (here: `null`) and the scope for the search have to be specified. This offers another possibility for expert knowledge integration. After each step the criteria of all candidates are reported and can be used to decide among several variable candidates of similar performance for the one that is most appropriate from a business point of view. The corresponding variable can be manually added to the formula of a new initial model in a subsequent variable selection step.

The function `smbinning.logitrank()` of package `smbinning` runs all possible combinations of a specified set of variables, ranks them according to AIC and returns the corresponding model formulas in the result data frame. Depending on the size of the preselected set of variables (cf. Section 4), this can be time-consuming.

As an alternative to AIC and BIC, Scallan (2011) presents how variables can be selected in line with the concept of information values (cf. Section 3) using so-called marginal information values, but currently none of the presented packages offers an implementation of this strategy.

It is also common to consider the variance inflation factor of the explanatory variables of a final model given by:

$$VIF(X_i) = \frac{1}{1 - R_i^2} \tag{3}$$

where R_i^2 is the R^2 of a linear regression model with X_i as dependent variable and all other explanatory variables except X_i as regressors. Large values of $VIF(X_i)$ denote that this variable can be explained by the other regessors and are an indication of multicollinearity. Both the packages `car` (Fox and Weisberg 2019; Fox et al. 2021) and `scorecard` offer a function `vif()` that can be used for this purpose as well as the functions `vif.calc()` and `lr_vif()` of the packages `creditR` and `creditmodel` (cf. Example 13).

```
### Example 13: VIF (based on Example 11)
car::vif(bicglm)
scorecard::vif(bicglm)
creditR::vif.calc(bicglm)
creditmodel::lr_vif(bicglm)
```

Not only variable selection during the scorecard development but also the question of segmentation may arise, i.e., whether one single model or several separate models should be used for different subsets of the population. For this purpose, the package `glmtree` offers a function `glmtree()` that computes a potential segmentation scheme according to a tree of recursive binary splits where each leaf of the tree consists in a logistic regression model. The resulting segmentation optimizes AIC, BIC or alternatively the likelihood or the Gini coefficient on validation data. Note that this optimization does not account for variable selection as described above.

5.2. Turning Logistic Regression Models into Scorecard Points

>From the coefficients of the logistic regression model, the historical shape of a scorecard is obtained by assigning the corresponding effect (aka points) to each bin (such that the score of a customer is the sum over all applicable bins and can easily be calculated by hand). Typically, the effects are scaled to obtain some predefined points to double the odds (pdo, cf. e.g., Siddiqi 2006) and rounded to integers.

The package `scorecard` offers a function `scorecard()` that translates a `glm` object into scorecard points as described above and in addition returns key figures such as frequencies, default rates and WoE for all bins. A function `scorecard_ply()` is available that can be used to assign scores to new data. In addition to the `glm` object, the `bins` as created by `scorecard`'s `woebin()` (cf. Section 3) have to be passed as an input argument. Further arguments do specify the (`pdo`) as well as a fixed number of points `points0` that corresponds to odds of `odds0` and whether the scorecard should contain an intercept or whether the intercept should be redistributed to all variables (`basepoints_eq0`). The function requires WoEs (not just the binned `factors`) and the variable names in the `coef(glm)` to match the convention of variable renaming as it is done by `scorecard`'s `woebin_ply()` function (i.e., a postfix `_woe`)[21].

Alternatively, a function `scorecard2()` is available that directly computes a scorecard based on `bins` and a data frame of the original variables. Here, in addition, the name of the target variable (`y`) and a named vector (`x`) of the desired input variables have to be passed[22]. Example 14 illustrates the usage of `scorecard2()` and its application to new data (here represented by the validation set) as well as its output for the variable `duration.in.month`.

```
### Example 14: calculation of scores (based on Example 2)
# note: variable 20 (foreign.worker) not used (cf. also Example 12)
sc <- scorecard2(bins, train, y = 'creditability', x = names(train)[1:19])
# scorecard points table for the variable 'duration.in.month'
sc$duration.in.month[,c(1,2,4,5,6,7,8,13)]
```

```
##              variable       bin count_distr neg pos posprob     woe points
## 1: duration.in.month  [-Inf,8)     0.08062  51   6  0.1053 -1.2988     65
## 2: duration.in.month    [8,16)     0.35785 194  59  0.2332 -0.3491     18
## 3: duration.in.month   [16,34)     0.37907 179  89  0.3321  0.1425     -7
## 4: duration.in.month   [34,44)     0.10467  44  30  0.4054  0.4583    -23
## 5: duration.in.month [44, Inf)     0.07779  26  29  0.5273  0.9504    -48
```

```
train_scored <- scorecard_ply(train, sc, only_total_score = FALSE)
valid_scored <- scorecard_ply(valid, sc, only_total_score = FALSE)
```

In addition, the package further contains a function `report()` that takes the data, the (original) names of all variables in the final scorecard model and a breaks list (cf. Section 3 that can be obtained from the bins) as input arguments and generates an excel report summary of the scorecard model. Different sheets are reported with information and figures on the data, model, scorecard points, model performance and the binning figures for all variables of the model which can be used for model development documentation in practice.

To translate a glm based on `factor` variables (bins instead of WoEs) into scorecard points, the package `scorecardModelUtils` provides a function `scalling()`. Its output can be used to predict scores for new data by function `scoring()` (cf. Example 15).

```
### Example 15: score points for a model based on bins, not WoEs
(based on Example 4)
library(scorecardModelUtils)
# create glm using factor variables -- foreign worker excluded (cf. above)
full_bins <- glm(creditability~.,data = train_bins[,-21], family = binomial)
# calculate scorecard points from effects
sc2 <- scalling(train_bins, ‘‘creditability’’, full_bins, point = 15,
factor = 2)
sc2
# apply scorecard to new data
scoring(valid_bins, target = ‘‘creditability’’, sc2)
```

The package `creditmodel` transforms a `glm` object into scorecard points via a function `get_score_card()`, which requires a bin table created by `creditmodel::get_bins_table_all()` and thus is restricted to application within its own universe. In addition, if a table of scorecard points is not required, it offers a function `score_transfer()` that directly applies the `glm` object to data and scales the resulting points accordingly (cf. Example 16) and another function `p_to_score` to turn posterior probabilities into score points.

```
### Example 16: directly predict score points from a glm object
(based on Example 12)
library(creditmodel)
train_scored_3 <- score_transfer(bicglm, train_woes, a = 500, b = 20)
valid_scored_3 <- score_transfer(bicglm, valid_woes, a = 500, b = 20)
```

Another implementation of calculating scorecard points from a glm object based on bins and not WoEs is given by the function `smbinning.scaling()`, which comes with a predict function `smbinning.scoring.gen()` that can be used to score new observations but that requires the binned variables have been generated with `smbinning.gen()` or `smbinning.factor.gen()` (cf. Section 3). A function `smbinning.scoring.sql()` is available that transforms the resulting scorecard into SQL code.

The package `Rprofet` also contains a function `ScorecardProfet()` for this purpose, which calculates a glm with corresponding scorecard points but only based on binning and WoEs as calculated by functions from the package itself (cf. Section 3), and no function is available for application of the scorecard points to new data. The function `scaled.score()` of the package `creditR` transforms posterior default probabilities into scores where any `increase` points double the odds (of nondefault), and odds of `increase` correspond to `ceiling_score` points. In addition, the package `creditR` offers a function that can be used to recalibrate an existing glm on calibration data. A simple logistic regression is fit on the `calibration_data` with only one input variable: the predicted log odds by the current model.

5.3. Class Imbalance

In credit scorecard modelling, the class typically is highly unbalanced in the training sample. This issue has been addressed in several papers (Brown and Mues 2012; Crone and Finlay 2012; Vincotti and Hand 2002). Usual remedies are oversampling, undersampling, synthetic minority over-sampling (SMOTE, Chawla et al. 2002) or simply reweighing observations. A comprehensive benchmark study of these techniques as well as overbagging is undertaken in Bischl et al. (2016), and it turns out that logistic regression is less sensitive to class imbalance than tree-based classifiers. Furthermore, note that different from, e.g., the accuracy of the two most commonly used performance measures in credit scorecard modelling, the Gini coefficient and the KS statistic (cf. Section 6) do not depend on the class imbalance ratio.

The package `klaR` allows for specifying observation weights for WoE computation (see Section 3.7). Within the `mlr3` framework, imbalance correction can be performed using `mlr3pipelines` (Binder et al. 2021). Several resampling algorithms are implemented in the packages `imbalance` (Cordón et al. 2020, 2018) and `unbalanced` (Pozzolo et al. 2015). The SMOTE algorithm is also implemented in the `smotefamily` package (Siriseriwan 2019).

6. Performance Evaluation

6.1. Overview

In credit scoring modelling, performance evaluation is used not only for model selection but also for third-party assessments of an existing model by auditors or regulators and to drive future management decisions about whether an existing model should be kept in place or whether it should be replaced by a new one. Note that, as opposed to common practice in machine learning, hyperparameter tuning typically has no separate validation data used for model selection (cf. e.g., Bischl et al. (2012), Bischl et al. (2021)), but in credit scorecard modelling, the validation data serves for independent model validation (corresponding to test data in frameworks such as `mlr3`). While this is less critical in the case of simple models such as logistic regression, it should still be kept in mind, especially if the model is benchmarked against more flexible machine learning models such as support vector machines, random forests or gradient boosting (cf. e.g., Hastie et al. (2009)).

6.2. Discrimination

The two most popular performance metrics for credit scorecards are the Gini coefficient, $Gini = 2(AUC - 0.5)$ and the Kolmogorov–Smirnov test statistic. While for the latter, R provides the function `ks.test()`, one of the most popular ways to compute the AUC in R is given by the package `ROCR` (Sing et al. 2005, 2020). Nonetheless, for the purpose of credit scorecard modelling, it is referred to the package `pROC` at this point for the following three reasons:

1. Different from standard binary classification problems, credit scores are typically supposed to be increasing if the event (= default-) probability decreases. The function `roc()` of the package `pROC` has an argument `direction` that allows for specifying this.
2. In credit scoring applications, it may be given that not all observations of a data set are of equal importance, e.g., it may not be as important to distinguish which of two customers with small default probabilities has the higher score if his or her application will be accepted anyway. The package's function `auc()` has an additional argument `partial.auc` to compute partial area under the curve (Robin et al. 2011).
3. Finally, its function `ci()` can be used to compute confidence intervals for the AUC using either bootstrap or the method of DeLong (DeLong et al. 1988; Sun and Xu 2014), e.g., to support the comparison of two models.

Example 17 demonstrates how pROC can be used for performance analysis.

```
### Example 17: Gini coefficient using {pROC} (based on Example 13)
library(pROC)
curve <- roc(valid$creditability, valid_scored$score,
levels = c(‘‘good’’,‘‘bad’’), direction = ‘‘>’’)
# levels = c(‘‘controls’’, ‘‘cases’’),
# direction = controls > cases
plot(curve)
```

```
auc(curve)
# gini coefficient:
2 * (auc(curve) - 0.5)
# confidence limits for the auc:
ci(auc(curve), method = ‘‘bootstrap’’)
```

Among the packages enumerated above, creditR offers a function Kolmogorov-Smirnov(), and riskr has two functions, ks() and ks2(), for computation of the Kolmogorov–Smirnov test statistic. In addition, riskr provides a function divergence() to compute the divergence between two empirical distributions as well as gg_dists() and gg_cum() to visualize the score densities for defaults and nondefaults and their empirical cumulative distribution functions. To compute the Gini coefficient, the package riskr provides functions aucroc (AUC), gini (Gini coefficient), gg_roc() (visualization of the ROC curve), gain() (gains table for specified values on the x-axis) and gg_gain() /gg_lift() (for visualization of the gains-/lift-chart).

In the package creditmodel, two functions ks_value() and auc_value() are available as well as a model_result_plot() to visualize the ROC curve, cumulative score distributions of defaults vs. nondefaults, lift chart and the default rate over equal-sized score bins. A table with respective underlying numbers can be obtained via perf_table().

The package InformationValue contains two functions, ks_stat() and ks_plot(), for Kolmogorov-Smirnov analysis and several functions: AUROC(), plot_ROC(), Concordance() and SomersD() (Gini coefficient) to support analyses with regard to the Gini coefficient. Additionally, the confusionMatrix() and derivative performance measures misClassError(), sensitivity(), specificity(), precision(), npv(), kappaCohen() and youdensIndex() (cf. e.g., Zumel and Mount (2014) chp. 5 for an overview) can be computed for a given cut off by the corresponding functions. Note that these measures are computed with respect to the nondefault target level (supposed to be coded as ‘1’ in the target variable) as well as a cut off optimization w.r.t. the misclassification error, Youden’s Index or the minimum (/maximum) score such that no misclassified defaults (/non-defaults) occur in the data (function optimalCutoff()).

Similar measures (accuracy, precision, recall, sensitivity, specificity, F1) are computed by the function fn_conf_mat() of the scorecardModelUtils package. Numeric differences between the (0/1-coded) target and the model’s predictions in terms of MSE, MAE and RMSE can be computed by its fn_error() function. The package boottol contains a function boottol() to compute bootstrap confidence intervals for Gini, AUC and KS, where subsets of the data above different cut off values are also considered. It may be desirable to analyze the (cumulative) frequencies of the binned scores. A table of such frequencies is returned by the function gini_table() in the scorecardModelUtils package. Example 18 shows selected columns for a binned score using the function gains_table() from the scorecard package.

```
### Example 18: score bin frequencies (...for valid_scored from Example 14)
library(scorecard)
gt <- gains_table(valid_scored$score, valid$creditability, bin_num = 8)
gt[,c(2,4,5,6,7,8,10,11)]
```

```
##            bin cum_count neg cum_neg pos cum_pos posprob approval_rate
## 1: [628, Inf)        37  37      37   0       0 0.00000        0.1263
## 2:  [575,628)        76  36      73   3       3 0.07692        0.2594
## 3:  [529,575)       112  34     107   2       5 0.05556        0.3823
## 4:  [492,529)       148  30     137   6      11 0.16667        0.5051
## 5:  [448,492)       185  26     163  11      22 0.29730        0.6314
## 6:  [399,448)       222  21     184  16      38 0.43243        0.7577
## 7:  [353,399)       257  14     198  21      59 0.60000        0.8771
## 8: [-Inf,353)       293   8     206  28      87 0.77778        1.0000
```

Note that although the Gini coefficient is generally bounded by -1 and 1, the value it can take for a specific model strongly depends on the discriminability of the data. For this reason, it is suitable to compare performance on different models on the same data rather than comparing performance across different data sets. Consequently, for the purpose of an out-of-time monitoring of a scorecard, it is advisable to compare an existing scorecard's performance against a recalibrated version of it rather than to compare it with its performance on the original (development) data. Drawbacks of the Gini coefficient as a performance measure for binary classification are discussed in (Hand 2009), and the H-measure is proposed as an alternative which is implemented in the package `hmeasure` (Anagnostopoulos and Hand 2019). The expected maximum profit measure (Verbraken et al. 2014) as implemented in the package `EMP` (Bravo et al. 2019) further takes into account the profitability of a model.

6.3. Performance Summary

Many of the functionalities as provided by the packages for scorecard modelling in the previous subsection already exist in other packages and are thus not indispensable. In addition to these, however, some of the package provide performance summary reports of several performance measures. These functions are listed in the following table.

In Example 19, computation of a scorecard performance summary is demonstrated using the package `smbinning` (which returns the largest number of performance measures of the four functions from Table 5) as well the function `riskr::gg_perf()` that can be used to produce several graphs on the scorecard's performance (cf. Figure 4). Note that although ROC curves are one of the most popular tools for performance visualization of binary classifiers, they are hardly suited to visualize the performance difference of several competitive models. One reason for this is that large areas of the TPR-FPR plane (e.g., everything below the main diagonal) are typically of no interest given a specific data situation. For this reason, in practice, ROC curves are not very useful for model selection.

```
### Example 19: scorecard performance summary (based on Example 13)
library(smbinning)
perf_dat <- data.frame(‘‘creditability’’ = as.integer
(valid$creditability == ‘‘good’’), ‘‘score’’ = valid_scored$score)
smbinning.metrics(perf_dat, ‘‘score’’, ‘‘creditability’’, cutoff = 450)

# roc curve, ecdf, score distribution and gain chart
library(riskr)
gg_perf(as.integer(valid$creditability == ‘‘good’’), valid_scored$score)
```

Table 5. Overview of scorecard performance summary functions.

Package Function	Riskr Perf()	Scorecard Perf_Eva()	ScorecardModelUtils Gini_Table()	Smbinning Smbinning.Metrics()
KS	✓	✓	✓	✓
AUC	✓	✓		✓
Gini	✓	✓	✓	
Divergence	✓			
Bin table			✓	
Confusion matrix		✓		✓
Accuracy		✓		✓
Good rate				✓
Bad rate				✓
TPR				✓
FNR		✓		✓
TNR				✓
FPR		✓		✓
PPV				✓
FDR				✓
FOR				✓
NPV				✓
ROC curve	✓	✓	✓	✓
Score densities | y	✓			
ECDF	✓	✓		✓
Gain chart	✓			

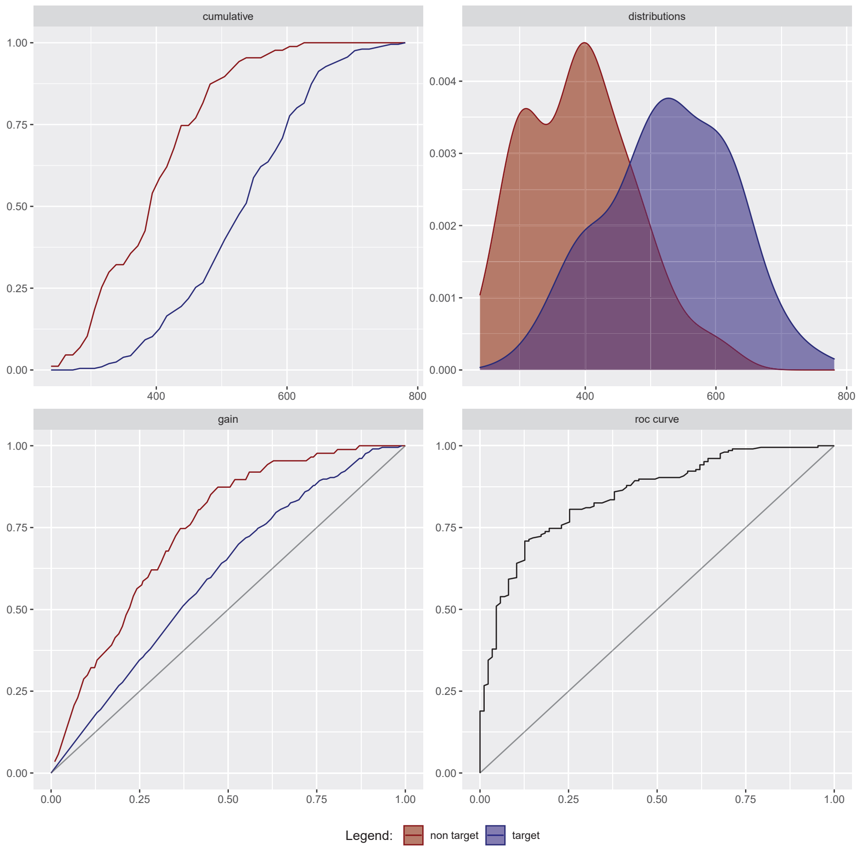


Figure 4. Scorecard performance graphs: ECDF (**top left**); score densities (**top right**); gains (**bottom left**); ROC (**bottom right**).

6.4. Rating Calibration and Concentration

>From a practical point of view, it is often desirable to aggregate scorecard points into classes (rating grades) of similar risk, which is once again a binning task (cf. Section 3). The package `creditR` contains a function `master.scale()` that takes a data frame with scores and corresponding default probabilities as input and uses the function `woeBinning::woe.binning()` to group scores of similar WoE (cf. Example 20). The function `odds_table()` of the `riskr` package allows setting a `breaks` argument with arbitrary bins.

Rating classes should be appropriately calibrated in the sense that the predicted and observed default probabilities match for all rating grades. In order to check this, the package `creditR` contains three functions (`chisquare.test()`, `binomial.test()` and `adjusted.binomial.test()`) that provide a table with indicators for each rating grade (cf. Example 19). Another function, `binomial.point()`, compares the observed average predicted default probability on the data with prespecified boundaries around some desired central tendency default probability. Bootstrap confidence intervals for default probabilities of rating grades can be computed using the function `vas.test()` of the package `boottol`. A Hosmer–Lemeshow goodness-of-fit test (Hosmer and Lemeshow 2000) is, e.g., implemented by the function `hoslem.test()` in the `resourceselection` package (Lele et al. 2019).

```
### Example 20: rating calibration analysis (based on Example 14)
library(creditR)
# calculate PDs from scores
odds   <- 1/19 * 2^(-(valid_scored$score - 600)/50)
pd     <- odds / (1 + odds)
pd.dat <- data.frame(pd = pd,
creditability = as.integer(valid$creditability == ''bad''))
# aggregate scores to rating grades
mscale <- creditR::master.scale(pd.dat, ''creditability'', ''pd'')
# transform $Bad.Rate into numeric
mscale$Bad.Rate <- as.numeric(gsub(''%'',''',mscale$Bad.Rate))/100
# test calibration of the rating grades
# chisquare.test(mscale, ''PD'', ''Bad.Count'', ''Total.Observations'')
bintest <- binomial.test(mscale,''Total.Observations'',''PD'',''Bad.Rate'')
bintest[,c(1,3,8,9,14)]
```

```
##    Final.PD.Range Total.Distr Bad.Rate      PD          Test_Result
## 1 <= 0.0692759267       25.9%    0.039 0.03835 Target Value Correct
## 2 <= 0.1477666759       17.4%    0.078 0.11149 Target Value Correct
## 3 <= 0.1904265313        7.2%    0.190 0.17482 Target Value Correct
## 4  <= 0.275937974        9.6%    0.321 0.23297 Target Value Correct
## 5 <= 0.3709582356        7.5%    0.364 0.32236 Target Value Correct
## 6 <= 0.4365863463        5.1%    0.467 0.41036 Target Value Correct
## 7 <= 0.4605851205        3.1%    0.333 0.45143 Target Value Correct
## 8 <= 0.5695614029        9.2%    0.630 0.51257 Target Value Correct
## 9             <= Inf    15.0%    0.727 0.72768 Target Value Correct
```

According to regulation, ratings must avoid risk concentration (i.e., a majority of the observations being assigned to only a few grades). The Herfindahl–Hirschman index ($HHI = \sum_j \hat{f}(j)^2$, with the empirical distribution $\hat{f}$ of the rating grades j) can be considered to verify this, as e.g., implemented by `creditR`'s `Herfindahl.Hirschman.Index()` or `Adjusted.Herfindahl.Hirschman.Index()`. Small values of HHI indicate low risk concentration.

6.5. Cross Validation

Some of the mentioned packages also provide functions for cross-validation. As both binning and variable selection are interactive, they are not suited for cross-validation (cf.

Sections 3 and 4). For this reason it should be used on the training data and restricted to analyzing overfitting of the logistic regression model. There are already several packages available that provide general functionalities for execution of cross-validation analyses (e.g., `mlr3` or `caret`). The function `k.fold.cross.validation.glm()` of the `creditR` package computes cross-validated Gini coefficients, while the function `perf_cv()` of the `scorecard` package offers an argument to specify different performance measures such as ''`auc`'', ''`gini`'' and ''`ks`''. Both functions allow setting seeds to guarantee reproducibility of the results. The function `fn_cross_index()` somewhat more generally returns a list of training observation indices that can be used to implement a cross-validation and compare models using identical folds.

7. Reject Inference

7.1. Overview

Typically, the final stage of scorecard development consists of reject inference. The scorecard model is based on historical data but already in the past, credit applications of customers that were assumed to be high risk were rejected, and thus for these data only, the predictor variables are available from the application but not the target variable. The use of these observations with unknown performance is commonly referred to as reject inference.

The benefits of using reject inference in practice still remains questionable. It has been investigated by several authors (cf. e.g., Crook and Banasik (2004), Banasik and Crook (2007), Verstraeten and den Poel (2005), Bücker et al. (2013), Ehrhardt et al. (2019)) and is nicely discussed in Hand and Henley (1993). The appropriateness of different suggested algorithms for reject inference depends on the way the probability of being rejected can be modelled, i.e., whether it is solely a function of the scorecard variables (MAR) or not (MNAR) (for further details cf. also Little and Rubin (2002)). A major issue is that, especially for the most relevant MNAR situation, the inference entirely relies on expert judgments. For this reason the appropriateness of the model cannot be tested anymore. In consequence, reject inference should be used with care.

In R, the only package that offers functions for reject inference is the package `scoring Tools`, which is available on Github but not on CRAN. It provides five functions for reject inference: `augmentation()`, `fuzzy_augmentation()`, `parcelling()`, `reclassification()` and `twins()`, which correspond to common reject inference strategies of the same name (cf. e.g., Finlay (2012)). In the following, two of the most popular strategies, namely augmentation and parcelling are briefly explained as they are implemented within the package, completed by an example of their usage.

7.2. Augmentation

An initial logistic regression model is trained on the observed data of approved credits (using all variables, i.e., variable selection has to be done in a preceding step). Afterward, weights are assigned to all observations of this sample of accepted credits, according to their probability of being approved. For this purpose, all observations (accepted and rejected) are scored by the initial model. Then, score-bands are defined and within each band[23] the probability of having been approved is computed by the proportion of observations with known performance in the combined sample from both accepted and rejected credits. Finally, the logistic regression model is fitted again on the sample of the accepted loans with only observed performance but reweighted observations[24].

7.3. Parcelling

Based on an initial logistic regression model which is trained on the observed data of approved credits, only score-bands are defined, and the observed default rate $\widehat{PD}_j$ of each score-band j is derived. The observations of the rejected subsample are then scored by the initial model and assigned to each score-band. Labels are randomly assigned to the rejected observations such that they will have a default probability of $\widehat{PD}_j \times \alpha_j$[25] in each band where α_j are user-defined factors to increase the score-bands' default rates which have to be specified by expert experience. Typically the α_j are set to be increasing for score-bands with larger default probabilities. Note that accepting these credit applications in the past might have happened for reasons beyond those that were reflected by the score variables but which led to a reduced risk for these observations in the observed sample compared to observations with a similar score in the total population. For this reason, parcelling is suitable for the MNAR situation.

Example 21 illustrates parcelling using the `scoringTools` package. Note that all other functions of this package are of similar syntax and output. For parcelling in particular, the `probs` argument specifies quantiles w.r.t. the predicted default probabilities (i.e., from low risk to high risk). Although in the example the factor vector `alpha` is constantly set to 1 for all bands, in practice it will be chosen to be increasing, at least for quantiles of high PDs.

```
### Example 21: reject inference using parcelling (based on Example 4)
library(scoringTools)
# use validation data as 'rejects' for this example
# ...remove target variable and constant variable foreign.worker_bin
reject_woes <-  valid_woes[,-(1:2)]
# apply parcelling
set.seed(42) # reproducibility
ri_parc <- parcelling(xf = train_woes[,-(1:2)], xnf = reject_woes,
yf = ifelse(train_woes[,1] == ‘‘bad’’, 1, 0),
probs = c(0, 0.25, 0.5, 0.7, 0.8, 0.9, 1),
alpha = rep(1, 6))
# final model after reject inference
class(ri_parc@infered_model)
# observations weights
ri_parc@infered_model$weights
# combined sample after parcelling (note automatically renamed variables)
str(ri_parc@infered_model$data)

# recompute WoEs on combined sample using weight (cf. also Example 4)
combined_bins               <- rbind(train_bins, valid_bins)
combined_bins$creditability <- ifelse(ri_parc@infered_model$data$labels==1,
‘‘bad’’, ‘‘good’’)
combined_bins$creditability <- as.factor(combined_bins$creditability)

library(klaR)
woe_model_after_ri <- woe(creditability ~ ., data = combined_bins,
weights = ri_parc@infered_model$weights)
combined_woes     <- data.frame(creditability = combined_bins$creditability,
                                woe_model_after_ri$xnew)
```

The initial model and the final model are stored in the result object’s slots `financed_model` and `infered_model`. Both are of class `glm`. Note that both models are automatically calculated without any further options of parameterization such as variable selection or a recomputation of the WoEs based on the combined sample of accepted applications and rejected applications with inferred target. For this purpose, the `woe()` function of the `klaR` package can be used, which supports the specification of observation weights as the only one among all presented packages. Finally, the combined sample can be used to rebuild the scorecard model as described in Sections 4–6.

8. Summary and Discussion

For a long time in the R universe, no packages were available that were explicitly dedicated to the credit risk scorecard development process, while during the last few years a simultaneous growth of several packages on this task has been observed. Some of these packages are available on CRAN, while some are only available on Github.

This paper aims to give a comparative overview on the different functionalities of currently available packages guided by the sequence of steps along a typical scorecard development process. At the same time, any required functionality is available, which makes it easy to develop scorecards using R. As a conclusion of this systematic review, currently the most comprehensive implementations are given by the packages `scorecard`, `scorcardModelUtils`, `smbinning` and `creditmodel`. With regard to the important modelling step of variable binning and WoE computation, the package `woeBinning` provides an implementation that reflects a broad range of practical issues (cf. Section 3). The package `creditmodel` comes with a whole set of additional functionalities such as cohort analysis, correlation based variable preselection or Cramer's V. It further allows for an easy development of challenging models using xgboost (Chen et al. 2021), gradient boosting (Greenwell et al. 2020) or random forests (Liaw and Wiener 2002). In turn, it does not support manual modification of the bins but rather claims to make the development of binary classification models simple and fast. Unfortunately, its functions are poorly documented, and for the user it is not clear what exactly many of the functions do without looking into the source code. While it seems based on individual experiences, the package `scorecard` is close to the methodology as described in literature (Siddiqi 2006).

Thanks to its large developing community and the huge amount of freely available packages, developers have access to many additional packages that are not explicitly designed for the purpose under investigation but that still provide valuable tools and functions to facilitate and improve the analyst's life, making R a serious alternative to commercial software on this topic.

An investigation of the functionalities provided by the different packages concludes that the packages seem to have been developed quite independently of one other. Some steps of the developments are addressed in many packages, especially the important one of binning variables. However, links between the packages are mostly missing,[26] and many packages are not flexibly designed in the sense that their functions require input arguments and variable naming conventions restricted to results from functions of the same package, which makes it somewhat difficult to benefit from advantages of different packages at the same time. The paper's supplementary code provides several remedies for this issue[27]. Some of the packages are missing predictive functionalities to apply the results of the modelling to new data. It would be desirable, if package developers in the future would check thoroughly for existing implementations and take these into account before generating new code. In particular, respecting existing naming conventions and output objects of other packages may help users simultaneously use different packages and maximally profit from the advantages provided by the R package system.

To summarize the results as they have been worked out in the previous sections, Table 6 lists the presented packages with an explicit scope of scorecard modelling together with the stages of the development process that are addressed.

Finally, and with regard to the title of the paper, Figure 5 aims to visualize the 'landscape' of R packages dedicated to scorecard development using logistic principal component analysis (Landgraf and Lee 2015) as implemented in the `logisticPCA` package (Landgraf 2016) on the binary data given by Table 6.

Table 6. Overview of R packages with the explicit scope of scorecard modelling and addressed stages of the development process.

Package	Binning & WoEs	Preselection	Scorecard	Performance	Reject Inference
boottol				✓	
creditmodel	✓	✓	✓	✓	
creditR	✓	✓	✓	✓	
glmdisc	✓		✓		
glmtree			✓		
Information		✓			
InformationValue		✓			
riskr	✓	✓		✓	
Rprofet	✓		✓		
scorecard	✓	✓	✓	✓	
scoringTools	✓				✓
scorecardModelUtils	✓	✓	✓	✓	
smbinning	✓	✓	✓	✓	
woe	✓	✓			
woeBinning	✓				

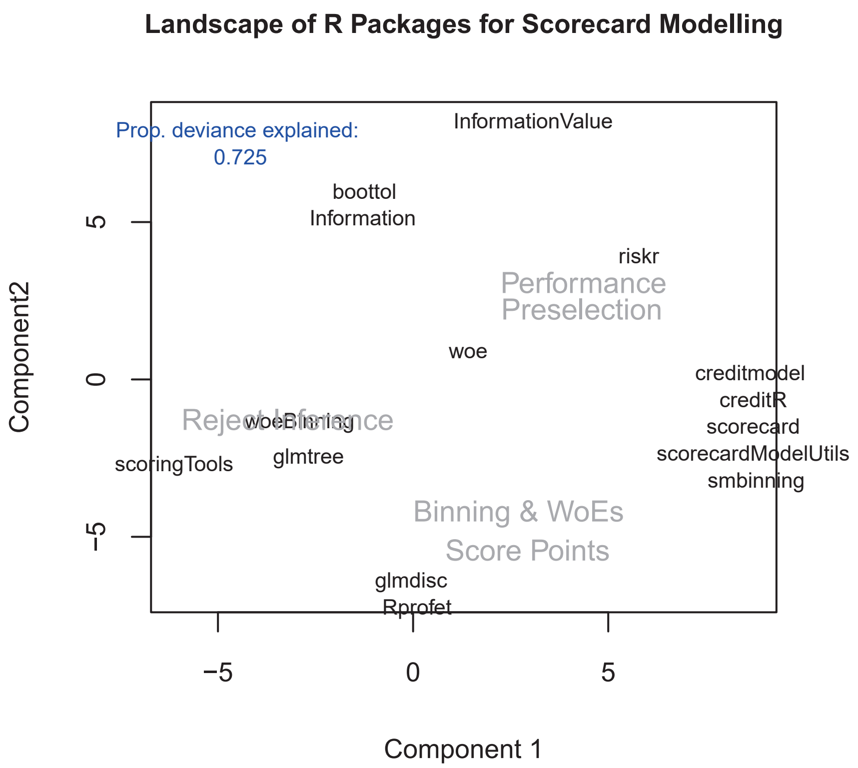


Figure 5. Landscape of R packages for scorecard modelling using logistic PCA.

The future will show to what degree the traditional process of credit risk scorecard development will stay as it is or whether or up to what extent the use of logistic regression will be replaced by more recent machine learning algorithms such as those offered by the recent powerful `mlr3` framework in combination with explainable ML methodology to fulfill regulatory requirements (Bücker et al. 2021). The availability of open source frameworks for scorecard modelling as described above may help bridge the gap between academic advances in machine learning research and the traditional modelling process in the financial industry.

Funding: The APC was funded by Institute of Applied Computer Science, Stralsund.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Reproducible code is available on the GitHub repository https://github.com/g-rho/CSwR (accessed on 15 February 2022).

Acknowledgments: Thanks to Rabea Aschenbruck and Alexander Frahm for support in writing this paper and to the forfunding of this work.

Conflicts of Interest: The author declares no conflict of interest.

Notes

1. https://www.sas.com/en_us/software/credit-scoring.html (accessed on 15 February 2022).
2. https://cran.r-project.org/web/views/Finance.html (accessed on 15 February 2022).
3. https://www.openriskmanual.org/wiki/Credit_Scoring_with_Python (accessed on 15 February 2022).
4. https://towardsdatascience.com/how-to-develop-a-credit-risk-model-and-scorecard-91335fc01f03 (accessed on 15 February 2022).
5. https://github.com/ShichenXie/scorecardpy (accessed on 15 February 2022).
6. https://archive.ics.uci.edu/ml/datasets/South+German+Credit+%28UPDATE%29 (accessed on 15 February 2022).
7. https://www.lendingclub.com/ (accessed on 15 February 2022).
8. Note that the package `creditmodel` supports a `pos_flag` to define the level of the positive class which currently does not work for Binning and Weights of Evidence.
9. Argument `equal_bins = FALSE` or initial bins of equal sample size otherwise.
10. An example code for the package `riskr` is given in Snippet 2 of the supplementary code.
11. An example using a lookup table for the variable `purpose` is given in Snippet 3 of the supplementary code.
12. A code example of looping through all (numeric) variables for the package `smbinning` is given in Snippet 4 of the supplementary code.
13. An example code for application of this mapping to new data is given in Snippet 5 of the supplementary code. The names of the resulting new levels are the concatenated old levels, separated by commas. Note that the function cannot deal with commas in the original level names: a new level `<NA>` will be assigned
14. Using `method = ‘‘chimerge’’`.
15. Using `best = TRUE`.
16. This can be easily checked using the variable `purpose`, cf. e.g., Snippet 6 of the supplementary code.
17. A code snippet for creating a `breaks_list` (cf. above) from a binning result using the package `woeBinning` that can be imported for further use within the package `scorecard`, e.g., for manual manipulation of the bins is given by the function `woeBins2breakslist()` in Snippet 7 of the supplementary code
18. See footnote 10.
19. Note that the call of `glmdisc()` ran in an internal error (`incorrect number of subscripts on matrix`) for more than 10 iterations. For this reason the number of iterations has been reduced to 10 which is much smaller than the default of 1000 iterations and the reported Gini coefficient does still strongly vary among subsequent iterations. For larger numbers of iterations better results might have been possible.
20. Its argument `data` denotes the training data, `output` is a data frame with two variables specifying the variable names of the training data (`character`) and the corresponding cluster index, as given, e.g., by the result from `variable.clustering()`. Finally, its arguments `variables` and `clusters` denote the names of these two variables in the data frame from the `output` argument where the clustering results are stored.
21. A remedy how it can be used in combination with WoE assignment using the package `klaR` as shown in Example 4 is given in Snippet 9 of the supplementary code.
22. Snippet 10 of the supplementary code illustrates how the vector `x` of the names of the input variables in the original data frame can be extracted from the `bicglm` model after variable selection from Example 12.
23. For the function `augmentation()`, this is obtained by rounding the posterior probabilities to the first digit.
24. Here, the augmented weights within each score-band are computed by $1 + \frac{n_{rejected}}{n_{accepted}}$.
25. Within the function `parcelling()` this is done by sampling the labels from a binomial distribution.
26. As an exception, the package `creditR` has been developed as an extension of the package `woeBinning`.

[27] Cf. corresponding footnotes in the paper. Supplementary code is available under https://github.com/g-rho/CSwR (accessed on 15 February 2022).

References

Anagnostopoulos, Christoforos, and David J. Hand. 2019. *Hmeasure: The H-Measure and Other Scalar Classification Performance Metrics*. R Package Version 1.0-2. Available online: https://CRAN.R-project.org/package=hmeasure (accessed on 15 February 2022).

Anderson, Raymond. 2007. *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford: Oxford University Press.

Anderson, Raymond. 2019. *Credit Intelligence & Modelling: Many Paths through the Forest*. Oxford: Oxford University Press.

Azevedo, Ana, and Manuel F. Santos. 2008. KDD, SEMMA and CRISP-DM: A parallel overview. Paper presented at IADIS European Conference on Data Mining 2008, Amsterdam, The Netherlands, July 24–26. pp. 182–85.

Baesens, Bart, Tony Van Gestel, Stijn Viaene, Maria Stepanova, Johan Suykens, and Jan Vanthienen. 2002. Benchmarking state-of-the-art classification algorithms for credit scoring. *JORS* 54: 627–35. [CrossRef]

Banasik, John, and Jonathan Crook. 2007. Reject inference, augmentation and sample selection. *European Journal of Operational Research* 183: 1582–94. [CrossRef]

Biecek, Przemyslaw. 2018. DALEX: Explainers for complex predictive models. *Journal of Machine Learning Research* 19: 1–5.

Binder, Martin, Florian Pfisterer, Michel Lang, Lennart Schneider, Lars Kotthoff, and Bernd Bischl. 2021. mlr3pipelines—Flexible machine learning pipelines in r. *Journal of Machine Learning Research* 22: 1–7.

Bischl, Bernd, Martin Binder, Michel Lang, Tobias Pielok, Jakob Richter, Stefan Coors, Janek Thomas, Theresa Ullmann, Marc Becker, Anne-Laure Boulesteix, and et al. 2021. Hyperparameter optimization: Foundations, algorithms, best practices and open challenges. *arXiv* arXiv:2107.05847.

Bischl, Bernd, Tobias Kühn, and Gero Szepannek. 2016. On class imbalance correction for classification algorithms in credit scoring. In *Proceedings Operations Research 2014*. Edited by Marco Lübbecke, Arie Koster, Peter Lethmathe, Reinhard Madlener, Britta Peis, and Grit Walther. Heidelberg: Springer, pp. 37–43.

Bischl, Bernd, Michel Lang, Lars Kotthoff, Julia Schiffner, Jakob Richter, Erich Studerus, Giuseppe Casalicchio, and Zachary Jones. 2016. mlr: Machine learning in R. *Journal of Machine Learning Research* 17: 1–5.

Bischl, Bernd, Michel Lang, Lars Kotthoff, Patrick Schratz, Julia Schiffner, Jakob Richter, Zachary Jones, Giuseppe Casalicchio, Mason Gallo, Jakob Bossek, and et al. 2020. *mlr: Machine Learning in R*. R Package Version 2.17.1. Available online: https://CRAN.R-project.org/package=mlr (accessed on 15 February 2022).

Bischl, Bernd, Olaf Mersmann, Heike Trautmann, and Claus Weihs. 2012. Resampling methods for meta-model validation with recommendations for evolutionary computation. *Evolutionary Computation* 20: 249–75. [CrossRef]

Bravo, Cristian, Seppe van den Broucke, and Thomas Verbraken. 2019. *EMP: Expected Maximum Profit Classification Performance Measure*. R Package Version 2.0.5. Available online: https://CRAN.R-project.org/package=EMP (accessed on 15 February 2022).

Breiman, Leo, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. 1984. *Classification and Regression Trees*. New York: Wadsworth and Brooks.

Brown, Iain, and Christophe Mues. 2012. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert Systems with Applications* 39: 3446–53. [CrossRef]

Bücker, Michael, Maarten van Kampen, and Walter Krämer. 2013. Reject inference in consumer credit scoring with nonignorable missing data. *Journal of Banking & Finance* 37: 1040–45.

Bücker, Michael, Gero Szepannek, Alicja Gosiewska, and Przemyslaw Biecek. 2021. Transparency, Auditability and explainability of machine learning models in credit scoring. *Journal of the Operational Research Society* 73: 1–21. [CrossRef]

Chavent, Marie, Vanessa Kuentz, Benoit Liquet, and Jerome Saracco. 2017. *ClustOfVar: Clustering of Variables*. R Package Version 1.1. Available online: https://CRAN.R-project.org/package=ClustOfVar (accessed on 15 February 2022).

Chavent, Marie, Vanessa Kuentz-Simonet, Benoît Liquet, and Jerome Saracco. 2012. Clustofvar: An r package for the clustering of variables. *Journal of Statistical Software* 50: 1–6. [CrossRef]

Chawla, Nitesh V., Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. 2002. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* 16: 321–57. [CrossRef]

Chen, Chaofan, Kangcheng Lin, Cynthia Rudin, Yaron Shaposhnik, Sijia Wang, and Tong Wang. 2018. An interpretable model with globally consistent explanations for credit risk. *arXiv* arXiv:1811.12615.

Chen, Tianqi, and Carlos Guestrin. 2016. Xgboost: A scalable tree boosting system. Paper presented at the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13–17. pp. 785–94. [CrossRef]

Chen, Tianqi, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, Rory Mitchell, Ignacio Cano, Tianyi Zhou, and et al. 2021. *xgboost: Extreme Gradient Boosting*. R Package Version 1.5.0.2. Available online: https://CRAN.R-project.org/package=xgboost (accessed on 15 February 2022).

Cordón, Ignacio, Salvador García, Alberto Fernández, and Francisco Herrera. 2018. Imbalance: Oversampling algorithms for imbalanced classification in r. *Knowledge-Based Systems* 161: 329–41. [CrossRef]

Cordón, Ignacio, Salvador García, Alberto Fernández, and Francisco Herrera. 2020. *imbalance: Preprocessing Algorithms for Imbalanced Datasets*. R Package Version 1.0.2.1. Available online: https://CRAN.R-project.org/package=imbalance (accessed on 15 February 2022).

Crone, Sven F., and Steven Finlay. 2012. Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting* 28: 224–38. [CrossRef]

Crook, Jonathan, and John Banasik. 2004. Does reject inference really improve the performance of application scoring models? *Journal of Banking & Finance* 28: 857–74. [CrossRef]

Csárdi, Gábor. 2019. *cranlogs: Download Logs from the 'RStudio' 'CRAN' Mirror*. R Package Version 2.1.1. Available online: https://CRAN.R-project.org/package=cranlogs (accessed on 15 February 2022).

DeLong, Elizabeth R., David M. DeLong, and Daniel L. Clarke-Pearson. 1988. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics* 44: 837–45. [CrossRef]

Dis, Ayhan. 2020. *creditR: A Credit Risk Scoring and Validation Package*. R Package Version 0.1.0. Available online: https://github.com/ayhandis/creditR (accessed on 15 February 2022).

Dua, Dheeru, and Casey Graff. 2019. *UCI Machine Learning Repository*. Irvine: University of California.

Ehrhardt, Adrien. 2018. *scoringTools: Credit Scoring Tools*. R Package Version 0.1.2 Available online: https://CRAN.R-project.org/package=scoringTools (accessed on 15 February 2022).

Ehrhardt, Adrien. 2020. *glmtree: Logistic Regression Trees*. R Package Version 0.2. Available online: https://CRAN.R-project.org/package=glmtree (accessed on 15 February 2022).

Ehrhardt, Adrien, Christophe Biernacki, Vincent Vandewalle, and Philippe Heinrich. 2019. Feature quantization for parsimonious and interpretable predictive models. *arXiv* arXiv:1903.08920.

Ehrhardt, Adrien, Christophe Biernacki, Vincent Vandewalle, Philippe Heinrich, and Sébastien Beben. 2019. Réintégration des refusés en credit scoring. *arXiv* arXiv:1903.10855.

Ehrhardt, Adrien, and Vincent Vandewalle. 2020. *glmdisc: Discretization and Grouping for Logistic Regression*. R Package Version 0.6. Available online: https://CRAN.R-project.org/package=glmdisc (accessed on 15 February 2022).

Eichenberg, Thilo. 2018. *woeBinning: Supervised Weight of Evidence Binning of Numeric Variables and Factors*. R Package Version 0.1.6. Available online: https://CRAN.R-project.org/package=woeBinning (accessed on 15 February 2022).

Fan, Dongping. 2022. *creditmodel Toolkit for Credit Modeling, Analysis and Visualization*. R Package Version 1.3.1. Available online: https://CRAN.R-project.org/package=creditmodel (accessed on 15 February 2022).

Financial Stability Board. 2017. Artificial Intelligence and Machine Learning in Financial Services—Market Developments and Financial Stability Implications. Available online: https://www.fsb.org/2017/11/artificial-intelligence-and-machine-learning-in-financial-service/ (accessed on 15 February 2022).

Finlay, Steven. 2012. *Credit Scoring, Response Modelling and Insurance Rating*. London: Palgarve MacMillan.

Fox, John, and Sanford Weisberg. 2019. *An R Companion to Applied Regression*, 3rd ed. Thousand Oaks: Sage.

Fox, John, Sanford Weisberg, Brad Price, Daniel Adler, Douglas Bates, Gabriel Baud-Bovy, Ben Bolker, Steve Ellison, David Firth, Michael Friendly, and et al. 2021. *car: Companion to Applied Regression*. R Package Version 3.0-12. Available online: https://CRAN.R-project.org/package=car (accessed on 15 February 2022).

Goodman, Bryce, and Seth Flaxman. 2017. European union regulations on algorithmic decision-making and a "right to explanation". *AI Magazine* 38: 50–57. [CrossRef]

Greenwell, Brandon, Bradley Boehmke, Jay Cunningham, and GBM Developers. 2020. *gbm: Generalized Boosted Regression Models*. R Package Version 2.1.8. Available online: https://CRAN.R-project.org/package=gbm (accessed on 15 February 2022).

Groemping, Ulrike. 2019. *South German Credit Data: Correcting a Widely Used Data Set*. Technical Report 4/2019. Berlin: Department II, Beuth University of Applied Sciences Berlin.

Hand, David. 2009. Measuring classifier performance: A coherent alternative to the area under the roc curve. *Machine Learning* 77: 103–23. [CrossRef]

Hand, David, and William Henley. 1993. Can reject inference ever work? *IMA Journal of Management Matehmatics* 5: 45–55. [CrossRef]

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning*, 2nd ed. New York: Springer.

Hoffmann, Hans. 1994. German credit data set (statlog). Available online: https://archive.ics.uci.edu/ml/datasets/statlog+(german+credit+data) (accessed on 15 February 2022).

Hosmer, David W., and Stanley Lemeshow. 2000. *Applied Logistic Regression*. Hoboken: Wiley.

Hothorn, Thorsten, Kurt Hornik, and Achim Zeileis. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics* 15: 651–74. [CrossRef]

Hothorn, Torsten, and Achim Zeileis. 2015. partykit: A modular toolkit for recursive partytioning in R. *Journal of Machine Learning Research* 16: 3905–9.

Izrailev, Sergei. 2015. *Binr: Cut Numeric Values into Evenly Distributed Groups*. R Package Version 1.1. Available online: https://CRAN.R-project.org/package=binr (accessed on 15 February 2022).

Jopia, Herman. 2019. *smbinning: Optimal Binning for Scoring Modeling*. R Package Version 0.9. Available online: https://CRAN.R-project.org/package=smbinning (accessed on 15 February 2022).

Kaszynski, Daniel. 2020. Background of credit scoring. In *Credit Scoring in Context of Interpretable Machine Learning*. Edited by Daniel Kaszynski, Bogumil Kaminski and Tomasz Szapiro. Warsaw: SGH, pp. 17–26.

Kaszynski, Daniel, Bogumil Kaminski, and Tomasz Szapiro. 2020. *Credit Scoring in Context of Interpretable Machine Learning*. Warsaw: SGH.
Kim, HyunJi. 2012. *Discretization: Data Preprocessing, Discretization for Classification*. R Package Version 1.0-1. Available online: https://CRAN.R-project.org/package=discretization (accessed on 15 February 2022).
Kuhn, Max. 2008. Building predictive models in r using the caret package. *Journal of Statistical Software* 28: 1–26. [CrossRef]
Kuhn, Max. 2021. *Caret: Classification and Regression Training*. R Package Version 6.0-90. Available online: https://CRAN.R-project.org/package=caret (accessed on 15 February 2022).
Kunst, Joshua. 2020. *riskr: Functions to Facilitate the Evaluation, Monitoring and Modeling process*. R Package Version 1.0. Available online: https://github.com/jbkunst/riskr (accessed on 15 February 2022).
Landgraf, Andrew J. 2016. *logisticPCA: Binary Dimensionality Reduction*. R Package Version 0.2. Available online: https://CRAN.R-project.org/package=logisticPCA (accessed on 15 February 2022).
Landgraf, Andrew J., and Yoonkyung Lee. 2015. Dimensionality reduction for binary data through the projection of natural parameters. *arXiv* arXiv:1510.06112.
Lang, Michel, Martin Binder, Jakob Richter, Patrick Schratz, Florian Pfisterer, Stefan Coors, Quay Au, Giuseppe Casalicchio, Lars Kotthoff, and Bernd Bischl. 2019. mlr3: A modern object-oriented machine learning framework in R. *Journal of Open Source Software* 4: 1903. [CrossRef]
Lang, Michel, Bernd Bischl, Jakob Richter, Patrick Schratz, Giuseppe Casalicchio, Stefan Coors, Quay Au, and Martin Binder. 2021. *mlr3: Machine Learning in R—Next Generation*. R Package Version 0.13.0. Available online: https://CRAN.R-project.org/package=mlr3 (accessed on 15 February 2022).
Larsen, Kim. 2016. *Information: Data Exploration with Information Theory*. R Package Version 0.0.9. Available online: https://CRAN.R-project.org/package=Information (accessed on 15 February 2022).
Lele, Subhash R., Jonah L. Keim, and Peter Solymos. 2019. *ResourceSelection: Resource Selection (Probability) Functions for Use-Availability Data*. R Package Version 0.3-5. Available online: https://CRAN.R-project.org/package=ResourceSelection (accessed on 15 February 2022).
Lessmann, Stefan, Bart Baesens, Hsin-Vonn Seow, and Lyn Thomas. 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European Journal of Operational Research* 247: 124–36. [CrossRef]
Liaw, Andy, and Matthew Wiener. 2002. Classification and regression by randomforest. *R News* 2: 18–22.
Ligges, Uwe. 2009. *Programmieren Mit R*, 3rd ed. Heidelberg: Springer.
Little, Roderick, and Donald Rubin. 2002. *Statistical Analysis with Missing Data*. Hoboken: Wiley.
Louzada, Francisco, Anderson Ara, and Guilherme Fernandes. 2016. Classification methods applied to credit scoring: A systematic review and overall comparison. *Surveys in OR and Management Science* 21: 117–34. [CrossRef]
Maechler, Martin, Peter Rousseeuw, Anja Struyf, Mia Hubert, and Kurt Hornik. 2021. *Cluster: Cluster Analysis Basics and Extensions*. R Package Version 2.1.2. Available online: https://CRAN.R-project.org/package=cluster (accessed on 15 February 2022).
Molnar, Christoph, Bernd Bischl, and Giuseppe Casalicchio. 2018. iml: An R package for Interpretable Machine Learning. *JOSS* 3: 786. [CrossRef]
Paradis, Emmanuel, Simon Blomberg, Ben Bolker, Joseph Brown, Julien Claude, Hoa Sien Cuong, Richard Desper, Gilles Didier, Benoit Durand, Julien Dutheil, and et al. 2021. *ape: Analyses of Phylogenetics and Evolution*. R Package Version 5.6.1. Available online: https://CRAN.R-project.org/package=ape (accessed on 15 February 2022).
Paradis, Emmanuel and Klaus Schliep. 2018. ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35: 526–28. [CrossRef]
Poddar, Arya. 2019. *scorecardModelUtils: Credit Scorecard Modelling Utils*. R Package Version 0.0.1.0. Available online: https://CRAN.R-project.org/package=scorecardModelUtils (accessed on 15 February 2022).
Pozzolo, Andrea Dal, Olivier Caelen, and Gianluca Bontempi. 2015. *unbalanced: Racing for Unbalanced Methods Selection*. R Package Version 2.0. Available online: https://CRAN.R-project.org/package=unbalanced (accessed on 15 February 2022).
Prabhakaran, Selva. 2016. *InformationValue: Performance Analysis and Companion Functions for Binary Classification Models*. R Package Version 1.2.3. Available online: https://CRAN.R-project.org/package=InformationValue (accessed on 15 February 2022).
Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalie Tiberti, Frederique Lisacek, Jean Sanchez, and Markus Müller. 2011. pRoc: An open-source package for r and s+ to analyze and compare roc curves. *BMC Bioinformatics* 12: 1–8. [CrossRef]
Robin, Xavier, Natacha Turck, Alexandre Hainard, Natalia Tiberti, Frederique Lisacek, Jean-Charles Sanchez, Markus Müller, Stefan Siegert, and Matthias Doering. 2021. *pROC: Display and Analyze ROC Curves*. R Package Version 1.18.0. Available online: https://CRAN.R-project.org/package=pROC (accessed on 15 February 2022).
Roever, Christian, Nils Raabe, Karsten Luebke, Uwe Ligges, Gero Szepannek, and Marc Zentgraf. 2020. *klaR: Classification and visualization*. R Package Version 0.6-15. Available online: https://CRAN.R-project.org/package=klaR (accessed on 15 February 2022).
Rudin, Cynthia. 2019. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* 1: 206–15. [CrossRef]
Ryu, Choonghyun. 2021. *dlookr: Tools for Data Diagnosis, Exploration, Transformation*. R Package Version 0.5.4. Available online: https://CRAN.R-project.org/package=dlookr (accessed on 15 February 2022).

Scallan, Gerard. 2011. Class(ic) scorecards—Selecting attributes in logistic regression. In *Credit Scoring and Credit Control XIII*. Available online: https://www.scoreplus.com/papers/paper (accessed on 15 February 2022).
Schiltgen, Garrett. 2015. *boottol: Bootstrap Tolerance Levels for Credit Scoring Validation Statistics*. R Package Version 2.0. Available online: https://CRAN.R-project.org/package=boottol (accessed on 15 February 2022).
Sharma, Dhuv. 2009. *Guide to Credit Scoring in R.* CRAN Documentation Contribution. Available online: https://cran.r-project.org/ (accessed on 15 February 2022).
Siddiqi, Naeem. 2006. *Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*, 2nd ed. Hoboken: Wiley.
Sing, Tobias, Oliver Sander, Niko Beerenwinkel, and Thomas Lengauer. 2005. ROCR: Visualizing classifier performance in R. *Bioinformatics* 21: 7881. [CrossRef]
Sing, Tobias, Oliver Sander, Niko Beerenwinkel, and Thomas Lengauer. 2020. *ROCR: Visualizing the Performance of Scoring Classifiers*. R Package Version 1.0-11. Available online: https://CRAN.R-project.org/package=ROCR (accessed on 15 February 2022).
Siriseriwan, Wacharasak. 2019. *smotefamily: A Collection of Oversampling Techniques for Class Imbalance Problem Based on SMOTE*. R Package Version 1.3.1. Available online: https://CRAN.R-project.org/package=smotefamily (accessed on 15 February 2022).
Staniak, Mateusz, and Przemysław Biecek. 2019. The Landscape of R Packages for Automated Exploratory Data Analysis. *The R Journal* 11: 347–69. [CrossRef]
Stratman, Eric, Riaz Khan, and Allison Lempola. 2020. *Rprofet: WOE Transformation and Scorecard Builder*. R Package Version 2.2.1. Available online: https://CRAN.R-project.org/package=Rprofet (accessed on 15 February 2022).
Sun, Xu, and Weichao Xu. 2014. Fast implementation of delong's algorithm for comparing the areas under correlated receiver operating characteristic curves. *IEEE Signal Processing Letters* 21: 1389–93. [CrossRef]
Szepannek, Gero. 2017. On the practical relevance of modern machine learning algorithms for credit scoring applications. *WIAS Report Series* 29: 88–96.
Szepannek, Gero. 2019. How much can we see? A note on quantifying explainability of machine learning models. *arXiv* arXiv:1910.13376.
Szepannek, Gero, and Karsten Lübke. 2021. Facing the challenges of developing fair risk scoring models. *Frontiers in Artificial Intelligence* 4: 117. [CrossRef] [PubMed]
Therneau, Terry, and Beth Atkinson. 2019. *rpart: Recursive Partitioning and Regression Trees*. R Package Version 4.1-15. Available online: https://CRAN.R-project.org/package=rpart (accessed on 15 February 2022).
Thomas, Lynn C., Jonathan N. Crook, and David B. Edelman. 2019. *Credit Scoring and its Applications*, 2nd ed. Philadelphia: SIAM.
Thoppay, Sudarson. 2015. *woe: Computes Weight of Evidence and Information Values*. R Package Version 0.2. Available online: https://CRAN.R-project.org/package=woe (accessed on 15 February 2022).
Verbraken, Thomas, Christian Bravo, Weber Richard, and Baesens Baesens. 2014. Development and application of consumer credit scoring models using profit-based classification measures. *European Journal of Operational Research* 238: 505–13. [CrossRef]
Verstraeten, Geert, and Dirk Van den Poel. 2005. The impact of sample bias on consumer credit scoring performance and profitability. *JORS* 56: 981–92. [CrossRef]
Vigneau, Evelyne, Mingkun Chen, and Veronique Cariou. 2020. *ClustVarLV: Clustering of Variables Around Latent Variables*. R Package Version 2.0.1. Available online: https://CRAN.R-project.org/package=ClustVarLV (accessed on 15 February 2022).
Vigneau, Evelyne, Mingkun Chen, and El Mostafa Qannari. 2015. ClustVarLV: An R Package for the Clustering of Variables Around Latent Variables. *The R Journal* 7: 134–48. [CrossRef]
Vincotti, Veronica, and David Hand. 2002. Scorecard construction with unbalanced class sizes. *Journal of The Iranian Statistical Society* 2: 189–205.
Wrzosek, Malgorzata, Daniel Kaszynski, Karol Przanowski, and Sebastian Zajac. 2020. Selected machine learning methods used for credit scoring. In *Credit Scoring in Context of Interpretable Machine Learning*. Edited by Daniel Kaszynski, Bogumil Kaminski and Tomasz Szapiro. Warsaw: SGH, pp. 83–146.
Xie, Shichen. 2021. *scorecard: Credit Risk Scorecard*. R Package Version 0.3.6. Available online: https://CRAN.R-project.org/package=scorecard (accessed on 15 February 2022).
Zumel, Nina, and John Mount. 2014. *Practical Data Science with R.* New York: Manning.

risks

Article

Risk Factors Affecting Bancassurance Development in Poland

Adam Śliwiński [1,*], Joanna Dropia [2] and Norbert Duczkowski [2]

[1] Risk and Insurance Department, Warsaw School of Economics, 02-554 Warszawa, Poland
[2] Doctoral School, Warsaw School of Economics, 02-554 Warszawa, Poland; jd110752@doktorant.sgh.waw.pl (J.D.); nd110753@doktorant.sgh.waw.pl (N.D.)
* Correspondence: asliwin@sgh.waw.pl

Abstract: The aim of the article is to identify the risk factors affecting bancassurance development in Poland. The development is understood here as a change of gross written premiums obtained through banks in Poland. The group of risk factors selected in a survey conducted among financial sector employees was subject to statistical verification. The analysis used both variables directly related to the insurance product (e.g., a regulatory restriction of insurance acquisition costs) as well as those resulting from the specificity of the bancassurance channel, such as the sales of banking products, i.e., cash loans, housing loans and the value of funds placed by customers on deposits. The study was conducted on the basis of data on the gross premiums written in Poland in the years 2004–2019. The result of the applied model confirms the assumptions and the importance of insurance distribution in banks. Significant risk factors (statistically significant) which determine gross premiums written in the bancassurance channel are: the size of policyholder's family (number of children, dependants) represented by the average number of people in a household in Poland, demand on mortgage loans represents by bank housing loans for households and agent's commission, represented by the ratio of acquisition costs to gross written premium. The results of the econometric model obtained are consistent with expectations arising from the principles and practice of cooperation between banks and insurers as well as the specificity of insurance products distribution (also local) in the bancassurance channel.

Keywords: bancassurance; insurance; risk factors

JEL Classification: G22; G52

Citation: Śliwiński, Adam, Joanna Dropia, and Norbert Duczkowski. 2021. Risk Factors Affecting Bancassurance Development in Poland. *Risks* 9: 130. https://doi.org/10.3390/risks9070130

Academic Editors: Krzysztof Jajuga and Józef Dziechciarz

Received: 10 May 2021
Accepted: 21 June 2021
Published: 7 July 2021

1. Introduction

Bancassurance is a term used for the first time in Poland in the 1990s and means the cooperation of banks and insurance companies in order to distribute protection products to retail customers. As part of such cooperation, an insurance company prepares a product offer for bank customers, which is next distributed through bank branches, its hotline or internet banking platforms. Bancassurance plays an important role in insurance distribution from the perspective of customers, banks and insurance companies. Although the phenomenon of cooperation between banks and insurance companies is not new and has been described in many studies, it focuses on the general rules of this market functioning or legal and financial aspects. Publications place less emphasis on the practical aspects of cooperation between banks and insurers or on risk factors modelling (a research gap), including the specificities of insurance products distribution in bancassurance.

The aim of the article is to indicate risk factors affecting the level of gross premiums written obtained by insurers through banks and present the importance of bancassurance as a way of insurance distribution in Poland. The Polish market was selected for analysis because the Polish insurance market is the biggest among other CEE countries in the region. The Polish insurance market, as in other CEE countries, is still regarded as an emerging market but it is worth to pointed out that the financial sector in countries such as Poland plays a crucial role in overall macroeconomic development and capital flow in the region (Śliwiński 2019). Poland is also a good example of a market that has passed through economic transformations. It is also worth adding that Poland has significant shares in the bancassurance channel in the distribution of insurance in the region, and the biggest bancassurance market among other CEE.

A group of risk factors selected in a survey conducted among financial sector employees was subject to statistical verification. The analysis used both variables directly related to an insurance product (e.g., insurance acquisition costs) and risk factors resulting from the specificity of bancassurance, such as the sales of banking products, i.e., cash loans, housing loans and the value of funds placed by customers on deposits.

This publication is divided into six parts. In the Introduction (Section 1), the aim of the article is stated, and the content of its sections is described. In Section 2, an overview of publications concerning bancassurance is presented, as well as an attempt at their synthesis and a discussion of the studies that are key, in the authors' opinion. Next, in Section 3, the importance of insurance distribution by banks is presented. Section 4 of this publication includes characteristics of the scope of the research and its stages with a description of the research method used. The group of risk factors selected in the survey conducted among financial sector professionals was subjected to statistical verification. The survey results and discussion are presented in Section 5. The last section (Section 6) contains a conclusions summary.

2. Bancassurance—Literature Overview

Bancassurance, although defined fairly recently, has already been the subject of a number of scientific studies. In the international literature, publications dedicated to the problems of a given country (Dharmaraj 2019; Saha and Dutta 2019; Chen 2019) or attempts of cross-border comparisons of relationships between banks and insurers (Kramaric 2019; Preckova 2017) prevail, with the account given to the specificities of a given markets and the differences between them. Such studies are most often conducted for developed markets (Kramaric 2019; Liang 2015), where the history of bancassurance is relatively long or for markets that have a population potential manifested by a huge number of inhabitants, e.g., in India, over 200 million people belong to the middle class (Puja et al. 2019). They also include attempts of summarising the previous research. An exceptionally interesting attempt at a synthesising study in the bancassurance area, in the opinion of the authors, was presented in (Ricci and Fiordelisi 2012). This publication discusses articles on the diversification hypothesis, assuming that the turning of banks to profitable business, which insurance undoubtedly is, offers banks a possibility of improving their risk profile and the rate of return on the banking portfolio. The research also presented a comparison of the performance of institutions diversifying their business together with specialised companies and analysed an increase in efficiency through cost-effectiveness and profit synergies among several types of financial activities. Also particularly interesting are the insights and conclusions drawn from studies on markets' responses to the announcement of certain events related to bancassurance. A separate group is a summary of articles based on empirical analyses evaluating the wealth effect as well as mergers and acquisitions involving banks and insurance companies. Ricci and Fiordelisi (2012) also take note of other empirical studies on bancassurance, particularly focusing on the combination of banking and insurance activities, but they were not included in any other group. This division is visible in other studies conducted by authors from around the world.

Let us take a closer look at the research on economic efficiency. It is natural that bancassurance business is conducted due to the financial benefits. These benefits may be as follows: profitability increasing, cost-effectiveness and risk reduction (from the bank perspective). Bancassurance research is dominated by studies confirming this observation (Kramaric 2019; Karimian 2017; Hota 2016; Sreesha and Joseph 2011). On the other hand, there are works that indicate the complexity of the topic and emphasise the ambiguity of such a statement. An example of such research can be (Ricci and Fiordelisi 2010). They analysed bancassurance profit and cost efficiency for different ownership types of bank and insurance relation and took into account the impact of firm-specific factors on the attained performance. As a result, the distribution of premiums through bank branches appears positively related to profit efficiency, but the relationship is not statistically significant. What is more is that they do not find strong evidence in favour of revenue synergies for insurance companies cooperating with banks (Ricci and Fiordelisi 2010). As an explanation of this fact was the indication that there are no price advantages in selling policies through bank branches, or that banks are able to benefit most from this cooperation. Our observations concerning the Polish market show that the second explanation may be more valid. The ambiguity of the statement that bancassurance posits as a profitable complement for both banks and insurance companies was also confirmed by research (Ricci and Fiordelisi 2011). In this work, they try to assess Bancassurance performance gains (from both the banking and the insurance standpoints) by estimating cost and profit efficiency using stochastic frontier analysis. With regard to the banking industry, they do not observe any strong evidence in favour of entering the life insurance business. The investigation into the insurance industry highlights the competitive viability of Bancassurance as a distribution channel, especially in terms of cost-efficiency. In terms of profitability, findings suggest that the mix of products should be continuously revised to adapt to customer needs and the evolution of financial markets (Ricci and Fiordelisi 2011). As a consequence, more flexible forms of cooperation, such as cross-selling agreements, should be taken into account as bancassurance cooperation.

It is also worth paying attention to another trend observed in the world literature related to bancassurance. Customer experience has attracted a lot of attention in scientific research on the relationship between banks and insurers. It is believed that customer experience adds a competitive advantage to service firms (Choudhury and Singh 2021). However, the competitive advantage in bancassurance must be viewed in a broader sense, because of the fact that banks face competition not only within the banking industry but also with the insurance industry (Choudhury and Singh 2021). Key factors influencing customer experience in bancassurance in the 7p marketing mix dimensions are presented in Table 1. As can be seen from the table below, the key factors influencing the customer experience are rather related to the processes and specificity of the insurance distribution channel in banks (Place).

Table 1. Key factors influencing customer experience in bancassurance in the 7p marketing mix dimensions.

7p Marketing Mix Dimension	Key Factors Influencing Customer Experience in Bancassurance
Product	Product variety, receiving maturity benefit, the facility to get the claim payments, designing of service offering, ease of buying, after-sale services, past experiences.
Price	Price, designing of service offering.
Promotion	Price, designing of service offering.
Place	Image, reliability of the channel, responsiveness of the channel, accuracy of the channel, opinion about staff, multiple service delivery channels, service quality (service convenience),
People	Interpersonal relationships, trust, satisfaction on service, friendliness of the bank personnel, recommendation for business, past experiences.
Process	Ease of buying, after-sale services, digitisation of the process, receiving stock market-related information, recommendation for business, past experiences.
Physical evidence	Image, brand, trust, pleasant and welcoming branch environment, friendliness of the bank personnel.

Source: own study based on (Choudhury and Singh 2021; Fan et al. 2011).

During the analysis of international literature, with particular emphasis on the latest works published in 2021, we confirm the trends mention at the beginning of this section. In the recent international literature, publications dedicated to problems of a given country

(Zharikova and Cherkesenko 2021) or attempts of cross-border comparisons of relationships between banks and insurers prevail, with the account given to the specificities of given markets and differences between them, including digital (Ganapathy 2021). This confirmed our belief that it is impossible to conduct research in the field of bancassurance without taking into account the local specificity of this business.

Among Polish scientists, the topic was eagerly described in the first decade of the Twenty-first century. An increase in the number of publications was also observed in 2011 and 2015. Recently, the topic of bancassurance has lost some of its popularity, which can be proven by the number of publications in a given year available in the EKON base. The smaller number of publications was not in line with dynamic changes in the legal environment, which had a significant impact on the cooperation of banks and insurers, such as: Recommendation U, amendment of the Act on Insurance and Reinsurance Activity or introduction of the Act on Insurance Distribution, implementing to the Polish law the provisions of the EU Insurance Distribution Directive (IDD). In the authors' opinion, all these changes have constituted significant risk factors affecting the level of gross premiums written through bancassurance.

Having analysed the available scientific literature, it can be stated that the issues covered so far concern mainly relationships between banking and insurance sectors and a description of the general characteristics of the bancassurance market functioning. The research and studies prepared by (Monkiewicz 2000; Pajewska 2000), were of great importance for this publication and constituted a foundation of this research. Thus far, studies on bancassurance have included an analysis of market data on the structure of gross written premiums and the scope of insurance. Publications also refer to the matter of credit risk taken by banks and a possibility of its mitigation using relevant life and property insurance. The topic of Insurtech appears in the newest studies as an opportunity for further development of the bancassurance market in Poland (e.g., Polish Insurtech hiPRO and Braintri in Lisowski and Chojan 2020). Studies also include topics related to the legal or consumer aspects and also those related to product and finance.

This thesis can be confirmed by an analysis of articles included in the ECON base (Figure 1). As of 12 February 2021, this database contained 129 publications with bancassurance as a keyword, of which 64 were published in sector magazines and 65 in scientific magazines. The 65 publications from the scientific magazines were assigned to one (45) or two (20) key categories of the thematic area of cooperation between banks and insurers. They were categorised after an analysis of the summaries and keywords. The results are presented in Table 2.

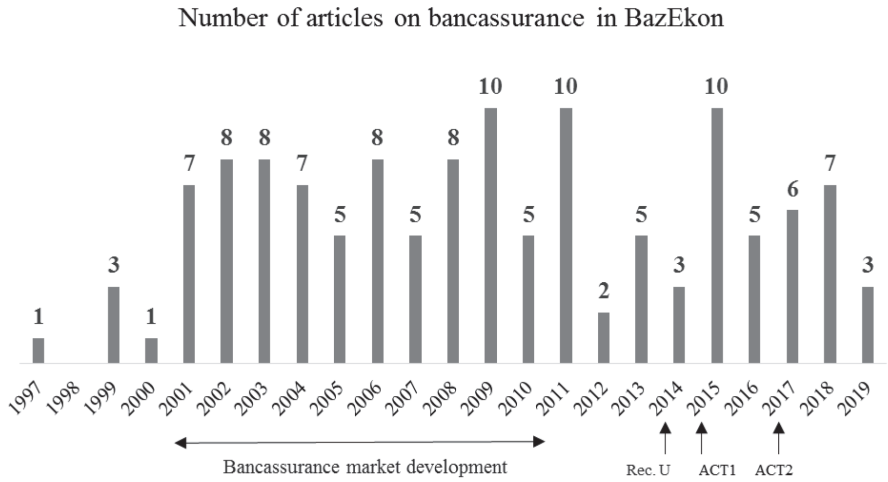


Figure 1. Number of publications in the EKON database with bancassurance as a keyword in the years 1997–2019, source: own study based on data from EKON.

Table 2. Classification of articles with bancassurance as a keyword.

Area of Bancassurance	Number of Publications
General rules of functioning	39
Legal aspects	9
Financial aspects	8
Consumer aspects	2
Sales analysis	8
Product analyses	8
Development premise	11

Source: own study based on data from the EKON database.

The largest number of studies in the EKON Útabase, which fulfil the above mentioned criteria, pertain to the general rules of the bancassurance market functioning. These studies are of various character and level of generality, though a vast majority of them contain characteristics of the existing models of cooperation between banks and insurers (Sadurska 2019; Malinowski 2011; Gajdek 2016; Wierzbicka 2009). These studies also point out the source and factors influencing the development of banking and insurance cooperation (Gajdek 2016; Gwizdała 2018, Swacha-Lech 2006) that could be used to its evaluation (Breś-Błażejczyk 2005). Polish researchers have often analysed trends in risk development by presenting and describing sales analyses and product analyses. In these studies, data show the number and structure of insurance contracts concluded as part of bancassurance in various contexts and timeframes (Košík and Poliak 2011; Ociepa-Kicińska 2019; Pajewska-Kwaśny and Tomaszewska 2009; Pisarewicz 2013, 2014; Staszczyk 2013).

Many studies on the bancassurance sector focus on the changing legal environment and consequences of implementing new regulations such as Recommendation U of the Polish Financial Supervision Authority (Recommendation U on Best Practice in Bancassurnace 2014). These studies point out to legal requirements arising from the legislative acts and the consequences of the newly implemented regulations and their evaluation as well as their impact on the correct functioning of the market and providing better protection to all its participants (Kowalski and Kowalska 2015; Łosiewicz-Dniestrzańska 2016; Sereda 2015).

Almost as often as with legal aspects, studies and papers on bancassurance have dealt with broadly understood financial aspects. They have analysed the principles of settling remuneration received by banks for offering insurance products and assessed an impact of changes in recognition of insurance commissions in banking books on their results, treatment of revenues, costs and products cash flows (Łada and Białas 2017; Pielichaty 2014, 2015, 2017; Szewieczek 2011).

Polish researchers have dedicated much less space to the consumer aspects and customer experience during insurance distribution by banks. Nevertheless, there appeared studies (Wierzbicka 2016) that referred to the instances and consequences of mis-selling, the unfair sales of financial products not matching customer needs. Some of these works are a loose reference to Gigerenzer (2014).

Last but not least is the category called by the authors of this article a development premise. In these papers, authors refer to market development factors in Poland (Gostomski 2011; Dobrucka 2004), by comparing them to the foreign markets (Swacha-Lech 2003), and analyse modern customer service which uses the Insurtech technology (Gwizdała 2018).

An analysis of the Polish and worldwide literature on bancassurance, taking into account its pieces analysing the demand on insurance (Jaspersen 2016; Zietz 2003; Śliwiński 2016, 2019; Śliwiński et al. 2013), has allowed the authors of this paper to find that the area of modelling the risk factors which determine the level of gross written premiums generated as part of bancassurance had not been widely covered in research and, in part 4, they indicated a list of variables that were subject to testing and used for the model. Not without significance was an analysis of the market of insurances offered through banks, included in part 3 of this article.

3. Meaning and Specificity of Bancassurance in Insurance Distribution in Poland

Bancassurance is a commonly used term, although not defined in the Polish provisions of law regulating the insurance distribution market. In a broad sense, bancassurance is a cooperation between banks and insurance companies. Additionally, although bancassurance definitions in the literature have evolved along with the market development, it should be stated that one of the first definitions of the cooperation between banks and insurers in Poland (taking into account local conditions), which was the basis for further studies, was formulated in Pajewska (2000). It most often consists of offering insurance with the use of a bank's distribution network. Usually, the role of banks is to act as an intermediary in concluding individual insurance contracts or offer to join an insurance contract for another party's account concluded by the bank (the so-called group insurances). Insurance offered to bank customers may be bundled with a banking product (e.g., loan, payment card) or not bundled directly with a banking product. Cooperation between banks and insurance companies encompasses life and property insurances, also including insurance-based investment or savings products. It is commonly believed that the cooperation of financial institutions in the scope of bancassurance is beneficial to all stakeholders, i.e., banks, insurance companies and customers (Table 3).

Table 3. Benefits for bancassurance stakeholders.

Bank	Insurance Company	Customer
• Increase in customer loyalty • Products matching customer requirements and expectations • Creation of an attractive and comprehensive offer for customers • Additional source of income • Credit risk reduction (policy as repayment collateral) • Positive impact on marketing activities, cooperation in preparing a new product, conducting market analyses, using brand reputation for own image improvement) • Cross-sell of complementary products and services • Standardisation (offer is clear and simple, leads to operational cost reduction)	• Expanded distribution network • Acquisition cost reduction • Positive impact on marketing activities, brand recognition by customers • Product adjustment to a specific, defined group of bank customers	• Banking and insurance products available in one place (bank branch) • Insurance scope matching a given bank's customer needs • Favourable price of insurance (negotiated by a bank)

Source: own study based on: (Białas 2015; Śliperski 1998).

According to the Act (Insurance Distribution Act 2017), an insurance distributor can be a direct insurance company, but it can also act through an insurance agent, a supplementary insurance agent or an insurance broker. The Act states directly that an insurance intermediary distributes insurances in exchange for remuneration. For the full picture, one should also bear in mind a possibility for the insured to join an insurance contract concluded for the other party's account, particularly group insurance. However, according to the Act (Act on Insurance and Reinsurance Activities 2015), in this relationship, the policyholder may not receive remuneration or other benefits in connection with offering an opportunity to benefit from insurance coverage or activities related to the performance of an insurance contract. The role of banks in insurance distribution in Poland is presented in Appendix B. According to the Act (Act on Insurance and Reinsurance Activities 2015), as in other European countries, risks in insurance are divided into two sections: Section I is life insurances, whereas Section II is other personal insurances and property insurances.

3.1. Life Insurances in Bancassurance

For more than a decade, bancassurance has changed its significance in the distribution of life insurances. As one can see in the chart below (Figure 2) and Figures A1–A3 in Appendix C, the value of gross premiums written and the share in its total value were systematically growing until 2007. In 2008, the value of gross premiums written in bancassurance considerably increased as part of group contracts. The significant share of group insurances maintained until 2014 and, when Recommendation U (Recommendation U on Best Practice in Bancassurnace 2014) and the Act on Insurance and Reassurance Activity (Act on Insurance and Reinsurance Activities 2015) entered into force, their share fell below 10% of the gross premiums written in bancassurance. Such a major change in bank offers was related mainly to regulations restricting the possibility of charging remuneration by policyholders in connection with activities performed under group contracts. The new regulations also had an impact on the distribution of insurances of the investment character because the Act (Act on Insurance and Reinsurance Activities 2015) introduced a cap on the cost of terminating an insurance contract to a maximum of 4% of the premiums paid or the current value of the units of insurance capital funds. The last strong legislative impulse was the entry into force in 2017 of the Act on Insurance Distribution (Insurance Distribution Act 2017), implementing the Insurance Distribution Directive (Directive of the European Parliament and of the Council (EU) 2016) to the Polish law. It imposes a number of requirements on insurance intermediaries, connected with organising insurance distribution, matching an offer to customer needs, as well as information obligations.

Figure 2. Value of gross premiums written in Section I, source: own study based on KNF's data (Polish Financial Supervision Authority 2001–2019).

Life insurances with the highest value of gross premiums written in bancassurance are the products being credit collateral (Payment Protection Insurance, PPI) and savings and investment products. PPI is insurance aimed at facilitating the repayment of a consumer loan, mortgage loan or credit card in the situation of an insured's death or when the borrower becomes ill, has an accident or loses a job. It is considered to be the product that gave rise to bancassurance. This type of insurance has been offered on the British market since the 1990s and its greatest development took place in the years 2001–2009. Already in 2004, the British supervisory institution Financial Services Authority diagnosed that offering PPI involves a risk of violating consumer rights and issued a number of guidelines regulating this market. In Poland, the key risk factor affecting gross premium development, among the others mentioned above, were the restrictions on the level of commission paid to banks, set by the Polish Financial Supervision Authority under Recommendation U. Products of the savings and investment nature can take a form of life insurances from Section I of the Act (Act on Insurance and Reinsurance Activities 2015), in particular, life insurances from group 1 or insurances from group 3. In this category, an offer includes PRIIPs, the assets of which are fully invested in investment funds (the so-called unit-linked) and structured products where assets are partly invested in low-risk investment instruments (bank deposits, treasury bonds) and partly in derivatives. This market started to develop in Poland in the 1990s and was very successful. In the peak moment, the annual value of the gross premiums written under group 3 insurances was PLN 10–12 bn.

The Act (Act on Insurance and Reinsurance Activities 2015) has introduced significant changes in the construction of this type of insurances, in particular a cap on the so-called liquidation fees, namely, the costs incurred by the customer when terminating an insurance contract, to a maximum of 4% of the premiums paid or the current value of insurance capital fund units. It was the legislator's response to the allegations raised by the regulator and institutions representing consumers that customers are misled as to product features and costs related to this type of policy, including too high of costs burdening policyholders in the case of terminating a contract before the end of an insurance period. This change, apart from strengthening customer protection, has also had an impact on the profitability of products offered, from the point of view of an insurance company and a distributor. In practice, it has caused a portion of investment insurances to withdraw from offering this product, in particular the long-term ones with a regular premium paid (period of insurance 10 years or more).

3.2. Other Personal and Property Insurances

It is worth noting that life insurance is dominant in banking distribution, and other personal and property insurances are rather marginal (Figure 3). Other personal and property insurance sales volumes by banks are not significant in relation to life insurance, described in Section 3.1. Out of them, on a visible scale, banks only offer property insurances linked with mortgage loans. Banks' attempts can be observed to open up to the distribution of other Section II insurances as well, e.g., car insurances, civil liability, financial risks, legal protection and travel insurance, but to date, the level of gross premiums written in these groups, generated via bancassurance, has been negligible.

The purpose of a property insurance is to protect the customer against the risk of losing the ability to repay the contracted loan and the bank against the credit risk consisting of no repayment of the liability by the debtor. When insurance contracts are entered into in connection with a mortgage loan, a policy is prepared, along with its assignment to the bank. It means that if the real property becomes destroyed (for instance, by a fire, flood, etc.), the bank will obtain money as compensation from an insurance company and will dedicate it first to the payment of the debt.

From the bank's point of view, insuring the real property is obligatory when a mortgage loan agreement is concluded. This arises from the provisions of the act on the mortgage loan and on the supervision over mortgage loan intermediaries and agents (Act on Mortgage Loan and Supervision over Mortgage Loan Intermediaries and Agents 2017): 'a lender may require from a consumer to conclude or have an insurance contract related to the mortgage loan agreement or an assignment of receivables from this insurance contract to the lender, informing the customer, at the same time, about a possibility of choosing an offer of any insurer who covers the minimum scope of insurance accepted by the lender'. For a borrower, it means that concluding the contract is a condition for being granted a mortgage loan, though they are free to decide whether such a policy is to be bought through the bank to which they apply for a loan or at another insurance intermediary or directly in an insurance company.

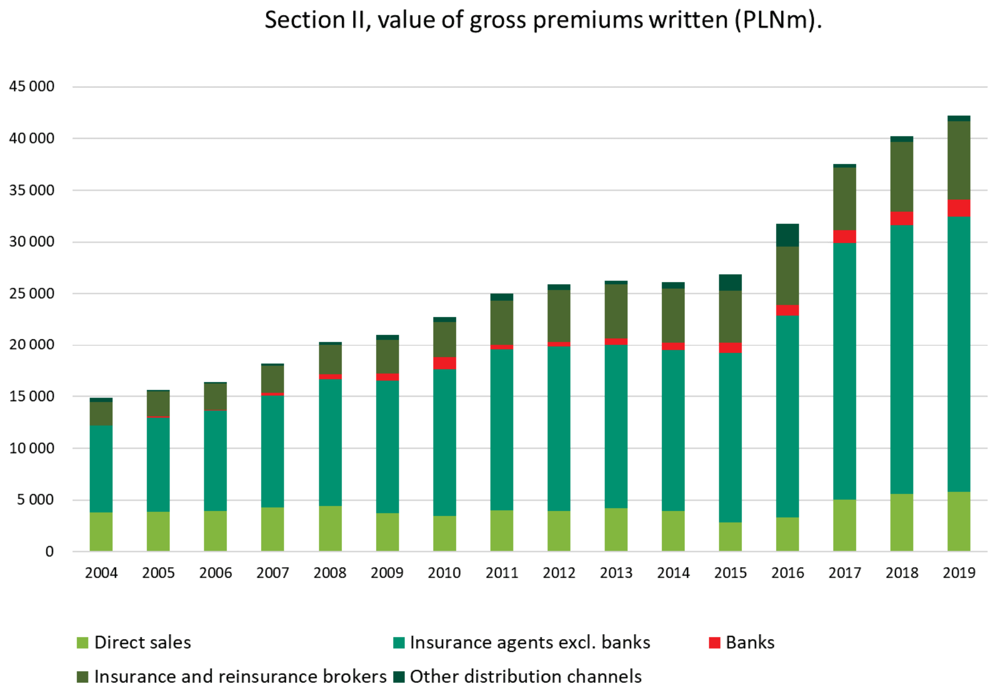


Figure 3. Value of gross premiums written in Section II, source: own study based on KNF's data: (Polish Financial Supervision Authority 2001–2019).

4. The Research Method and Database

The determinants of premiums written in bancassurance were tested in two stages. In the first stage, based on the literature studies (see page 2) and market analysis (page 6), the authors of this study prepared a list of factors that might determine the level of gross written premiums generated by banks. These factors were subject to verification in the form of a survey (Appendix A) among professionals taking part in the process of insurance products preparation and distribution, both in banks and insurance companies. The research was conducted between 15 February 2021 and 26 February 2021. Twenty-seven fully completed questionnaires were obtained as part of the survey out of 39 requests for filling in (the response rate was 69.23%). In the opinion of the research team, the survey encompassed a significant part of experts dealing with bancassurance in Poland and working as professionals in financial institutions. Each answer was given the following score: I strongly agree, 3; I rather agree, 2, I neither agree nor disagree, 1; I rather disagree, −1; I strongly disagree, −2; I have no opinion, 0. Next, the scores were summed for each question and the results obtained are presented in the table below.

Answers obtained as part of the survey should be considered reliable (Cronbach's $\alpha = 0.709 > 0.05$). The first quartile was determined based on the sum of the answers to each question, and only those variables for which the sum of the scores was greater than the first quartile of all analysed variables were included in the model (marked in grey in Table 4). The first quartile method allowed for the selection of the variables according to their significance (indicating the factors about which the experts were most consistent). Moreover, this method allows minimising the probability of omitting important variables while optimising the efforts in model preparation.

Table 4. Results of a survey among financial institution employees.

No.	Variable Name Used for the Model	Variable	Detailed Data Description	Total Score from Survey among Professionals
1	Age	Policyholder's age	Average life expectancy for women and men in Poland, data source: own calculation based on Statistic of Poland (GUS)	49
2		Policyholder's gender		7
3	Fam	Size of policyholder's family (number of children, dependants)	Average number of people in a household in Poland, data source: Statistic of Poland (GUS)	49
4	Sal	Policyholder's earnings	Average disposable income, data source: Statistic of Poland (GUS)	69
5	Wealth	Policyholder's assets	Assets accumulated in investment funds, data source: chamber of fund and asset managers (IZFA)	54
6	RealEst	Real properties owned by policyholder	Average usable floor space per 1 person in Poland, data source: Statistic of Poland (GUS)	50
7	P	Insurance price	CPI for insurance, data source: Statistic of Poland (GUS)	64
8		Inflation		19
9	Unemp	Unemployment	Average unemployment rate in a given year, data source: own calculation based on Statistic of Poland (GUS)	40
10		Situation of the stock exchange (e.g., WIG20 quotations)		12
11		Exchange rate (EUR/PLN, CHF/PLN, etc.)		12
12		Interest rates (WIBOR, etc.)		26
13	Cash	Demand on cash loans	Consumer bank loans for households, data source: National Bank of Poland (NBP)	56
14	Mort	Demand on mortgage loans	Bank housing loans for households, data source: National Bank of Poland (NBP)	59
15	Dep	Level of bank deposits	Deposits and other liabilities of banks to households, data source: National Bank of Poland (NBP)	36
16		Link of insurance and banking product (loan, payment card, etc.)		53
17	Wka	Agent's commission	The ratio of acquisition costs to gross written premium, data source: own calculation based on Polish Financial Supervision Authority (PFSA)	54
18	DUM1	Caps in insurance products construction	Dummy variable, value −1 for every year since the restriction has been in force, and the 0 value for the earlier ones, data source: own calculation	58
19	DUM2	Changes in legal environment		39
		Quartile 1		31

Source: own study.

Data used for the research were taken from the most reliable sources in Poland and Europe. Annual bulletins on the insurance market published by the Polish Financial Supervision Authority in the years 2004–2019 were the basis for establishing the total gross premiums written in Poland (gross written premium, GWP) and the ratio of acquisition costs to gross written premium (Wka). The gross written premium in bancassurance (GWP_total) was established as the product of the premium and the banks' share in insurance distribution. Data on bancassurance were taken from the website of Insurance Europe (InsuranceData 2020). The share of the gross premiums written under Section II generated through bancassurance in the total gross premiums written in bancassurance is not significant. Data on bank real property loans to private individuals (in PLNm) and loans and other banking receivables from households, corresponding to cash loans, were taken from publications of the National Bank of Poland (NBP 2020). Moreover, data from the Polish Central Statistical Office (Statistics of Poland; GUS 2020) and Polish Chamber of Fund and Asset Managers (IZFA 2020) were used for the study (as it is in Table 4). A binary variable (DUM) corresponding to the caps on the so-called costs of exit from a PRIIP (life) product of up to 4% described in Section 3 was also introduced into the analysis. This variable takes the value −1 when the caps are introduced, and a value of 0 before the caps were introduced.

The research method used a linear regression between the dependent variable, which was, respectively, change of gross written premium in bancassurance (ΔGWP_total) in the years 2005–2019 and the explanatory variables shown in the table below (Table 5). The model was built on the increments of variables.

Table 5. Summary of the regression procedure used.

	Model
Dependent variable (Y)	Δ(GWP_total)
Explanatory variables (Xi)	Δ(Age), Δ(Fam), Δ(Sal), Δ (Wealth), Δ(RealEst), Δ(P), Δ(Unemp), Δ(Cash), Δ(Mort), Δ(Dep), Δ(Wka), DUM1, DUM2
Regression equation	
Additional explanations	Ai- estimated regression coefficients, C constant

Source: own study.

5. Results and Discussion

In the procedure for the model, some variables were found to be statistically insignificant; therefore, these variables were removed from the model. The final variables used for the risk model for bancassurance development are: the size of policyholder's family (number of children, dependants) represented by the average number of people in a household in Poland; demand on mortgage loans, represented by Bank housing loans for households and agent's commission, represented by the ratio of acquisition costs to gross written premium. The final form of the model of the gross written premium in bancassurance is shown in Table 6.

Table 6. Results of regression model.

	Model	
Final form of the model	Δ(GWP_total) = A1 × Δ(Fam) + A2 × Δ(Mort) + A3 × Δ(Wka) + C	
Values of model parameters	A1 = −118,491,660.79 A2 = 65.35 A3 = −221,466,185.71 C = −4,269,523.78	
Regression statistics	R2 = 0.846 Adjusted R2 = 0.804	
Regression parameter statistics	A1: Standard error: 30,981,328.09 t-stat: −3.83 *p*-value 0.0028 < 0.05 A2: Standard error: 25.52 t-stat: 2.56 *p*-value = 0.0265 < 0.05	A3: Standard error: 40,025,754.19 t-stat: −5.53 *p*-value = 0.0002 < 0.05 C: Standard error: 1,389,088.16256004 t-stat: −3.07361612839268 *p*-value = 0.0106 < 0.05
Variance analysis	F: 20.12 Significance F: < 0.05	
Stationarity of variables	KPSS Test for the Δ(GWP_total), Δ(Fam), Δ(Mort), Δ(Wka) variables shows *p*-value > 0.05, no basis for rejecting the series stationarity hypothesis	
Residuals randomness	α = 0.05, (+): 8, (-) 7, (#):8, critical values of series test: (D): 3; (G): 13, no basis for rejecting the residuals randomness hypothesis (although the value at the test limit)	

Source: own study.

The results obtained indicate the complexity of the modelling issue regarding risk factors that affected bancassurance market development. Statistically significant variables come from two different groups of factors determining its growth—general for all insurance and specific to bancassurance. They also take into account local conditions (the dominant position of Section I—life insurance). The size of the family of the insured person is a factor naturally determining the demand for life insurance, irrespective of the distribution channel. This observation confirms the research cited by Zietz (2003). Family size or the number of children is often indicated as an important personal and demographic determinant of insurance consumption. The positive relationship has been confirmed in the 'classical' studies of such authors as in (Zietz 2003; Berekson 1972; Ferber and Lee 1980; Burnett and Palmer 1984; Auerbach and Kotlikoff 1989; Lewis 1989; Bernheim 1991; Browne and Kim 1993; Showers and Shotick 1994). Of course, this relation has also been confirmed in a few more recent studies, such as Reddy et al. (2020) or Song et al. (2019). On the other hand, the sale of mortgages is already a bancassurance-specific factor due to the link between the insurance offered and the banking product. The last factor, named acquisition costs, may be classified in both groups, but it also has its own specificities regarding insurance distribution by banks. In further discussion, we will focus on bancassurance channel-specific variables.

The statistical significance of the variable associated with the distribution of loans by banks (Mort) is not surprising. It fully reflects the specific nature of that bancassurance market. According to data from the Polish Insurance Chamber, the share of gross premium written on the sale of stand-alone (not linked with banking products) life insurance in the bancassurance channel (other than investment insurance) was only 5% in 2019 (PIU 2019). This indicates the scale and significance of bundling as a major risk factor in the development of insurance distribution by banks. The statistical significance of this variable also underlines the need to integrate processes and adapt products to a banking product, which banks and insurers are already doing quite effectively. However, it is also worth pointing out additional risks factors that are gaining importance. This factor, which indirectly affects the development of the bancassurance market, is the credit risk of mortgage potential clients. This risk can be represented by two components: endogenous (internal) are located inside households, and exogenous (external) are independent of households. The following factors are also important for lending volumes and credit risk: risk appetite, collateral or diversification of the banking portfolio within target groups. All these factors indirectly become risk factors for the bancassurance sector and could be modelled as separate variables as part of further research.

It should also be noted here that, in order to give a complete picture of the situation, the variable directly linked to the cash loan was not statistically significant. Both individually (Cash variable itself) and as the sum of the variables Cash and Mort. However, the p-value of such variable was 0.051, so only 0.001 exceeded the limit level. An explanation for this can be a large correlation of the variable Cash and Mort (r- Pearson correlation coefficient equal to 0.96).

When analysing the results obtained, it is also worth pointing out the statistical significance of the acquisition cost index (Wka). Acquisition costs are the direct and indirect variable costs incurred by an insurer at the time of selling an insurance contract (both new and renewal). The costs may be in the form of agent or brokerage fee, underwriting costs or medical expenses. The largest share of acquisition costs in the bancassurance channel is the bank's commission. This result is also not surprising, and it is linked with limiting commissions on the Polish market described earlier and the results of the study by (Ricci and Fiordelisi 2010) cited in Chapter 3. The authors claimed that banks are able to benefit most from bancassurance cooperation. Nevertheless, this factor, in our view, may gain and lose importance. In the current macroeconomic situation, with historically low interest rates, its role is significant. Banks are looking for an additional source of revenue as the interest margin itself becomes insufficient (significant decrease in ROE in the banking sector). In a higher interest rate environment, this factor may play a smaller role. However, it is undoubtedly a significant risk factor.

It should be pointed out, with regard to the specification of the local polish market, that although Polish authors indicated the development of the banking sector as one of the elements of the development of the bancassurance market (Gostomski 2011; Dobrucka 2004; Swacha-Lech 2003), none of them explicitly indicated the factors identified in this research as significant (bancassurance-specific factors): sale of mortgages and acquisition cost. Results obtained in this paper also indirectly confirm the observations contained in (Choudhury and Singh 2021). Significant variables come from two different groups of factors determining its growth—general for all insurance and specific to the bancassurance market. These results are in line with the observation that bancassurance must be viewed in a broader sense because of the fact that banks face competition not only within the banking industry but also with the insurance industry (Choudhury and Singh 2021).

6. Conclusions

Risk factors influencing the development of bancassurance in Poland were verified and determined as part of this research. The group of factors selected in the survey among financial sector experts was statistically verified. Significant risk factors (statistically significant) which determine gross premiums written in the bancassurance channel are: the size of the policyholder's family (number of children, dependants), represented by the average number of people in a household in Poland; the demand on mortgage loans, represented by Bank housing loans for households and agent's commission, represented by the ratio of acquisition costs to gross written premium. These above-identified risk factors, through confirmation in the empirical model, constitute a significant added value to this study. Confirmation of development factors or risk factors in the discussion about bancassurance by empirical research is rather rare in the literature. Additionally, it is these empirical studies that decide this paper's contribution to bancassurance literature.

The results of the econometric model obtained are consistent with expectations arising from the principles and practice of cooperation between banks and insurers as well as the specificity of insurance product distribution (also local) in the bancassurance channel. These results confirm the bancassurance development factors indicated by experts in the studies referred to in part 3 of this publication and are in line with expectations. The mortgage sales volume mainly determines the level of gross premiums written as part of bancassurance. It means that the insurance sector, while distributing its products through banks, incorporates risks related to core banking activities and strongly depends on the credit policies of financial institutions.

Despite the statistically good results, in line with the expectations confirmed by the econometric model, the limitations of this research should not be forgotten. The main limitation is the study period (2004–2019). However, it should also be emphasised that this is the longest available time series in terms of gross premiums written in the bancassurance channel. The number of surveys completed by bancassurance experts may be considered as another limitation (27 fully completed questionnaires were obtained out of 39 requests). Although, in the opinion of the research team, the survey encompassed a significant part of experts dealing with bancassurance in Poland and working as professionals in financial institutions; subsequent research may include some improvement of selecting risk factors based on empirical studies.

Despite the limitations mentioned above subject, the findings provided by this paper are equally very important from the business sector perspective and from the point of view of insurance theory.

Author Contributions: Conceptualization, A.Ś., N.D. and J.D.; methodology, N.D.; validation, N.D.; formal analysis, N.D. and J.D.; resources, N.D. and J.D.; writing—original draft preparation, N.D. and J.D.; writing—review and editing, A.Ś.; supervision, A.Ś. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: https://www.knf.gov.pl/publikacje_i_opracowania, https://www.nbp.pl/home.aspx?f=/statystyka/pieniezna_i_bankowa/nal_zobow.html, https://stat.gov.pl/, https://www.insuranceeurope.eu/insurancedata (accessed on 23 December 2020).

Acknowledgments: The authors would like to thank the anonymous reviewers designated by the Journal for their valuable feedback and suggestions, which were helpful in further improving this paper.

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

Appendix A. Survey on Bancassurance

Information about the study: this survey is conducted as part of testing the determinants of the gross premiums written within bancassurance in Poland. The survey is anonymous, and participation in the study is fully voluntary. An invitation to taking part in the study was sent to professionals participating in the process of insurance products preparation and distribution, both in banks and insurance companies. The study will last from 22 February 2021 to 5 March 2021. The research pertains to all insurance sections and groups, i.e., both investment insurance (insurances with insurance capital funds) and insurance linked and not linked with a banking product, both property and life insurances. Researches: Norbert Duczkowski, Joanna Dropia We encourage you to participate in the survey. It will take not more than 5 min to fill it in.

1. Your current role relating to insurance distribution:
    - Natural person performing agency activities (in Polish: *OFWCA*)
    - Expert in Agent's Headquarters
    - Management staff in Agent's Headquarters
    - Expert in the Insurance Company Headquarters
    - Management staff in the Insurance Company Headquarters
2. How strongly do you agree or disagree with the following statements:

Key demographic, economic, market factors affecting the premiums written in bancassurance are:

	I Strongly Disagree	I Rather Disagree	I Neither Agree Nor Disagree	I Rather Agree	I Strongly Agree	I Have No Opinion
Policyholder's age						
Policyholder's gender						
Policyholder's family size (number of children, dependants)						
Policyholder's earnings						

	I Strongly Disagree	I Rather Disagree	I Neither Agree Nor Disagree	I Rather Agree	I Strongly Agree	I Have No Opinion
Policyholder's assets						
Real properties owned by policyholder						
Insurance price						
Inflation						
Unemployment						
Situation on the stock exchange (e.g., WIG20 quotations)						
Exchange rate (EUR/PLN, CHF/PLN, etc.)						
Interest rate (WIBOR, etc.)						

3. Other demographic, economic, market factors affecting premiums written in bancassurance (which ones—please list them).
4. How much do you agree or not agree with the following statements:

   Key factors specific for the bancassurance sector affecting premiums written are:

	I Strongly Disagree	I Rather Disagree	I Neither Agree Nor Disagree	I Rather Agree	I Strongly Agree	I Have No Opinion
Demand on cash loans						
Demand on mortgage loans						
Level of bank deposits						
Link of insurance with banking product (loan, payment card, etc.)						
Agent's commission						
Caps in insurance products construction (e.g., max cost of exit in PRIIPs 4%)						
Changes in legal environment (e.g., ECJ judgment on returning commissions in case of early loan repayment)						

5. Other factors specific for the bancassurance sector affecting premiums written in bancassurance (which ones—please list them):

Appendix B. Role of Banks in Insurance Distribution in Poland

Role of banks in insurance distribution in Poland

Type of Insurance Distribution Entity	Application of Banks' Role in Insurance Distribution in Poland
Insurance company	No
Insurance agent	Yes
Agent offering supplementary insurance	Yes
Insurance broker	No
Policyholder in insurance contract for other party's account	Yes

Source: own study.

Appendix C. Section I (Life Insurance) Polish Market and Bancassurance Specification

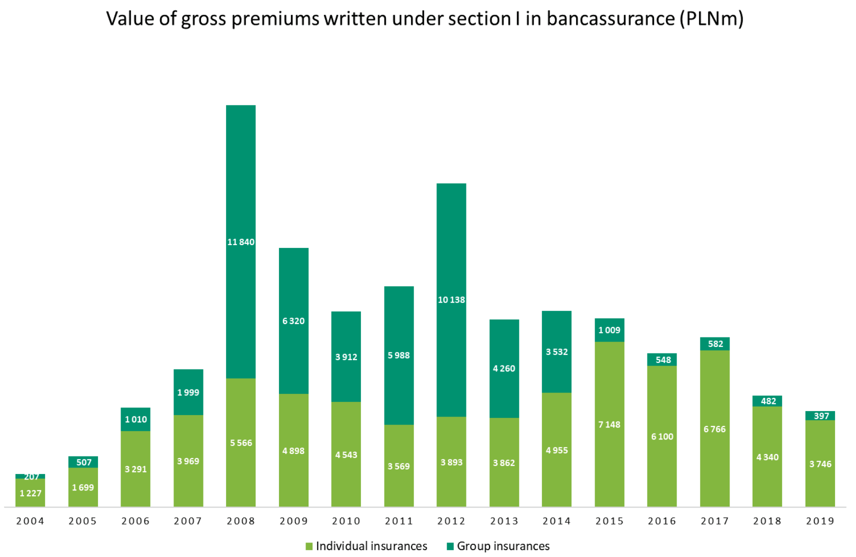


Figure A1. Value of gross premiums written in the bancassurance channel, source: own study based on KNF's data: (Polish Financial Supervision Authority 2001–2019).

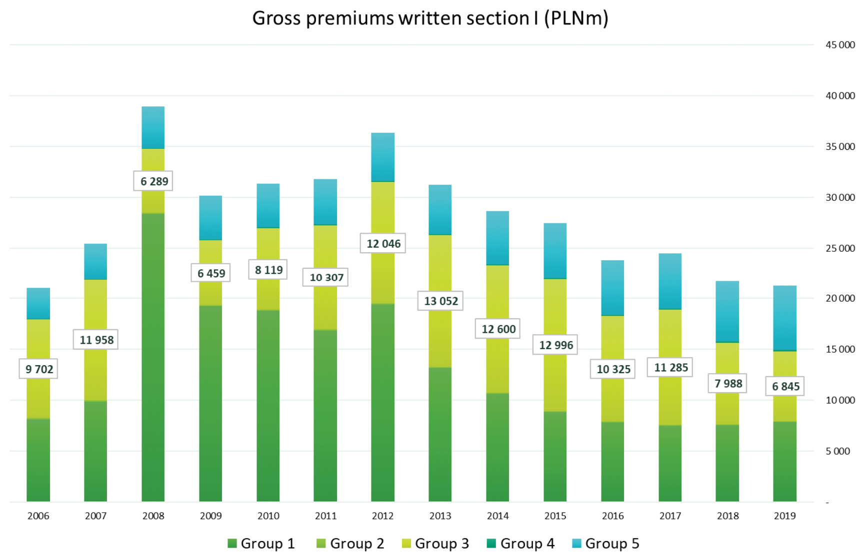


Figure A2. Gross premiums written in Section I, source: own study based on KNF's data: (Polish Financial Supervision Authority 2001–2019).

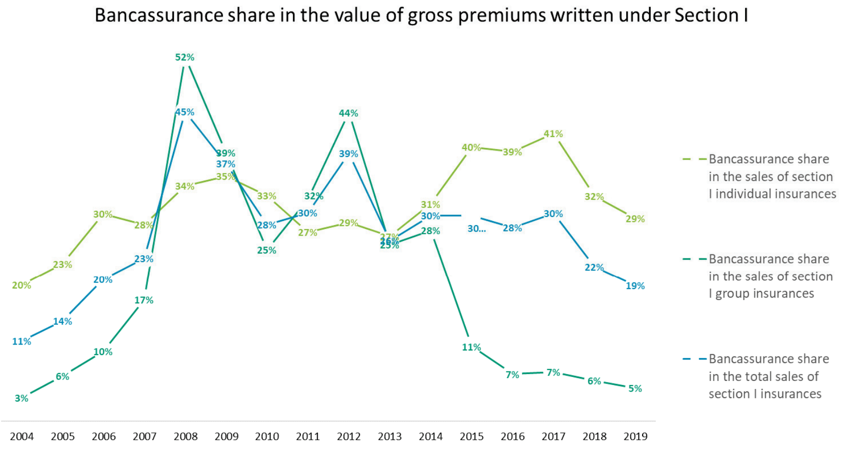


Figure A3. Bancassurance share in Section I gross premiums written, source: own study based on KNF's data: (Polish Financial Supervision Authority 2001–2019).

References

Act on Insurance and Reinsurance Activities. 2015. Act on Insurance and Reinsurance Activities of 11 September 2015. Polish Journal of Laws 2015, Item 1844. Available online: https://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20150001844/U/D20151844Lj.pdf (accessed on 14 February 2020).

Act on Mortgage Loan and Supervision over Mortgage Loan Intermediaries and Agents. 2017. Act on Mortgage Loan and Supervision over Mortgage Loan Intermediaries and Agents of 23 March 2017, Polish Journal of Laws 2017, Item 819. Available online: http://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20170000819/U/D20170819Lj.pdf (accessed on 14 February 2020).

Auerbach, Alan, and Laurence Kotlikoff. 1989. *How Rational Is the Purchase of Life Insurance?* Working Paper No. 3063. Cambridge: National Bureau of Economic Research.

Berekson, Leonard. 1972. Birth Order, Anxiety, Affiliation and the Purchase of Life Insurance. *Journal of Risk and Insurance* 39: 93–108. [CrossRef]

Bernheim, Douglas. 1991. How Strong Are Bequest Motives? Evidence Based on Estimates of the Demand for Life Insurance and Annuities. *Journal of Political Economy* 99: 899–927. [CrossRef]

Białas, Małgorzata. 2015. Polskie realia współpracy banków z ubezpieczycielami. *Studia Ekonomiczne/Uniwersytet Ekonomiczny w Katowicach* 225: 28–37.

Breś-Błażejczyk, Jadwiga. 2005. Bancassurance—Sukces czy porażka? *Studia i Prace Kolegium Zarządzania i Finansów, Szkoła Główna Handlowa* 62: 31–40.

Browne, Mark J., and Kihong Kim. 1993. An International Analysis of Life Insurance Demand. *Journal of Risk and Insurance* 60: 616–34. [CrossRef]

Burnett, John J., and Bruce A. Palmer. 1984. Examining Life Insurance Ownership Through Demographic and Psychographic Characteristics. *Journal of Risk and Insurance* 51: 453–67. [CrossRef]

Chen, Tsai-Jyh. 2019. Marketing channel, corporate reputation, and profitability of life insurers: Evidence of bancassurance in Taiwan. *Geneva Papers on Risk and Insurance-Issues and Practice* 44: 679–701. [CrossRef]

Choudhury, Mousumi, and Ranjit Singh. 2021. Identifying factors influencing customer experience in bancassurance: A literature review. *Journal of Commerce & Accounting Research* 10: 10–22.

Dharmaraj, Senthil. 2019. Customer perception towards bancassurance—A study of select banks in Tamilnadu. *Smart Journal of Business Management Studies* 15: 47–57. [CrossRef]

Directive of the European Parliament and of the Council (EU). 2016. Directive of the European Parliament and of the Council (EU) 2016/97 of 20 January 2016 on Insurance Distribution. Available online: https://eur-lex.europa.eu/legal-content/PL/TXT/PDF/?uri=CELEX:32016L0097&from=en (accessed on 14 February 2021).

Dobrucka, Marta. 2004. Bancassurance jako czynnik rozwoju sektora finansowego. *Zeszyty Naukowe/Akademia Ekonomiczna w Krakowie* 637: 127–39.

Fan, Chiang Ku, Li-Tze Lee, Yu-Chieh Tang, and Yu Hsuang Lee. 2011. Factors of cross-buying intention—Bancassurance evidence. *African Journal of Business Management* 5: 7511–15.

Ferber, Robert, and Lucy Chao Lee. 1980. Acquisition and Accumulation of Life Insurance in Early Married Life. *Journal of Risk and Insurance* 47: 713–34. [CrossRef]

Gajdek, Magdalena. 2016. Bancassurance, nowe zjawisko we współczesnej bankowości. *Journal of Modern Management Process* 1: 18–26.

Ganapathy, Venkatesh. 2021. Digital Bancassurance Business Models. *IBM Pune Research Journal* 21: 48–62.

Gigerenzer, Greg. 2014. *Risk Savvy. How to Make Good Decisions*. London: Penguin Books, pp. pp. 97–99, 255–56.

Gostomski, Eugeniusz. 2011. Perspektywy rozwoju bancassurance w Europie. *Prace Naukowe Wyższej Szkoły Bankowej w Gdańsku* 9: 119–30.

GUS. 2020. Statistics of Poland. Available online: https://stat.gov.pl/ (accessed on 23 December 2020).
Gwizdała, Jerzy. 2018. Perspektywy funkcjonowania koncepcji bancassurance w Polsce. *Annales Universitatis Mariae Curie-Skłodowska, Sectio H Oeconomia* 52: 61–69. [CrossRef]
Hota, Sweta Leena. 2016. Bancassurance: Convergence of banking and insurance—A saga. *International Journal of Research and Development—A Management Review* 5: 11–15.
Insurance Distribution Act. 2017. Insurance Distribution Act of 15 December 2017, Polish Journal of Laws 2017, Item 2486. Available online: http://isap.sejm.gov.pl/isap.nsf/download.xsp/WDU20170002486/U/D20172486Lj.pdf (accessed on 14 February 2020).
InsuranceData. 2020. Available online: https://www.insuranceeurope.eu/insurancedata (accessed on 23 December 2020).
IZFA. 2020. Chamber of Fund and Asset Managers. Available online: https://www.izfa.pl/ (accessed on 23 December 2020).
Jaspersen, Johannes. 2016. Hypothetical Surveys and Experimental Studies of Insurance Demand: A Review. *The Journal of Risk and Insurance* 83: 217–55. [CrossRef]
Karimian, Paniz Haji. 2017. The effect of bancassurance on bank productivity and profitability, ARDL approach (evidences from banking industry in Iran). *American Journal of Economics* 7: 177–85.
Kowalski, Jacek, and Elżbieta Kowalska. 2015. Ubezpieczenia grupowe a Rekomendacja U. *Zeszyty Naukowe Wyższej Szkoły Bankowej w Poznaniu* 59: 99–115.
Košík, Oto, and Peter Poliak. 2011. Sprzedaż produktów bankowo-ubezpieczeniowych w kontekście kryzysu finansowego i postępującej globalizacji. *Zeszyty Naukowe Wyższej Szkoły Bankowej w Poznaniu* 35: 159–72.
Kramaric, Tomislava Pavicz. 2019. Does bancassurance affect performance of non-life insurance sector-case of EU Countries. *International Journal of Economic Sciences* 8: 96–108. [CrossRef]
Łada, Monika, and Małgorzata Białas. 2017. Financial Settlements as A Core of Inter-Organizational Management Accounting—Case Study of Bancassurance Cooperation. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 474: 77–84. [CrossRef]
Lewis, Frank. 1989. Dependents and the Demand for Life Insurance. *American Economic Review* 79: 452–67.
Liang, Hsin-Yu. 2015. Suggestions for Bancassurance Markets in China: Implications from European Countries. *Geneva Papers on Risk and Insurance-Issues and Practice* 40: 279–94. [CrossRef]
Lisowski, Jacek, and Anna Chojan. 2020. InsurTech in CEE Region—Where Are We? In *Global, Regional and Local Perspectives on the Economies of Southeastern Europe*. Edited by Alexandra Horobet, Lucian Belascu, Persefoni Polychronidou and Anastasios Karasavvoglou. Springer Proceedings in Business and Economics. Berlin: Springer.
Łosiewicz-Dniestrzańska, Ewa. 2016. Rola jednostki zapewnienia zgodności z regulacjami w procesie wprowadzania nowego produktu bankowego na rynek. *Studia Ekonomiczne/Uniwersytet Ekonomiczny w Katowicach. Współczesne Finanse* 275: 107–18.
Malinowski, Artur. 2011. Zastosowanie bancassurance w Polsce. Zeszyty Naukowe Uniwersytetu Przyrodniczo-Humanistycznego w Siedlcach. *Administracja i Zarządzanie* 89: 137–46.
Monkiewicz, Jan. 2000. *Podstawy ubezpieczeń*. Warszawa: Wydawnictwo POLTEXT.
NBP (National Bank of Poland). 2020. Available online: https://www.nbp.pl/ (accessed on 23 December 2020).
Recommendation U on Best Practice in Bancassurnace. 2014. Recommendation U on Best Practice in Bancassurnace. Available online: https://www.knf.gov.pl/knf/pl/komponenty/img/Rekomendacja_U_38338.pdf (accessed on 14 February 2020).
Ociepa-Kicińska, Elżbieta. 2019. Produkty bancassurance w bankowości detalicznej w Polsce. *Rozprawy Ubezpieczeniowe. Konsument na rynku usług finansowych* 33: 53–67.
Pajewska, Renata. 2000. *Sojusze bankowo-ubezpieczeniowe. w: J., Monkiewicz, (red), Podstawy ubezpieczeń, tom I—Mechanizmy i funkcje*. Poltext: Warszawa, pp. 306–8.
Pajewska-Kwaśny, Renata, and Ilona Tomaszewska. 2009. Tendencje rozwoju bancassurance na rynku polskim. *Myśl Ekonomiczna i Prawna* 4: 174–83.
Pielichaty, Edward. 2014. Zasady rachunkowości Banku w Świetle Rekomendacji U. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 373: 208–16. [CrossRef]
Pielichaty, Edward. 2015. Rozpoznawanie przychodów ze sprzedaży produktów ubezpieczeniowych w księgach rachunkowych banków. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 390: 191–202. [CrossRef]
Pielichaty, Edward. 2017. Accounting Rules for Revenues Arising from Insurance Products Offered in Banks. *Central European Review of Economics and Management* 1: 85–105.
Pisarewicz, Piotr. 2013. Struktura sprzedaży ubezpieczeń na krajowym rynku bancassurance. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Finanse, Rynki Finansowe, Ubezpieczenia* 62: 419–31.
Pisarewicz, Piotr. 2014. Nowe standardy rynku bancassurance w zakresie ubezpieczeń z elementem inwestycyjnym lub oszczędnościowym. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 342: 206–15. [CrossRef]
PIU. 2019. PIU Report—Polish Market Bancassurance 4th Quarter 2019. Available online: https://piu.org.pl/wp-content/uploads/2020/04/Prezentacja-Rynek-BA-2019Q4.pdf (accessed on 9 May 2020).
Polish Financial Supervision Authority. *Annual Bulletin. Insurance Market*. Warsaw: Polish Financial Supervision Authority, 2001–2019.
Preckova, Lenka. 2017. Evaluation of Bancassurance Functioning in Selected Countries of the Financial Group KBC Group. In *European Financial Systems 2017: Proceedings of the 14th International Scientific Conference*. Edited by Josef Nesleha, Tomáš Plihal and Karel Urbanovsky. Brno: Masaryk University, pp. 192–98.
Puja, Dua, Namita Sahay, and Onkar Singh Deol. 2019. Bancassurance model and its impact on Financial Inclusion: Review and Analysis. *International Journal of Research and Analytical Reviews* 6: 2.

Reddy, Venkatanarsi Narsi, Seelam Marulu Reddy, and Peela Appala Naidu. 2020. Determinants of Life Insurance at Household Level: An Empirical Analysis of Andhra Pradesh, India. *Waffen-Und Kostumkunde Journal* XI: 49–61.

Ricci, Ornella, and Franco Fiordelisi. 2010. Efficiency in the Life Insurance Industry: What Are the Efficiency Gains from Bancassurance? EMFI Working Paper No. 2. *SSRN Electronic Journal*. [CrossRef]

Ricci, Ornella, and Franco Fiordelisi. 2011. Bancassurance Efficiency Gains: Evidence from the Italian Banking and Insurance Industries. *The European Journal of Finance* 17: 789–810.

Ricci, Ornella, and Franco Fiordelisi. 2012. *Bancassurance in Europe. Past, Present and Future*. London: Palgrave Macmillan.

Sadurska, Małgorzata. 2019. Bancassurance jako sposób generowania efektów synergicznych na rynku bankowo-ubezpieczeniowym. *Zeszyty Naukowe Wyższej Szkoły Bankowej w Poznaniu* 84: 75–84.

Saha, Shyamasree, and Anirban Dutta. 2019. Factors Influencing Service Quality Perception in Indian Life Insurance Sector. *Global Business Review* 20: 1010–25. [CrossRef]

Sereda, Paweł. 2015. Wpływ Rekomendacji U na funkcjonowanie rynku bancassurance. *Zeszyty Naukowe Uniwersytetu Szczecińskiego*. *Finanse, Rynki Finansowe, Ubezpieczenia* 2: 647–55.

Showers, Vince, and Joyce Shotick. 1994. The Effects of Household Characteristics on Demand for Insurance: A Tobit Analysis. *Journal of Risk and Insurance* 61: 492–502. [CrossRef]

Śliperski, Marek. 1998. Związki banków z firmami ubezpieczeniowymi i perspektywy ich rozwoju w Polsce. *Ruch Prawniczy, Ekonomiczny i Socjologiczny* 60: 213–34.

Śliwiński, Adam. 2016. Popyt na ubezpieczenia na życie- przegląd badań światowych. In *Polski Rynek Ubezpieczeń na tle kryzysów społeczno-gospodarczych*. Edited by Stanisław Nowak, Alojzy Nowak and Andrzej Sopoćko. Warsaw: Wydawnictwo naukowe Wydziału Zarządzania Uniwersytetu Warszawskiego.

Śliwiński, Adam. 2019. *Rola ubezpieczeń w gospodarce*. Wydawnictwo: Oficyna Wydawnicza SGH.

Śliwiński, Adam, Tomasz Michalski, and Małgorzata Rószkiewicz. 2013. Demand for Life Insurance—An Empirical Analysis in the Case of Poland. *The Geneva Papers on Risk and Insurance-Issues and Practice* 38: 62–87. [CrossRef]

Song, In Jung, Heejung Park, Narang Park, and Wookjae Heo. 2019. The effect of experiencing a death on life insurance ownership. *Journal of Behavioral and Experimental Finance* 22: 170–76. [CrossRef]

Sreesha, C. H., and M. A. Joseph. 2011. Financial performance of banks in bancassurance: A study with special reference to State Bank of India. *Indian Journal of Finance* 5: 10–17.

Staszczyk, Mateusz. 2013. Ubezpieczenie kart płatniczych jako sposób ochrony klienta banku przed finansowymi skutkami nagłych zdarzeń. *Acta Universitatis Lodziensis*. *Folia Oeconomica* 284: 195–207.

Swacha-Lech, Magdalena. 2003. Społeczno-ekonomiczne determinanty atrakcyjności działalności bancassurance na rynku finansowym w Polsce (z punktu widzenia banków komercyjnych). *Prace Naukowe Akademii Ekonomicznej we Wrocławiu* 982: 144–55.

Swacha-Lech, Magdalena. 2006. Bancassurance w mBanku—przyczyny podjęcia współpracy z firmami ubezpieczeń, sposoby realizacji strategii oraz jej pierwsze efekty. *Prace Naukowe, Akademia Ekonomiczna w Katowicach*. *Współczesne problemy finansów, bankowości i ubezpieczeń w teorii i praktyce* 1: 243–52.

Szewieczek, Daniel. 2011. Finansowe konsekwencje oferty produktowej w warunkach dynamicznych zmian na rynku bancassurance. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu* 175: 138–49.

Wierzbicka, Ewa. 2009. Kierunki ewolucji bancassurance. *Zeszyty Naukowe, Uniwersytet Ekonomiczny w Poznaniu* 127: 737–43.

Wierzbicka, Ewa. 2016. Misselling barierą rozwoju ubezpieczeń w Polsce. Zeszyty Naukowe Wyższej Szkoły Humanitas. *Zarządzanie* 2: 315–27.

Zharikova, Olena, and Katerina Cherkesenko. 2021. Integration of banks and insurance companies activities in Ukraine. *Journal of Scientific Papers 'Social Development and Security'* 11: 42–57. [CrossRef]

Zietz, Emily Norman. 2003. An examinantion of demand for life insurance. *Risk Management and Insurance Review* 6: 159–91. [CrossRef]

risks

MDPI

Article

Which Curve Fits Best: Fitting ROC Curve Models to Empirical Credit-Scoring Data

Błażej Kochański

Faculty of Management and Economics, Gdańsk University of Technology, 80-233 Gdańsk, Poland; bkochanski@zie.pg.gda.pl

Abstract: In the practice of credit-risk management, the models for receiver operating characteristic (ROC) curves are helpful in describing the shape of an ROC curve, estimating the discriminatory power of a scorecard, and generating ROC curves without underlying data. The primary purpose of this study is to review the ROC curve models proposed in the literature, primarily in biostatistics, and to fit them to actual credit-scoring ROC data in order to determine which models could be used in credit-risk-management practice. We list several theoretical models for an ROC curve and describe them in the credit-scoring context. The model list includes the binormal, bigamma, bibeta, bilogistic, power, and bifractal curves. The models are then tested against empirical credit-scoring ROC data from publicly available presentations and papers, as well as from European retail lending institutions. Except for the power curve, all the presented models fit the data quite well. However, based on the results and other favourable properties, it is suggested that the binormal curve is the preferred choice for modelling credit-scoring ROC curves.

Keywords: credit scoring; ROC curve; Gini coefficient

Citation: Kochański, Błażej. 2022. Which Curve Fits Best: Fitting ROC Curve Models to Empirical Credit-Scoring Data. *Risks* 10: 184. https://doi.org/10.3390/risks10100184

Academic Editor: Montserrat Guillén

Received: 30 June 2022
Accepted: 14 September 2022
Published: 20 September 2022

1. Introduction

In this paper, we discuss the modelling of receiver operating characteristic (ROC) curves in the domain of credit scoring. Over the last three decades, the retail lending world has transformed into a sophisticated, data-driven industry with automated processes and statistical tools. In a challenging environment where the danger of rising inflation, impending recession, repercussions of the COVID-19 pandemic and the Russo-Ukrainian war (to name only the most apparent risk factors) threaten the stability of the financial sector, a solid foundation of accurate data and proper tools seems even more needed.

Credit scoring is one of the most successful implementations of machine learning in finance. It is used to produce a numerical assessment of a customer's probability of not repaying the loan. Various tools are used to build credit-scoring models (scorecards), including logistic regression, decision trees and neural networks (Anderson 2007). The discriminatory power of a scoring model is usually assessed by the measures summarising its ROC curve.

ROC curve models are discussed in the literature primarily in the context of biostatistical applications. However, they may also prove to be a valuable tool in credit-risk management. As discussed in the literature section, credit-risk practitioners and researchers might need such models, for example, to estimate the AUROC (the area under the ROC curve) on a sample basis or to better understand the shape of the curve and its implications for credit decisions. In credit management practice, the need to model ROC curves also arises when one wants to assess the impact of the credit scorecard that is yet to be built. For example, a lender knows that an analyst team can produce a new scorecard with a Gini coefficient 15% higher than the current one. This new scorecard may be based on newly available data or may be developed using a new classification methodology. The question arises: to what extent introducing the scorecard can reduce credit losses or increase the profitability of the loan portfolio. Using the ROC curve models presented in the following sections, one can easily draw the expected ROC curve and perform the appropriate calculations.

The main goal and contribution of our work is to review the ROC curve models proposed in the literature, primarily in medical sciences and biostatistics, to reformulate them, if necessary, so that they better reflect credit-risk-management needs, and to fit them to real-life credit-scoring ROC data in order to determine which of the curve models fits the empirical data best and could be used in the practice of credit-risk management. The novelty of this paper is the discussion of theoretical ROC curves from the perspective of credit scoring. Not many authors mention theoretical ROC curve models within the domain of credit-risk management. Satchell and Xia (2008), Kürüm et al. (2012) and Kochański (2021) are notable exceptions. To the best of our knowledge, this article constitutes the first attempt to systematically test a set of ROC curve models against empirical data related to credit scoring. We not only perform the quantitative analysis, but also provide arguments that are not purely quantitative for choosing ROC curve models that best serve credit-risk management.

The next section provides a brief literature background. Then, we present several theoretical models for an ROC curve and describe them in the credit-scoring context. The model list includes the binormal, bigamma, bibeta, bilogistic, power, and bifractal curves. The models are then tested against empirical credit-scoring ROC data found in publicly available publications or obtained from European credit institutions. We show that except for the power curve, all the presented models fit the data quite well. However, based on the results and the discussion of the additional properties of the models, we suggest that the binormal curve should be treated as the preferred choice for modelling credit-scoring ROC curves.

2. Literature Background

The literature on ROC curves is extensive, but most of it comes from areas not related to credit risk. This is not surprising as ROC curves derive from signal-detection theory (Wichchukit and O'Mahony 2010; Swets 2014). They are currently popular in many domains as a method to graphically present the separation power of binary classifiers (Fawcett 2006). Examples of such binary classifiers include diagnostic tests and biomarkers in biostatistics (Hajian-Tilaki 2013; Mandrekar 2010; Faraggi et al. 2003; Cook 2007), detectors in signal processing (Bowyer et al. 2001; Atapattu et al. 2010), credit scores in banking (Blöchlinger and Leippold 2006; Thomas 2009; Anderson 2007), and, generally, binary classification models in machine-learning applications (Hamel 2009; Guido et al. 2020).

An ROC curve plots the true-positive fraction against the false-positive fraction as the cut-off point varies (Metz 1978; Krzanowski and Hand 2009). Depending on the domain, the true-positive and the false-positive fractions may be referred to as a detect rate and false-alarm rate in signal processing (Levy 2008; Chang 2010), as sensitivity and 1-specificity in biostatistics (Park et al. 2004), or as a cumulative bad and good proportion in credit scoring (Thomas 2009).

An ROC curve can be summarised with its "area under the curve" (AUC or AUROC) index (Bamber 1975; Bradley 1997; Bewick et al. 2004), also called a *c*-statistic or a *c* index (Cook 2008). The AUROC measures the discrimination power of the binary classifier. Generally, with some reservations related to uncertainty (Pencina et al. 2008), the shape of the curve (Idczak 2019; Řezáč and Koláček 2012) as well as to the cost of misclassification (Cook 2007; Hand 2009), the higher the AUROC, the better. Plotting the ROC curve for a scorecard is not a necessity. Some researchers propose alternative approaches (e.g., Hand 2009), but the ROC curve is considered standard practice by credit-risk professionals (Anderson 2007; Thomas et al. 2017; Siddiqi 2017). Credit-scoring researchers and practitioners frequently use the AUROC to assess, improve, and compare scorecards (Djeundje et al. 2021; Tripathi et al. 2020; Xia et al. 2020; Shen et al. 2021; Lappas and Yannacopoulos 2021).

In the credit-scoring domain, the Gini coefficient, which, in fact, is a rescaled AUROC, is often used by the practitioners (Thomas et al. 2017, p. 191; Řezáč and Řezáč 2011):

$$Gini = 2 \cdot AUROC - 1 \tag{1}$$

It should be noted that the Gini coefficient is equivalent to a special case of Somers' D statistic where one of the associated variables (the target/response) is binary (0/1) and the other one (the classifier) is at least ordinal (Somers 1962; Thomas et al. 2017, p. 189).

If the classifier function and the data set are given, then the empirical ROC curve is uniquely determined. However, an ROC curve model, i.e., a mathematical formula approximating the curve, may be needed in certain circumstances. The literature examples of such situations include (1) the sample-based inference and estimation of AUROC, (2) the ROC curve shape description, and (3) the simulation of the ROC curve when data are absent or scarce.

(1) The use of ROC curve models to estimate the AUROC is widespread, primarily in biostatistics. With the ROC curve model, the path of the curve can be estimated based on sample results, and the confidence intervals for the AUROC may be computed (Lahiri and Yang 2018; Gonçalves et al. 2014; Hanley 1996; Faraggi and Reiser 2002). Satchell and Xia (2008) discussed using the analytic models for ROC curves in the credit-scoring context. They claimed that the theoretical ROC curves are helpful, especially when the sample size of bad customers is small, as the models help increase the accuracy of the AUROC estimation. Indeed, the scarcity of data usually drives the need for such models. Some of the formulas described in this paper were derived in this context and serve as the basis for such an estimation (Hanley 1996; Bandos et al. 2017; Mossman and Peng 2016).

(2) Knowing the area under the curve turns out to be insufficient, especially when one takes no account of its shape. Janssens and Martens (2020) discussed the importance of ROC curve shapes in medical diagnostics. Hautus et al. (2008) utilised the binormal model to demonstrate how the shape of the ROC curve relates to the same–different sensory tests in the food industry. Omar and Ivrissimtzis (2019) fitted the theoretical ROC bibeta curve to the results of machine-learning models in order to show that such an approach provides additional insight into the behaviour at the ends of the ROC curves. Řezáč and Koláček (2012) and Idczak (2019) showed that, depending on the shape, some credit-scoring models excel at distinguishing the best customers from good ones, while others are preferred if a lender intends to exclude the worst loan applicants. A practical example of such an analysis was provided by Tang and Chi (2005); the model with a much lower AUROC still outperformed the competing, high-AUROC model in terms of the accuracy in classifying the best customers.

Another example when information about the shape of an ROC curve is needed is a situation when the data are censored or "truncated" (Scallan 2013), i.e., a lender does not have good/bad information on rejected applicants. Models for ROC curves could enable the modelling of the unknown portion of the curve.

Not taking the shape of the ROC curve into account when summarising it with the AUROC was raised as one of the major deficiencies of an ROC analysis, which led to proposing alternative measures such as Hand's h-measure. (Hand 2009; Hand and Anagnostopoulos 2013).

(3) Occasionally, a credit-risk manager might need to draw ROC curves when data are absent or scarce (for example, when a scorecard is yet to be built). In this context, Kürüm et al. (2012) showed how to use the binormal approximation to optimise the AUROC of the model for a set of corporate loans where the number of bad customers was limited. Outside the domain of credit scoring, Lloyd (2000) provided an extreme example of an inference from limited data. He estimated ROC curves from one data point per curve, assuming that the ranking function (equivalent to credit scoring in finance) was an unobserved latent variable. Of course, such an inference would not be possible without the ROC curve models (binormal model in this case).

In the literature, many models for ROC curves have been suggested. Most of the propositions come from medical statistics. The most frequently used model is the binormal curve (Hanley 1996; Metz 1978) and its modifications (Metz and Pan 1999). Other models include the bibeta (Chen and Hu 2016; Gneiting and Vogel 2022; Mossman and Peng 2016; Omar and Ivrissimtzis 2019), bigamma (Dorfman et al. 1997), bilogistic (Ogilvie and Creelman 1968), power curve (Birdsall 1973) and exponential/bifractal model (England 1988; Kochański 2021). These models will be discussed in more detail in the following section. We found no systematic review where multiple ROC curve models were fitted to the same data, be it medical data or data from any other field. However, a few papers discussed the empirical goodness of fit of particular models. For example, Swets (1986) claimed that in many empirical cases the binormal model turns out to have a good fit, and Gneiting and Vogel (2022) showed that the bibeta model fits better than the binormal one, especially under the assumption of a concave ROC curve.

Note that in the following text, the credit-scoring nomenclature is used. The binary classifier in question is the credit scoring; the observed values of the classifier are referred to as the credit scores, positives (signal, hits, cases) are "bads" or "bad customers'", negatives (noise, false alarms, controls) are "goods" or "good customers", etc. In line with the practice in credit-risk management, the Gini coefficient is preferred over the AUROC as the summary statistic of an ROC curve.

Before we go any further, one thing needs to be clarified to avoid confusion. Despite the similarity of some names, ROC curve models and classification models are completely different animals. The latter are basically classification functions that return predictions or rankings (such as credit scores). When applied to data sets, the ranking functions generate empirical ROC curves. The ROC curve models, on the other hand, are approximations of the shape of ROC curves. If we find that, for example, the bilogistic curve best approximates the ROC curve of the scoring model, then it does not mean that the logistic regression was, or should be, used to develop the model. On the contrary, the ROC curve generated by logistic regression may be best approximated by the bibeta function, and the bilogistic model may prove to have the best fit when a neural network or support vector machine is used.

3. ROC Curve Models

One way to look at an ROC curve is to view it as a function [0; 1]→[0; 1] built using two cumulative distribution functions. In the context of credit scoring, these two CDFs are those of the credit scores for good and bad customers. The general formula for an ROC curve is then (Gönen and Heller 2010):

$$y = F_B\left(F_G^{-1}(x)\right), \tag{2}$$

where F_B is a CDF of the scores of bad customers, F_G is a CDF of the scores of good customers, and F_G^{-1} is its inverse.

$$\begin{aligned} y &= F_B(s) \\ x &= F_G(s), \end{aligned} \tag{3}$$

where s denotes the value of the test, that is, the score or its monotone transformation. Several ROC curve models proposed below take advantage of this simple observation: it may be assumed that the two CDFs follow specific probability distributions.

Equation (2) shows that an ROC curve is invariant to monotone transformations of the underlying scores; the score does not go directly to the ROC equation, only the CDFs do. Therefore, in the following text, the "scores" may refer to the scores themselves or to their monotone transformations.

3.1. Bibeta and Simplified Bibeta Models

If the scores for bad borrowers are distributed according to a beta distribution with parameters α_B and β_B and the scores for good customers follow $Beta(\alpha_G,\beta_G)$, then the formula for the ROC curve is

$$y = F_{\alpha_B,\beta_B}\left(F^{-1}_{\alpha_G,\beta_G}(x)\right), \tag{4}$$

where $F_{\alpha B,\beta B}$ and $F_{\alpha G,\beta G}$ are CDFs of the two beta distributions (Gneiting and Vogel 2022; Omar and Ivrissimtzis 2019). Such a model could be called a "bibeta" ROC curve. The bibeta model that has been used in several articles (Chen and Hu 2016; Mossman and Peng 2016) is a simplified version of the above equation, where $\alpha_B = 1$ and $\beta_G = 1$. Then:

$$F_{\alpha_B,\beta_B} = 1 - (1-x)^{\beta_B} \tag{5}$$

and

$$F^{-1}_{\alpha_G,\beta_G}(x) = x^{\frac{1}{\alpha_G}} \tag{6}$$

so, the formula for the bibeta ROC curve is reduced to the following:

$$y = 1 - \left(1 - x^{\frac{1}{\alpha_G}}\right)^{\beta_B}. \tag{7}$$

In this paper, the curve generated by (7) will be called the "simplified" bibeta.

3.2. Bigamma Model

The "bigamma" model (Dorfman et al. 1997), by analogy, assumes that the scores, or some monotone transformation of them, follow two gamma distributions:

$$y = G_{\alpha_B,\beta_B}\left(G^{-1}_{\alpha_G,\beta_G}(x)\right). \tag{8}$$

Note that Dorfman et al. (1997) proposed that $\alpha_B = \alpha_G$, $\beta_B = 1$ and $0 < \beta_G \leq 1$, but these restrictions are not included in this article, because they resulted in a much worse fit in the preliminary calculations.

3.3. Binormal Model

The same analogy could be used to build the binormal model:

$$y = F_{\mu_B,\sigma_B}\left(F^{-1}_{\mu_G,\sigma_G}(x)\right). \tag{9}$$

Equation (9) has four parameters, but it can be rewritten in an equivalent form with just two parameters:

$$y = \Phi\left(a + b\Phi^{-1}(x)\right). \tag{10}$$

where Φ is a CDF of a standard normal variable, Φ^{-1} is its inverse and a equals the distance between the mean scores of goods and bads measured in terms of units of s.d. of the good scores:

$$a = \frac{\mu_G - \mu_B}{\sigma_G} \tag{11}$$

and b is the ratio between the standard deviations of the bad and good scores:

$$b = \frac{\sigma_B}{\sigma_G} \tag{12}$$

3.4. Bilogistic Model

There is also another perspective of the binormal model. Φ may be viewed as a (probit) link function in the more general equation for an ROC curve:

$$y = g\Big(\alpha_0 + \alpha_1 g^{-1}(x)\Big), \tag{13}$$

where $g(\ )$ is a link function. Taking such a perspective, we can select another link function. If a logit function is taken:

$$g(\cdot) = \exp(\cdot)/(1 + \exp(\cdot)), \tag{14}$$

the bilogistic ROC curve model is derived. After several transformations, the formula for the bilogistic curve looks as follows:

$$y = \left(1 + \exp\left(\alpha_1 ln\left(\frac{1}{x} - 1\right) - \alpha_0\right)\right)^{-1}. \tag{15}$$

The Formula (15) will be used as the bilogistic model in the next section.

All the curves discussed before the bilogistic model are examples of parametric ROC curve models, where one starts with distributions and then arrives at the formula. The bilogistic curve can be viewed as an "algebraic" ROC curve model, where the underlying distribution is somewhat "secondary" to the formula itself.

3.5. Power Function

A simple example of a purely algebraic model is the "power function" (Birdsall 1973, p. 138):

$$y = x^{\theta},\ \theta < 1. \tag{16}$$

The power function can also be derived as the "Lehmann ROC curve" based on proportional hazards specification (Gönen and Heller 2010). From a credit-scoring perspective, the power ROC curve has an attractive property, which could be called "fractal" (Kochański 2021). If the shape of an ROC curve follows Equation (16), and if we take any fraction of the lowest-scored customers and graph the ROC curve for this group, then the shape of the ROC curve remains the same. The AUROC also remains constant: AUROC = $1/(1 + \theta)$, as well as the Gini coefficient: Gini = $(1 - \theta)/(1 + \theta)$. If one would like to make the Gini coefficient an explicit function parameter, one could reformulate Equation (16) in the following way:

$$y = x^{\frac{1-\gamma}{1+\gamma}},\ \ 0 < \gamma < 1 \tag{17}$$

where γ is a parameter for the Gini coefficient.

3.6. Bifractal Model

Kochański (2021) proposed a function that keeps the shape (and the AUROC/Gini) when plotted for any fraction of the highest-scored customers:

$$y = 1 - (1 - x)^{\frac{1+\gamma}{1-\gamma}}, \tag{18}$$

and showed that the empirical ROC curves lie somewhere between those two extremes. This observation prompted the development of the "bifractal" model:

$$y = \beta\left(1 - (1 - x)^{\frac{1+\gamma}{1-\gamma}}\right) + (1 - \beta)x^{\frac{1-\gamma}{1+\gamma}}, \tag{19}$$

as a linear combination of the two "fractal" curves. The analogous formula was also experimentally found to provide a good fit and was referred to as "exponential ROC" or EROC by England (1988).

Figure 1 illustrates the bifractal ROC curves for γ (Gini) = 0.5 and four levels of β.

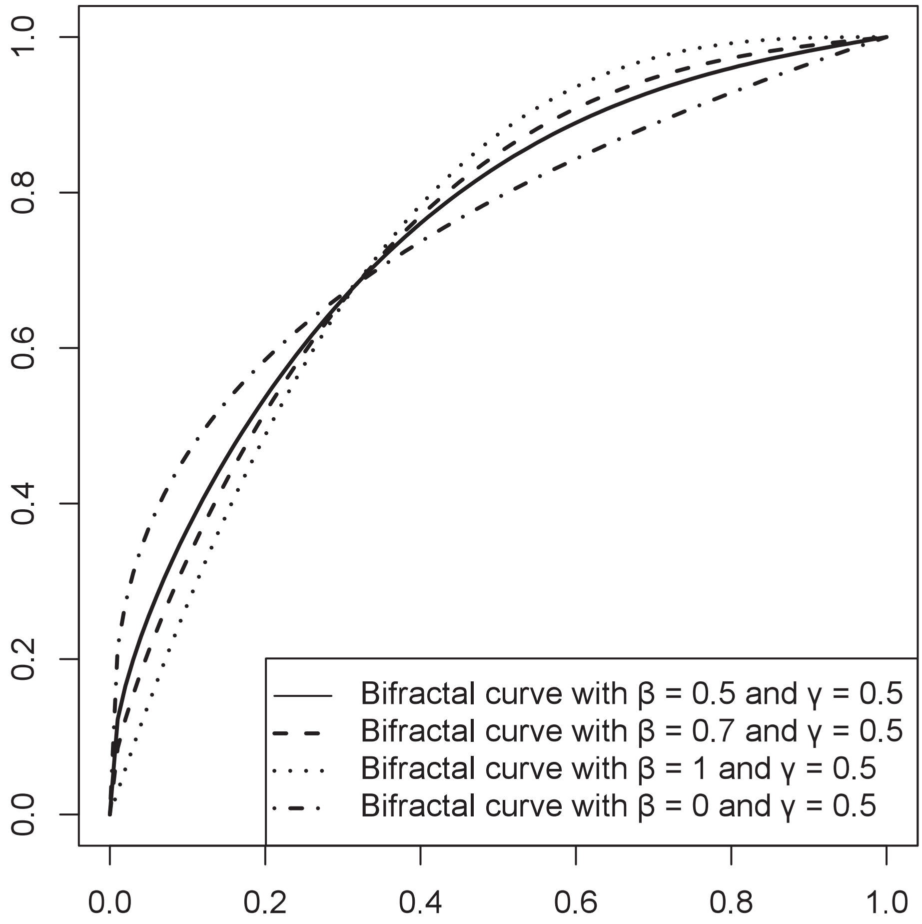


Figure 1. Bifractal curves.

The apparent advantage of the bifractal function in Equation (19) is that it contains the Gini coefficient as its explicit parameter. Moreover, the other parameter of the bifractal ROC curve has a meaningful and intuitive interpretation as a distance between the two fractal curves.

3.7. Reformulated Binormal and Midnormal Models

An explicit Gini coefficient could be a decisive advantage of the bifractal compared to other models. However, as it turns out, the binormal curve function may also be reformulated, and we may obtain a function with the Gini coefficient as a parameter. Thanks to a simple analytic formula for the area under the ROC curve (Bandos et al. 2017):

$$AUROC = \Phi\left(\frac{a}{\sqrt{1+b^2}}\right), \tag{20}$$

Equation (10) transforms so that now the formula has the Gini coefficient as its parameter (γ):

$$y = \Phi\left(\Phi^{-1}\left(\frac{\gamma+1}{2}\right)\sqrt{1+b^2} + b\Phi^{-1}(x)\right). \tag{21}$$

The explicit Gini coefficient in the formula seems essential for the modelling. Consequently, the parametric form of the binormal model described by Equation (21) will be used for empirical curve fitting in the next section.

Figure 2 illustrates the binormal ROC curves for γ (Gini) = 0.5 and various levels of b.

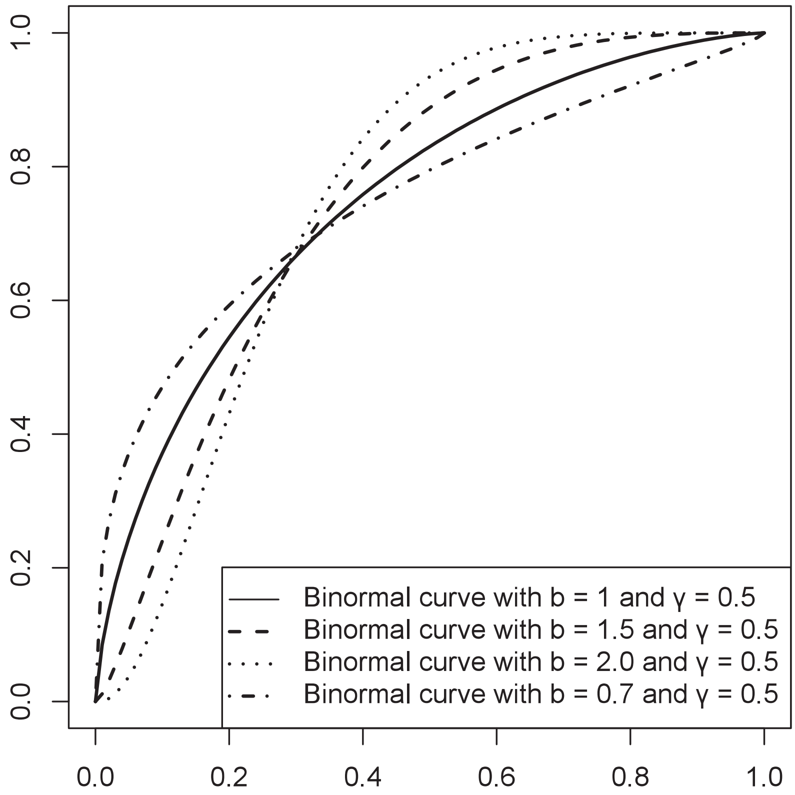


Figure 2. Binormal curves.

Older works have also described a simplified version of the binormal model (this simplified version is sometimes referred to as the "normal" curve). This model is based on the assumption of the equal variances of the underlying score distributions of good and bad observations (Swets 1986). As a consequence, the b parameter equals 1, and Equation (21) is transformed to:

$$y = \Phi\left(\Phi^{-1}\left(\frac{\gamma+1}{2}\right)\sqrt{2} + \Phi^{-1}(x)\right). \tag{22}$$

Such a model has only one parameter, the Gini coefficient, and will here be referred to as the "midnormal" model.

3.8. Midfractal Model

Similarly, one can reduce the bifractal model (Equation (19)) to a one-parameter curve by setting β to 0.5 (in the middle of the two fractal curves):

$$y = \frac{1}{2}\left(\left(1 - (1-x)^{\frac{1+\gamma}{1-\gamma}}\right) + x^{\frac{1-\gamma}{1+\gamma}}\right). \tag{23}$$

Again, this curve has only one parameter γ, which stands for the Gini coefficient. This curve will here be referred to as the "midfractal" model.

4. Fitting ROC Curve Models to Empirical Data

The theoretical models for an ROC curve presented in the previous section can be fitted to the empirical ROC data of real-life scoring models in credit institutions. For the empirical analysis presented below, two types of sources of ROC curve data were used: (1) research/industry articles and presentations and (2) data from credit institutions obtained under an anonymity condition.

(1) We used the following papers containing data or at least graphs of empirical ROC curves related to credit scoring: Řezáč and Řezáč (2011), Wójcicki and Migut (2010),

Hahm and Lee (2011), Iyer et al. (2016), Tobback and Martens (2019), and Berg et al. (2020). Additionally, presentations by Jennings (2015) and Conolly (2017) were used. To obtain the numbers (x and y coordinates of the points that make up the empirical ROC) in some cases, it was necessary to read the data from the graph itself; therefore, an online tool was used to transform graphs into numbers by pointing and clicking.

(2) Four retail lenders in Europe shared the empirical ROC curves of their credit-scoring models. The data were provided under the condition of anonymity. These models are presented in this article under the symbols A1, A2, B1, B2, B3, C1, and D1.

All the empirical curves in the analysis reflect the discrimination characteristic of some form of a credit-scoring model; the only exception is the antifraud model from Wójcicki and Migut (2010), where the target variable is a fraudulent loan, not a default. The empirical ROC curves describe scoring models that are created using various methods, including support vector machines (Tobback and Martens 2019) or a proprietary methodology (Jennings 2015), but the dominant approach is logistic regression in various forms (Hahm and Lee 2011; Řezáč and Řezáč 2011; Wójcicki and Migut 2010; Berg et al. 2020). In the case of the data coming from the four credit institutions, we do not have information about the methods used to develop the scoring models.

Once the data are available, the question emerges: What is the adequate procedure for fitting the curve? The binormal and bilogistic curves may be fitted quite intuitively through a probit/logit transformation and a simple linear-regression fitting (Swets 1986), and the parametric ROC curve models may be fitted with maximum-likelihood procedures (Metz and Pan 1999; Ogilvie and Creelman 1968). Such a procedure is not available for algebraic models (such as the bifractal curve or power function). As it is reasonable to use the same fitting method for all the ROC curve models, we applied the minimum distance estimation (MDE) method as developed by Hsieh and Turnbull (1996) and Davidov and Nov (2012), and described by Jokiel-Rokita and Topolnicki (2019). We used the numerical optimisation in R (*optim* function from the R *stats* package). The "objective function" to be minimised is the L_2-distance measure between the empirical ROC curve and the theoretical ROC curve function (Jokiel-Rokita and Topolnicki 2019):

$$f_{obj}(\boldsymbol{\theta}) = \int_0^1 (ROC_e(x) - ROC_t(\boldsymbol{\theta},\ x))^2 dx \tag{24}$$

where $ROC_e(x)$ is the empirical ROC curve (piecewise linear interpolation of empirical ROC data points) and $ROC_t(\boldsymbol{\theta},\ x)$ is the ROC curve model with $\boldsymbol{\theta}$ as a vector of parameters (1–4 parameters depending on the function). The minimal distance estimator $\hat{\boldsymbol{\theta}}$ of the parameter vector is defined by:

$$f_{obj}(\hat{\boldsymbol{\theta}}) = \inf_{\boldsymbol{\theta} \in \Theta} f_{obj}(\boldsymbol{\theta}) \tag{25}$$

As a result of the choice of such an objective function, the average vertical root-mean-square distance between the empirical curve and the theoretical curve was minimised.

Before we provide the summary results based on all the curve models and all the ROC data sets gathered, let us introduce an illustrative example. The example results of fitting four curves (bifractal, binormal, bilogistic, and power) to the empirical ROC curve obtained for the purposes of this study from an anonymous credit institution (D1) are presented in Figure 3 and Table 1. As it can be seen, the binormal and bifractal models fit the data quite well. At the same time, the highest deviation was observed in the case of the power curve. Moreover, the Gini coefficient implied by the power curve differed slightly from the Gini coefficient implied by the first two models (and also differed from the actual underlying Gini coefficient, which was circa 0.43).

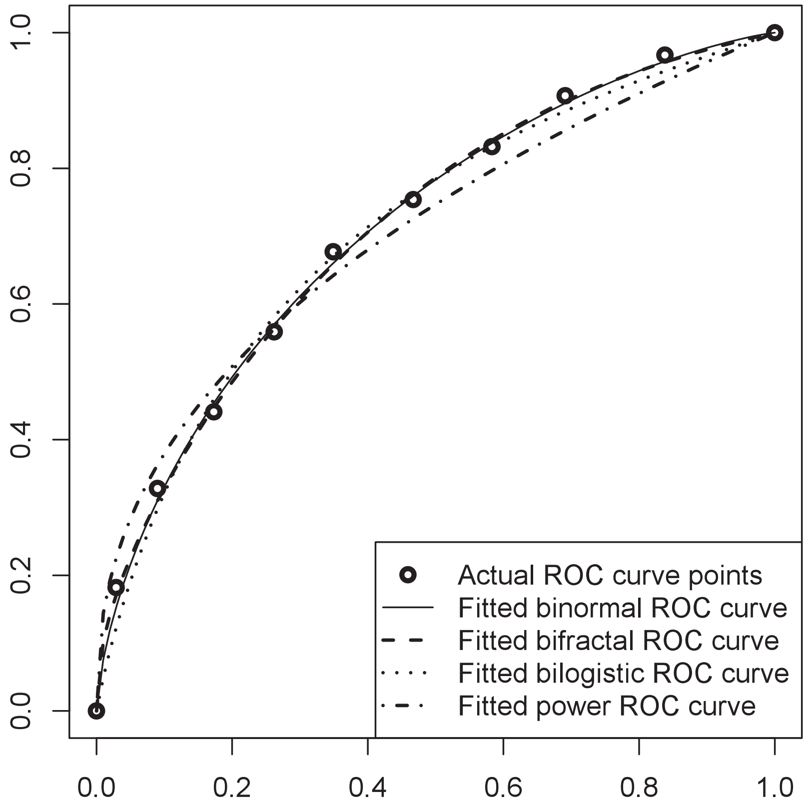


Figure 3. Results of the fitting of ROC models to the D1 data.

Table 1. Results of the fitting of four ROC models to the D1 data.

Model	Parameters	Fitting Objective (f_{obj})
Binormal (Equation (21))	b = 0.9539; γ = 0.4290	f_{obj} = 8.09 × 10^{-5}
Bifractal (Equation (19))	β = 0.4239; γ = 0.4298	f_{obj} = 8.40 × 10^{-5}
Bilogistic (Equation (12))	α_0 = 1.2884; α_1 = 0.9279	f_{obj} = 3.31 × 10^{-4}
Power (Equations (16) or (17))	θ = 0.4072 or γ = 0.4212	f_{obj} = 1.27 × 10^{-3}

Table 2 presents the goodness of fit for all the data sets gathered. For reasons of clarity, the square root of the minimised objective (f_{obj}) was multiplied by 100; note that it can then be interpreted as the average (root mean squared) vertical distance between the empirical ROC curve and the fitted ROC curve, expressed in percentage points. As it turns out, the binormal model was the best in terms of goodness of fit in 10 cases. On average, the vertical distance between the empirical and binormal ROC curves was less than one percentage point and in the worst case it did not exceed two points, which could be considered an excellent fit. The bibeta model won in five instances. The bilogistic model, which on average fit slightly worse than the competing models, was the best in four cases. The bifractal model also showed a good fit, but it was worse than the binormal in all but one case. The power curve showed the worst fit, with one exception.

Figure 4 summarises findings from Table 2 and presents the average goodness of fit (square root of $f_{obj}(\hat{\theta})$) for each model. The summarised data showed that all the models presented, except for the power curve, were comparable in terms of the average goodness of fit. However, both four-parameter models (bibeta and bigamma) and the binormal model were, on average, the ones matching the data best.

Table 2. ROC model curve fitting—the goodness of fit ($\sqrt{f_{obj}} \times 100$).

	Binormal	Midnormal	Bifractal	Midfractal	Bilogistic	Bibeta	Simplified Bibeta	Bigamma	Power
Berg et al. (2020), Credit Bureau model	0.79	1.40	1.14	1.58	0.53	0.80	1.07	0.80	2.75
Berg et al. (2020), Digital Footprint model2	**0.53**	2.14	1.00	2.25	1.35	0.53	0.65	0.64	2.66
Conolly (2017), curve I	**0.84**	1.40	1.19	1.71	1.22	0.90	0.94	0.95	5.15
Conolly (2017), curve II	0.82	0.94	1.06	1.14	**0.80**	0.83	0.96	0.94	2.46
Jennings (2015)	**0.75**	1.09	1.38	1.64	0.78	0.77	1.11	0.80	5.76
Hahm and Lee (2011), model A	1.38	3.14	2.13	3.59	**1.06**	1.45	1.70	1.39	3.39
Hahm and Lee (2011), model B	**0.68**	0.95	1.65	1.79	1.11	0.91	1.04	0.69	5.24
Iyer et al. (2016)	**0.42**	1.42	1.04	1.77	1.29	0.43	0.61	0.42	6.13
Řezáč and Řezáč (2011)	**0.82**	1.11	1.34	1.58	1.11	0.90	1.05	1.05	5.21
Řezáč and Řezáč (2011)—additional data points read from the graph	0.53	0.53	0.71	0.72	1.50	**0.50**	0.53	0.52	4.45
Tobback and Martens (2019)	1.95	2.52	1.89	2.51	3.24	**1.08**	1.64	1.60	7.55
Wójcicki and Migut (2010)	**0.67**	0.98	1.66	1.80	1.63	0.68	0.85	0.74	6.54
Model A1	**0.50**	0.55	0.68	0.73	0.88	0.51	0.62	0.60	3.36
Model A2	**0.68**	0.81	0.80	0.88	0.88	0.68	0.80	0.78	3.00
Model B1	0.90	0.90	1.34	1.40	2.04	**0.82**	0.83	0.84	5.41
Model B2	1.66	3.32	1.69	3.32	**1.55**	1.67	1.81	1.83	8.61
Model B3	1.71	2.27	2.08	2.63	2.61	**1.55**	1.60	1.57	4.03
Model C1	**0.46**	0.71	0.95	1.13	0.72	0.49	0.79	0.78	4.78
Model D1	0.90	1.18	0.92	1.13	1.82	**0.78**	0.79	0.79	3.57

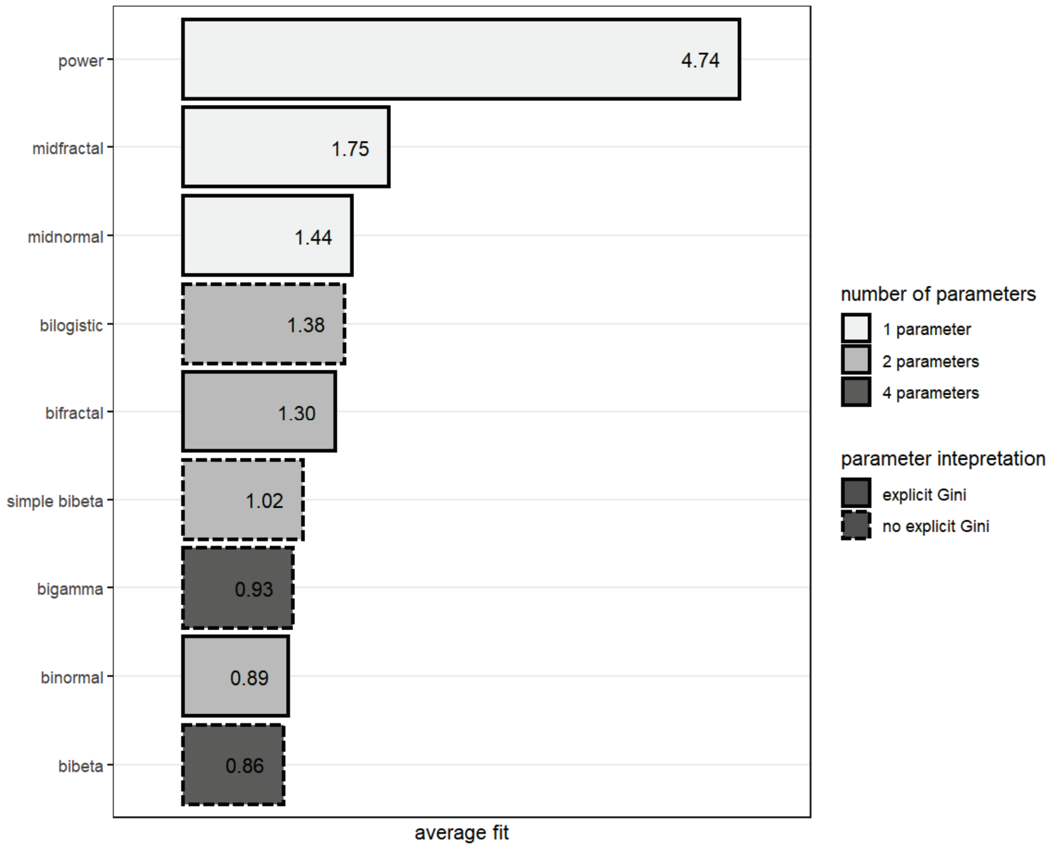


Figure 4. Average goodness of fit ($\sqrt{f_{obj}} \times 100$) for particular models.

5. Discussion and Conclusions

The empirical results presented in the previous section constitute the first—as far as we know—test of various ROC curve models against empirical credit-scoring data. They show that the bibeta model, on average, fits the empirical data best. The goodness of fit for the binormal model is at a comparable level. The binormal model is also the model that proved to be the best fit for the largest number of data sets. Obviously, the obtained results can be reinforced or undermined by further research based on more data on empirical ROC curves. This is one of the reasons why we are sharing the code used in this article that would allow anyone interested to perform such tests.

It is worth noting that ROC curve models are not predictive statistical models, as understood by, e.g., Hastie et al. (2016). ROC curve models are descriptive mathematical models; some are parametric (they use assumptions about probability distributions), whereas others are purely algebraic. Their primary purpose is to approximately describe empirical ROC curves. Therefore, the methods of selecting and assessing predictive models (Hastie et al. 2016; Ramspek et al. 2021) are not directly applicable in the exercise presented in the preceding section, as it is not about finding the best predictive model. Instead, it is an exercise in finding the best theoretical curve approximating the empirical one. It is somewhat similar to the curve-fitting tasks in computer science or engineering (Fang and Gossard 1995; Frisken 2008; Guest 2012). The minimum distance approach proposed and described by Hsieh and Turnbull (1996), Davidov and Nov (2012), and Jokiel-Rokita and Topolnicki (2019) seems to be the optimal method for such an exercise as it allows for fitting both parametric and algebraic curves in the same, unified way.

When assessing the appropriateness of a particular ROC curve model, not only the goodness of fit counts. One should consider some other aspects, including the number of parameters and the possibility of interpreting them. Intuitively, the fewer parameters in the formula, other things being equal, the better. As illustrated in Figure 4, there were three models with only one parameter (the Gini coefficient of a given scorecard): the midnormal, midfractal, and power models. The midnormal curve was the best-fitting model in this category. The binormal, bifractal, bilogistic and simple bibeta models require two parameters. The binormal model won over all the other two-parameter models. The best-fitting (on average) model, the bibeta curve, has four parameters.

From the perspective of the credit-scoring modelling practice, it is vital to have an explicit Gini/AUC parameter in the ROC formula (Kochański 2021). Models based on fractal curves (the bifractal and midfractal models) fulfil this postulate. When Equation (21) is the basis of a binormal (or midnormal) curve, the explicit Gini parameter is also available. A power curve is another example of a curve that can be defined so that its (one and only) parameter is the Gini coefficient. Still, as shown in the previous section, its fit with the empirical data was much worse than that of the competing models. In consequence, there were five models with an explicit Gini parameter on our list; other models did not allow for simple reformulation aimed at obtaining the Gini coefficient as the input.

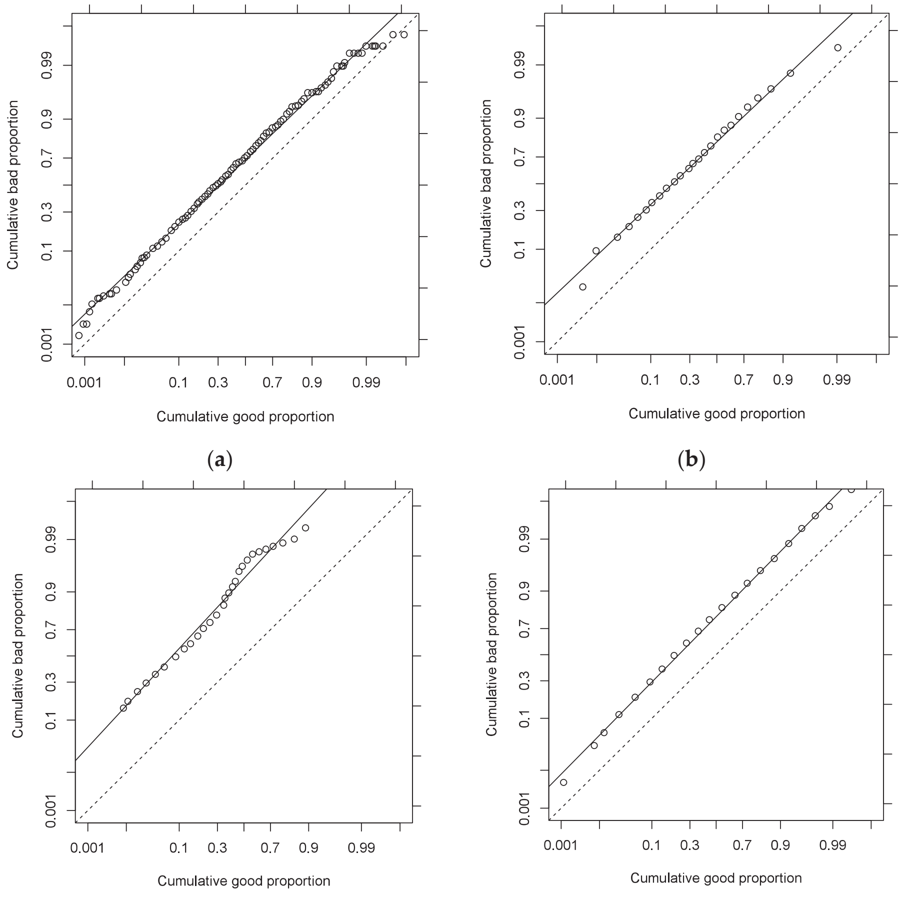
The other parameter of the bifractal model, responsible for the shape of the curve, also has an apparent meaning. However, the bifractal model lacks a theoretical foundation, which may be considered a substantial disadvantage of this approach.

A potential shortcoming of the binormal model is the presence of "hooks", i.e., non-concave regions that are irrational for the ROC curve. Such a "hook" is a portion of the curve below the 45º diagonal, which makes random guessing in these regions a better option than making decisions based on the ROC. Curves with such "hooks" are referred to as "improper" curves. To address this shortcoming, "proper" ROC curves have been suggested (Chen and Hu 2016; Dorfman et al. 1997; Metz and Pan 1999).

It seems that in practice, in the credit-scoring context, the "improperness" does not constitute a problem. The binormal curve demonstrates no hooks for $b = 1$, and if b is close to 1, the size of the hook regions is negligible from a practical standpoint. For the empirical data sets from Section 3, the maximum b of the fitted binormal curve was 1.29, and the minimum was 0.79. Visual inspection confirmed that the hooks were not visible (yet they were present; for example, for the fitted binormal curve with $b = 1.29$ and $\gamma = 0.74$, there was a hook region for $x < 10^{-9}$).

Another argument against the binormal/midnormal curves (as well as against the bibeta and bigamma) is that these models require quite complex mathematical operations (Birdsall 1973, pp. 100–8). Such an argument would support the bilogistic and bifractal model; however, thanks to the availability of specialised computer software and cheap computing power, it is not as important as it probably would have been half a century ago.

Hanley (1988) summarised the arguments in favour of the binormal model. The claims included mathematical tractability and convenience, and theoretical considerations. Empirical results (Swets 1986) have also shown that the model fits the data quite well. Additionally, as shown by Hanley (1988), because of the relative scarcity of medical data, the random noise is much more visible than the deviations driven by the differences between the models. The scarcity of data is a less frequent problem in the credit-scoring domain, but the empirical argument (the fit turns out to be the best for many real-life instances) is even more vital. Figure 5 shows the results of a visual inspection as proposed by Swets (1986). The plots are "binormal" in the sense that the cumulative bad and good proportions were rescaled according to their standard normal distribution deviates ($\Phi^{-1}(x)$ and $\Phi^{-1}(y)$). If the binormal model is adequate, then the empirical data points should gather along a straight line. For most of the empirical curves presented Section 3, this was true. The data from Tobback and Martens (2019) had the most pronounced deviations (see the irregular shape in Figure 5c), but none of the other ROC curve models could explain this anomaly.



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

(**c**) (**d**)

Figure 5. Empirical ROC curves on binormal plots. (**a**) A1 model; (**b**) Data from Řezáč and Řezáč (2011); (**c**) Data from Tobback and Martens (2019); (**d**) C1 model.

Table 3 brings together the advantages and disadvantages of the ROC curve models discussed in this article.

Table 3. ROC curve models—pros and cons.

Model	Pros	Cons
Bigamma/bibeta/simple bibeta	Good fit. "Proper" with some restrictions.	4 parameters (2 in case of simple bibeta), no explicit AUROC parameter, complicated implementation (requires beta and gamma distribution functions).
Bilogistic	2 parameters, simple mathematical operations.	On average, the worst among the models with more than one parameter, presence of non-concave regions.
Bifractal/midfractal	2 parameters (or 1 in case of midfractal), explicit Gini parameter, interpretable shape parameter, only the simplest mathematical operations needed, monotone in the whole domain.	Lack of theoretical background, clearly "algebraic".
Binormal/midnormal	2 parameters (or 1 in case of midnormal), the model may be reformulated to produce an explicit Gini parameter. Mathematical tractability, convenience, good empirical fit in case of credit scoring and in other domains.	Presence of "hooks": non-concave regions of the curve if $b \neq 1$.
Power	One parameter	Very poor fit.

So far, the literature on the subject of credit scoring has not devoted much attention to ROC curve models. With the exception of Satchell and Xia (2008), Kürüm et al. (2012), and Kochański (2021), we did not find articles in this area that directly refer to curve models, binormal or others. This study helps to fill that gap and provides credit-risk managers and researchers with a useful set of ROC curve modelling tools. As demonstrated in the literature section, ROC curve models are valuable for credit-risk management. They provide methods for determining confidence intervals and inferring the AUROC from a sample. Thanks to them, it is possible to model the impact of scoring models that have not yet been built. They also allow for a concise description of the curve shape: a scorecard with the same AUROC but a different shape of the curve may be used differently (cutting off the worst customers versus selecting the best of the best). Each of these applications deserves separate research; the results provided in this paper provide a good starting point for such studies.

Concluding, the binormal model seems to be the optimal approach to modelling credit-scoring ROC curves. When a one-parameter ROC curve model is needed, the midnormal (the binormal model with equal variances assumption) seems to be the right choice. The binormal model can be accommodated to have the Gini coefficient as a parameter. This feature is quite essential from a credit-risk-management perspective. Additionally, the mathematical tractability of the model, as well as convenience and theoretical considerations provide arguments in favour of this approach. "Improperness" of the binormal model (presence of nonconcave "hook" regions if variances are not equal) seems to have little practical importance.

Funding: This research received no external funding.

Data Availability Statement: The data and R code performing the calculations as well as generating the graphs are made available in the script file roc_curves_fitting.R at https://github.com/roccurves/scoringROCcurves/blob/main/roc_curves_fitting.R.

Conflicts of Interest: The author declares no conflict of interest.

References

Anderson, Raymond. 2007. *The Credit Scoring Toolkit: Theory and Practice for Retail Credit Risk Management and Decision Automation*. Oxford: Oxford University Press.

Atapattu, Saman, Chintha Tellambura, and Hai Jiang. 2010. Analysis of area under the ROC curve of energy detection. *IEEE Transactions on Wireless Communications* 9: 1216–25. [CrossRef]

Bamber, Donald. 1975. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *Journal of Mathematical Psychology* 12: 387–415. [CrossRef]

Bandos, Andriy I., Ben Guo, and David Gur. 2017. Estimating the Area Under ROC Curve When the Fitted Binormal Curves Demonstrate Improper Shape. *Academic Radiology* 24: 209–19. [CrossRef] [PubMed]

Berg, Tobias, Valentin Burg, Ana Gombović, and Manju Puri. 2020. On the Rise of FinTechs: Credit Scoring Using Digital Footprints. *The Review of Financial Studies* 33: 2845–97. [CrossRef]

Bewick, Viv, Liz Cheek, and Jonathan Ball. 2004. Statistics review 13: Receiver operating characteristic curves. *Critical Care* 8: 508. [CrossRef]

Birdsall, Theodore G. 1973. *The Theory of Signal Detectability: ROC Curves and Their Character*. Ann Arbor: Cooley Electronics Laboratory, Department of Electrical and Computer Engineering, The University of Michigan.

Blöchlinger, Andreas, and Markus Leippold. 2006. Economic benefit of powerful credit scoring. *Journal of Banking & Finance* 30: 851–73. [CrossRef]

Bowyer, Kevin, Christine Kranenburg, and Sean Dougherty. 2001. Edge Detector Evaluation Using Empirical ROC Curves. *Computer Vision and Image Understanding* 84: 77–103. [CrossRef]

Bradley, Andrew P. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30: 1145–59. [CrossRef]

Chang, Chein-I. 2010. Multiparameter Receiver Operating Characteristic Analysis for Signal Detection and Classification. *IEEE Sensors Journal* 10: 423–42. [CrossRef]

Chen, Weijie, and Nan Hu. 2016. Proper Bibeta ROC Model: Algorithm, Software, and Performance Evaluation. In Medical Imaging 2016: Image Perception, Observer Performance, and Technology Assessment. Presented at the Medical Imaging 2016: Image Perception, Observer Performance, and Technology Assessment, SPIE, San Diego, CA, USA, March 2; pp. 97–104. [CrossRef]

Conolly, Stephen. 2017. Personality and Risk: A New Chapter for Credit Assessment. Presented at the Credit Scoring and Credit Control XV Conference, Edinburgh, UK, August 30–September 1; Available online: https://www.business-school.ed.ac.uk/crc-conference/accepted-papers (accessed on 27 April 2018).

Cook, Nancy R. 2007. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation* 115: 928–35. [CrossRef]

Cook, Nancy R. 2008. Statistical evaluation of prognostic versus diagnostic models: Beyond the ROC curve. *Clinical Chemistry* 54: 17–23. [CrossRef] [PubMed]

Davidov, Ori, and Yuval Nov. 2012. Improving an estimator of Hsieh and Turnbull for the binormal ROC curve. *Journal of Statistical Planning and Inference* 142: 872–77. [CrossRef]

Djeundje, Viani B., Jonathan Crook, Raffaella Calabrese, and Mona Hamid. 2021. Enhancing credit scoring with alternative data. *Expert Systems with Applications* 163: 113766. [CrossRef]

Dorfman, Donald D., Kevin S. Berbaum, Charles E. Metz, Russell V. Lenth, James A. Hanley, and Hatem Abu Dagga. 1997. Proper Receiver Operating Characteristic Analysis: The Bigamma Model. *Academic Radiology* 4: 138–49. [CrossRef]

England, William L. 1988. An Exponential Model Used for Optimal Threshold Selection on ROC Curves. *Medical Decision Making* 8: 120–31. [CrossRef]

Fang, Lian, and David C. Gossard. 1995. Multidimensional curve fitting to unorganized data points by nonlinear minimization. *Computer-Aided Design* 27: 48–58. [CrossRef]

Faraggi, David, and Benjamin Reiser. 2002. Estimation of the area under the ROC curve. *Statistics in Medicine* 21: 3093–106. [CrossRef]

Faraggi, David, Benjamin Reiser, and Enrique F. Schisterman. 2003. ROC curve analysis for biomarkers based on pooled assessments. *Statistics in Medicine* 22: 2515–27. [CrossRef]

Fawcett, Tom. 2006. An Introduction to ROC Analysis. *Pattern Recognition Letters* 27: 861–74. [CrossRef]

Frisken, Sarah F. 2008. Efficient Curve Fitting. *Journal of Graphics Tools* 13: 37–54. [CrossRef]

Gneiting, Tilmann, and Peter Vogel. 2022. Receiver operating characteristic (ROC) curves: Equivalences, beta model, and minimum distance estimation. *Machine Learning* 111: 2147–59. [CrossRef]

Gonçalves, Luzia, Ana Subtil, M. Rosário Oliveira, and Patricia de Zea Bermudez. 2014. ROC Curve Estimation: An Overview. *REVSTAT-Statistical Journal* 12: 1–20. [CrossRef]

Gönen, Mithat, and Glenn Heller. 2010. Lehmann Family of ROC Curves. *Medical Decision Making* 30: 509–17. [CrossRef]

Guest, Philip George. 2012. *Numerical Methods of Curve Fitting*. Cambridge: Cambridge University Press.
Guido, Giuseppe, Sina Shaffiee Haghshenas, Sami Shaffiee Haghshenas, Alessandro Vitale, Vincenzo Gallelli, and Vittorio Astarita. 2020. Development of a Binary Classification Model to Assess Safety in Transportation Systems Using GMDH-Type Neural Network Algorithm. *Sustainability* 12: 6735. [CrossRef]
Hahm, Joon-Ho, and Sangche Lee. 2011. Economic Effects of Positive Credit Information Sharing: The Case of Korea. *Applied Economics* 43: 4879–90. [CrossRef]
Hajian-Tilaki, Karimollah. 2013. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine* 4: 627–35.
Hamel, Lutz. 2009. Model Assessment with ROC Curves. In *Encyclopedia of Data Warehousing and Mining*, 2nd ed. Pennsylvania: IGI Global. [CrossRef]
Hand, David J. 2009. Measuring classifier performance: A coherent alternative to the area under the ROC curve. *Machine Learning* 77: 103–23. [CrossRef]
Hand, David J., and Christoforos Anagnostopoulos. 2013. When is the area under the receiver operating characteristic curve an appropriate measure of classifier performance? *Pattern Recognition Letters* 34: 492–95. [CrossRef]
Hanley, James A. 1988. The Robustness of the "Binormal" Assumptions Used in Fitting ROC Curves. *Medical Decision Making* 8: 197–203. [CrossRef] [PubMed]
Hanley, James A. 1996. The Use of the 'Binormal' Model for Parametric ROC Analysis of Quantitative Diagnostic Tests. *Statistics in Medicine* 15: 1575–85. [CrossRef]
Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2016. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. New York: Springer.
Hautus, Michael J., Michael O'Mahony, and Hye-Seong Lee. 2008. Decision Strategies Determined from the Shape of the Same–Different ROC Curve: What Are the Effects of Incorrect Assumptions? *Journal of Sensory Studies* 23: 743–64. [CrossRef]
Hsieh, Fushing, and Bruce W. Turnbull. 1996. Nonparametric and semiparametric estimation of the receiver operating characteristic curve. *The Annals of Statistics* 24: 25–40. [CrossRef]
Idczak, Adam Piotr. 2019. Remarks on Statistical Measures for Assessing Quality of Scoring Models. *Acta Universitatis Lodziensis. Folia Oeconomica* 4: 21–38. [CrossRef]
Iyer, Rajkamal, Asim Ijaz Khwaja, Erzo F. P. Luttmer, and Kelly Shue. 2016. Screening Peers Softly: Inferring the Quality of Small Borrowers. *Management Science* 62: 1554–77. [CrossRef]
Janssens, A. Cecile J. W., and Forike K. Martens. 2020. Reflection on modern methods: Revisiting the area under the ROC Curve. *International Journal of Epidemiology* 49: 1397–403. [CrossRef]
Jennings, Andrew. 2015. Expanding the Credit Eligible Population in the USA. Presented at the Credit Scoring and Credit Control XIV Conference—Conference Papers, Edinburgh, UK, August 26–28. Available online: https://www.business-school.ed.ac.uk/crc/category/conference-papers/2015/ (accessed on 27 April 2018).
Jokiel-Rokita, Alicja, and Rafał Topolnicki. 2019. Minimum distance estimation of the binormal ROC curve. *Statistical Papers* 60: 2161–83. [CrossRef]
Kochański, Błażej. 2021. Bifractal Receiver Operating Characteristic Curves: A Formula for Generating Receiver Operating Characteristic Curves in Credit-Scoring Contexts. *Journal of Risk Model Validation* 15: 1–18. [CrossRef]
Krzanowski, Wojtek J., and David J. Hand. 2009. *ROC Curves for Continuous Data*, 1st ed. London: Chapman and Hall/CRC.
Kürüm, Efsun, Kasirga Yildirak, and Gerhard-Wilhelm Weber. 2012. A classification problem of credit risk rating investigated and solved by optimisation of the ROC curve. *Central European Journal of Operations Research* 20: 529–57. [CrossRef]
Lahiri, Kajal, and Liu Yang. 2018. Confidence Bands for ROC Curves With Serially Dependent Data. *Journal of Business & Economic Statistics* 36: 115–30. [CrossRef]
Lappas, Pantelis Z., and Athanasios N. Yannacopoulos. 2021. A machine learning approach combining expert knowledge with genetic algorithms in feature selection for credit risk assessment. *Applied Soft Computing* 107: 107391. [CrossRef]
Levy, Bernard C. 2008. *Principles of Signal Detection and Parameter Estimation*, 2008th ed. Berlin: Springer.
Lloyd, Chris J. 2000. Fitting ROC Curves Using Non-linear Binomial Regression. *Australian & New Zealand Journal of Statistics* 42: 193–204. [CrossRef]
Mandrekar, Jayawant N. 2010. Receiver Operating Characteristic Curve in Diagnostic Test Assessment. *Journal of Thoracic Oncology* 5: 1315–16. [CrossRef] [PubMed]
Metz, Charles E. 1978. Basic principles of ROC analysis. *Seminars in Nuclear Medicine* 8: 283–98. [CrossRef]
Metz, Charles E., and Xiaochuan Pan. 1999. "Proper" Binormal ROC Curves: Theory and Maximum-Likelihood Estimation. *Journal of Mathematical Psychology* 43: 1–33. [CrossRef]
Mossman, Douglas, and Hongying Peng. 2016. Using Dual Beta Distributions to Create "Proper" ROC Curves Based on Rating Category Data. *Medical Decision Making* 36: 349–65. [CrossRef]
Ogilvie, John C., and C. Douglas Creelman. 1968. Maximum-likelihood estimation of receiver operating characteristic curve parameters. *Journal of Mathematical Psychology* 5: 377–91. [CrossRef]
Omar, Luma, and Ioannis Ivrissimtzis. 2019. Using theoretical ROC curves for analysing machine learning binary classifiers. *Pattern Recognition Letters* 128: 447–51. [CrossRef]

Park, Seong Ho, Jin Mo Goo, and Chan-Hee Jo. 2004. Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists. *Korean Journal of Radiology* 5: 11–18. [CrossRef]

Pencina, Michael J., Ralph B. D'Agostino Sr., Ralph B. D'Agostino Jr., and Ramachandran S. Vasan. 2008. Evaluating the added predictive ability of a new marker: From area under the ROC curve to reclassification and beyond. *Statistics in Medicine* 27: 157–72; discussion 207–12. [CrossRef]

Ramspek, Chava L., Kitty J. Jager, Friedo W. Dekker, Carmine Zoccali, and Merel van Diepen. 2021. External validation of prognostic models: What, why, how, when and where? *Clinical Kidney Journal* 14: 49–58. [CrossRef]

Řezáč, Martin, and František Řezáč. 2011. How to Measure the Quality of Credit Scoring Models. *Czech Journal of Economics and Finance (Finance a Úvěr)* 61: 486–507.

Řezáč, Martin, and Jan Koláček. 2012. Lift-Based Quality Indexes for Credit Scoring Models as an Alternative to Gini and KS. *Journal of Statistics: Advances in Theory and Applications* 7: 1–23.

Satchell, Stephen, and Wei Xia. 2008. 8—Analytic models of the ROC Curve: Applications to credit rating model validation. In *The Analytics of Risk Model Validation*. Edited by George Christodoulakis and Stephen Satchell. Cambridge: Academic Press, pp. 113–33. [CrossRef]

Scallan, Gerard. 2013. *Why You Shouldn't Use the Gini. ARCA Retail Credit Conference, Leura, Australia*. Available online: https://www.scoreplus.com/assets/files/Whats-Wrong-with-Gini-why-you-shouldnt-use-it-ARCA-Retail-Credit-Conference-Nov-2013.pdf (accessed on 29 June 2022).

Shen, Feng, Xingchao Zhao, Gang Kou, and Fawaz E. Alsaadi. 2021. A new deep learning ensemble credit risk evaluation model with an improved synthetic minority oversampling technique. *Applied Soft Computing* 98: 106852. [CrossRef]

Siddiqi, Naeem. 2017. *Intelligent Credit Scoring: Building and Implementing Better Credit Risk Scorecards*. Hoboken: John Wiley & Sons.

Somers, Robert H. 1962. A New Asymmetric Measure of Association for Ordinal Variables. *American Sociological Review* 27: 799–811. [CrossRef]

Swets, John A. 1986. Form of Empirical ROCs in Discrimination and Diagnostic Tasks: Implications for Theory and Measurement of Performance. *Psychological Bulletin* 99: 181–98. [CrossRef]

Swets, John A. 2014. *Signal Detection Theory and ROC Analysis in Psychology and Diagnostics: Collected Papers*. London: Psychology Press. [CrossRef]

Tang, Tseng-Chung, and Li-chiu Chi. 2005. Predicting multilateral trade credit risks: Comparisons of Logit and Fuzzy Logic models using ROC curve analysis. *Expert Systems with Applications* 28: 547–56. [CrossRef]

Thomas, Lyn C. 2009. *Consumer Credit Models: Pricing, Profit and Portfolios*. Oxford: Oxford University Press.

Thomas, Lyn, Jonathan Crook, and David Edelman. 2017. *Credit Scoring and Its Applications*, 2nd ed. Philadelphia: Society for Industrial and Applied Mathematics. [CrossRef]

Tobback, Ellen, and David Martens. 2019. Retail Credit Scoring Using Fine-Grained Payment Data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182: 1227–46. [CrossRef]

Tripathi, Diwakar, Damodar R. Edla, Venkatanareshbabu Kuppili, and Annushree Bablani. 2020. Evolutionary Extreme Learning Machine with novel activation function for credit scoring. *Engineering Applications of Artificial Intelligence* 96: 103980. [CrossRef]

Wichchukit, Sukanya, and Michael O'Mahony. 2010. A Transfer of Technology from Engineering: Use of ROC Curves from Signal Detection Theory to Investigate Information Processing in the Brain during Sensory Difference Testing. *Journal of Food Science* 75: R183–R193. [CrossRef] [PubMed]

Wójcicki, Bartosz, and Grzegorz Migut. 2010. Wykorzystanie skoringu do przewidywania wyłudzeń kredytów w Invest-Banku. In *Skoring w Zarządzaniu Ryzykiem*. Kraków: Statsoft, pp. 47–57.

Xia, Yufei, Junhao Zhao, Lingyun He, Yinguo Li, and Mengyi Niu. 2020. A novel tree-based dynamic heterogeneous ensemble method for credit scoring. *Expert Systems with Applications* 159: 113615. [CrossRef]

MDPI
St. Alban-Anlage 66
4052 Basel
Switzerland
www.mdpi.com

Risks Editorial Office
E-mail: risks@mdpi.com
www.mdpi.com/journal/risks

Academic Open
Access Publishing

mdpi.com

ISBN 978-3-7258-1415-2