Advanced Studies in Emerging Markets Finance

Alexander M. Karminsky
Paolo Emilio Mistrulli
Mikhail I. Stolbov
Yong Shi *Editors*

Risk Assessment and Financial Regulation in Emerging Markets' Banking

Trends and Prospects

Springer

Advanced Studies in Emerging Markets Finance

The series offers insights into the broad range of finance concepts applied to the specific environment in emerging capital markets. The series presents a broad range of theoretical concepts, empirical analyses and policy conclusions regarding financial markets, financial institutions, and corporate finance. Giving a voice to scholars from emerging and developed countries, comparative studies analyze different financial markets as well as firms' performance in emerging and developed economies. There is a particular focus on the largest emerging economies, namely the BRICS countries.

More information about this series at http://www.springer.com/series/15606

Alexander M. Karminsky • Paolo Emilio Mistrulli • Mikhail I. Stolbov • Yong Shi
Editors

Risk Assessment and Financial Regulation in Emerging Markets' Banking

Trends and Prospects

Springer

Editors
Alexander M. Karminsky
NRU HSE
Moscow, Russia

Paolo Emilio Mistrulli
Bank of Italy
Rome, Italy

Mikhail I. Stolbov
MGIMO University
Moscow, Russia

Yong Shi
University of Nebraska Omaha
Omaha, NE, USA

ISSN 2662-4281 ISSN 2662-429X (electronic)
Advanced Studies in Emerging Markets Finance
ISBN 978-3-030-69747-1 ISBN 978-3-030-69748-8 (eBook)
https://doi.org/10.1007/978-3-030-69748-8

This Springer imprint is published by the registered company Springer Nature Switzerland AG.
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Series Editors' Preface

The Springer series on Advanced Emerged Markets Finance promotes leading-edge research on corporate finance and asset pricing in emerging markets. Given the initiative by the National Research University Higher School of Economics (HSE), Russia, to present the advances in trends, processes, and performance in different areas of finance in the specific framework of emerging markets, the Springer Series will include empirical research by leading scholars from HSE and around the world.

The scope of the series is to present comparative and interdisciplinary research which focuses on studies about BRICS, about BRICS and developed economies, as well as new research advances in emerging markets finance. The series will be relevant to a number of social science disciplines including economics, finance, and management. It is also relevant to a wide variety of professionals in financial, business, and governmental institutions.

One major difference between developed and emerging markets is the need for emerging markets to adapt and modify some methods and models of risk assessment. In this volume, original models are explored that target estimation of the probability of default and of expected loss for emerging market financial institutions, as well as adopting various rating models. The theoretical and empirical work will show how regulators in Russia and BRICS have transformed these risk models and the respective rating systems at the regional and country level. This analysis also uncovers new methods of assessment of systemic risk and stress testing of financial institutions.

The volumes in this series also broadly address important topics in corporate finance, from mergers and acquisitions to asset pricing anomalies in emerging capital markets, and have a specific focus on interdisciplinary aspects differently from other researches. We would expect that this series will be able to provide new and useful insights into these important problems and further suggestions to policymakers.

It is obvious that even an encyclopedic volume cannot cover all aspects of this rapidly developing area of knowledge. Nevertheless, it appears the authors have dealt with many issues which are most vital for the current stage of development of that particular field. This book contains the material, which is consistently set forth and well organized to facilitate the process of perceiving and learning, and thus it can be used as a manual for a specialized course and research seminars in finance.

Finally, mature professionals, who apply their knowledge for developing and implementing risk management systems based on their specialized expertise and skills in various applied areas, are another category of potential readers for whom the book will also be a useful tool from a methodological point of view.

Moscow, Russia Irina Ivashkovskaya
Bologna, Italy Elettra Agliardi

Preface

We analyze the features of risk measurement in developing countries and demonstrate a number of approaches to financial risk analysis for emerging financial markets. In particular, the experience of Russia, Belarus, China, Brazil, and some other countries is compared.

The main objective of the book is to distill current trends in financial risk assessment and measurement in emerging markets. To achieve this goal, we focus on the following research directions:

- figuring out the contemporary features of banking in emerging markets;
- considering the peculiarities of formation and use of ratings in banking;
- exploring the assessment and modeling of risks in banking;
- studying systemic risk and stress testing;
- assessing the role of innovations in financial risk estimation and management.

The advantages and disadvantages of each approach to assessing and modeling financial risks in the banking sector in developing countries and their applicability in Russia naturally lie within the framework of this study.

In part I of the book, we consider the features of the banking system development by conducting a comparative analysis across different countries and regions. We uncover topical trends in the twenty-first century banking systems, showing the evolution of key banking indicators and financial services. This part also includes comprehensive studies of domestic and global financial risk regulation in emerging markets.

Part II is dedicated to modeling ratings and assessing their application in commercial banks. First, the principles of rating process in emerging markets are discussed. Second, rating typologies are presented. Third, rating methodologies as well as external and internal rating comparisons in emerging markets are carried out. Special attention is attached to rating system development for practical aims in commercial banks. Ratings determine the class to which this or that business entity or financial asset should be assigned, implicitly assessing the probability of a company's default.

One of the key tasks of banks is the efficient allocation of capital in the economy, which is not attainable without sound risk management. Part III deals with the estimation and modeling of credit, liquidity, payment, and other risks with a primary focus on the BRICS experience. The benefits and shortcomings of each methodology and its applicability in Russia are identified. This part covers both theoretical and empirical works. Also, in this part, an integrated system for risk assessment in a commercial bank is considered.

Part IV examines different aspects of systemic risk, which is a big challenge to financial stability. The book analyzes in more detail the methods for systemic risk identification and measurement. To a great extent, this part of the book is based on the presentations and discussions, which took place at two international workshops at Higher School of Economics in November 2018 and 2019.

Finally, in Part V, in light of the rapid development of technologies entailing business transformation, the book discusses innovations in the financial sector and their impact on financial risk estimation and management in emerging countries. For example, a dynamic fractal model of asset pricing and a network model for payment risk assessment are discussed. Peculiarities of risk management in Islamic banks are also covered in this part.

The authors' team consists of the researchers from public and private banks, rating agencies, consulting companies, and leading universities all around the world. Thus, it synthesizes the academic rigor with the practitioners' views and experience.

As a result, the book will be of interest not only to researchers, PhD students, and undergraduate students with a financial background, but also to practitioners from banking as well as from other economic and financial fields.

Moscow, Russia — Alexander M. Karminsky
Rome, Italy — Paolo Emilio Mistrulli
Moscow, Russia — Mikhail I. Stolbov
Omaha, NE — Yong Shi

Acknowledgements

It is a great pleasure to acknowledge the efforts of many people who have contributed either directly or indirectly to our work on this book. The ideas presented in this book have been shaped and influenced by the practical implementation of risk management and topics raised at academic research conferences and seminars at the National Research University Higher School of Economics, New Economic School in Moscow. We are also indebted to MA and PhD students whom we supervised. We also acknowledge those, who helped us design the volume, especially Natalia Dyachkova, Andrey Egorov, Ekaterina Khromenko.

Special thanks to Series Editors Irina Ivashkovskaya and Joseph McCahery for their ideas and useful discussion about the structure of the volume as well as for the tips on the language proofreading of the book.

We would like to thank the contributors, who despite their busy schedules have jointly worked with us, especially Dean of the Economic Science Faculty of NRU HSE, Sergey E. Pekarski.

We are thankful to the Springer publishing team for their eager assistance and expert advice. In particular, we thank Dr. Prashanth Mahagaonkar, and the Springer editorial team, including Ms. Rocio Torregrosa.

We express our heartiest thanks to the administration of the National Research University and the School of Finance at this University for its longtime support. We are also grateful to our partners, i.e. universities, banks, and financial companies where the contributing authors and editors of this volume work, in particular, Moscow State Institute of International Relations (MGIMO University), the University of Nebraska at Omaha, the Bank of Russia, and the Bank of Italy.

Abstract

This book focuses on topical issues connected with the evaluation and modeling of financial risks in emerging markets. The contributing authors present results and methodologies related to credit and market risks, integrated and global payment risks, as well as systemic risks and innovations in risk management and banking.

The objective of the book is to characterize and provide evidence on current trends in financial risks assessment and measurement in emerging markets. To achieve this goal, a number of contributions document the features of risk measurement in developing countries and also apply relevant up-dated approaches to financial risk analysis for emerging financial markets. In particular, the experience of Russia, China, Brazil, Belarus, and some other countries is put into the spotlight. Besides, issues such as financial contagion and system risk are discussed in this context, which present a significant challenge to contemporary risk management.

All these aspects are investigated at the level of financial institutions and at the aggregate macro-financial level, making the proposed approach comprehensive and the book truly unique in comparison with other existing books on similar topics.

The authors' team consists of specialists from public and private banks, rating and consulting agencies, and leading universities around the world. Each of them makes a huge impact on the results of the book, which builds on the original materials presented at international conferences and workshops where some of the above-mentioned topics were discussed. The book will be interesting for students, researchers, and practitioners in banking and other financial activities.

Contents

Part I Banks in Emerging Markets

Peculiarities and Trends of Banking Systems Development 3
Artem Arkhipov, Natalia Arkhipova, and Alexander Karminsky

Regulation of Financial Risks in Emerging Markets: Past, Present, and Future 33
Artem Arkhipov, Natalia Arkhipova, and Alexander Karminsky

Part II Ratings and Risk Measuring

Principles of Rating Estimation in Emerging Countries 67
Sergei Grishunin, Natalia Dyachkova, and Alexander Karminsky

Aggregation of Rating Systems for Emerging Financial Markets 93
Sergei Grishunin, Natalia Dyachkova, and Alexander Karminsky

Part III Estimating and Modeling Credit and Market Risks in Banking

Bank Credit Risk Modeling in Emerging Capital Markets 123
Alexander Karminsky and Alexei Morgunov

Loss Given Default Estimations in Emerging Capital Markets 145
Mikhail Pomazanov

Comparing Bankruptcy Prediction Models in Emerging Markets 177
Roman Burekhin

Measures and Assessment of ALM Risks in Banks: Case of Russia 189
Ekaterina Seryakova

Forecasting and Backtesting of Market Risks in Emerging Markets 199
Dean Fantazzini

Integrated Risk Measurement System in Commercial Bank 225
Alexander Zhevaga and Alexei Morgunov

Economic Capital Structure and Banking Financial Risks Aggregation 251
Marina Pomorina

Part IV Systemic Risks Modeling and Stress Testing

Exploring the Interplay Between Early Warning Systems' Usefulness and Basel III Regulation 277
Elena Deryugina, Maria Guseva, and Alexey Ponomarenko

Does Only Volume Matter? A Stress Test for the Adequacy of International Currency Reserves for Russia 287
Renat Akhmetov, Vera Pankova, Oleg Solntsev, and Elizaveta Orlova

Regulatory Measures Against Systemic Risk in Banking Sector: The Evidence for the Republic of Belarus 309
Svetlana Malykhina

Real Effects of Financial Shocks in Russia 329
Vasilisa Baranova

Part V Estimating and Managing Financial Risks: Topical Trends in Emerging Capital Markets

Innovation in Developing Countries' Risk Estimation and Management 343
Andrey Egorov and Dmitry Pomazkin

Dynamic Fractal Asset Pricing Model for Financial Risk Evaluation . . . 355
Bruno de Conti, Vladimir Gisin, and Irina Yarygina

Network Effects in Retail Payments Market: Evidence from Individuals 369
Egor Krivosheya

Conclusion: Instruments of Financial Sustainability in Emerging Markets 395
Alexander Karminsky, Paulo Emilio Mistrulli, Mikhail Stolbov, and Yong Shi

Part I
Banks in Emerging Markets

Peculiarities and Trends of Banking Systems Development

Artem Arkhipov, Natalia Arkhipova, and Alexander Karminsky

Abstract Comparison of trends and peculiarities of financial systems in different countries, especially, in emerging markets, should start with setting the global context. The study identifies several periods of development of world financial institutions in the twenty-first century—deregulation (global optimism regarding financial development), re-regulation (change in the paradigm following the Global Financial Crisis), and de-globalization (growing divergence between conditions for doing banking in different countries). Progress of banking systems in EMs was additionally shaped by local peculiarities at the turn of the century, most of which were related to their location (e.g. European banks penetrated in CEE countries, Russian financial system was dominating in CIS). The common features of emerging markets were low banking services' penetration and high promised returns. Through time, higher market saturation, technological advances, and trends in regulation and supervision increased degree of convergence in financial systems in developed and developing countries in what concerns main KPIs. That caused revision of focus towards greater attention to risk management and local needs. Global macroeconomic risks, related to countries with high debt, technological risks, and changing clients demands will become the drivers of banking systems development in emerging markets.

Keywords Development trends · Financial systems · Emerging markets

JEL Classification B26 · E44 · F38 · G21

A. Arkhipov · N. Arkhipova · A. Karminsky (✉)
National Research University Higher School of Economics, Moscow, Russia
e-mail: narkhipova@hse.ru

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_1

1 Introductory Comments

Unprecedented shock happened to the world economy in 1Q20—the pandemic spread of coronavirus disease CoViD-19. At the moment of completing the paper, the harm of this virus has not been realized in full: there is still no ultimate information on how many people will suffer (and doubtfully it is possible to estimate from this moment in time). However, what is clear is that this awful event will definitely become a "black swan" for global financial system as well, and it will substantially sharpen its future for many years in terms of functions and institutions.

However, before turning to the future of financial systems, it is necessary to have a retrospective look over last decades of banking sector development. This might allow to discover common trends in different markets and regions, as well as differentiate them properly on the basis of various characteristics, including institutional, economic, and others. Provided relative sizes of emerging and developed economies in the world, and taking into account that financial globalization has been a dominating trend for several decades, one should start with identifying trends, which are relevant not only to banking sectors in emerging economies, but also to all financial systems. While global trends are set by the largest economies and their financial institutions, the above-mentioned analysis would create a context for better understanding of peculiarities and general trends in emerging banking systems. The next stage should be an investigation into specific regional trends and aspects which are relevant to evolution of banking sectors in particular locations. This might cover both institutional characteristics of economy and international developments.

The next stage is to follow the progress of financial systems in emerging markets on the basis of sector's key performance indicators. This should be approached in terms of characteristics which are important for stakeholders in developed and emerging markets. Among the most objective approaches to such a goal is to consider viewpoint of an investor. However, upon completion of analysis of trends which are peculiar to banking systems in emerging economies one should pay attention to new challenges which have appeared recently. Those challenges might become important factors for the next decades, and they might sharpen development of individual banking institutions and, more generally, financial systems.

Presumably, such an analysis will not result in a complete picture of all peculiarities of banking systems in emerging economies, and might leave many aspects unattended. However, it might provide a general context of issues that banks in emerging economies face, and also become an additional contribution to better understanding of a pretty diverse world in which, in spite of everything, common laws seem to dominate.

2 Outline of Global Banking Trends in Twenty-First Century: Ups and Downs

2.1 The Main Periods of Banking Business in Twenty-First Century

Traditional banking business in twenty-first century has passed through several stages of development. While each researcher might distinguish and classify them into various periods, the following classification criterion seems the most objective—the attractiveness of the banking business to investors. The most important reason for this is straightforward: when a business segment seems attractive, one can infer that the global financial community believes in the opportunities and perspectives in this segment. Otherwise, investors see upcoming problems for getting returns on investments, and so the cost of capital for this segment rises, thus creating pressure on businesses. Hence underinvestment either brings a new era of growth (if the segment proves to be strong and needed by economic agents) or opens the door to peer businesses.

Based on this criterion, one might divide the years of the twenty-first century for banks into three time periods. All three are well-observed in Fig. 1 and Fig. 2, describing world market indicators related to the financial sector.

The first is the period of high optimism in the world economy, world financial markets, and the future of globalization (2000–2007). This period was characterized by a substantial increase in the volume of financial operations, banking sector profits, and, among others, by a rapid growth of the capitalization of financial companies.

Fig. 1 World stocks performance in 2000–2019 (index, 01.01.2000 = 100). Source: Bloomberg, authors' calculations

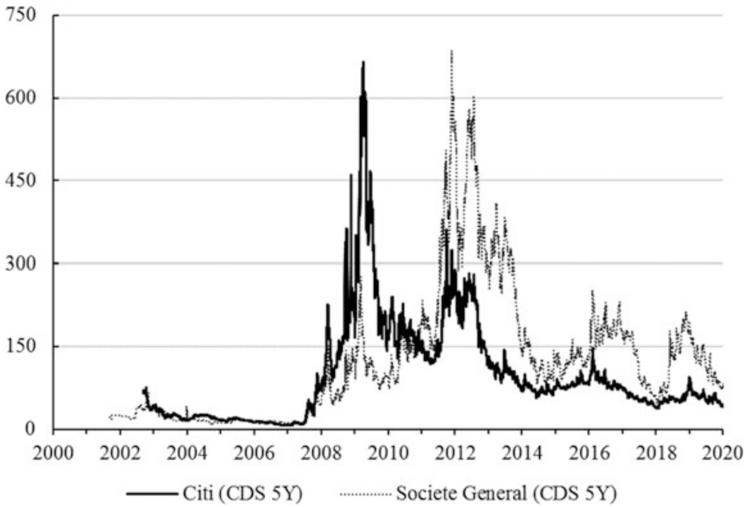


Fig. 2 CDS of selected global banks (basic points)

The second period is (2009–2014) a time of a severe hangover, which followed the Global Financial Crisis (2007–2009). During this period, investors were concerned about the future of banks and banking, and also there were different views on different regions—e.g. more skepticism about European banks due to PIIGS problems, and more optimism about US financial institutions due to the active support that the sector received from the regulator. Global cooperation was required to resolve this crisis, and new international institutions started to impose global rules over the financial sector. These rules were by far more restrictive than previous regulations, and that created substantial difficulties for banks all over the world to comply with them.

The most recent phase (2014–2019) is substantially different from the previous one. It generally characterizes a new era of more fragmented financial markets—at least when compared to how we used to define traditional banking services. On the one hand, macroeconomic and geopolitical trends increased costs of doing international banking, creating incentives to slow down the process of establishing a truly global financial market. On the other hand, thanks to technological advances, new players started to intervene in traditional banking activities, providing the same type of services with lower costs (partly due to lack of regulation).

2.2 Global Financial Rally Prior to the Global Financial Crisis

While the global financial system significantly suffered from the East Asian and Russian crises in 1997–1998, the new century brought an unusual optimism. The following reasons might have been behind it:

- IT developments helped to decrease the cost of acquiring and processing information for market participants, thus inviting them to serve a wider range of clients;
- Unification of Europe in a currency union brought expectations that future transaction costs of trade and investment would fall further, and the years after the creation of EUR, other regions in the world regularly discussed the introduction of a common currency (e.g., ACU—Asian Currency Unit), etc.;
- A policy of easing the rates in the USA was caused by two factors—a need to overcome the consequences of the stock market turbulence in the developed markets caused by the "dot com" crisis, and a need to re-build confidence in the US economy following the 9/11 tragedy.

In some sense, this was a continuation of the policy paradigm called "financial liberalization" which was an attempt to propose an economic solution to macroeconomic rigidities and inefficiencies. More importantly, this time the financial systems that were the focus of attention were different from EMs, which investors liked before the crisis of 1997.

Altogether this created two important trends: firstly, a decrease in the cost of funds for the world largest banks, and secondly, a trend to build up global financial institutions. This caused the largest banking institutions to participate in cross-border mergers, open branches and establish representative offices (including legal entities) in various countries. For instance, the share of foreign ownership in the Russian banking system jumped in the 2000s from 6% to 28%. Although a substantial part of this increase was driven by IPOs and SPOs of Sberbank and VTB, and also by the interest of foreign investors in these stocks, multiple new foreign names appeared or widely expanded in Russia and similar countries during this period—Société Générale, Nordea, Barclays, etc.

Similar trends were observed in other EM countries during this period: many financial institutions tried to build up their global presence to ensure their status as a global player. During this period many of them had to take risks, and some of these risks were excessive.

2.3 Zero-Rate Period: Times for Regulators and Tighter Regulations and Supervision

The optimistic trend occurred during most of the 2000s, yet in 2007 financial risks started to grow. The "risk-originating" country this time was the USA—the insurer Ambac, and a number of others, investment banks (e.g. Bear Stearns, Lehman Brothers), and other financial institutions (including federal agencies for mortgage-back securities FNMA, FHLMC) started to experience problems. To cope with these, the financial authorities in the USA cut the federal fund rate to almost zero. The other largest economic region—Europe—had to join the club to avoid financial conditions' tightening. Mechanics of this is well described in Stiglitz (2010). Hence the world moved into an era of key (targeted) money market rates being close to zero. A few years later the need to continue coping with crisis trends brought a powerful paradigm change to banking: the European central bank introduced negative interest rates. This has never been considered as a policy tool before.

Apart from cutting interest rates, the central banks in key world economic regions (USA, EU, UK, and Japan) had to establish recovery programs, including providing local banks with liquidity under the pledge of assets, and injecting fresh equity into their capital. Hence the cost of liquidity fell down dramatically (see Fig. 3), but altogether these measures worked as a wake-up call for regulators to increase their role. Having been heavily criticized for missing the turning point of risk accumulation, central banks of key world economies started to establish new regulatory regimes which would avoid similar situations in the future. To avoid regulatory arbitrage politicians in key world economies agreed to establish the Financial Stability Board, which became responsible for the creation of new rules for financial

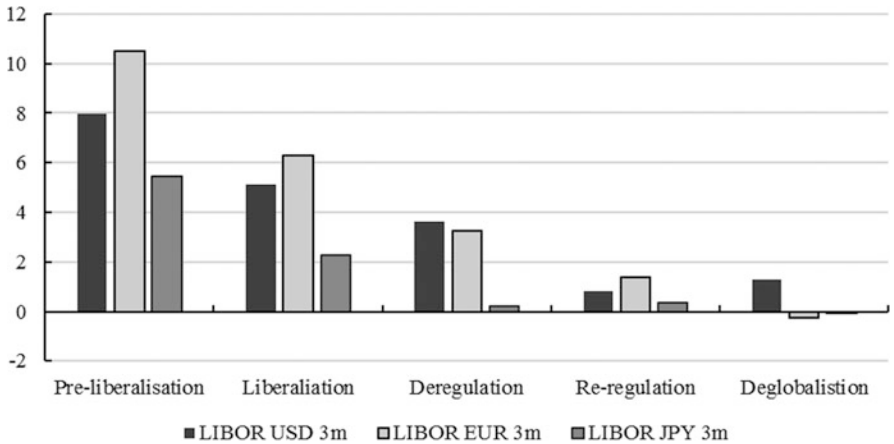


Fig. 3 Cost of liquidity. Pre-liberalization period covers data prior to 1991; Liberalization period—from 1991 to 1999; Deregulation covers 2000–2007; Re-regulation covers both GFC and period of resolution, including 2008–04.2014; Deglobalization trend covers from 05.2014 till now

institutions which made taking the excessive risks by financial institutions more costly (for more details see PWC' banking regulations navigator 2016, 2018).

During this period the authorities agreed to establish new amendments to the Basel agreement, build up new measures of financial stability and introduced the role of the so-called Globally important financial banks (GSIBs), which was supposed to become a subject of special (more severe) regulation even under "normal" market conditions. In exchange for this additional capital buffer,[1] the regulators agreed to provide embedded "insurance" to clients of such institutions. But the role of financial institutions in that period was rather subordinate—they had to agree on the rules of the regulators. Yet, only a few years after the crisis, some of the financial institutions started repaying their debts to governments, and so their independence from regulators started to grow.

Hence following the first "repayment" of the equity injection back to the authorities, there was a continuing fight over the influence and the agenda between regulators and financial institutions. The bottom-line might be that the former generally defended their positions, although banks managed to restore a part of their influence, which was previously lost following 2007–2008 crisis.

2.4 Deglobalization Trends: New Challenges for Doing Banking in the World

In 2014, the financial world faced a new reality, generally characterized by several new trends. The first was the beginning of the Fed's policy normalization: it signaled an intention to initiate rate hikes in the foreseeable future. This created a wave of risk-off in the emerging markets, and this additionally increased cost of capital for financial institutions all over the world. While the situation was more pronounced in the emerging economies, the valuations of banks started to decline over the years.

The second was an imposition of economic and financial sanctions on Russia—the first time economic restrictions were applied against a country of that size since the fall of the Soviet Union. Although it was not supposed to create an effect on financial institutions except for the targeted entities, applied measures established a new and more risky set-up for global financial institutions. The three following types of restriction measures were imposed:

- Firstly, a denial to continue the process of signing intergovernmental agreements, including FATCA. At the same time the responsibility of financial institutions for non-compliance was not canceled, creating grounds for fines and penalties.
- Secondly, a downgrade of the credit rating of the sovereign debt from investment to junk level, which also caused downgrades of the credit quality outlook for financial institutions in the country. This measure on its own might not be

[1] Amounting up to 2.5% of the risk-weighted assets.

considered as initiated by the authorities, more likely it reflected the credibility of the risk of the potential harm on financial institutions that foreign authorities threatened to bring.

- Finally, direct prohibition of the US and EU authorities to provide funding to a list of Russian entities. Initially it was prohibited to provide funding for more than 90 days to the sanctioned entities, later on the maximum term was decreased to 30 days, now it is only 14 days. There were even attempts to cut Russia out of the SWIFT international payment network and to disconnect its access to a payment system in USD.

New technologies substantially decreased the cost of collecting and analyzing data, including past data. That became the reason why breaching sanctions become costly, and authorities become highly credible in their threats to punish wrong-doings. For instance, large financial institutions were severely fined for participation in earlier sanctioned activities even before 2014, although in some cases banks had exited from those operations long ago (see Fig. 4).

The third trend does not have an exact start date, as it covers not a legal action or a policy decision, but a set of changes in approaches towards doing business. This trend is de-globalization. While this is still an arguable point how noticeable this will be, however, a set of effects cannot be ignored anymore—especially after new agreements are in place. The two most prominent aspects of this world-scale trend are Brexit and US–China trade conflict.

Brexit is a name of process of United Kingdom's leave of European Union. It started in 2016 with a referendum, which unexpectedly showed that the UK citizens want to leave the EU. This caused a period of turbulence in political and economic spheres, however, one of the most important aspects of that was the future of London as a financial center. After decades of integration many European banks had their

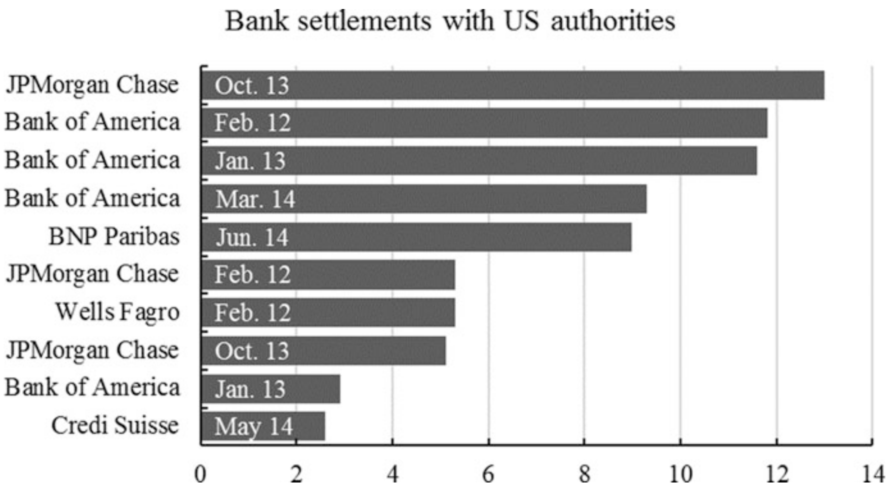


Fig. 4 Fines of selected financial institutions during 2006–2018

representative offices and divisions (mostly related to investment banking and capital markets) in London, but with a new regime of relations between the UK and EU lots of organizational issues required completely new approaches, including tax, legal, immigration, etc. Obviously, this will have a pro-longed effect on both financial and other businesses in Europe and many other jurisdictions, as (1) competition between UK and Frankfurt-on-Main for provision of funding for European financial institutions and emerging markets will increase and (2) this has demonstrated that integration process might be reversed.

The other prominent economic conflict of the recent years is the US–China trade conflict, which started in July 2018 with imposition of tariffs for imported goods. While in the beginning of the conflict the amount of goods under tariffs was rather small as compared to total trade turnover, through time and escalation the majority of goods traded became under new, higher tariffs. Many factors were named among reasons behind this conflict, but it increased uncertainty in global financial markets tremendously.

Apart from mentioned conflicts, there were a number of other tensions between countries in economic and political areas resulting in a change of regime of investment and trade, thus affecting the financial markets and their participants (signing and break of the Iranian nuclear deal, conflicts in Syria and Lybia, etc.). Altogether that marked the appearance of a new era for global financial institutions: an era of credit rationing, open and hidden restrictions. All this will shape the face of the global banking system in the 2020s.

3 Banking Systems in Emerging Markets at the Start of Twenty-First Century

3.1 CEE Region

At the start of the century, the banking sector in the developing CEE countries hits the fairway of the European integration process. This process had two aspects: one is the imposition of the Euro in the selected countries which satisfied Maastricht's criteria, and the second—the embedded economic and financial integration of the CEE countries with Western Europe.

On January 1 2002, the Euro as a currency became the official currency in the EU. Dermine (2002) reported that "A single currency in Europe changes fundamentally the competitive structure of the corporate bond and equity markets, since one key-source of competitive advantage, namely home currency, disappears. Indeed, savers will diversify their portfolio across European markets." He suggested that consolidation trends will accelerate in many sectors, including the financial sector. Large-scale businesses will rely on top level expertise and will require a large-scale banks which can compete globally (and so might have an access to the best-in-class financial technologies) and also might ensure optimal in terms of price and efforts ways to attract and place funds. In this case the key competitive advantages for local financial institutions will be knowledge of local client base—established relationship and deep understanding of business risks. This will be directly related to long-term exposure into the accounting, legal, tax, language, and cultural environment. Similar conclusions might be also found in Buch and Heinrich (2003).

Creation of the monetary union required certain unification of the economic parameters across member states, including inflation. This might create a significant pressure on banks in countries where prices growth and relatively high interest rates have created significant interest margins. Doing banking business in a low inflation environment would require from banks higher efficiency in both operational and risk management.

The proposed Financial Services Action Plan (1999–2005) contained initiatives which will ensure integration of banking and capital markets. A dozen of initiatives for the wholesale market, and five actions for the retail market were approved and monitored to fulfill the plan, whose main objectives were as follows:

- to ensure competitive and secure retail banking & insurance markets;
- to establish a single EU wholesale market for funding; and,
- to develop proper prudential regulation and supervision ensuring adequate risk culture.

Obviously, European financial institutions gained the most from these trends, especially provided that the cost of funding was declining. These effects started to appear in eastern European countries, one of the most illustrative examples is the banking sector evolution in Poland.

Its banking sector survived a dramatic transformation prior to the start of the century. A good description of this period and developments in the Polish banking sector are in Wiesiołek and Tymoczko (2015). In the mid-1990s, more than 60% of the banking sector was owned by the state, domestic private investors were quite active (increasing the share to about 30% of the sector in late 1990s), and foreign banks entered the Polish market to gain a substantial market share. This plan was successfully realized in the early 2000s, so the market share of foreign-owned banks reached about 2/3 of the sector (see Table 1). What is highly specific about Poland is that most foreign investors were from Europe, which at that period was directed to unify and develop further the European Union.

Table 1 Evolution of ownership in the banking sector in Croatia, mid-1990–2000s. Source: Kraft et al. (2002)

	Type	1994–1995	1996–1997	1998–1999	2000
Number of banks	State	20	8	9	3
	Private	31	44	36	20
	Foreign	1	6	12	20
Share in total assets	State	52.0	36.2	38.8	6.1
	Private	47.0	61.3	37.9	10.2
	Foreign	1.0	2.5	23.3	83.7

Another prominent example of the stance and development of the banking sector in that region is Croatia. Being much smaller than the Polish banking system, it still had many common features of evolution. Indeed, at that time the dominant paradigm was that state-owned banks are overstaffed, poorly equipped technically and reluctant to adopt banking innovations (EBRD 1998), Hungary was among the first to privatize its banking system via selling large portions of its assets to foreign banks. In particular, while in 1994, out of 50 banks operating in Croatia, only one was owned by a foreign investor and 26 belonged to the state. In the early 2000s, there were only three state-owned banks and the total number of foreign-owned credit institutions reached 20 (out of 43, implying that local players were under severe pressure and had difficulties competing with foreigners). Even more important is that the ownership of banking sector assets evolved more dramatically: foreign investors held more than 80% of the assets in 2000. One of the conclusions of the paper that summarizes the development of the Croatian banking sector over that period was that “liberalization in the form of opening the banking market to all comers is not an especially productive exercise” (Kraft et al. 2002), although one of the findings was that “reputable foreign banks do seem to have strong efficiency advantages.”

Overall, one might conclude that the dominating trend in the region at that time was the trend to invite foreign banks in order to increase the provision of financial services.

3.2 CIS Region

The main trend in the CIS countries in 1990s, which predetermined the structure and perspectives of the financial sector, was a transformation of the newly established economies from centrally planned into market-driven. In the Soviet Union the number of banks in the country was limited—only the so-called specialized banks existed, each of which concentrated on performing its function, e.g. servicing transactions of foreign trade, providing funds for construction or agricultural sector, or working capital to industrial sector, or pooling savings of the population. However, when the USSR collapsed, economies were in need to build up their own financial systems. So depending on a country-specific culture, legislation and local economic conditions entrepreneurs received an opportunity to establish new financial institutions. For instance, in Russia the number of credit institutions in early 1990s jumped to 2,5 thousand, in many other CIS countries total number of institutions was by far lower—only several dozens. However, the collapse of the economic structure also changed the real sector of the economy: it became almost paralyzed. Low payment discipline, inappropriate skills to do marketing and sell products, broken trade links, and low competitiveness of the production—all that strongly discouraged financial institutions from traditional banking operations, e.g. lending. Hence activities of banking institutions were more focused on operations in financial markets.

Trofimova (2005) reported that the relatively small size of banking systems and the growing demand for credit resources from both legal entities and individuals predefined the fact that in CIS countries banking systems growth rates were generally higher than in their respective economies.

World leading think tanks and development institutions reported that after the crisis of 1998 in Russia the overall resilience of CIS economies has increased, albeit mostly driven by developments in the energy sector. However, in what concerns banking sector “An improved investment environment needs to be accompanied by stronger banking systems—the main source of investment finance—which remain generally weak and underdeveloped” (International Monetary Fund 2004).

In the CIS the largest economy is Russia. This explains why its financial system is by far the largest among the post-USSR countries. Yet at the turn of the century Russian banking system was in a very difficult situation. First, the country has just survived the sovereign debt, banking, and currency crisis of 1998. Depreciation created new opportunities for business, and economic growth was accelerating, but the trust to the financial system was low. Second, the interest rates in the economy were enormously high: even when the CBR rate decreased from their crisis peaks, it still exceeded 50% year-on-year, which was driven by high double-digit inflation. In addition, the number of credit institutions in late 1990s was sharply declining—more than 100 licenses per year (see Table 2). This was driven by two main trends: in late 1990s Bank of Russia was regularly revoking licenses of credit institutions, and in early 2000s the regulations started to improve and so attractiveness of providing financial services declined for those institutions which did not have sustainable and diversified client base.

Crisis of 1998 brought a significant change in the paradigm of the banking sector development in Russia. The government’s support following the crisis increased the role of the state-owned banks in the sector. State participation became significant in Russian banking system, and has been increasing further in the subsequent years. Worth mentioning that the foreign ownership in last years of 1990s was rather low, however, improvement of the regulations started to bring additional attractiveness for foreign investors. Concentration in the sector became large already at the turn of the century. The banking sector was also highly concentrated with five banks holding about 73% of the total assets of the sector.

The situation in other CIS economies was, in certain aspects, similar, but in others—pretty different from that in Russia. For instance, while in Russia the total number of banks exceeded 2000 at the beginning of 2000s, in Belarus the number of credit institutions was 28. However, like in Russia, the concentration was large—seven banks dominated the financial system, with share in the capital of the banking

Table 2 Number of banking institutions in Russia at the turn of the century

Years	1995	1996	1997	1998	1999	2000	2001	2002
Number of credit institutions registered by CBR	n.a.	2589	2562	2481	2376	2124	2004	1826
Change	n.a.	n.a.	−27	−81	−105	−252	−120	−178
Revoked licenses	216	275	329	227	127	33	12	10
Credit institutions liquidated	n.a.	n.a.	52	73	100	258	144	216

sector of ca. 80%, and in assets—at ca. 65%. Share of foreign banks in Belarus was also rather small—at 4%. Belarus initiated an improvement in banking supervision following crisis of 1997–1998 in Eastern Asia and Russia as well.

Banking system of Ukraine looked rather solid at early 2000s: with capital adequacy at above 16% (in 1999 it exceeded 20%[2]), and growth rate of loans at ca. 45–60% year-on-year. Yet concentration was also pretty high—for top 3–5 institutions it exceeded 80–90%, despite the fact that number of banks exceeded 150 in early 2000s. Worth mentioning that then the share started to fall, and competition has increased, which was presumably driven by foreign banks—they has a significant presence in Ukraine already at that times—about 15% of total banks registered in the country.

Kazakhstan's banking system was also pretty similar. While sector concentration was very high in the last years of 1990s (about 70–80%), this is likely to be related to small number of credit institutions—about few dozens. However, at that time a distinctive feature of Kazakh banking system was a relatively large share of foreign participation. Both the number of foreign banks was large—up to 35%, and their role in the sector was also noticeable, while the government influence was rather limited (Federal Reserve Bank of St Louis, Economic Research 2017; Kapparov 2018).

3.3 Banking Sector KPIs in EMs at a Glance

Banking sector ROE in CIS and CEE countries offered beneficial opportunities to investors (both private domestic and foreign): the indicator rarely fell below 15–18%. Returns on capital in developed economies were also rather high, and so high ROEs in EMs were considered as a sufficient condition to invest into the banking sector in these economies, assuming that market potential is large enough. Moreover, risk-reward ratio of doing business in these regions was considered as reasonably attractive due to high ROEs. At the same time, early stages of development of financial services could bring enormous opportunities to pioneering banks: for instance, those financial institutions which first entered retail lending market in Russia not only gained market share in the system for at least 15 years, but also could earn supernormal return, exceeding 40–50% a year.

According to Trofimova (2005), due to the infancy of capital markets in this region, banks in late 1990s remained the key financial intermediaries, but the level of financial intermediation in the CIS was low in comparison with some other developing economies, for example, with Central and Eastern Europe (CEE). Data suggests that while credit to GDP in CIS economies was in the range of 15–20%, the figure for CEE countries on average exceeded 30%. This implied that there had been a considerable potential for growth.

[2]Source: https://fred.stlouisfed.org/series/DDSI03UAA156NWDB, authors' calculations.

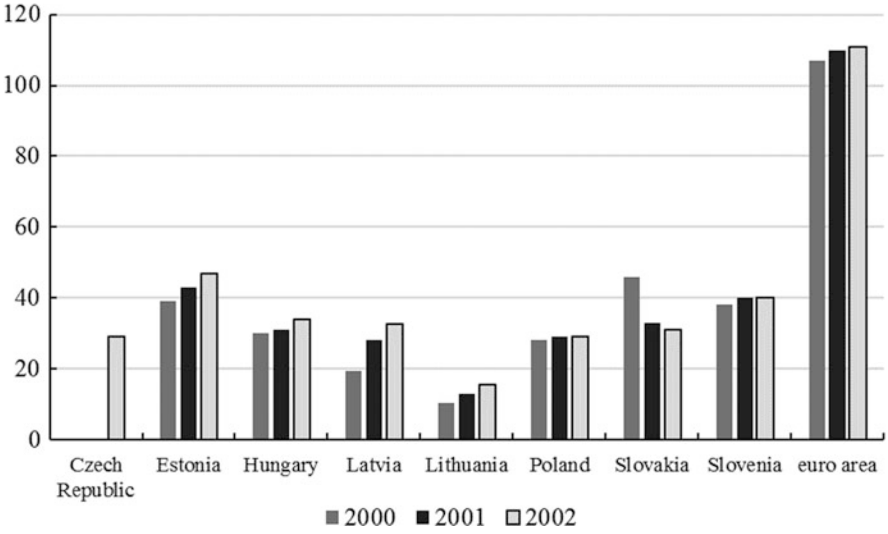


Fig. 5 Private sector Credit-to-GDP ratios, 1999–2004

At the same time, credit-to-GDP in CEE countries was, in turn, by far lower than that in Western Europe (see Fig. 5). Provided that Baltic countries were attempting to integrate into the European financial system as fast as possible, these economies had experienced higher growth rates of loans: credit volume had increased more than 1,5×–2,5× during early 2000s, but even then it was undoubtedly underpenetrated region for banking services.

Hence at the start of the twenty-first century markets in emerging economies promised good opportunities for investing into their financial sectors.

4 Evolution of Key Banking Metrics and Financial Services' Saturation

European Banking Federation for assessment of the banking sector performance relies on a set of variables, which include in the first place bank capital adequacy and funding sustainability, while assigning secondary importance to assets volume and bank profitability. Yet this approach is valid when the evaluation is based on today's knowledge and regulation, which focuses attention at the financial stability due to the accumulated experience from the Global Financial Crisis. In this respect, authorities now behave more like debt holders.

However, if bank management emphasizes interests of shareholders, then the assessment of the financial sector should be based on a different set of indicators. Shareholders are interested in an upside in their stake; hence they are much more interested in growth, rather than in sustainability. In some sense, authorities in emerging economies are more likely to favor shareholder-like approach, rather than approach of debt holders. This is mostly because (according to one of the Nobel Prize winners), in emerging economies banks lead economic development. Hence high growth rates of financial sector might create hopes for acceleration in economy's growth rates as well. This approach requires a favorable investment regime for strategic investors.

Many strategic investors assume that financial control and risk management policies will be imported from the holding company, hence they focus on returns and potential for future growth (market share, etc.). Having this in minds, the following characteristics seem to be the most important to describe the banking sector performance in the twenty-first century in emerging countries:

- return on capital or equity;
- market size growth rates;
- various metrics of bank services penetration, out of which the most important are market size relative to GDP and number of branches per population: the first grasps the financial depth and ability to fulfill the market with loans, while the second used to be a strong proxy to measure market capacity to generate funding.

These measures do not provide a full picture of the sector, especially in times of difficulties, however, the globalization trend, that has prevailed during the last decades, highlights the need to concentrate on potential of the market, as convergence theory suggests.

4.1 Return on Equity

Return on equity is the most general measure of the performance of any institution. For banks it summarizes not only operational revenues and costs, but also includes the cost of risk, which is embedded into the banking business.

Kohlscheen (2018) reported dynamics and determinants of ROE in emerging markets for the period of 2002–2014. This period covers more than a half of the time horizon of this study. Figure 6 summarizes developments in ROA and ROE in emerging markets during this period. It clearly shows that from the beginning of the century the average profitability of banks in emerging markets was declining.

The main reason for this was increasing market competition and elevated pressure from regulations, and this is fully in line with the trends in the developed markets as well. Worth mentioning that some papers resulted in different conclusions (e.g. Căpraru and Ihnatov 2014) upon investigation into determinants of banks' profitability in selected CEE countries over the similar period (2004–2011). The sample contained 143 commercial banks from Romania, Hungary, Poland, Czech Republic, and Bulgaria. Authors concluded that management efficiency and capital adequacy growth influence bank profitability for all performance proxies. A policy recommendation for authorities was to improve supervision for credit risk and capital adequacy, which would drive profitability up.

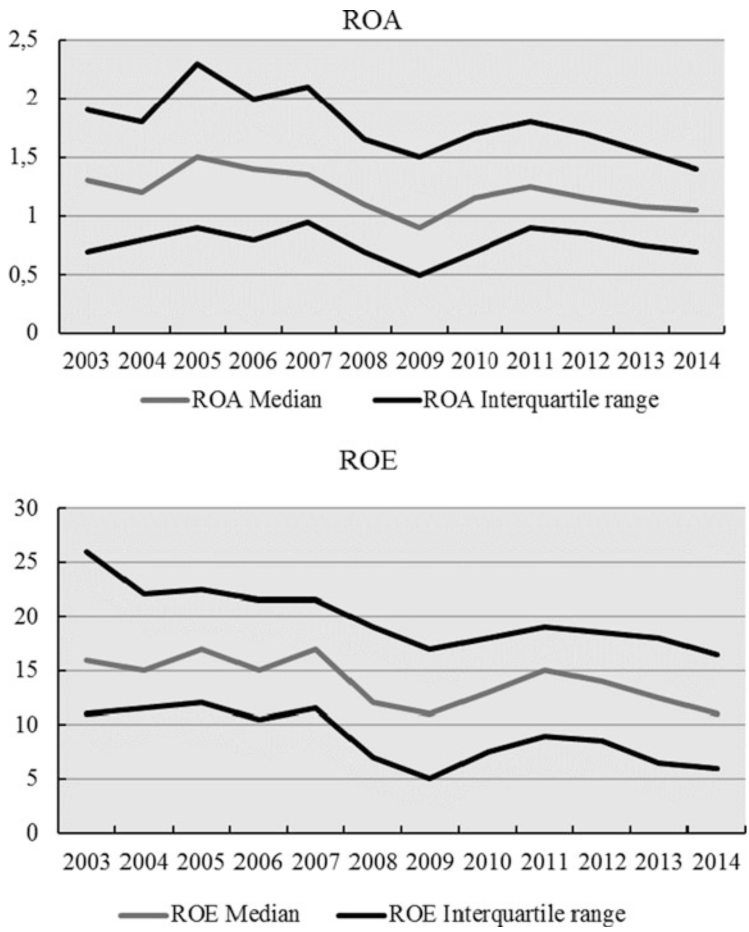


Fig. 6 Return on assets and on equity in emerging markets in 2003–2014. Source: Kohlscheen et al. (2018)

However, although the overall profitability in the sector following the period of re-regulation stayed under pressure, a very different situation was in CEE economies. In this region ROE was improving in recent years (see Fig. 7) thanks to very positive economic conditions: high growth rates across the region, low inflation, and improving ties with trading partners amid growing global demand. Worth mentioning that to some extent such high ROEs were driven by a recovery from the 2011–2012 crisis in Western Europe, which cost CEE countries a lot due to FX devaluations and other reasons.

This suggests that banking sector's profitability in different countries might become less uniform than in the period of re-regulation and before. At the same time, future financial attractiveness of banks in these countries will more and more depend on individual business models.

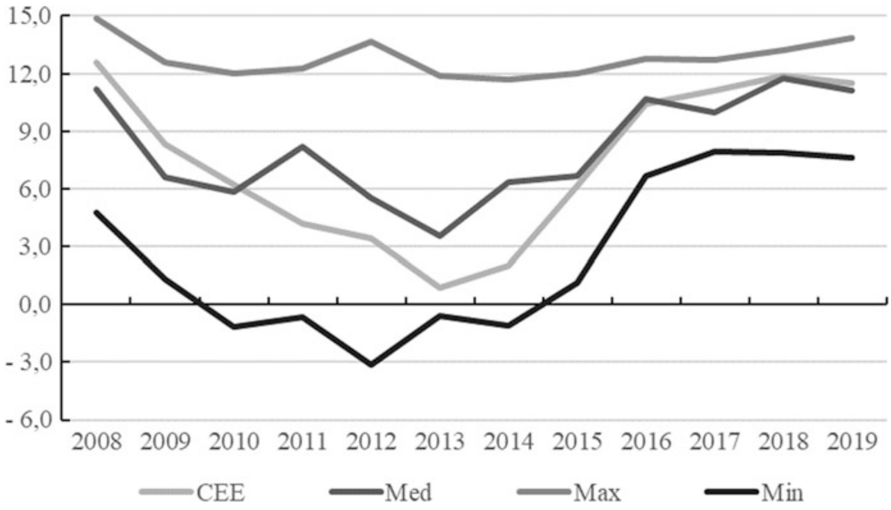


Fig. 7 Return on equity in CEE following GFC. Source: Bloomberg, authors' calculations

4.2 Banking Services Penetration

Banking sector assets growth in emerging countries in twenty-first century was impressive during the first decade. This had a number or reasons behind, including low cost of external financing, cross-border mergers, convergence with the developed economies, high optimism about globalization of banking services, etc. Provided a rather dynamic growth of economies and low saturation of markets for financial services at the turn of the century, growth of banking business was impressive, sometimes showing double-digits.

Following recession in key economies in times of the Global Financial Crisis and sometimes afterwards, lending fell down in many countries, including emerging markets. However, the recovery was uneven. Indeed, while the weighted average in most of the years exceeded 10%, the median growth was in the range of 3–7% per annum. While the impact of Russian and Turkish markets had the largest effect on the aggregated average, almost each year there was a market in the region where loans portfolio was shrinking (see Fig. 8).

However, generally in EMs growth of traditional banking products slowed down for several reasons. First, investors' appetite to add more capital for their banking business expansion has declined following the crisis of 2008–2009. This was additionally fueled by the increased regulatory pressure. Secondly, after a decade of decent growth the inconsistency between economic and credit developments became too high for too long, and natural deceleration became reasonable: high-quality clients had already been supplied with banking services. Also, a growing impact of technological developments on the banking business created a new wave of

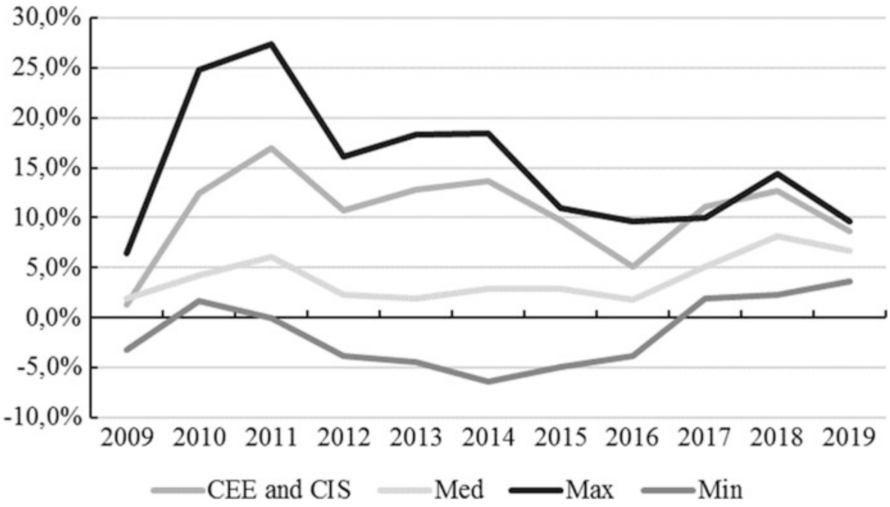


Fig. 8 Growth rates of loans portfolio in CEE and CIS countries following GFC. Data was cleaned from outliers in each of the year, authors' calculations based on central banks data

disintermediation due to a direct access of borrowers to crowd-funding platforms, capital markets, etc.

4.3 Assets-to-GDP in Different Countries in 2000, 2010, and the Most Recent

Another important measure of market potential is the degree of penetration of banking services. This is usually measured by assets-to-GDP ratio. One of the pros of this measure is that all banking systems might be compared on its basis—both through time and between the peering countries. Over the last decades, according to the International Monetary Fund, this ratio has considerably increased: from 45.75% for all the countries in the world in 2000 to about 58.6% in 2010. Despite all restrictions and regulations, the indicator has increased further to reach 65.6% in 2016. Moreover, in some countries this ratio by far exceeds country's GDP: sometimes this is due to a country's role as the international financial or trade center, in other cases—this is a result of improper regulations. The list of countries with the highest assets-to-GDP ratio consists of both developed and developing economies from all the continents.

Market saturation has increased in all the regions during twenty-first century (see Fig. 9), however, there is no uniformity in that development. For instance, in "emerging economies" all the groups of countries have demonstrated improvements during both decades. However, economic growth in Asian and Latin American

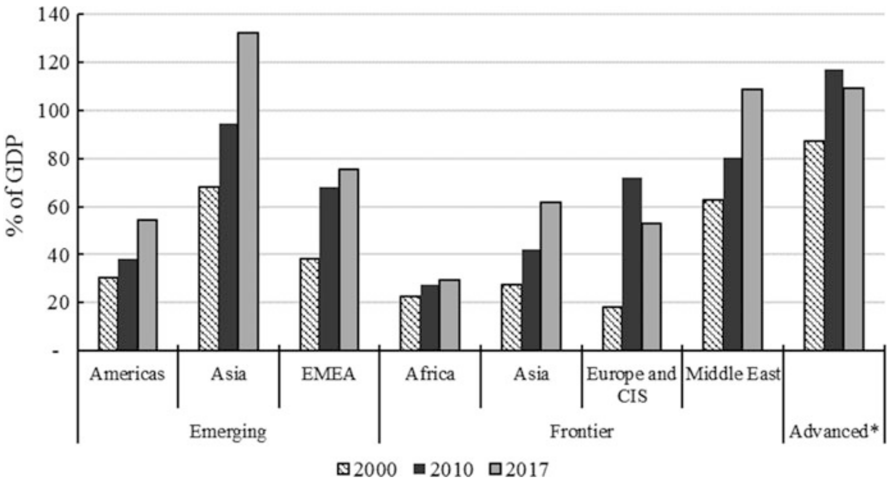


Fig. 9 Banking markets saturation in different types of economies (median)

countries following GFC exceeded growth in EMEA region, while attention to risk management in the latter was relatively higher given the influence of Western banks. In "frontier economies" the two outliers are African countries and Europe and CIS. The former shows satiation is increasing too slow presumably due to low literacy, while in the latter region balance sheets of the banks are downscaling.

Summing up the results of development of banking systems in the world in terms of bank assets to GDP ratio, the regions are very different in what concerns degree of reliance on the intermediation (see Fig. 10). First, there are obviously underbanked economies (mostly in Africa), frontier markets[3] in CIS and Europe also might provide some opportunities for growth in the future. However, both types of the market presumably lack institutional environment for doing banking business. Second, most of the markets are likely to stabilize in the range between 75% and 100% of GDP, hence completing the convergence. Notable exceptions in Asia (China, S. Korea, others) might be explained with reference to specific institutional factors of the region.

4.4 Bank Branches Per 100K Population

The financial access survey (IMF data base, n.d.) revealed that higher-income countries are much better equipped with banking branches than poorer counties. For instance, banks in OECD countries have 1.33× more branches per 100K population than even upper-middle income countries. This obviously reflects

[3]Frontier markets are the least advanced economies in the developing world.

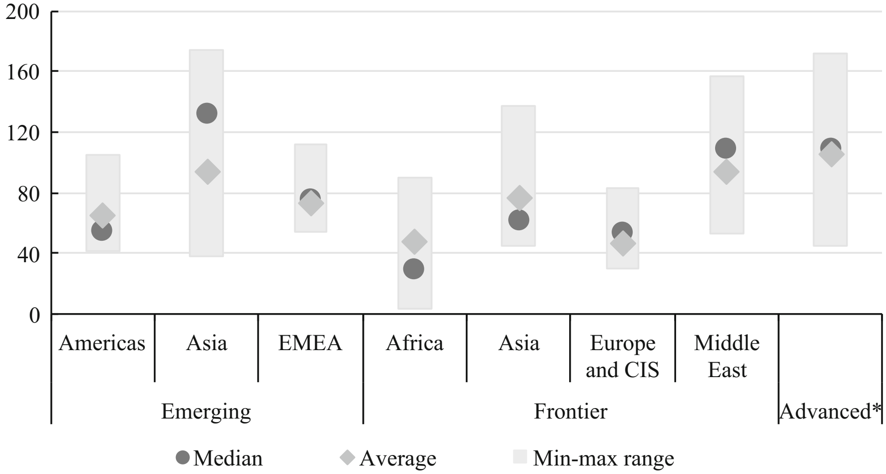


Fig. 10 Bank assets to GDP in different regions in 2017. Source: MSCI classification, data from TheGlobalEconomy.com, authors' calculations. *excluding Hong Kong, where Assets-to-GDP ratio exceeds 250% (The Global Economy n.d.)

differences in how bank branch networks in those countries had been developing, which includes, among others, not only the degree of market maturity, but also amount of investments into physical premises.

However, the trends in those markets are different in nature: while the developing countries are trying to increase the number of physical branches per capita, the most developed countries seem to rely more and more on the modern technological shifts (see Fig. 11). Compared to the mid-2000s, the number of physical branches in OECD countries has decreased by more than 25%. However, it seems doubtful that this has caused major issues for their clients, which seem to be used to operating with mobile and on-line banking services (see Box 1).

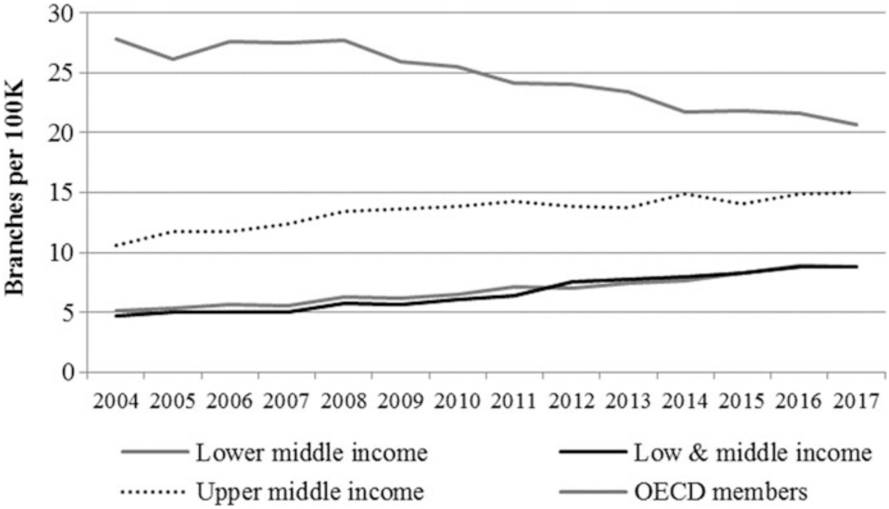


Fig. 11 Development of financial services provisioning in different countries: branches per 100K. Data are shown as the number of branches of commercial banks for every 100,000 adults in the reporting country. It is calculated as (number of institutions + number of branches)*100,000/adult population in the reporting country. Source: IMF's "Financial Access Survey", author's calculations

Box 1 Case of Tinkoff Credit Systems

Oleg Tinkov establishes Tinkoff Credit Systems Bank in late 2006, its business model for a banking institution was new. TCS' approaches to risk management, product marketing and delivery, and other business aspects differed substantially from peers from the very beginning of the project. Yet, the most important feature is huge investments in IT that the founder and co-investors directed to build up a highly differentiable bank. Its Internet bank was launched in 2008 (almost pioneering the market), mobile bank—in 2011, thus allowing to acquire majority of customers via on-line channels.

Having successfully launched on-line lending program this bank managed to persuade its customers that its on-line deposits are also safe and very convenient. It had not stopped its development and gaining market share even in the most difficult times of the last decade, and now this is the second largest bank in Russia in terms of credit cards issued. It is the only fully on-line pledged credit institution in emerging markets and one of the most IT developed companies in the world. On the top of that, worth mentioning that Tinkoff Bank was named among The Banker's Most Profitable CEE Banks in 2019.

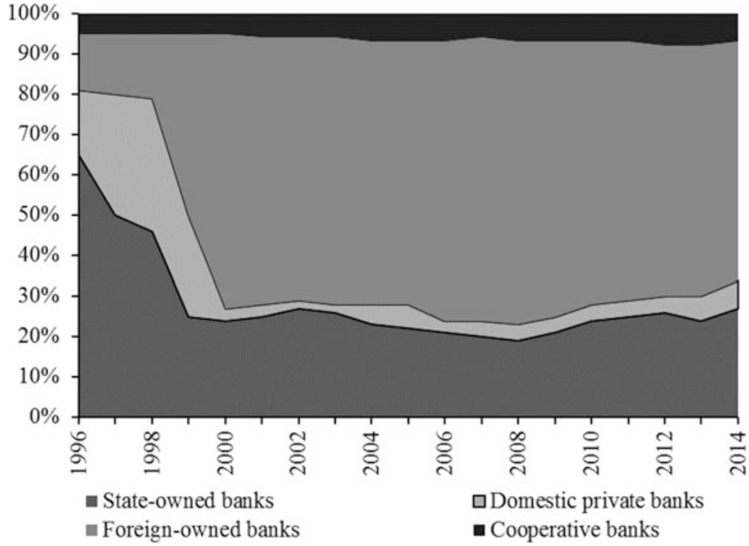


Fig. 12 Ownership structure of the Polish banking sector. Source: Wiesiołek and Tymoczko (2015)

4.5 *Cooperation with the Foreign Strategic Investors*

However, this era of cooperation between local players and global banks did not last too long. One of the most prominent examples is Eastern Europe. While at the turn of the century EU member states expected to receive multiple benefits from consolidation in Europe, thus fully supporting the cross-border mergers, the Global Financial Crisis changed the pattern of behaviors. For instance, provided a substantial hit that the largest international banks got from the regulations following losses during GFC, local banks increased their degree of autonomy—in cases they proved to be financially viable and independent despite affiliation with GSIBs. Also, local monetary authorities learnt from the ECB's responses to crisis that centralized response to an external shock might have asymmetrical impacts on member states. This substantially undermined incentives for closer cross-border integration of financial systems in countries where such integration was at relatively early stage, e.g. Poland (see Fig. 12). The country decided to keep its currency and started to protect national financial system with new legislation. So the role of international banks in Poland started to decline.

Box 2 Foreign Banks Exit S. Korean Market[4]

Indian Overseas Bank (IOB) entered the South Korean market in 1977, but recently announced that it closes its branch in this country. Although it is a state-owned entity, it operates just like a private institution. So, during the period of rapid growth of the Korean economy, the bank used to demonstrate solid returns due to the high interest margin. However, a decrease in profitability due to tightening regulation, low growth and low interest rates after the Global Financial Crisis of 2008–2009 brought IOB to leave South Korea after four decades of business.

Despite intentions of the Korean government to develop a financial center in the country, many foreign banks are closing their branches and representative offices, and the IOB is just one example of many. In 2017 Goldman Sachs closed its branch, Royal Bank of Scotland, Barclays, UBS, and others followed. Moreover, it is not only Western banks, but also Asian banks (e.g. Macquarie Bank) are leaving the market.

Foreign banks used to earn in corporate lending and derivatives, but recently have lost competitive advantages in these segments due to financial innovations, development of the global capital markets, and tightened regulation. Banks are increasingly viewed by regulators as local utilities, and their products and services similar to "public sector goods," while accumulated expertise and high value added services are not enough rewarded anymore, thus discouraging financial institutions to do business in multiple locations (Yoon 2019).

4.6 Priority Reversal: From Market Potential to Risk

Given the trends mentioned above there should be no surprise that the focus of attention in what concerns doing banking in emerging markets started to shift from highlighting market potential to detailed evaluation of risks. At first glance the portfolio quality in EMEA countries should not be a cause for concern: share of non-performing loans in total portfolio seems is stable, moreover, in last years of 2010s there is a clear trend to resolve the worst cases (see Fig. 13). However, management of the banking institutions still evaluates cost of risk at levels similar to periods of tensions (see Fig. 14). Provided that technological innovations help to calculate risks better than before, this highlights declining tolerance towards financial losses. Hence unlike in the past, risk culture and instruments to measure and manage risks will drive appetite to invest in banking institutions in EMs.

[4]Source: http://www.businesskorea.co.kr/news/articleView.html?idxno=33055. Accessed 15.03.2020.

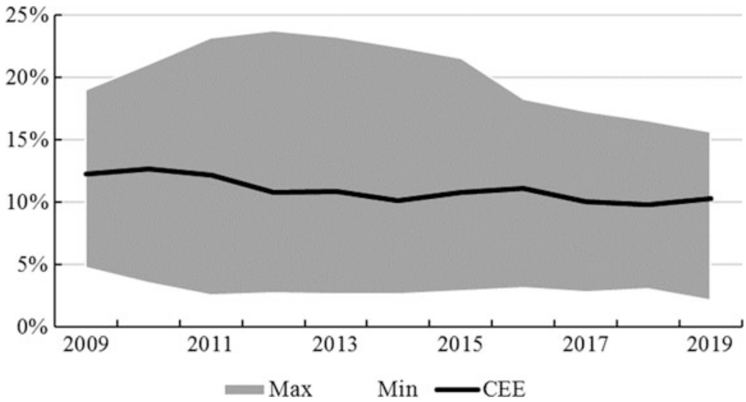


Fig. 13 Non-performing loans in EMEA economies. Source: national central banks, authors' calculations

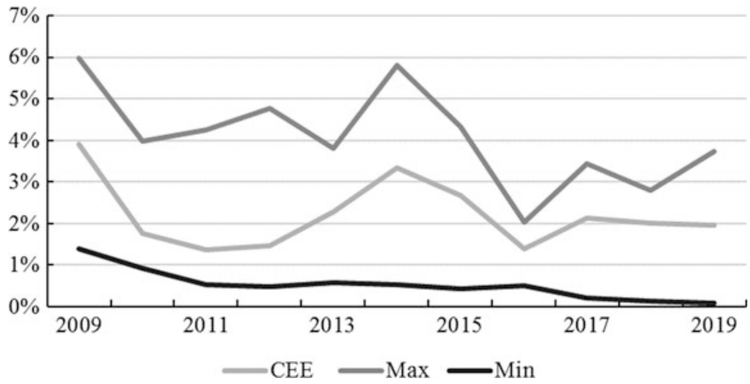


Fig. 14 Cost of Risk in CEE banking systems. Source: national central banks, authors' calculations

5 Doing Banking in Emerging Markets: New Challenges

5.1 Growth of Macroeconomic Risks

One of the most significant challenges for the banking industry in the upcoming years lies in the area of macroeconomic developments. In particular, the following trends seem to be important: firstly, a slowdown in global growth, which will impact both emerging and developed economies, and secondly, further increase in probability of elevated tensions between countries.

A global recession in 2020 was caused by a "black swan"—pandemic of coronavirus disease (CoViD-19) all over the world. In order to contain its outbreak, authorities tried to limit economic activity banning tourism, international Avia transportation, public events, etc., which caused severe discontinuities in economic

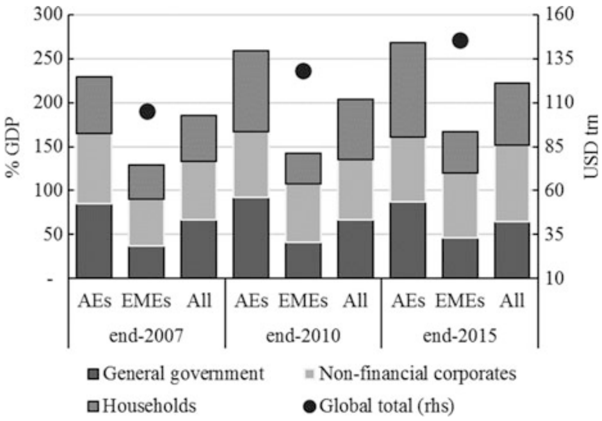


Fig. 15 Global debt metrics. Source: Bank for International Settlements

links between subjects. However, a slowdown in global growth had been expected in advance by leading think tanks on the basis of the following considerations.

Firstly, elevated global credit risks. During the last decades, the growth rate in key economic regions—USA, Europe, and Asia—was driven by relatively cheap and available credit resources (see Fig. 15) as well as fiscal stimulation. Currently, the overall indebtedness is high. For instance, no European country currently satisfies the Maastricht criteria. Or, in the USA, companies have doubled their debt since 2008, and the risky segment (with rating at BBB- or lower) has added the most. A similar problem is in China, where the debt overhang is significant—both in the public and private sectors.

Secondly, limited potential to stimulate the economic development with monetary policy instruments. Despite promises to normalize monetary policy, in during the second half of 2010s world rates were still much lower than those before the GFC and deregulation period. In the European Monetary Union the deposit rates were even negative. Hence, in the case of growth deceleration below potential it is doubtful that monetary authorities would be capable of resolving the issues with ordinary instruments. Worth mentioning that zero growth rates in Europe and Japan did not result in acceleration of growth, which might imply that a substantial part of the world economy is in liquidity traps.

Thirdly, the world growth drivers start experiencing a very low return on additional resources employed. For instance, since 1981 the Chinese economy has grown 26 times, and most of the easily available resources have already been employed. The structural shifts in the Chinese labor force from rural to urban have largely been completed, and so demography will become the issue for the economy, and the eldering of the population will also add pressure. In the USA the problem is of the same nature: the available fixed capital no longer generates the return which is appropriate to shareholders, and so the attractiveness of new capital investments is low. Hence fixed investments are not enough to cover depreciation, so future growth is also questioned.

Fourthly, this situation of falling growth perspectives makes the world's leading economies intensify the tensions between countries. This trend can be observed in multiple different locations and situations, e.g. Brexit, the US–China trade conflict. The latter became much more evident in 2018 when trade tariffs were increased for a substantial share of traded goods and services. Even if an agreement is reached, this might not end this story: value chains are complicated, and some researches estimated Europe to become the biggest looser from the US–China trade conflict. One might add to this list conflicts between Russia and Ukraine, or Venezuela and its neighboring countries, India and Pakistan, etc. Hence the chances are growing that the world trade growth will slow, which is very negative for the financial sector in the world.

Finally, those conflicts would take specific forms: most likely, they would take form of a ban on certain activities, or at least a severe restriction. During past 30 years economic restrictions were applied to relatively small countries, i.e. Iran, or Iraq, but recently large emerging market countries started to face this situation. With IT developments authorities can control not only movements of physical goods, but also all financial flows. In case of tension this might create a threat to proper functioning of a supply chain. Such a risk would call for a restructuring of the entire system of trade and investment ties, thus reinforcing risks for banks.

5.2 Appearance of Technological Risks: Evolution of Banking Sector Functions

One might subdivide the developments of technological innovations for the banking sector into several eras. In the 1990s there was a shift from first web pages to e-commerce and payment processors. Standard business models were enhanced by representations in the Internet, and this caused the appearance of the first business-to-client models. For bank customers, it ultimately resulted in the appearance of the PayPal service, and Wells Fargo's on-line banking services in 1995.

In the 2000s, subjects started communicating with each other around the globe, and this caused the appearance of new media content and effect of sharing. This new type of impact was related to feed-back from consumers and bloggers which might ultimately impact production. At that stage, the level of integration of the businesses was rather limited and slow, mostly with supply-chains, although companies started to attempt integrating their ERP systems with each other via clouds. Eventually, it helped to establish a LEGO-type approach to doing business, including traditional. For instance, in the financial sector, new scoring models appeared which took information from social networks, payment processors, etc., and with a support of data scientists creating new business opportunities.

The 2010s resulted in a jump in customer expectations. Wide usage of smartphones created both a need and an opportunity to implement a One Window concept. Consumers prefer not to see intermediaries, they like mutually integrated eco-systems which provide services to clients in one interface via aggregating the microservices of their partners. This might be called the "Uberisation" of business. In the financial industry this started to erase the border between financial and non-financial services, and also between core and non-core banking expertise. Open API allows to outsource each and every function of a financial intermediary, and also to enrich available set of services with almost infinite number of microservices via nexus of partners.

The 2020s could herald the appearance and proliferation of artificial intelligence (AI). It might transform the financial services business model: for instance, new payment models will be required, including Machine-to-Machine payments based on the digital footprint of the owner; real time monitoring of pledged assets and gathering data from leased assets; consumer demand and overall business performance monitoring.

5.3 Changing Client Demands as Key Drivers for Banking Sector Developments

A detailed investigation of client needs for the next decade implies that the overall demands and required characteristics will be the best of two experiences: the ease of using retail banking, and the complexity and sophistication of digital banking. Changes in demands are well discussed in Kalara and Zhang (2018). However, the analysis of prospective clients' needs to imply that the bank counterparties will require the following set of characteristics:

- 24/7 availability of online services;
- Seamless digital servicing;
- Instant or very high speed of operations;
- High quality of service provided;
- Clarity and simplicity of products;
- A user-friendly, secure interface of Internet and mobile banking solutions;
- Cheapest banking products.

Communication with a client is moving to remote channels, and hence interaction becomes completely digitalized. Thus interface and IT solutions should be seamless for existing customer and easy to use. Moreover, bank-client communication channels should additionally provide the opportunity to discuss complex or confidential issues in person, by phone or in the office. The IT systems should be well-prepared to efficiently process large amounts of data, extract valuable insights from it, and predict customer behavior, as well as.

One of the possible aspects should be a close integration between Banks' and Clients'/Partners' IT systems that will bring new business opportunities. In particular, clients themselves will participate in the development of new banking products via marketplace with clients.

With that in mind, the borderline between the interface and the product is diffusing, while both are important for bank clients. Moreover, client's emotional responses to products may be more important, than the quality of the interface itself. Financial institutions might take the lead in developing various already existing, or emerging eco-systems (e.g.: Government-related transactions, personal disposable income transactions, or social media interactions). But the increasing number of such services will urge banks to arrange partnerships with service pillars. With partnerships growing, banks will be providing almost the same product range; however, the winner will be the one with most comfortable and user-friendly interface. On the one hand, this will allow the ecosystem owner to serve its clients exactly via a "One-window" approach, but on the other hand, bank will then become a "utility" company, while its customers will not bother on the service provider unless it works properly.

Thus banks of the future might change their role from a license holder to a financial advisor, with a best-in-class IT solution for clients' needs. Summing up, these challenges suggest that risks for financial stability are increasing, while regulation might become more and more complicated.

6 Concluding Comments

Global trends are likely to dominate development of financial systems in emerging economies. This is likely to be driven by their relative scale, although through time situation might change, as the example of China and its financial system teaches us. However, despite the global regime of financial intermediation—deregulation, re-regulation, de-globalization, etc.—banking systems in every country are likely to keep their own peculiar aspects and trends, which can help to differentiate it from others. For instance, banking systems in CEE economies' were sharpened by creation of European monetary union, while in CIS for decades the trends were related to the stance of Russian economy. Such developments create a strong case for localities in banking systems while keeping linkages with the rest of the world.

During the period of global financial system deregulation banking systems in the emerging economies were likely to be judged on the basis of growth potential and profitability. Cheap funding and friendly regulation in the developed markets created a strong basis for earning high ROEs, especially in low-competition markets, most of which were in developing countries. Upon market saturation and changes in regulation following GFC, priorities of investors might start changing, with greater focus on risk profile of economies. At the same time, technological shifts and deeper understanding on how top-class world banks operate increased confidence of local players. They started to catch the market share more aggressively. This predetermined the slowdown in convergence of banking technologies between emerging and developed markets.

So the future of the doing banking business in developing economies is less clear now, mostly thanks to technological developments (which are easy to copy, but difficult to implement in local markets) and evolving demand of the clients. These two trends exist in the context of macroeconomic developments which are characterized with an unprecedented level of uncertainty. It is rooted in the scale of imbalances in economies and unexpected occurrences which might change the whole landscape of business—both in financial and real sectors in an economy.

These developments might change the role that banks used play in economies over years—for instance, from financial intermediaries to financial advisors. To a large extent this will be determined with the role of regulation of managing financial risks. This factor of legal restrictions seems to be a key determinant of short-term future in this area.

References

Buch, C. M., & Heinrich, R. P. (2003). Financial integration in Europe and banking sector performance. In P. Cecchini, F. Heinemann, & M. Jopp (Eds.), *The incomplete European market for financial services. ZEW economic studies (publication series of the Centre for European Economic Research (ZEW), Mannheim, Germany)* (Vol. 19). Heidelberg: Physica.

Căpraru, B., & Ihnatov, I. (2014). *Banks' profitability in selected central and eastern European countries*. https://core.ac.uk/download/pdf/82771271.pdf

Dermine, J. (2002). *European banking: Past, present and future*. Second ECB central banking conference Frankfurt am Main 24 and 25 October 2002.

EBRD. (1998). *Transition report 1998: Financial sector in transition*. London: EBRD.

Federal Reserve Bank of St Louis, Economic Research. (2017). *Bank capital to total assets for Ukraine*. https://fred.stlouisfed.org/series/DDSI03UAA156NWDB

International Monetary Fund. (2004). *World economic outlook. Advancing structural reforms*. https://www.imf.org/external/pubs/ft/weo/2004/01/

International Monetary Fund. (n.d.). *Financial access survey*. https://data.imf.org/?sk=E5DCAB7E-A5CA-4892-A6EA-598B5463A34C

Kalara, N., & Zhang, L. (2018). *The changing landscape of firm financing in Europe, the United States and Japan, CPB Netherlands Bureau for economic policy analysis*.

Kapparov, K. (2018). *Financial inclusion and financial literacy in Kazakhstan*. Asian Development Bank Institute.

Kohlscheen, E., Murcia, A., & Contreras, J. (2018) *Determinants of bank profitability in emerging markets*. Bank for International Settlement Working Papers No 686.

Kraft, E., Hofler, R., & Payne, J. (2002). *Privatization, foreign bank entry and bank efficiency in Croatia: A fourier-flexible frontier cost function analysis*. Croatian National Bank, Working Paper №9/2002.

PWC. (2016). *Banking regulation navigator*.

PWC. (2018). *Banking regulation navigator*.

Stiglitz, J. E. (2010). *Freefall. Free markets and the sinking of the global economy*. Penguin Group.

The Global Economy. (n.d.). *Business and economic data for 200 countries*. https://www.TheGlobalEconomy.com

Trofimova, E. (2005). *CIS Banking systems: Different perspectives, common risks*. Kazakhstan Security Market.

Wiesiołek, P., & Tymoczko, D. (2015). *The evolution of banking sectors in Central and Eastern Europe – the case of Poland*. BIS Paper No. 83r.

Yoon, Y.-S. (2019). *Foreign banks leaving South Korea due to stronger financial regulations*. http://www.businesskorea.co.kr/news/articleView.html?idxno=33055

Regulation of Financial Risks in Emerging Markets: Past, Present, and Future

Artem Arkhipov, Natalia Arkhipova, and Alexander Karminsky

Abstract Regulation of risks in banking is driven by evolution of financial intermediation and markets, and vice versa. The study analyzes a changing nature of financial institutions' regulatory and supervisory trends in emerging markets over last 20 years, providing outlook for the future. Although the principles of the Basel Accord have long been the cornerstone of banking regulation in the world, precise requirements and scope were reformed and implemented in response to crises and global trends. At the turn of the century, the regulatory themes in EMs were focused on ensuring financial stability which was closely associated with regulatory and supervisory independence. However, the global financial crisis of 2008–2009 has changed the paradigm from partial improvements under financial liberalization regime to a world-wide regulation tightening on the basis of close coordination between regulators and supervisors in the world. The role of the G-20's Financial Stability Board was to ensure that initiatives are implemented globally, which further enhanced convergence of financial risks regulation in EMs and DMs. In recent years, that uniformity started to decline as the number of local peculiarities and initiatives impacting banking business increases: some countries eased or lifted certain globally accepted restrictions, yet imposing local regulations (including financial sanctions). Functioning of financial institutions in emerging markets becomes more and more complicated. Modern technological innovations enter spheres of compliance and supervision via RegTechs and SupTechs as a solution to this growing number of such inconsistences.

Keywords Regulation of financial risks · Banking regulation · Emerging markets

Jel Codes B26 · G21 · G28 · G32

A. Arkhipov · N. Arkhipova (✉) · A. Karminsky
National Research University Higher School of Economics, Faculty of Economic Science, Moscow, Russia
e-mail: narkhipova@hse.ru

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_2

1 Introductory Comments

One of the most important things in banking is trust. Common sense suggests that trust is difficult to earn, easy to lose, and even more difficult to regain. The issue is simple: it is about always doing the right things in the right way. It is about being committed to doing those right things, and here is the place for regulation. It works as an incentive-compatible constraint (unless it fails to prevent the loss of trust), aiming at banker conduct in such a way that the sector keeps the trust of its customers, thus ensuring a smooth development of the economy.

However, the problem is that in any given country, financial sector is a strategically important sector and this substantially limits political will for conducting reforms. Moreover, the diversity in bank definitions is gigantic: some countries do not prevent their banks from diversification, while others are severely restrictive. It was concluded that the regulatory environment, not just unfettered market forces, largely determined what banks in different countries around the world might do (Barth et al. 2001). Currently, in the age of increasingly growing financial innovation defining financial activities will become even more problematic.

This paper covers the evolution of regulatory themes over last 20 years, providing some outlook for the future. First, it describes the aspects of regulatory framework and topics at the turn of twenty-first century, effectively after the financial crisis in East Asia and Russia in 1997–1998. Then, it turns to main impact of the Global Financial Crisis of 2008 for regulation. In fact, this crisis effectively demonstrated what happens when there is a lack of coordination between different authorities in what concerns the regulation and supervision of the financial institutions. Then, the new Basel agreements are discussed as a response to the GFC, as is the creation of the supranational institution—Financial Stability Board, which was set to align regulatory bodies in the world in such a way that will help preventing regulatory arbitrage. Following these initiatives, the cost of capital and liquidity has changed, reflecting a higher regulatory burden and more lackluster business expectations. However, in the 2010s the world faced a new reality—the number of various trade and investment restrictions started to grow and the cost of breaching those sanctions started to increase exponentially. Hence this topic requires special attention, as it will sharpen the legal environment of the financial institutions in the future. The existence of those restrictions has created a substantial difference between the commonly recognized trends of globalizing the financial sector and the aligning regulations all over the world, and ensuring that restrictions imposed in certain countries become global practice without being adopted as part of the global regulation. And finally, we provide a discussion on modern themes related to the impacts that technological trends have on banking regulation, namely, regulatory and supervisory technologies, or “RegTech” and “SupTech.” These have already become a mainstream of approaching the regulatory frameworks for the doing banking globally.

2 Key Regulatory Themes in Banking at the Turn of Century

At the turn of the century concerns about safe banking had been widely shared by the financial community and the regulators. The laws had been changing, and efforts to step up prudential supervision had been under way. This was the legacy from two decades of banking and financial crises. Indeed, the harshness of the crisis in Eastern Asia and in Russia affected not only the financials of some institutions, but also posed a number of fundamental questions. For instance, a failure of Long-Term Capital Management company, which used the most advanced (at the times) techniques of investments, made regulators start thinking that in certain circumstances the room for market failures is bigger than one could expect. Another open issue was whether abovementioned crises were caused by local mismanagement in the respective countries, or they have more in common than it seems.

After having studied various literature, the three regulatory themes were identified as dominating at that times. The first is the application and role of Basel Accord in avoiding banking crises. The other is the regulatory independence: how to organize and operate regulatory and supervisory bodies to minimize adverse effects. And third is emerging issues of financial stability: how to design incentive-compatible constraints to institutions to ensure smooth functioning of financial systems.

2.1 *Basel Accord*

The standards of banking supervision started to change in late 1980s, and the 1992 Basel Committee propositions for the supervision of international banking groups and their cross-border establishments helped provide breakthroughs. The Basel Committee for Banking Supervision (BCBS) had two main goals while proposing a single capital standard for internationally active banks: to strengthen the soundness and stability of the international banking system, and to improve competition among internationally active banks. If capital is enough to absorb shocks at the level of any individual bank, then the financial system is also solid. Hence, financial institutions should be encouraged to increase their capital. Second, a standard approach to institutions from different countries would. Provided the trade-off between solidity and development in banking services provisioning, a single set of rules might help to ensure reasonable level of competition.

Acceleration of banking reforms in emerging economies was driven by increasing globalization of financial services' industry. Incentives to upgrade banking supervision and comply with international standards were rooted in arising opportunities to attract foreign capital from international financial markets. Leading banks in developing countries perceived that their business would be at risk unless national banking supervision was upgraded as otherwise they would not be allowed to compete globally.

The easiest way to upgrade local supervision and regulation in emerging economies was an adoption of the Basel Capital Accord, which have already been applied in many developed countries. That was a clearly positive change as it caused numerous necessary reforms in financial systems of developing economies via inclusion of banks from various institutional environments into the global financial network. More than 100 countries adopted Basel Accord in twentieth century. "The 1988 Basel capital accord has played an important role in strengthening the banking systems in many emerging market economies. Most countries have adopted risk-weighted capital adequacy rules, often after hard-fought battles with vested interest groups. The new capital adequacy framework submitted for worldwide consultation by the Basel Committee on Banking Supervision was intended to better align regulatory capital requirements with underlying banking risks and to recognize new risk management and control techniques." (De Krivoy 2000).

As per the results of the growing role of the Basel standards, at the turn of the century the experts of the BCBS reported that "the introduction of formal minimum capital requirements across the G-10 appears to have induced relatively weakly capitalized institutions to maintain higher capital ratios. . . . a common structure of formal regulatory capital requirements across countries may have enabled financial markets to exert greater market discipline on undercapitalized banks than would otherwise have been the case." (Jackson et al. 1999).

Researches of the early 2000s registered a very remarkable and contradicting trend: regulations in different countries were different, but looked harmonized. For instance, of all countries in the survey by Barth et al. (2001), 60% mentioned that the minimum capital adequacy requirement is 8% and another 14% of countries set it at exactly 10%. Hence, the vast majority or regulators (93%) believed that their minimum capital requirement was aligned with Basel guidelines. However, bank risk exposures had very different legal definitions across the countries. In particular, Fig. 1 illustrates that almost 40% of surveyed countries responded that operations with real estate are prohibited (while only 15% did not restrict those operations in any way), almost 80% of regulators did not restrict operations with the other type of risky asset—securities. Such a differentiation in approaches set a trend for future considerations in the BCBS.

The Basel Accord was evolving through time: while imposition of minimal capital standards on lending started to constrain credit exposures, the other types of risks appeared and became reasons for bank failures. So, the BCBS initiated discussions and amendments to the Accord in order to deal with market and operations risks. Having adopted the Basel in some of its form, emerging economies do not have an option to avoid complying with the global rules. Hence, regulators in EMs further contribute to globalization of risk regulation and management practices.

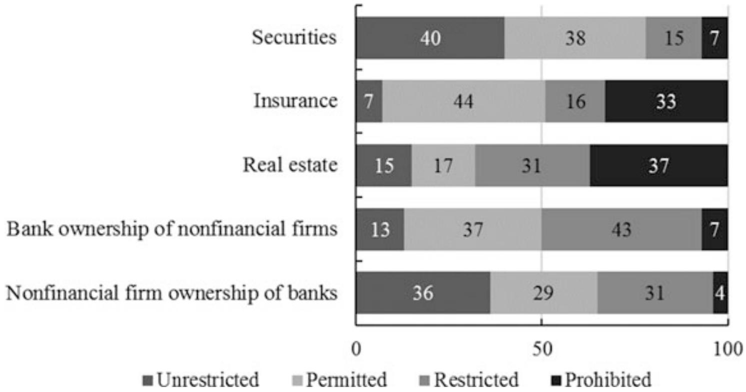


Fig. 1 Regulatory restrictions on bank activities and the mixing of banking and commerce (Source: Barth et al. (2001))

2.2 *Regulatory Independence and Other Institutional Characteristics*

Crises during the last decade of 20th century revealed that local regulators were not fully capable of dealing with the changing nature of banking activities: financial liberalization created new risks which local banking supervision simply ignored in the absence of shocks. While nowadays supervisory independence seems normal, the situation at the turn of the century was very much different: "Despite its importance, the issue of the independence for financial sector regulatory and supervisory authorities has only received a marginal attention in the literature and in the practice" (Taylor and Quintyn 2002).

The vast majority of policy papers highlight that financial stability will be achieved if regulatory independence is in place, differentiating at least four types. First, the regulatory independence per se, i.e. the ability of the agency to have a degree of autonomy in setting technical rules and regulations for the sector. Second, the supervisory independence in what concerns the on-site inspections and off-site monitoring, sanctioning, and the enforcement of sanctions—including revoking licenses—are the supervisor's main tools to ensure the stability of the system. The third aspect is institutional independence, which implies, among others, the status of the agency as an institution separated from the executive and legislative branches of the government. An agency that forms part of the executive branch, such as the ministry of finance, typically lacks independence. In almost all of the systemic financial sector crises of the 1990s, the lack of independence of supervisory authorities from political influence was cited as one of the contributing factors to the deepening of the crisis. And finally, budgetary independence refers to capability of the agency to determine the size of the agency's budget and its use, including the staffing of agencies and salary levels.

Based on the practice of managing regulatory and supervisory agencies in the emerging markets, the following recommendations were highlighted. First, one should ensure incentives for the executive authorities to have an independent regulation of the financial sector. At the beginning of the century, financial globalization provided the most powerful incentive. In a world of global financial markets, proper transactions pricing is more and more reliant on correct assessment of the counterparty risk, which requires the implementation of common rules. International bodies had already endorsed a variety of standards, including those of the Basel Committee on Banking Supervision, the International Organization of Securities Commissions (IOSCO), the International Monetary Fund (IMF), and the World Bank. But if the common rules are not obeyed, the financial community will not consider investments transparent. Lack of trust would result in lack of investments. Hence, the more dependent a country is on international finance, the more likely it would have independent financial sector supervision.

Second, the regulators should have enough resources to conduct their routine properly. Supervisory agencies at that time in developing countries were often represented by units within a ministry of finance, without sufficient funds, information technology, and skilled human resources to perform the job.[1] It became apparent that supervisors had to know as much as the supervised institutions about all aspects of bank business, e.g. products pricing, risk management, and market trends. Provided that financial innovation has always been unleashing new risks, regulators should allot funds to analyze those risks in order to develop instruments of mitigation. Hence governments or market participants should fund the supervisory agency. Alternatively, the agency might be allowed to earn income, and use these proceeds to fund itself.

The next point is that the regulators should be active. Efficiency of prudential regulation (which is creating rules and incentives that encourage banks to be prudent) is higher if the supervisors act timely. With rapid changes in the environment, supervisors should be able to recognize new risks and approach them before they are realized in a massive and non-controllable way.

Researchers (e.g. De Krivoy 2000) also believed that central banks in emerging economies should unite functions of both types of regulation—monetary policy issues and banking supervision. The two main reasons for that were: first, to ensure policy independence (which is easier to ensure for one agency than for a number of smaller-scale institutions), and second, to improve policy response (in terms of resolving scarcity of skilled human resources).

Additional argument in favor of consolidation of powers is the role of money market in emerging economies. In countries with mature capital markets, prices of bank shares or bonds are reliable signals of the financial health of banks. In emerging market economies ownership of banks is usually concentrated while capital markets

[1] In fact this was the case for many years ago, now the situation is rather different: central banks are perceived as a very financially independent, solid and well-equipped bodies, and so they are in a good position to hire, motivate, and keep skilled staff.

are small, the money market becomes the main indicator of the current banking sector's stance. With that in mind, regulators in the emerging economies should be highly concerned with ensuring liquidity provisioning: liquidity is often a primary proof of financial stability.

Uniting authorities under one body also helps to deal with a problem of undercapitalization, which is a frequent source of bank weakness in emerging market economies. If the environment is adverse, e.g. the investment climate is not favorable, or the economy is entering a recession, or financial market sentiment is weak, then banking institutions would not be capable of accumulating enough capital even if required by the regulators. If functions of monetary policy and banking supervision are split, bank recapitalization requirements might become a source of systemic risk rather than a solution.

2.3 Issues of Financial Stability at the Macroeconomic Level

At the turn of the century regulators in emerging markets were also concerned with several other issues, which negatively impact macroeconomic financial stability. One of such problems was the problem of connected lending. When the financial system is highly concentrated and dependent on few banks, and institutional environment is weak, any conflicts of interest including lending to related parties easily translate into systemic banking crises. Consolidated supervision over banks and holding companies was considered as a possible solution, yet it took more than a decade to start implementing that.

Other aspects of banking regulation and supervision that did attract attention in early 2000s were the problems that might have become a source of moral hazard: deposit insurance, crisis management and preparations for a crisis, and bailing-out rules. Again, it took years to develop clear rules, and make them work, be respected and complied with, but important peculiarities were highlighted already 20 years ago. First, the contingency planning tool for the preparation for and avoidance of bank failures is necessary, as it is the bank top management which best knows the true situation in the bank, and so they should be closely involved into exercises of contingency planning. And, second, authorities should be ready to step-in to ensure the stability of the financial system in absence of private capital to avoid larger-scale consequences. One of the proposed solutions in this case was to establish a special bank rescuing body for emergency cases.

A retrospective analysis shows that while systemic risks and related topics were under discussion around the turn of the century, regulators and supervisors from all over the world (including emerging economies) were mainly interested in mitigating risks at the level of individual banks and ensuring proper institutional design of supervision in order to ensure that the financial system is global. That global nature of the financial system seemed to be the main instrument for ensuring financial sustainability at the level of individual countries.

3 Crisis of 2008: A Wake-Up Call and a Turning Point for Regulators

The following two regulatory frameworks had dominated before the Global Financial Crisis of 2008. First, regulatory and supervisory strategies should promote private sector forces as the latter seem to work better than regulators might ensure on their own. And second, diversification of income streams and loan portfolios improves the performance and stability of financial institutions and the sector in general. It was a mainstream view that countries in which banks diversify their portfolios domestically and internationally suffer fewer crises (Barth et al. 2001). Yet such a mixture of views brought to a life a brand new phenomenon, which substantially affected the mindset of both regulators and banking sector participants. This was a phenomenon of systemic risk.

3.1 The Appearance of Systemic Risk Notion

In fact, the notion of systemic risk had appeared long ago, e.g. Stolz (2002) mentioned that there was no generally accepted definition of systemic risk, and so it was defined as the probability that the failure of one single bank leads to successive losses along the chain of institutions with a negative impact on the whole economy. In part this is similar and is related to domino (or knock-on) effect, with the two important differences: one is that initially by systemic risk one suggested to deal with problems localized in a single jurisdiction, and second is that financial systems susceptible to contagion were supposed to be small and weak.

However, with creation of EMU in early twenty-first century researchers started to highlight risks related to contagion effect for European countries. The highly integrated interbank market was named to serve as a transmission mechanism for cross-border spillovers, which was intensified with substantial amounts of FDIs that largest European economies directed into new member states. Moreover, the share of foreign ownership in European banking systems might intensify the problem of contagion, thus adding to the systemic risks.

Probability of contagion in the European Union was correctly linked with the regulatory system, in which the "home country" supervisors' mandate was limited to guarantee prudential behavior of home banks in their home markets only. Yet, such a "national" approach did not achieve efficient supervision from the perspective of an overall EU optimum: the probability of failure would be inefficiently high as in the conditions of the integrated financial system, a failure might be related to activities in a non-home jurisdiction. With a national mandate, supervisors do not take these spillovers into account. This is what many has underestimated before the crisis of 2008.

Formally, EU legislation has eliminated legal barriers to the exchange of information between national supervisory authorities, opening up the opportunity of cooperation. “However, countries are always hesitant to transfer rights to the supranational level. Hence, in order to find political approval, such a new institution could build on existing EU instances which may have already established a reputation in the field of supervision.” (Stolz 2002).

While in the USA the supervision did not face the same problems as the banking landscape was much more unified, than in Europe, the level of systemic risk has also been underestimated due to the complexity of products, especially in the investment banking industry.

3.2 New Paradigm: Financial Repression and Tighter Regulation and Supervision

The global financial crisis of 2008–2009 revived the discussion on the role of regulation and the government. The new mainstream was to prevent similar crises at any price, as the cost was huge: authorities all over the world injected billions of US dollars into their local financial systems to keep it afloat and ensure stability of payments. The surge in government debt as a source of funding for the anti-crisis measures has relaunched a debate on financial repression as a solution to the Global Financial Crisis.

Jafarov (2019) argues that repression has a long history in finance. Defined as direct government intervention that alters the equilibrium reached in the financial sector, it usually aims at providing cheap loans to companies and governments, reducing their burden of repayments by lowering returns to savers below the rate that would otherwise prevail. It has been applied in numerous forms such as ceilings on interest rates, directed credits to certain industries, or constraints on the composition of bank portfolios. Financial repression is typically accompanied by additional restrictions on financial activity, such as controls on international capital movements aimed at reducing the alternative investment opportunities available to savers. Unconventional monetary policies that keep the interest rate curve artificially flat should also be treated as a form of financial repression.

Despite short-term political benefits, financial repression and restrictions come at a macroeconomic cost by creating market distortions, which should have a tangible negative effect on long-term development. However, national governments and supervisors and regulators suggested that financial repression should reduce the probability of a debt crisis in a given period (which indirectly affected future growth perspectives positively), and this created a case for strengthening financial sector supervision before launching systemic reforms. The latter were necessarily provided the global economic cycle phase, but dealing with global economic slowdown was both more difficult and less obvious at the time of resolution of crisis effects, and so the authorities stocked to amending regulations in the financial sector which was responsible for triggering global financial crisis.

4 G-20 and Financial Stability Board: Initiatives and Implementation

Following Lehman Brothers' failure in 2008, the world financial sector lost confidence. This had tremendous effects of a complete loss of liquidity in the interbank market, causing massive sell-offs in almost all types of asset classes. Risks of the world economic collapse became real, and so France and the United Kingdom initiated a G-20 summit. This was a meeting of the heads of state and government of the Group of Twenty, who met in November 2008 to discuss possible common response to the global financial crisis.

This event helped to enhance coordination between governments and thus improved the sentiment in financial markets. Although this meeting did not stop sell-offs on its own, the financial stabilization became real. Yet it had taken a decade years to create a set-up which helped a lot to calm down the markets globally and contain crisis effects.

4.1 History of Creating G20

The Group of Twenty was founded at a conference in Berlin in late 1999. The main cause for the establishment of the G20 was the financial crisis that erupted in Asia in previous years and significantly affected the world economy. This crisis had clearly demonstrated that the era of financial liberalization brought not only opportunities but also high risks stemming from high interconnectedness of the institutions all over the world.

To resolve the issue, the USA and other members of the Asia-Pacific Economic Cooperation (APEC) at their meeting in Vancouver announced the creation of a group of twenty-two. The G22 included finance ministers and central bank governors from the industrialized G7 countries and 15 other countries.[2] The Group first met on April 1998 to discuss issues related to the stability of the international financial system and the effective functioning of global capital markets. However, it was not the only format considered in these years: the other one was a group of thirty-three (G33).[3]

However, G33 did not prove to be stable organization: some countries had to leave the Club due to political considerations, others—because of creation of the European Monetary Union. Size of financial systems of some member countries was considered too small to continue to rely on them in reforming the global financial

[2]Argentina, Australia, Brazil, Canada, China, France, Germany, Hong Kong, India, Indonesia, Italy, Japan, Republic of Korea, Malaysia, Mexico, Poland, Russia, Singapore, South Africa, Thailand, the United Kingdom, and the USA.

[3]In addition to countries from G22 it also included following members: Belgium, Chile, Cote d'Ivoire, Egypt, Morocco, Netherlands, Saudi Arabia, Spain, Sweden, Switzerland, Turkey.

system. Eventually G20[4] replaced all the other clubs that had been previously created to coordinate the financial regulations all over the world.[5].

However, after the founding conference in December 1999, the G20 did not hold summits until 2008. Its main form of activity was annual meetings of finance ministers and heads of central banks. So the fact that the G20 summits took place more than once within a year highlights severity of the 2008 crisis. The first summit took place in November 2008, the next two—in April and September 2009.

4.2 Coordination During 2008–2009

During the first Anti-Crisis Summit in November 2008, G-20 countries' leaders reached agreements on cooperation in key areas: not only to mitigate the effects of the financial crisis, but also to establish general principles for reforming financial institutions in order to prevent similar crises in the future.

In the declaration, they named the following two main reasons of the Global Financial Crisis. First, market participants failed to ensure an adequate evaluation of the risks while hunting for a return during a period of strong global growth. Policy-makers, regulators, and supervisors also overlooked the risks mounting in financial markets. Second, macroeconomic policies were inconsistent and insufficiently coordinated, which contributed to those risks and market failures. Altogether it resulted in severe market disruption.

While the list of actions in the declaration is rather long and contains both short- and long-term measures (see Declaration of the Summit), the following are of the major importance for the sake of this article. The first was an intention to strengthen financial markets and regulatory regimes via intensified international cooperation among regulators and wider implementation of international standards. Regulators were prescribed to diminish adverse impacts on other countries, including regulatory arbitrage, while financial institutions had recognized losses, improve their governance and risk management. The second was a more political rather than practical: it was a "victory" of emerging economies which called to higher representation in decision making procedures. The G20 leaders committed to reform the international financial institutions so that they can better reflect new economic realities: higher weights of emerging and developing economies.

The main directions for creation of additional recommendations were specified at mitigating against pro-cyclicality in regulatory policy and financial institutions' behavior; aligning global accounting standards; reviewing compensation practices

[4]Argentina, Australia, Brazil, Canada, China, the European Union, France, Germany, India, Indonesia, Italy, Japan, Mexico, Russia, Saudi Arabia, South Africa, South Korea, Turkey, the United Kingdom, and the USA.

[5]According to G20 official site, the members of the organization represent ca. 80% of the world's GDP, two-thirds of global population and three-quarters of international trade.

(which influence incentives for risk taking and innovation); defining the scope of systemically important institutions and determining their appropriate supervision, as well as many other measures concerning sound regulations, strengthening transparency, risk management, etc.

While short-term coordination was not as productive as expected in the first summit of G20, in the longer term, on the contrary, results were rather high, at least in what concerns regulation of financial markets. Presumably, this is due to creation of the Financial Stability Board.

4.3 Financial Stability Board as the Post-Crisis Reforms Coordination Body

To strengthen financial market oversight, in April 2009 G20 reformed the Financial Stability Forum, expanded its membership and renamed it the Financial Stability Board (FSB). The FSB includes not only G20 countries, but also Hong Kong SAR, Singapore, Netherlands, Switzerland, and Spain. This organization was designed to improve the functioning of financial markets and reduce systemic risk by expanding international cooperation between authorities responsible for maintaining financial stability. The launch of the institution is dated back to June 2009 meeting during which the FSB set up its organizational structure required to address its mandate (Washington Summit of G20 2008; Financial Stability Board 2009).

The FSB decided to focus on several main themes, including international cooperation, prudential regulation of banking institutions, increasing the scope of regulation on non-bank financial institutions and products, management compensation practices, issues related to credit rating agencies and accounting standards. The FSB has offered many prudential regulatory initiatives during the last decade, yet most important were agreed upon at September 2009 meeting: leverage ratio introduction, Tier I capital base quality improvement, countercyclical capital buffers, other measures aimed to deal with "too-big-to-fail" problem (most of them are discussed in detail in next part of the paper). Trough time G20 agenda was amending to address new and emerging vulnerabilities in the global financial system. In the most recent release (Financial Stability Board 2020) the Chair of the FSB named the following challenges and reforms to concentrate on:

- Issue of interest rate benchmark transition ("LIBOR");
- Adverse effects of technological innovation;
- Digital currencies and payment systems, including cross-border payment systems; and
- Growing non-bank financial intermediation.

According to the mandate and practice, the FSB reports to the G20 annually, covering general trends in financial intermediation, and also implementation and effects of reforms. But in addition, the FSB monitors and evaluates consequences of reforms in the financial sector on clients and general economy. The two reports examining effects of the G20 regulatory reforms on financial intermediation have already been published: the effects on the financing of infrastructure investment and on the financing of small and medium-sized enterprises.

4.4 Results of G20 Reforms Following Global Financial Crisis

G20 reforms were expected to reduce the probability of a crisis and, should a crisis occur, soften its impact. The goals should be achieved via smoothing provision of credit to the real sector, and, thus implies that in positive economic situation financial institutions will be limited in their risk appetite. This implies that while the impact on current GDP and other indicators of economic activity might be somewhat negative, net through-the-cycle (or, long term) effects will be positive. Although ex ante analysis suggests that the long-term economic benefits of reforms should overweight short-term costs, the former are more difficult to quantify since they are often less evident and take longer to unfold.

A complete empirical analysis of the benefits of the reforms would only be possible after a full financial cycle, when data shows how financial institutions have performed during both stressed and normal market conditions.[6] The available findings from the empirical analysis (Financial Stability Board 2019a) indicate that the introduction of risk-based capital requirements negatively impacted lending: it temporarily slowed down, the conditions were tightened. However, "there is some evidence of reallocation of credit towards more creditworthy enterprises and improved access to finance for financially stronger companies: after the reforms were introduced, better capitalized and more profitable firms increased their long-term borrowing more than other firms, and they invested more." By contrast, the liquidity reforms were found not to exert significant effects.

The November 2018 FSB report highlighted that "higher financial system resilience is being achieved without impeding the supply of credit to the real economy." This seems to be a positive result of a decade-long activity to improve financial regulation and supervision after the global financial crisis.

At the same time, findings suggest that the strength of the effects of the reforms on lending (which is a good proxy for macroeconomic efficiency) depends on country-specific factors. For instance, the effects were milder in jurisdictions where the

[6] In early 2020 spread of CoViD-19 all over the world and sharp reaction of businesses and authorities on that caused massive lockdowns in China and some European countries. This increases the probability that the stressed market conditions might be observed much sooner than expected. It seems that risks of global recession are high, and at least recessions will be severe in many large economies, while slowdown in China might be noticeable as well. Altogether this creates a strong risk that the economic developments will create a formal credit crisis, which will test efficiency of reforms in the past decade.

financial system started from a stronger basis and favorable economic conditions, and vice versa. Indirectly this implies that emerging economies might benefit somewhat more than developed economies from implementing reforms.

The most recent annual review of the Implementation and Effects of the G20 Financial Regulatory Reforms summarizes the progress in implementing reforms (see Table 1). First, emerging and developing economies are much more compliant to implement Basel III reforms, at least in what concerns Risk-Based Capital and Liquidity Coverage ratio, while developed economies (i.e. European Union and the USA) are either partially compliant, or even non-compliant to the new Basel rules. EM countries are also ahead of advanced economies in introducing and implementing other Basel aspects, including leverage ratio, net stable funding ratio, etc. Second, in coping with OTC derivatives, advanced economies have implemented reforms to a larger degree than emerging markets. Yet, here one should take into account that markets for derivatives are liquid and large only in developed economies, while emerging economies do not rely on local derivative markets—mostly due to high counterparty risk and low liquidity. In what concerns reforms of resolution and non-bank financial intermediation, the evidence is more mixed, yet on average developing economies are at least not worse than developed ones.

The global reforms made emerging counties follow the rules imposed for all largest economies. This is one of the two important drivers for the future of regulation of financial risks in emerging markets. The other is the degree of globalization. As long as both are changing in the same direction, the EMs will have to obey the rules of the BCBS and other international bodies.

5 Post-Crisis Basel Agreements: Timeline and Effects

Evolution of Basel requirements is remarkable when taking into account the history of the issue. Indeed, Basel Accord started with a simple regulation of capital adequacy in 1988 under conditions of absence of standardized rules in the world. At that time, capital regulation depended on the local definition in each particular country, and in some countries there were no formal rules. The weakness of the Basel Capital Accord became apparent quite soon: it limited the scope of the rules only to the regulation of credit risks. Due to bankruptcy of large banks in the early 1990s, the Basel Committee on Banking Supervision in 1996 amended the rules on capital adequacy, discouraging market risk exposure. However, the widespread crisis in East Asia required a deeper revision of the regulatory approach, and in 2004 BCBS developed a revised Basel II. The target was to implement all the requirements before 2007. Basel II contained additional capital requirements to protect against operational risks, this became one of the key elements of the new capital adequacy standard.

The Global Financial Crisis has led to the next stage of reforming how much capital banks should have prepared for various shocks. Moreover, the Basel III changed not only level of capital adequacy, but also the quality of capital, it

Table 1 Implementation of reforms in priority areas by FSB jurisdictions (as of September 2019)[a]

Reform Area	Basel III						Compensation	Over-the-counter (OTC) derivatives				Resolution				Non-bank financial intermediation	
	Risk-based capital	Liquidity Coverage Ration (LCR)	Require-ments for SIBs	Large exposures framework	Leverage ratio	Net Stable Funding Ratio (NSFR)		Trade reporting	Central clearing	Platform trading	Margin	Minimum external TLAC requirements for G-SIBs	Trasfer/ bail-in temporary stay powers for banks	Recovery and resolution planning for systemic banks	Transfer/ bridge/ run-off powers for insurers	Money market funds (MMFs)	Securitisation
Agreed phase-in (completed) date	2013 (2019)	2015 (2019)	2016 (2019)	2019	2018	20 18		end-2012	end-2012	end-2012	2016 (2021)	2019/ 2020 (2022/ 2028)					
Argentina	C	C					Δ										**
Australia	C	C		C	&	C										*	
Brazil	C	C		C		C	Δ										**
Canada	C	C		C		C										**	
China	C, Δ	C	C, &				Δ	R, F									
France	MNC	LC	C														*

24 Source: Implementation and Effects of the G20 Financial Regulatory Reforms: Fifth Annual Report, 2019

(continued)

Table 1 (continued)

Germany	MNC	LC	C														
Hong Kong	C	C														**	
India	C	LC		C		C											
Indonesia	LC	C														**	
Italy	MNC	LC	C														*
Japan	C	C	C														
Lorea	LC	C														**	
Mexico	C	C						R								**	*
Netherlands	MNC	LC	C														*
Russia	C	C					Δ									**	
Saudi Arabia	C	LC		C		C		R								**	
Singapore	C	C														**	
South Africa	C	C					Δ									**	
Spain	MNC	LC	C														*
Switzerland	C	C	C													**	
Turkey	C	C														**	
United Kingdom	MNC	LC	C														*
United States	LC	C	C, &				Δ										

Special notes: The colors and symbols in this snapshot indicate the timeliness of implementation. For Basel III, the letters indicate the extent to which implementation is consistent with the international standard (Regulatory Consistency Assessment Program—assessed "compliant" (C), "largely compliant" (LC), "materially non-compliant" (MNC) and "non-compliant" (NC) with Basel III rules). R/F: Further action required to remove barriers to full trade reporting (R) or to access trade repository data by foreign authority (F). &: Australia's implementation status on the leverage ratio is based on the revised (2017) exposure definition. China's G-SIB requirements are in force, while its D-SIB policy framework is under development. The US does not identify any additional D-SIBs beyond those designated as G-SIBs. * / ** Implementation is more advanced than the overall rating in one or more/all elements of at least one reform area or market sectors

[a]Source: Financial Stability Board (2019b)

introduced the credit valuation adjustment (CVA capital charge), new liquidity standards and mandatory leverage ratio requirement. One of the aims (as mentioned in the previous section) was a resolution of the “too-big-to-fail” problem, so largest banks were examined to have domestic or global importance, resulting in a specialized lists of credit institutions with additional loss absorbency requirements—special buffers against systemic shocks.

An important feature of Basel III is that this agreement is not a fixed set of rules, but its parts are flexible and evolving to reflect the market and financial sector developments. For instance, in 2010 the Basel III left the credit risk weights for different exposure types unchanged, but the finalized reforms package agreed in December 2017 changed the risk weights for some asset classes. It additionally limited the use of internal ratings-based (IRB) approaches for credit risk (e.g. by introducing input floors for Loss Given Default (LGD) estimates and changes to the recognition of eligible collateral), and introduced an “output floor” such that modeled outputs could not diverge too far in aggregate from Standardized Approaches.

As of today, Basel III is a comprehensive set of policy measures designed to strengthen the regulation, supervision, and risk management of the banking sector. Its standards are minimum requirements that apply to internationally active banks. Basel Committee for Banking Supervision member jurisdictions[7] commit to implementing them within the timeframe established by the Basel Committee (in stages, starting in January 2013 and ending by 2019, with some requirements having even longer term of implementation—till 2022), while non-member jurisdictions implement them on a voluntary basis.

Main pillars of Basel III are:

- change in the structure of capital with a higher proportion of the core, equity-based capital;
- increase of minimum capital requirements as compared to the pre-2008 level;
- introduction of capital conservation buffers—various buffers covering size of the business of a banking institution, the stage of the economic cycle, etc.;
- introduction of regulation of short- and long-term liquidity via specialized ratios.

Changes in the capital structure affected both core capital and additional capital. The main idea was to ensure that the core capital of a bank is high-quality, such as ordinary shares and retained earnings. Previously accepted hybrid instruments, e.g. preferred capital or “perpetual” subordinated bonds, were pushed to be converted into ordinary shares with a significant discount.

Basel III regulations intend to exclude those parts of the capital that cannot be used to cover losses, the goal was to minimize probability for authorities to intervene to absorb losses. So subordinated loans as a part of additional capital must be attracted at least for five years and convertible into ordinary shares; early repayment

[7]Largely coincide with G20 members, and also Belgium, Hong Kong SAR, Luxembourg, Netherlands, Singapore, Spain, Sweden, Switzerland.

Table 2 Capital adequacy requirements in accordance with the Basel III,% of RWA[a]

Year	2013	2014	2015	2016	2017	2018	2019
Common Equity Tier I capital	3.3	4.0	4.5	4.5	4.5	4.5	4.5
Capital Conservation buffers (CCB)	–	–	–	0.625	1.25	1.875	2.5
CET I + CCB	3.5	4	4.5	5.125	5.75	6.375	7
Tier I capital	4	5.5	6	6	6	6	6
Total (Tier I and Tier II) capital	8	8	8	8	8	8	8
Total capital + CCB	8	8	8	8.625	9.25	9.875	10.5

[a]Source: BCBS, Larionova (2018)

is possible only upon the consent of the regulator. Despite the total minimum capital adequacy requirements remained the same at 8%, the ratio of core capital to risk-weighted assets (RWA) was increased to at least 4.5%, and this ratio for Tier I capital should be at least 6% (see Table 2). Taking into account other aspects of capital regulation, this is not just a redistribution, it is equivalent to an increase in minimum requirements.

In addition, Basel III prescribes banks to create two capital buffers: a conservation buffer and a countercyclical buffer. The conservation buffer should amount to at least 2.5% of RWA, and is to be created from net profit during 2016–2018. This buffer should cover the losses of a financial institution in case of stress in the banking system, becoming an additional cushion. The resources needed to cover losses during crises should be accumulated in years of normal business conditions. The second buffer is to limit expansion of lending in order to avoid any credit bubbles. This countercyclical buffer should serve as an additional protection during the crisis, it should be created by banks in addition to the Core capital in the range 0 to 2.5% of RWA. It is the regulator who decides whether this buffer should be high or low, depending on the economic situation. During economic (or credit) overheating, authorities may increase the requirements for countercyclical buffer, and vice versa.

An important novel in Basel III is introduction of leverage metrics: the ratio of bank capital to the total assets of the financial institution (both on- and off-balance sheet accounts) should exceed 3% (this is mandatory after 2018). Provided that this ratio does not weight assets, it allows to eliminate possible manipulations with risk weights, which is especially important provided that large financial institutions tend to have larger financial markets operations which (prior to the crisis) were considered as more liquid, and, hence, requiring lower capital to absorb losses.

Another distinctive feature of Basel III is additional requirements on funding and liquidity position of a bank. The two ratios were defined: current liquidity ratio and long-term liquidity ratio. Banks were requested to fully cover short-term liabilities (maturing in less than 30 days) by liquid assets. The Liquidity Coverage Ratio is a mandatory metrics since 2019, while Net Stable Funding Ratio is mandatory but implementation ratio is still low.

While at the macroeconomic level, finalized Basel III implementation will become a great development which might increase soundness of the global financial system, the situation is less straightforward if one takes into account costs incurred

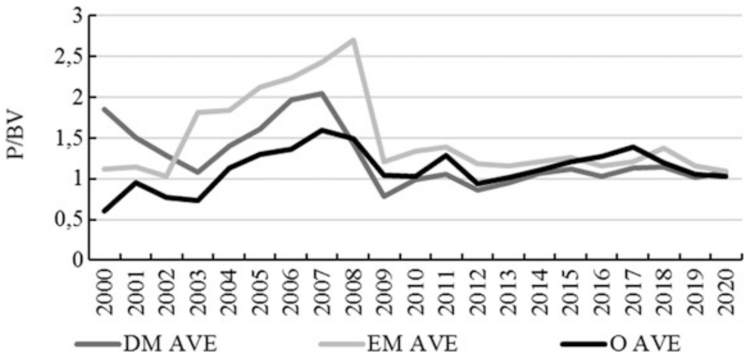


Fig. 2 Valuation of financial institutions in various markets (simple average over samples) (Source: Bloomberg, authors' calculations. *DM* developed markets, *EM* emerging markets, *O* other countries)

by financial institutions while complying with the rules. This is well illustrated with Fig. 2, which reflects the attractiveness of banks for investors (as measured by Price to Book value ratio). Implementation of the Basel III requirements will cost banks a lot—and not only in the form of direct charges to increase capital, but also in terms of valuation. Attractiveness of financial institutions before the Global Financial Crisis was much higher than that following the GFC and new Basel requirements. And what is specifically important for emerging economies is that they suffered more from that: prior to 2008 banks in emerging economies had much higher attractiveness for investors, presumably based on high expectations and opportunities that such markets might provide both for direct and portfolio investments due to non-satiation with financial services. However, following the adoption of Basel rules in G20, and provided higher discipline of regulatory and supervisory bodies in emerging economies to implement them, the valuation averages do not provide any premium for new markets.

What is more, convergence is also evident without averaging (Fig. 3): putting off all the outliers in terms of valuation at the level of countries, the difference between the range of II and II quartiles between the groups of developed and emerging economies is negligible. While in the past it was possible to find financial institutions valued substantially below or above its book value capital, now valuation is close to 1x in most of the markets. This implies that the cost of imposing restrictions on the capital and other aspects of banking business are high indeed: investors see only limited potential in investing, as future dividend flows might only cover already invested amounts. This reflects high burden that the sector bears because of regulations.

Worth also mentioning that this evidence contradicts to data reported by the FSB (see Table 3): their calculations of the pre-crisis and "updated" costs of various forms of funding imply that spreads to a risk-free rate have declined.

Complying with the Basel III rules has already changed the landscape and perspectives of the sector, but experts from Moody's (2018) suggest that complying to "finalized Basel III rules" (also named as Basel IV) will require additional efforts,

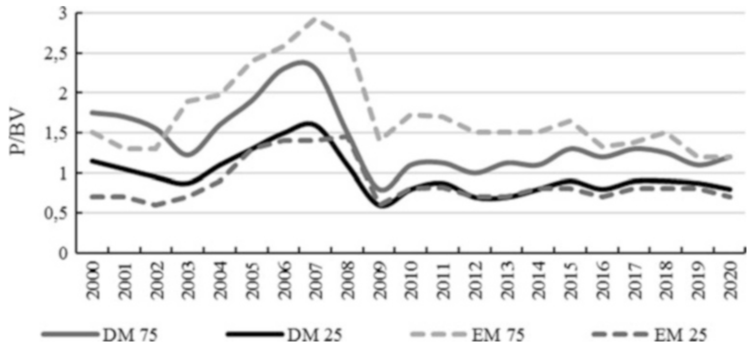


Fig. 3 Convergence of attractiveness of banks in EMs and DMs (I and III quartiles) (Source: Bloomberg, authors' calculations)

Table 3 Funding costs, relative to the spread to a risk-free rate[a]

Funding source	Pre-crisis cost	Updated cost
CET I	15%	11.8%
Additional Tier I	10%	6.84%
Tier II	7%	3.81%
Other funding (deposits, wholesale funding, etc.)	5%	2.8%
Risk-free rate proxy: 3 m US Libor	5%	2.8%

[a]Source: Evaluation of the effects of financial regulatory reforms on small and medium-sized enterprise (SME) financing (2019)

and it is already recommended to start preparing. Moody's suggests to rely more on clouds: migrating to the cloud will improve transparency of financial institutions' IT costs. Banks might become more agile thanks to the flexibility and scalability that clouds offer, and increase efficiency and outsource a significant part of their compliance burden. This indirectly suggests a higher reliance on the modern technological advances in Regulation and Supervision (discussed later in the paper).

6 The Contradictions in the Local and Global Legal Environment and Their Solutions

The Global Financial development report 2019/2020 (World Bank Group 2019) stipulates that "in today's interconnected global financial system, regulatory changes do not recognize national boundaries and affect advanced and developing countries alike."

As the active phase of the crisis ended, there was much talk about using the crisis to push through difficult but needed regulatory reforms. As we discussed in the previous sections, at the global level, the Financial Stability Board (FSB) promoted the coordinated reform agenda. Many countries have enacted or are still in the process of adopting new laws and regulations at their respective national levels in response to lessons from the crisis. Also, many countries have stepped up efforts in the area of macroprudential policy, and have put into effect better regimes for bank resolution and consumer protection.

However, after a decade of global regulatory reforms the legal environment for doing banking business started to change in a different direction. International consensus on general regulatory and supervisory reforms seems to be collapsing. The political appetite for globalization is retreating, and various tensions between countries are growing. Deloitte's paper on Banking Regulatory Outlook (2020) noted that "global standard-setting bodies (e.g. BCBS, or FSB) have less ambitious plans to introduce new standards than in previous years. Work to implement the remaining aspects of the G20 financial regulatory reforms has slowed, with many jurisdictions behind in implementing Basel III."

6.1 End of Regulatory Convergence Trends

While developed economies are not the focus of this research, it is not possible to omit developments in the EU for two reasons: firstly, this is a large economy which impact the world financial sector a lot, and secondly, developments in this area impact substantially regulatory trends in CEE countries.

The development of European banking legislation since 2011 should be analyzed having in minds the Eurozone debt crisis, which highlighted risks of contagion effects that arise in individual banks but might threat financial stability in the region. The legislative response to the EMU crisis included not only the urgent and necessary fire-fighting operations to revive economies and banks, but also a more fundamental restructuring of the basics of financial supervision as a whole in the region. The latter was considered as important to prevent a recurrence of the crisis, with more European integration in many areas being seen as the long-term solution to problems arising from European monetary union. The implementation of a Single Supervisory Mechanism (SSM) for banking institutions in the eurozone, and common bank recovery and resolution arrangements, are summarized in the Banking Regulation Review (Putnis et al. 2019).

Although recent changes to the European supervisory architecture and the commitment of the EC to introduce an EU-wide single rule book for financial services, the introduction of new EU rules is increasingly taking the form of directly applicable EU regulations. However, much of the EU banking regulation has traditionally been in the form of directives, which do not normally have legal effect in EU Member States until implemented by national laws.

Although non-eurozone member states do not participate in the European banking union, the regulation allows those countries to enter into close supervisory cooperation with the ECB. Till now, none of the nine non-EMU Member States has opted to do so, although in October 2017 Denmark, Sweden, and Bulgaria were considered as possibly joining the banking union. Moreover, the UK's decision to leave the European Union highlights that the regulations might not only unify, but also divide. It can be said with some degree of certainty, however, that the loss of the UK's voice from the conversation might have a significant effect on the shape of future EU banking legislation, although further detail should be added when the outcome of the Brexit negotiations is clearer.

Despite wide promotion of international cooperation and competition, the globally set rules become less demanded in emerging economies as well. For instance, in Poland laws limit foreign ownership of companies in selected strategic sectors, and restrict acquisition of real estate, especially agricultural and forest land. While some such restrictions are understandable, the government's willingness to increase the percentage of domestic ownership in certain industries (including banking and retail, which are currently dominated by foreign companies) is a clear signal that global approach to regulation of financial intuitions might no longer be global.

Provided increasing role of China in the world economy, its impact on financial sector might also grow. However, as reported in The Banking Regulations Review (Lovells 2019), prudential regulation in this emerging economy is not only applied to banking institutions, but also to their banking business products and services. Apart from traditional banking products (e.g. loans and deposits), special regulation appears to cover wealth management, structured deposits, etc. For instance, commercial banks issuing structured deposits must have the required derivative product trading business qualifications, and must comply with the local regulation of derivatives.

6.2 Sanctions as a New Reality for Doing Banking Business

Most of the regulations in the financial sector are characterized by, firstly, intention to improve risk profiles and/or the sustainability of legal entities in the sector against adverse scenarios, and secondly, similar applicability on all entities possessing similar characteristics. Those restrictions do not prohibit banking, and penalties mostly hit management.

However, restrictions might not only be imposed by local governments: under certain circumstances foreign bodies might also impose restrictions on financial companies. For many years the key source of such restrictions has served the United Nations' Security Council, acting under the Charter of the United Nations. That body had the right to adopt resolutions imposing sanctions against governments, persons, or entities: "Sanctions measures. . . encompass a broad range of enforcement options that do not involve the use of armed force. Since 1966, the Security Council has established 30 sanctions regimes."[8]

[8]https://www.un.org/securitycouncil/sanctions/information.

Restrictions have taken a number of different forms, depending on the goals: from comprehensive economic and trade sanctions to more targeted measures such as arms embargoes, travel bans, and financial or commodity restrictions. For years, the most severe versions were embargoes, mostly applied against governments which were caught in activities related to violating human rights, or to prevent proliferation of weapons of mass destruction or other violence. For those restrictions to be valid and executable, they had to be transposed into local laws via the adoption by local legislative bodies of governments.

However, over time the threats changed, and the fight against terrorism became more relevant. This created a new way of impacting the targeted entities: to cope with the financing terrorism and anti-money laundering activities. However, there is no a supranational body in financial sphere whose decision would be mandatory for all the countries in the world for financial sanctions. As a result, the largest economies started to impose restrictions without UN's mandate, insisting on the execution of such restrictions. Since the mid-2010s the number of restrictions in financial and other areas started to grow, thus complicating the legal environment for doing banking.

Box 1 The case of Venezuela

For more than a decade, the USA has imposed sanctions over the Venezuelan government and Venezuelan entities. The US government has imposed sanctions on Venezuela through executive orders which are not subject to consideration in the US Congress, thus substantially increasing the speed of the implementation of restrictions over the sanctioned entities and so for all banks doing business with them. Apart from freezing assets of individuals (most of them are officials), the US authorities closed access to US financial markets for the Venezuelan government.

In response, the Venezuelan government initiated a project to issue digital currency, which is more difficult to control by external bodies. Yet, in March 2018, another US executive order prohibited transactions involving the Venezuelan government's issuance of digital currency, coins, or tokens. Two months after that transactions related to purchasing Venezuelan debt were also banned. Joint ventures with Venezuelan government also became riskier, especially after inclusion of the Moscow-based Evrofinance Mosnarbank (owned by Russia and Venezuela) into the SDN list. The Venezuelan Economic and Social Development Bank (affiliated with the government) and its subsidiaries also joined the list, and finally, US Treasury sanctioned Venezuela's central bank.

Similar types of restrictions were imposed by many other economies, including Europe and the UK. As a result, Venezuelan financial system is effectively cut from the world financial markets.

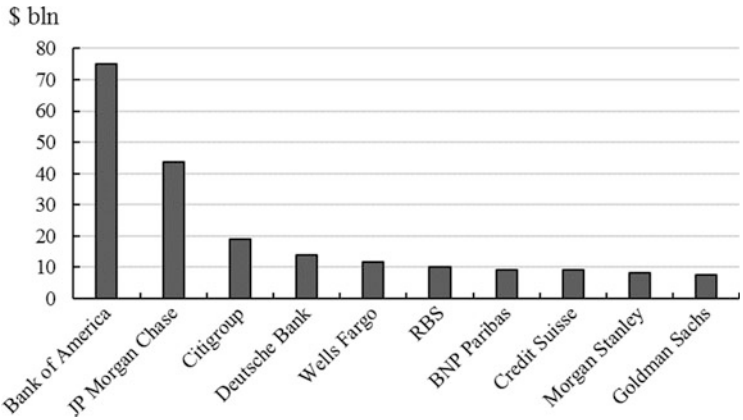


Fig. 4 Penalties and fees paid by banks since 2008 (Source: Keefe, Bruyette and Woods)

Cases of Venezuela (see Box 1), Russia, Iran, and others highlight that sanctions, although they are not a part of the formal banking supervisory and regulatory landscape, have become a very important and costly element of complying to the rules of doing business. Obviously, most of the mentioned restrictions are in a direct contradiction to the globally accepted banking rules, but they will stay as long as political matters prevail over business considerations. Worth mentioning that sanctions are more difficult to reverse than seems: the experience of the Jackson–Vanik amendment introduced in 1974, which restricted trade with the USSR, survived longer than the country. Indeed, while the Soviet Union collapsed in 1991, it took more than decade to abandon the amendment.

Also, the impact of sanctions on banks is higher than it seems: total amount of fees and penalties paid by banks and financial companies to the various authorities have exceeded USD 200 bn (see Fig. 4). The largest portion was paid by US banks—with the Bank of America paying more than USD 75 bn. JP Morgan Chase paid more than USD 43 bn. The largest fee imposed on a European bank was the one on Deutsche bank—which had to pay USD 14 bn. However, in recent years European banks seem to be under a special attention from the US regulators, with many of them trying to reach out-of-court settlements and reduce the fees to be paid.

6.3 Deregulation Is on the Agenda

Provided role of the largest economies in regulation trends for emerging markets, it is very important what are the dominating financial regulation trends in the USA. Deloitte's research highlights that political concerns grow as regulation impedes competition, new lending, and investment. US authorities might enter a deregulatory stance, because initial efforts might have gone too far and do not adequately balance the trade-off between safety and soundness and burden, especially for smaller, less complex banks. Some tailoring of the regulation already take place: smaller banks received modest to substantial relief, although the largest systemic financial institutions have only been granted only very small relief.

The first aspect is the new tailoring criteria on the Enhanced Prudential Standards: they are simple, intuitive, and transparent, yet still correlated with the risk posed by an institution's size, complexity of governance, and scope of operations. Other tailoring efforts resulting in regulation changes, resulted in a burden decrease, especially on regional and community banking organizations, include simplifying reporting requirements for qualifying community banks; prolonging preparatory and exam cycles for small banks, etc.

An important change lies in tailoring supervision versus regulation: examiners now focus on interpreting how well an institution is adhering to rules or guidance based on their own judgments about the bank's quality of management relative to its complexity and risk.

A similar trend is taking place in Russia: national regulatory and supervisory body—the Bank of Russia—in 2018 introduced a so-called "proportional" regulation approach, which is based on differentiating banks by their role, size and scope. All financial institutions holding banking licenses were divided into two broad groups on the basis of asset size: the ones with assets above RUB 3bn were not affected, but the rest were required to cut the scope of operations to only a few, traditional and "simple" activities. In response, the regulations of this latter group were substantially eased: they need to follow only 5 out of the dozen of main ratios. Their licenses are called "basic."

Other countries may follow, and there could become even a competitive deregulation. While deregulation might reduce some compliance costs, global firms will face more complexities and expenditure as regulatory standards across jurisdictions diverge in timing and substance. This will mean a new wave of differentiation between global and local requirements. Signals of the appearance of specific national forms of financial sector legal environment—both in terms of regulation and supervision—are increasing.

6.4 *Conduct as a Resolution of the Conflict Between Global and Local Legal Frameworks*

Over last 10 years market participants at all levels have made a clear shift in understanding the importance of conduct and culture. While authorities intend to ensure precise adherence to the rules, the regulations become less direct, and more relying on internal culture of the supervised organizations. Relying on positive behavioral patterns seems to be a good solution to the growing divergence between interests in different countries, although not a simple one.

For instance, as reported in the Group of Thirty paper on Banking conduct and culture (Kelly 2018), much the work has been done at the most senior levels of entities with the "tone from the top," however, the "tone from above" works much worse. For permanent positive shift, banks now need to focus on embedding culture awareness at all levels of the organization. Another area of positive change is performance management and incentives. The majority of banks have reviewed their remuneration schemes to integrate behavioral metrics into performance score-cards. For example, employee's performance is evaluated both with the "what" and the "how." But managing via a more balanced view might require management skills that need to be further developed, especially in middle management. That creates a special challenge for financial institutions, as this adds costs to the existing pressure from the authorities. And also more spending is needed to expand training and staff development programs to help employees better understand expectations of behavior in gray zones.

On the other hand, while over the last decade regulators and supervisors have significantly advanced in imposing rules, they are expected to provide impact on behavioral patterns of bankers as well, though regulators powers and skills in this area are limited.

The abovementioned paper by the Group of Thirty suggests that key lessons might be: firstly, managing culture is a continuous and ongoing effort that must be integrated into day-to-day business operations. Then, conduct is not just about purposeful misbehavior, but also about unintended consequences from decisions and/or lack of skills and knowledge. Hence an organization should have a proper set of instruments to differentiate between the two. Also "regulation can be an effective tool in outlining basic principles (especially related to good conduct), refocusing banks' attention on areas of persistent conduct failure, and providing insights and lessons learned from across the industry. Supervision can play a role in monitoring and providing feedback to banks that can aid the bank board and senior management in addressing culture and conduct issues" (Group of Thirty 2018).

Worth mentioning that another modern aspect that is driving the banking sector legal environment is accelerating digitalization. It will have a significant impact on the business and strategies in the sector in the foreseeable future. The share of transactions occurring in branches is continuously decreasing, in 2018 it was only 12%. At the same time, the deeper the digitalization enters into the financial companies, the more prolonged effects might take place if banks fail to comply to any of the rules imposed by more and more conflicting regulators. Reliance on technology as a solution to growing complexity of regulations is explained in more details in the next section.

The contradictions between the global agenda and local peculiarities, general rules and exceptions will grow. Financial firms must be prepared to respond to new, non-traditional regulatory trends: the regulators' agenda has changed its focus from establishing a fairground for international competition and macroeconomic stability to coping with technological change and social issues. Hence the future of banking behavior would not depend on the regulations, or on supervision, but by far will be determined by the mindset or culture of doing business in the financial sphere.

7 RegTech and SupTech: The Future of Banking Regulation

In the discussion above, the focus was mostly on how the sector is regulated or should be regulated from the institutional point of view. However, an important aspect was missed: technology, or, more precisely, the evolution of technology. In what concerns the relationships between financial institutions and their supervisory authorities, the two most important technological developments are "RegTech" and "SupTech."

The main driver behind the creation and development of these advancements is growing business and strategic risk, which originate in the environment and the decisions by authorities, based on the complication of the legal framework.[9] It became more and more evident that old-school legal back-up for doing banking based on human knowledge and skills is no longer valid and approaches the challenges of the twenty-first century.

7.1 RegTechs: A Tool to Improve and Ease Compliance

The starting point for the development of RegTech was the global financial crisis of 2008. As regulators have tightened requirements, various IT solutions appeared to ease and improve the efficiency of compliance for market participants.[10] They were named RegTechs (regulatory technologies), as they are any application or platform that makes regulatory compliance more efficient, through automatized processes and at a lower cost. Yet the term "RegTech" became widely used after 2015 when the first specialized companies in this area showed their first success. These days RegTech is actively developing in Western counties, especially in the USA and the UK. RegTech companies do not redraw the market, but integrated into the existing financial system. Most of them represent niche b-2-b products that work with large corporations.

According to the estimates of international associations of certified public accountants (ACCA), the number of innovations in the legal field after 2008–2009 increased fivefold. Now, only to comply with all the requirements, employees spend 10–15% of working time.

Based on the available technological developments, RegTech could help in resolving the following issues. First, customer identification and data verification in accordance with KYC policies and regulatory requirements. This simplifies user

[9] LexisNexis: RegTech: Navigating the jargon, the FCA sandbox and key initiatives.

[10] Feedback Statement. Call for Input on Supporting the Development and Adopters of RegTech. Financial Conduct Authority, 3 July 2016.

verification, helps track suspicious transactions and manage risks. Second, RegTech helps in the automation of data processing and compliance with standards.

Obviously, RegTech also help in improving data protection. Technologies help control data transfer, fight money laundering, and prevent fraud through transactional analysis. In this area, RegTech companies could even offer cyberattack insurance, provide employee behavior analysis, on test cybersecurity, etc. Also, it provides risk analysis and could even suggest possible solutions in areas of analysis of customer creditworthiness, reputation, and condition of companies, as well as ensuring compliance with legal requirements. To a lesser extent, RegTech is used in automation of control, verification of compliance of a financial product with regulatory requirements, stress testing, equity planning, etc.[11,12]

All these dimensions are to be realized via growing reliance on cloud computing, etc. However, these trends create a set of challenges. Putting the social part of them aside, the most important are regulatory and adoption risks. From the initial stage it is not fully clear whether RegTechs are capable of ensuring consumer protection, data privacy, and security while being flexible enough to support rapidly evolving innovation and growth. If the proposed frameworks are aligned to associated standards in other global markets, if they mitigate the regulatory risks of reliance on third-party systems and controls, and if the technology businesses (which might be new to the financial sector) understand their regulatory responsibilities. Many issues might arise during integrating new technology alongside legacy systems, effectiveness and efficiency might be limited by poor data quality of banking institutions and lack of budget to make additional investment in operational compliance areas and because regulated firms are often reluctant to be the first adopters and prefer investment in proven capabilities.

Though most RegTech solutions are capable of reading data from banks' legacy systems, the lack of standardization limits their capability for seamless integration with other newly implemented third-party or in-house applications.[13] This theme should be developed and investigated further—especially given the regulation is not based on exact norms, but also on the motivated judgments.

7.2 *SupTech: A Tool to Ensure Compliance*

As mentioned above, RegTech is not limited to participants in financial markets. In the field of regulation and supervision, its application is called "SupTech" (supervision technology or supervisory technology). SupTech solutions are designed to

[11] Feedback Statement. Call for Input on Supporting the Development and Adopters of RegTech. Financial Conduct Authority, 3 July 2016.

[12] RegTech Universe: Take a closer look at who is orbiting the RegTech space. Deloitte, 2017.

[13] RegTechReport 2018 Executive Summary by Medici (Signature Report | Vol: 3 IQ2 2018).

automate and streamline administrative and operational procedures, digitize data and work tools, and improve the analysis of loosely structured data. The technologies used are Big Data, machine learning, artificial intelligence, and cloud technologies. For example, Big Data and machine learning technologies allow the supervisor to analyze relationships, process unstructured data, including from external sources (media, the Internet), and use its results to detect illegal actions in the financial market and predict potential risks. The data-centric approach of the regulator's interaction with supervised organizations will contribute to further increasing the transparency of the financial sector and creating an effective supervisory environment. With its help, regulators can analyze the affiliation of borrowers, predict the demand for cash, determine the stability of credit organizations, conduct online data analysis, and identify cases of fraud.

In early 2018, the Central Bank of Ireland announced plans to launch a regulatory technology center to work with companies innovating in the financial services sector. The Russian Central Bank is also showing interest in RegTech. In the report "The main directions of the development of financial technologies for the period 2018-2020," the regulator named RegTech among the main vectors of the development of fintech in Russia (Central Bank of Russia 2018). To test new technologies in the financial network, the Central Bank launched its regulatory "sandbox." Other central banks also participate in projects of implementing SupTechs (see Box 2).

Box 2 The case of Austria

Austria is considered among the most well-known cases of the implementation of SupTech. In 2014 the National Bask of Austria launched a centralized data collection on the basis of platform "ABACUS," developed by BearingPoint. The platform is operated by Austrian Reporting Services (AuRep), which is a joint venture of 8 largest banking groups in the country, whose market share exceeds 85%.[14] Financial institutions supply micro-level data to ABACUS (including data on each financial agreement) in a standardized form almost in real-time regime (ordinary, next day after the transaction). Such an information is a so-called data cube, which might be aggregated into "smart cubes" in case of a specialized request by the supervisor. Such smart cubes reflect regulatory purposes and targets, and the platform is flexible enough to cover all requests.

The main advantage of this approach is that financial institutions are no longer required to provide the supervisory authority with the same data in different forms and for different goals, and banks might also save costs on data aggregation and the calculation of various analytical indicators.[15]

(continued)

[14] Austrian Reporting Services. AuRep (2018).

[15] Regulatory Utilities. BearingPoint (2018).

Box 2 (continued)

The National Bank of Austria says that the system allows to overcome the main limitation of data collection based on pre-set templates, as the latter approach might contain typos, doubling of data inputting, or insufficient detailing.[16] BearingPoint calculated that with launching ABACUS, the costs of financial institutions to provide regulatory reports declined by more than 30%.

The application and development of these technological advances to financial industry are even more important in developing countries than anywhere else. The primary reason is that, although contemporary trends imply lower coordination and the unification of legal environment between countries, the vast majority of emerging economies are not regulation-makers, but mostly regulation-takers. This is mostly related to the scale and size of their respective financial systems and also to the global interconnectedness of not only financial institutions, but also economies, and so for banks in emerging markets satisfying all the rules will become an increasingly difficult task. RegTechs and SupTechs are set to ease that.

8 Concluding Comments

At the turn of the century regulators and supervisors in the world were largely focused on mitigating risks at the level of individual banks and improving institutional design of supervision in such a way that would ensure that the financial system is global. Emerging economies expected to receive numerous benefits from complying with the global rules. And that helped to resolve a severe issue for regulators in developing countries when building stable financial systems—effective implementation. But the reforms made emerging counties follow the rules imposed for all largest economies—not once, but it became a continuous process, which is more and more costly for them, while benefits are limited due to a overregulation of the sector. This is illustrated the best with the Basel Accord evolution and impact on bank valuations.

However, the contradictions between the global agenda and local peculiarities grow. And financial companies must be prepared to respond to non-traditional regulatory trends, including coping with technological changes and social issues. They create all-time-changing regulatory environment.

With this in mind, regulation of financial risks in emerging markets will develop in accordance with the three drivers: one is the development of regulation in largest

[16]New Ways in Reporting for Austrian Banks. European Institute of Financial Regulation (20 September 2016).

economies, the other is the degree of globalization, and the third is development of technological solutions and social demands.

Hence the future of efficient banking would not depend on the regulations or on supervision, but by far will be determined by the mindset or culture of doing business in the financial sphere. In EMs, the role of technological advances applied to financial industry will be even more important than anywhere else. The primary reason is that such advances help to overcome problem of limited trust and weak institutional environment.

However, one should also take into account a very important aspect: in fact, past crises do not teach us well to forecast future ones. For instance, the supporting materials for the World Economic Forum in 2020 (for instance, Global Risk Report) did identify the following 5 risks as the most impacting: climate action failure, weapons of mass destruction, biodiversity loss, extreme weather, and water crisis. Few weeks later the world was highly concerned with epidemic of CoViD-19, and consequences of that risk were clearly underestimated. What is more, even provided that similar risks (although to a less extent) have realized several times since the beginning of the century (SARS, MERS, etc.), such a risk was not even considered as likely.

References

Barth, J. R., Caprio, G., & Levine, R. (2001). *The regulation and supervision of banks around the world - a new database* (Policy, Research working paper; no. WPS 2588). Washington, DC: World Bank. http://documents.worldbank.org/curated/en/981551468739510139/The-regulation-and-supervision-of-banks-around-the-world-a-new-database

Central Bank of Russia. (2018). *Issues and direction of development of regulatory and supervisory technologies* (Regtech and Suptech) on the financial market in Russia.

De Krivoy, R. (2000). *Reforming bank supervision in developing countries Conference Series*; [Proceedings], Federal Reserve Bank of Boston, issue jun, pages 113-143. Handle: RePEc:fip:fedbcp:y:2000:i:jun:p:113-143:n:44.

Deloitte. (2020). Banking regulatory outlook 2020.

Financial Stability Board. (2009). *Overview of progress in implementing the London Summit Recommendations for Strengthening financial stability*. Report of the Financial Stability Board to G20 leaders.

Financial Stability Board. (2019a). *Evaluation of the effects of financial regulatory reforms on small and medium-sized enterprise (SME) financing*.

Financial Stability Board. (2019b). *Implementation and effects of the G20 financial regulatory reforms*, 5th Annual report.

Financial Stability Board. (2020). To G20 Finance Ministers and Central Bank Governors.

Group of Thirty. (2018). Banking conduct and culture. A permanent mindset change.

Jackson, P., Furfine, C., Groeneveld, H., Hancock, D., Jones, D., Perraudin, W. et al. (1999). *"Capital Requirements and Bank Behaviour: The Impact of the Basle Accord," Basle Committee on Banking Supervision*. Working Paper No. 1.

Jafarov, E., Maino, R., & Pani, M. (2019). *Financial Repression is Knocking at the Door, Again. Should We Be Concerned?* IMF Working paper (WP/19/211).

Kelly, G. (2018). Banking Conduct and Culture: A Permanent Mindset Change, Group of 30 Steering Committee

Larionova, I. V. (2018). *Aligning banking regulation in volatile economy in accordance with the standards of Basel Committee on banking supervision (Basel III), Monography (in Russian)*. Moscow: Knorus.

Leslie-Bini, A. Suptech: More than just a new name for solving an old problem.

Lovells, H. (2019). *The banking regulatory regime in China*. The Law Reviews.

Moody's (2018) Regtech – enabler of the shift from compliance to performance.

Putnis, J., Bradley, E., & Raoufi, T. (2019). *The Banking Regulation Review. European Union*. Edition 10 (https://thelawreviews.co.uk/edition/the-banking-regulation-review-edition-10/1190826/european-union)

Stolz, S. (2002). *Banking Supervision in Integrated Financial Markets: Implications for the EU CESifo* (Working Paper Series 812). CESifo Group Munich.

Supervision of financial services in the OECD area. https://www.oecd.org/finance/financial-markets/1939320.pdf.

Taylor, M. W., & Quintyn, M. G. (2002). *Regulatory and supervisory independence and financial stability* (IMF Working Papers 02/46). International Monetary Fund.

Washington Summit of G20 (2008) Declaration of the Summit on financial markets and the world economy.

World Bank Group. (2019). *Bank regulation and supervision a decade after the global financial crisis*. Global financial development report 2019/2020.

Part II
Ratings and Risk Measuring

Principles of Rating Estimation in Emerging Countries

Sergei Grishunin, Natalia Dyachkova, and Alexander Karminsky

Abstract Ratings in emerging markets can serve as part of the early warning systems to reflect the weak signals of potential risks to the entity from the environment. Emerging markets have specific features that rating agencies usually consider in judgments of their credit ratings. They are underpinned by the higher volatility, exposure to sovereign issues, weaknesses in institutional governance, and lower rating transparency. Emerging markets are served by both international and national rating agencies. The latter assign national scale ratings which are the opinions of the relative creditworthiness of issuer or the entity relative to the national benchmark. National scale ratings primarily focus on niche markets where they draw on familiarity with specific domestic economic and political circumstances and thus cannot be directly compared to international scale ratings. In the field of the regulation of rating activities, emerging countries follow the regulatory trends that have been established in Europe and the USA. However, the quality and depth of regulation depends significantly on the maturity of the rating industry of the particular countries.

Keywords Emerging markets · Credit benchmark · Credit rating · International and domestic rating agencies · National scale rating · Early warning system · Regulation of ratings

JEL Codes G17 · G23 · G24 · K22

S. Grishunin
National Research University Higher School of Economics, Faculty of Economic Science, ZAO Deloitte & Touche CIS, Moscow, Russia

N. Dyachkova (✉) · A. Karminsky
National Research University Higher School of Economics, Faculty of Economic Science, Moscow, Russia
e-mail: ndyachkova@hse.ru

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_3

1 Rating as a Measure of Risk. Rating Agencies in Emerging Markets

Ratings in the economy—a comprehensive assessment of the risks of a company, bank, insurance company, mutual fund, country, region, issues of bonds, and other financial instruments on a discrete ordered scale called the rating scale (Karminsky and Polozov 2016). They determine the class (group) to which one or another business entity or financial instrument can be attributed. For instance, the credit rating of entity indirectly forms the estimates of the probability of failure to fulfill its obligations or, in other words, the assessment of investment potential of the entity.

Ratings are divided into the type of risks assessed (credit, market, liquidity, loss of management, technical, social, etc.) as well as by rating objects (corporate, banking, reliability of insurance companies, reliability of funds, sovereign (country), regional, bond, etc.). By the term period, ratings are divided into the short term (with a term of 1 year or less) and long term (with a term of more than 1 year).

The spread of ratings is underpinned by the growing complexity of the contemporary business environment, including financial markets. Business managers, investors, regulators, and other participants in financial markets do not have enough human resources, data, experience, and knowledge to properly assess risks in decision making. As a result, a rating industry has emerged which has sufficient resources and knowledge to assess the risks of entities in the market and present them in the form of ratings. The ratings thus become a basis of trust for investors. More details are provided in (Karminsky and Polozov 2016).

Depending on the purpose of the rating, the positioning of the rating in relation to users and objects of assessment, as well as the degree of the independence and objectivity of ratings, the following rating types can be distinguished.

Regulatory ratings. These ratings are used by various financial regulators for the remote supervision of participants of national banking sector. Examples of such ratings are the US rating system for evaluating US banks—CAMELS (or Uniform Financial Institutions Ratings System (UFIRS). Research concerning the application of regulatory ratings is limited due to the restricted nature of the ratings themselves. The study of (DeYoung 1998), using the management component of the rating, found that, when comparing well and poorly managed banks, well-managed banks had lower estimated unit costs and higher raw (accounting-based) unit costs, suggesting these costs involve bank management raise expenditures and this leads to bank failure.

Analyst recommendations on equity and debt instruments. These are “quasi ratings” reflecting the opinion of analysts on the purchase or sale of securities. They usually have the following grades: strong buy, buy, hold, sell or strong sell (Karminsky and Polozov 2016).

Public credit ratings (PCR). These ratings are assigned by either international credit rating agencies (IRA) or national rating agencies (NRA) and serve for assessment of creditworthiness of various entities. Hilscher and Wilson (2006) showed that ratings could be a good indicator of systematic risk and default probability of the company. PCR will be considered in detail later in this paper.

Internal ratings. The internal rating is not assigned by an external appraiser (rating agency), but directly by the company itself. For example, a bank may assign internal credit ratings to its borrowers. Both public and internal credit rating systems reflect the borrower's creditworthiness, however, the motivation for each of them is different (Boguslauskas et al. 2011). External credit ratings are used to increase market transparency and reduce information asymmetries between issuers and potential investors. Internal credit ratings are used for management purposes to make decisions quickly and efficiently or independently monitor changes, without waiting for rating agencies to update their ratings.

Business indices. This is the hierarchal structure of specifically interrelated indices; each one characterizes a certain quality of the internal or external environment of the company. The goal of business indices is the benchmarking of companies in the industry. The most widely used systems of business indices are DuPont formula or sustainable growth index (Curtis et al. 2015).

Market implied ratings. These ratings are constructed from information from markets (for example, the prices of traded assets) to directly infer the rating of the object with a minimal amount of subjective input demonstrated. They are applied when the information asymmetry is severe. The disadvantage of them is their volatility. More details are available in Jansen and Fabozzi (2017).

Rankings. Basically, the objects are ranked by certain economic phenomena (for example, efficiency of national governments). Many of these rankings, assigned by the institutions with long-standing reputations, are widely recognized and used by the investment community. These are, for example, the rankings of the World Bank or the global competitiveness indices of the world economic forum (WEF). The difference between ratings and rankings are the following: ratings compare the qualities of the objects using a common scale while rankings compare objects to one another. Rankings have a very wide application not only in economics but also in sports, social life, etc. (Davletshina et al. 2018).

There is a growing interest in the use of ratings in management accountings and controlling, as they allow to benchmark objects within the internal or external environment and signal to the management about the potential short and long-term trends. Thus, the ratings can serve as part of the early warning systems to reflect the weak signals of potential risks to the entity from the environment. Biglaiser et al. (2011) showed that models specified with bond ratings from the credit rating agencies were helpful for predicting economic crisis in late 1990s.

Let us consider PCR in detail as they are the common international measure of credit risk. PCR express a forward-looking opinion about the capacity and willingness of an entity to meet its financial obligations. They provide an efficient, widely recognized, and long-standing measure of relative credit risk. When making investment decisions, investors and other market participants can use ratings as a screening device to match the relative credit risk of an organization or individual debt issue with their own risk tolerance or credit risk guidelines (Karminsky and Polozov 2016).

PCR are assigned with the application of the fundamental analysis. These are the independent assessment of the qualities of all rated entities: financial statements/ratio analysis, financial modelling, knowledge of industry features, management strategy, corporate governance, etc. The rating agencies are regulated by national governments which set the standards of the rating process and rating methodologies.

In many countries PCR are also used as fixture of financial market regulations, the USA first introduced regulatory use of rating in 1931. For example, US financial institutions could satisfy certain regulations (for instance, how much capital they must have) by holding assets with the certain level of PCR. However, the recent financial crises revealed: rating agencies often inflate ratings to reduce the burden for regulated companies. These incentives contributed to an undercapitalized and fragile financial system (Nataf et al. 2018). That is why many international and national regulators called for the elimination of credit ratings in financial regulation and replacing credit ratings with alternative approaches.

These approaches come with challenges which results in still widespread application of PCR in regulation. The obligation to have and disclose credit rating information has been a feature of US macroprudential supervision since 2006. It has become mandatory for all registered and operating US banks to have a credit rating from a public rating agency. Similar obligations have also been introduced for most non-bank depositories in the USA, and this decision was extended to insurance companies. The similar pattern is observed in the regulation of emerging markets (Nataf et al. 2018).

Given all these challenges, investors should be aware of the limitations of PCRs. A credit rating does not reflect other types of risk, such as market or liquidity risks, which may also affect the value of a security. Nor does a credit rating consider the price at which the security is offered or sold. The investor should not interpret a credit rating as investment advice and should not view it as a recommendation to buy, sell, or hold securities. A credit rating is not a guarantee that a financial obligation will be repaid. Consequently, the investment decisions to finance the companies with a certain credit rating are a separate and independent decision, and rating agencies are not responsible for that decision. All other things being equal in the market, a low credit rating signals an increased credit risk or impending default and should cause the investor to demand larger collateral if these additional risks are recognized. Ultimately, if the level of the credit risk is too high for the investor, financing may be refusing.

Credit ratings express risk in relative rank order, which is to say they are ordinal measures of credit risk and are not predictive of a specific frequency of default or loss. The rating summarizes and synthesizes a wide range of risk factors. The process consideration of these factors is shown in Fig. 1. Rating agencies consider both quantitative and qualitative factors. Financial performance is a key component of any credit rating, but these ratings are also based on a number of economic, industrial, and business characteristics, including the assessment of the quality of the management and ownership structures.

The largest international rating agencies are Moody’s, Standard and Poor’s and Fitch Ratings (IRA). They together represent more than 90% of the global credit

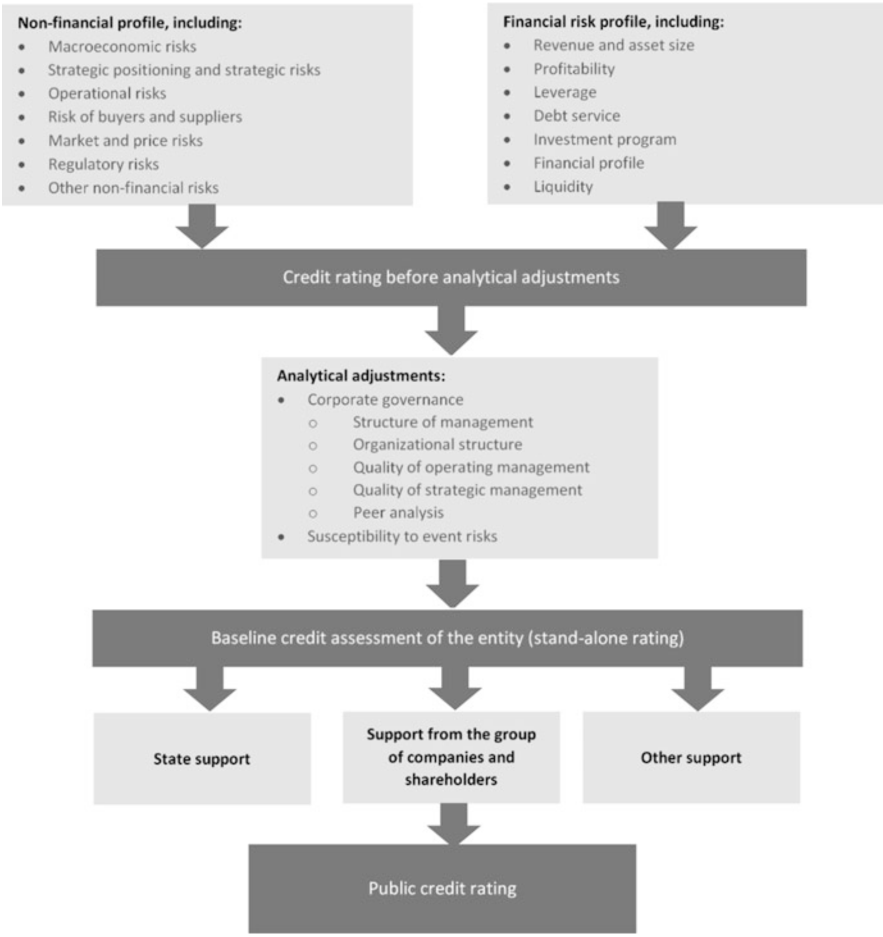


Fig. 1 The process of assigning of public credit rating by the rating agency

rating market. The public ratings assigned by these organizations are recognized by investors all over the world. Standard and Poor's commands about 50% of global rating services.[1] Moody's and Fitch have about 30% and 10% of the market, respectively. This oligopoly position gives IRA significant pricing power (Flynn and Ghent 2018). For example, Moody's operating margin in 2019 was 46% and net income margin was 23%.[2]

[1]Moody's annual report (2018) https://s21.q4cdn.com/431035000/files/doc_financials/annual/2018/MCO-2018-Annual-Report_FINAL.PDF.

[2]Moody's annual report (2018) https://s21.q4cdn.com/431035000/files/doc_financials/annual/2018/MCO-2018-Annual-Report_FINAL.PDF.

Table 1 National rating agencies accredited by the local regulators

Russia	China	India	South Africa
4	8	6	2

Source: Central banks of Russia, China, India and South Africa, 2019

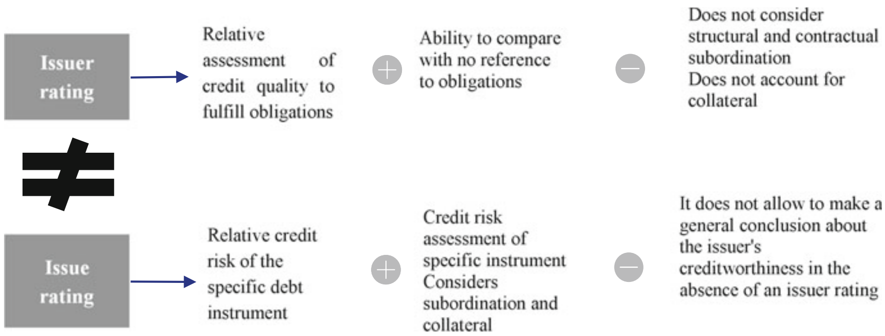


Fig. 2 The differences between issuer and issue rating

In additional to international rating agencies there are a number of NRA. They primarily focus on niche markets where they draw on familiarity with specific economic and political circumstances in specific countries. They are often created by sovereign governments to support the local credit rating assessment of the domestic bond market and maintain independence from IRA. To support NRA their ratings are often use for domestic regulatory purposes. However, the oligopoly position of IRA limits the ability of NRA to gain market share. The majority of NRA are registered in emerging markets such as India, China, and Russia (see Table 1).

Both national and international credit rating agencies can assign a wide variety of ratings. PCR are classified by (1) the type of rating entities; (2) the time horizon of rating (short term or long term); (3) the currency of the issue or the issuers (national scale rating, foreign, and local currency ratings); (4) the type of financial instruments (bond ratings, commercial paper rating, etc.); and (5) the type of entities (sovereign, corporate, financial institutions, etc.).

The two most well known and most common are the issuer's credit rating and the securities issue credit rating. The issuer's credit rating is the rating agency's opinion of the creditworthiness of the entire company which issues securities. The credit rating of the issue refers to a specific financial liability or a specific class of financial instruments and liabilities issued by the company (Bannier and Hirsch 2010). The credit rating of the issue integrates the credit rating of the main issuer, the credit-worthiness of any of the guarantors, insurance agents, other forms of financial liabilities, similar ratings, and the currency of liabilities. As a result, issuer credit rating is not equal to issue credit rating (Fig. 2).

Another important classification of PCR is point-of-time (PIT) and through the cycle (TTC). TTC ratings are the relative assessment of credit quality averaged over an economic business cycle. Credit ratings assigned by rating agencies are usually

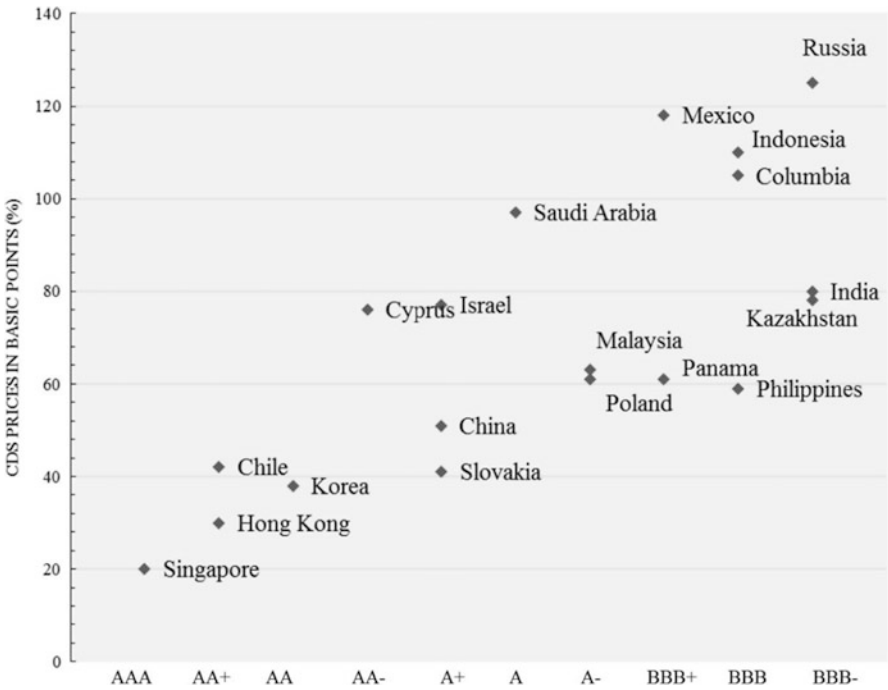


Fig. 3 Five years CDS spread and S and P credit ratings of the largest sovereign as of October 2019. Source: Portals Cbonds and World government bonds

considered TTC ratings even though research has demonstrated that these ratings are also vary with the business cycle (Topp and Perl 2010). In contrast, PIT ratings are the cardinal or ordinal assessment of credit quality over a short horizon, usually for 1 year. They are applied when investors need to estimate the current credit condition of the issuer and/or forecast future conditions. PIT ratings serve as early warning signals of downgrades or upgrades and usually complement the ratings of credit agencies.

The most widespread methodologies of obtaining PIT ratings are Merton structural model and KMV model (Crosbie and Bohn 2003). Another way to build the PIT rating is to infer the information from credit default swaps (CDS). For example, CDS price for sovereign bonds includes country risk which is also considered in the country's credit rating. Thus, fluctuations in CDS price (market approach) should coincide with credit ratings (fundamental approach) (Fig. 3).

Figure 3 shows the adequacy connection between sovereign credit rating and CDS market price. The prices of some are in line with rating (e.g. Poland and Malaysia). However, in some cases this connection does not hold. For instance, Saudi Arabia and Mexico have almost the same CDS price whereas Arabia's rating is higher (A vs. BBB+). The same is true for Cyprus. On the other hand, Philippine's CDS is underpriced comparing to its credit rating.

2 Rating Classification and the Comparison of Rating Methodologies Between Agencies for Banks

Rating agencies use rating scales, symbols, and definitions to express credit risk. Most use a scale of letters and/or numbers, and these symbols are defined by the particular credit rating agency issuing those ratings. The rating scale of IRA is presented in Table 2. Investment grade ratings are assigned to the assets with very low credit risk. Therefore, certain investors whose risk appetite is set low by regulators (banks, insurance companies or pension funds) must invest in them. The rest of the ratings is called sub-investment grade. The rating symbols have no direct association with the probability of default (Karminsky and Polozov 2016). An external user, if he needs estimates of the probability of default can rely on the observed default rates for each rating, taking into account the following: (1) statistics should cover the business cycle; (2) the higher the granularity of the scale, the more

Table 2 Credit rating scales of international rating agencies

Moody's	Standard and poors	Fitch	Description	Grade
Aaa	AAA	AAA	Prime	Investment grade
Aa1	Aa+	Aa+	High grade	
Aa2	AA	AA		
Aa3	AA−	AA−		
A1	A+	A+	Upper medium grade	
A2	A	A		
A3	A−	A−		
Baa1	BBB+	BBB+	Lower medium grade	
Baa2	BBB	BBB		
Baa3	BBB−	BBB−		
Ba1	Bb+	Bb+	Non-investment grade speculative	Sub-investment grade
Ba2	BB	BB		
Ba3	BB−	BB−		
B1	B+	B+	Highly speculative	
B2	B	B		
B3	B−	B−		
Caa1	CCC+	CCC+	Substantial risks	
Caa2	CCC	CCC		
Caa3	CCC−	CCC−		
Ca	CC	CC	Extremely speculative	
	C	C	Default imminent	
C	RD	DDD	In default	
/	SD	DD		
/	D	D		

Source: Websites of international rating agencies, 2019.

Table 3 Examples of observed probabilities of default of Standard and Poor's (for 1981–2018)

Ratings	1 year	2 years	3 years	4 years	5 years
AAA	0.00%	0.03%	0.13%	0.24%	0.35%
AA	0.02%	0.06%	0.12%	0.22%	0.32%
A	0.06%	0.14%	0.23%	0.35%	0.49%
BBB	0.17%	0.46%	0.80%	1.22%	1.64%
BB	0.65%	2.01%	3.63%	5.25%	6.78%
B	3.44%	7.94%	11.86%	14.95%	17.33%
CCC/C	26.89%	36.27%	41.13%	43.94%	46.06%
Investment grade	0.09%	0.25%	0.43%	0.66%	0.90%
Speculative grade	3.66%	7.13%	10.12%	12.56%	14.55%
All rated	1.48%	2.91%	4.16%	5.21%	6.08%

Source: Standard and Poors, 2018 Annual Global Corporate Default and Rating Transition Study.

data required; and (3) one should understand the overall stability of credit ratings over time (Tables 2 and 3).

For each rating the credit rating agency assigns the outlook. It is an opinion regarding the likely rating direction within the next 12 months (Bannier et al. 2010). A stable outlook indicates a low likelihood of a rating change. Around 80% of ratings globally have a stable outlook. Negative, positive, and developing outlooks indicate a higher likelihood of a rating change. A rating review (for upgrade or downgrade) indicates that the rating is under consideration for a change in 60–90 days. Rating are placed on review when further information or analysis is needed to reach a decision on the need of rating change and the magnitude of this change. The review may result in confirmation of the rating.

The regulators require IRA to publish their rating methodologies on the website and make methodologies transparent to investors (Nataf et al. 2018). Despite the certain similarities, these methodologies differ from each other. The comparative analysis of methodologies is necessary to identify the distinctive features of each rating agency. Such an analysis helps to better understand the procedure for assigning a rating and the class of its ratings, and to understand on the basis of what general factors the decision to assign the appropriate rating was made.

We performed such comparative analysis for financial institutions methodologies of IRA. This choice is underpinned by the importance of these institutions in national economics and that the majority of defaults in emerging markets historically happened in the financial industry (Karminsky et al. 2013). For this purpose, we obtained and analyzed information from the websites of Moody's,[3] Standard and Poor's,[4] and Fitch Ratings.[5]

[3]https://www.moodys.com/researchandratings/research-type/methodology/-/003006/003006|005000/-/0/0/-/0/-/-/-/-/-/-/-/-/global/pdf/-/rra|https://www.moodys.com/researchandratings/research-type/methodology/-/003006/003006|005000/-/0/0/-/0/-/-/-/-/-/-/-/-/global/pdf/-/rra (last accessed, January 2020).

[4]https://www.standardandpoors.com/en_US/web/guest/ratings/ratings-criteria (last accessed, January 2020).

[5]https://www.fitchratings.com/site/criteria (last accessed, January 2020).

Table 4 Key components of assignment of stand-alone rating of financial institution

The component	The description of the component
Macro profile	Captures the bank's operating and economic environment
Financial profile	Captures the bank's financial health, gauges key solvency and liquidity ratios and supplemental financial metrics and judgments
Strategic and operating profile	Qualitative judgment of business diversification, opacity and complexity and corporate behavior (strategy and quality of management)
Analytical adjustments	Support and structural analysis captures the affiliate support, liquidity structural analysis, government support and susceptibility to certain event risks

At the first stage IRA analyses the strategic, financial, and operating environment of the financial institution to capture its stand-alone probability of failure in the absence of extraordinary external support. The stand-alone analysis covers the following key components (Table 4).

To make the rating forward-looking, the agency develops forecasts of entity performance for the next 2–3 years in three scenarios: baseline, stress, and optimistic. According to TTC calibration the agency ratings averages the operating and financial profile of the entity for the 3–5 years preceding the analysis date and 2–3 years of forecast. Lastly, to account for adverse conditions, special stress tests are performed. The process of rating assignment is presented at Fig. 4.

However, the set of financial and non-financial coefficients and the methods of their calculation differ from agency to agency. Table 5 shows a comparative analysis of indicators that determine the creditworthiness of financial institutions by IRA.

This comparative analysis of the financial institutions rating methodologies of IRA reveals that despite similar methodologic frameworks applied by all three there are differences in the individual components of the ratings. They include: (1) some differences in qualitative and quantitative metrics and their calculations; (2) differences in design of scoring models; (3) differences in the boundaries of each rating class; (4) the degree of expert judgment in assigning the score; and (5) the agencies provide the different degrees of disclosure of their methodologies. All these drive the split of ratings of various rating agencies (Fig. 5).

The research outcomes of rating split phenomenon are controversary. Ederington (1986) concluded that most new issue splits reflect purely random intra-agency differences in judgment. A small minority appear to be due to systematic inter-agency disagreements regarding rating factors. While Morgan (2002) showed that the less transparent the issues the greater the rating difference. The difference was greatest for banks and financial institutions. Al-Sakka and Ap Gwilym (2012) reported that over 63% of sovereign foreign currency ratings of the nine major credit agencies are split within a range of 2 notches. Similar percentages apply for the two-notch range for both banks and corporates. The ratings assigned by credit rating agencies may differ for two reasons: the agencies have different opinions about the relative positioning of the rated entity (e.g., issuer or security) with respect to the universe of other rated entities; the agencies position the rated entity with respect to the universe of other rated entities in the same way, but they use different symbols to

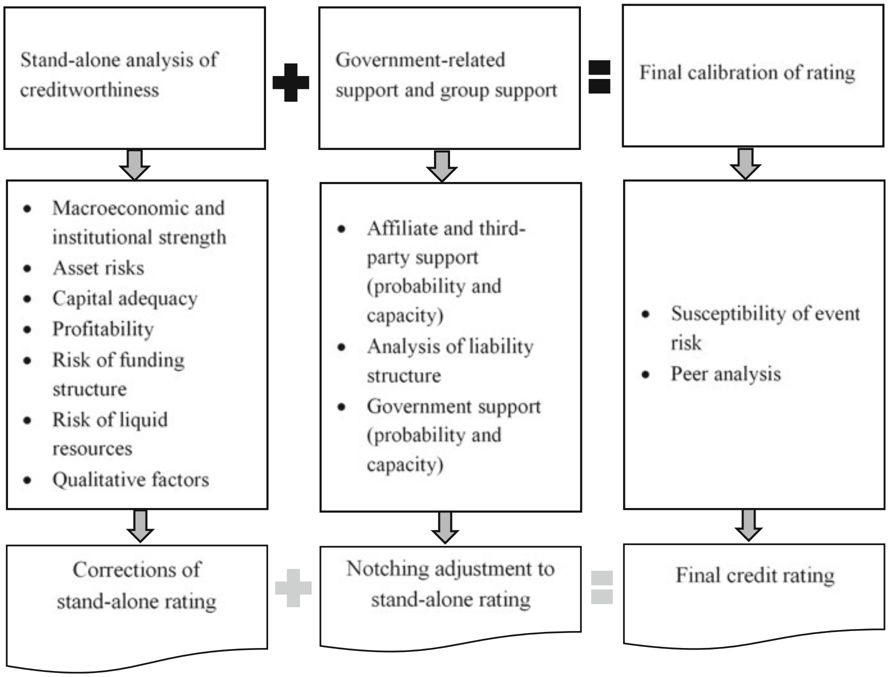


Fig. 4 Methodological framework of assigning ratings to the financial institutions of international rating agencies

represent this position. The split of more than two notches is unlikely assuming that major credit agencies are follow the strict standards of creditworthiness analysis and that the risk being measured by each credit rating agency is the same for a given rated entity at a given point in time. Type 1 differences should disappear when a large number of ratings are considered. The existence of type 2 differences requires a mapping from one rating scale to another. Studying international bank ratings for a five-year period shows that there are type 2 differences for the largest credit rating agencies (Karminsky et al. 2013). Given the rating split, the commonly accepted mapping from one credit rating agency to another by equalizing the ratings denoted with similar letter or letter descriptions is incorrect. The special mechanism of rating scale comparison will be described in the next section.

Table 5 The comparative analysis of methodologies of IRA for rating process of financial institutions

Factors	Moody's rating agency	Standard and poor's rating agency	Fitch rating agency
Macroeconomics and institutional strength	Identification of banking system macro profile, containing the following inputs: Economic variables, such as GDP growth, inflation, real interest rates References to the external sector (capital flows, reserves, and the exchange rate) Credit variables (private-sector credit relative to GDP and its growth rate) A set prices, especially real-estate values Regulatory, institutional and legal frameworks	Identification of banking system macroprofile with the following inputs: Economic policy flexibility Actual and potential economic imbalances Credit risk of economic participants Institutional framework and quality and effectiveness of bank regulation Competitive dynamics in the banking system The macro-profile (economic risk and industry risk) sets an anchor for standalone ratings. The anchor is further adjusted on the results of qualitative and quantitative (financial) analysis	Analysis of operating environment which includes: GDP per capital Ease of doing business ranking Size and structure of economy Economic performance Macroeconomic stability and level of growth in credit Development of financial market Regulatory and legal framework The level of sovereign rating and respective ceilings
Company profile (qualitative) analysis	Business diversification and business mix Opacity and complexity Corporate behavior Market shares and competitive position	Business stability (franchise stability, revenue stability) Concentration or diversity, business diversification The quality of management, strategy, and corporate governance	Franchise (market shares, competitive position) Business mix and earning volatility Organizational structure (complexity, opaqueness, intra-group transactions)
Risk appetite assessment	Not factored, partly considered as part of company profile	Efficiency of managing growth and changes in risk position Impact of risk concentration and risk diversification Complexity Risk governance Operational, credit risk, and market risk management	Underwriting factors (lending and credit standards, investment policy) Risk controls (control framework) Credit and balance sheet expansion Market risk
Asset quality	Impaired loans/gross loans	Not highlighted	Impaired loans/gross loans
Earnings and profitability	Net income/tangible assets	Normalized operating profit/risk weighted assets	Operating profit/risk weighted assets

(continued)

Table 5 (continued)

Factors	Moody's rating agency	Standard and poor's rating agency	Fitch rating agency
Capital and capitalization	Common equity/ adjusted risk weighted assets	Headroom over the prudential ratios Adjusted capital (common equity + hybrid instruments)/adjusted risk weighted assets	Adjusted core capital/adjusted risk weighted assets
Funding and liquidity	Market funds/tangible banking assets Liquid banking assets/ tangible banking asset	Loans/customers deposits Long-term funding ratio Stable funding ratio Net broad liquid assets/ short-term deposits Liquid assets/wholesales funding	Loans/customers deposits
Government support	Joint default correlation analysis between the ability and capacity of sovereign support The probability of support	Notching adjustment to stand-alone rating depending on: Likelihood of extraordinary sovereign support History of support Systemic importance of the bank Government interference	Notching adjustment (up to 2–3 notches) to stand-alone rating depending on: Sovereign ability to support Sovereign propensity to support the system and the bank Policy bank support
Affiliate support	Joint default correlation analysis between ability and capacity of third-party support Adjusted on support record	Notching adjustment to stand-alone rating depending on status of group members (based on strategic importance of group members) and assessment of group credit profile	Notching adjustment (up to 2–3 notches) to stand-alone rating depending on: Ability to support subsidiary and subsidiary's ability of using support Parent propensity to support Subsidiary performance and prospects Support records
Analysis of liability structure	Performed, qualitative methods	Performed, includes analysis of loss absorption capacity and applying notching	Performed, qualitative methods
Weights in scoring models	Weights are disclosed	Weights are not disclosed	Weights are not disclosed
Level of disclosure and details for the public	High	Medium	Medium

(continued)

Table 5 (continued)

Factors	Moody's rating agency	Standard and poor's rating agency	Fitch rating agency
The scoring model can be reproduced by third partner	Partially, with substantial certainty	Partially, with medium certainty	Partially, with medium certainty

Retrieved from: Moody's bank's rating methodology (https://www.moodys.com/research/Banks%2D%2DPBC_1128883) & Mining Rating Methodology (2018), S&P bank's rating methodology and assumptions (https://www.standardandpoors.com/en_US/web/guest/article/-/view/sourceId/6921376), Fitch's Bank Master Rating Criteria (https://www.fitchratings.com/site/re/10044408), designed by authors.

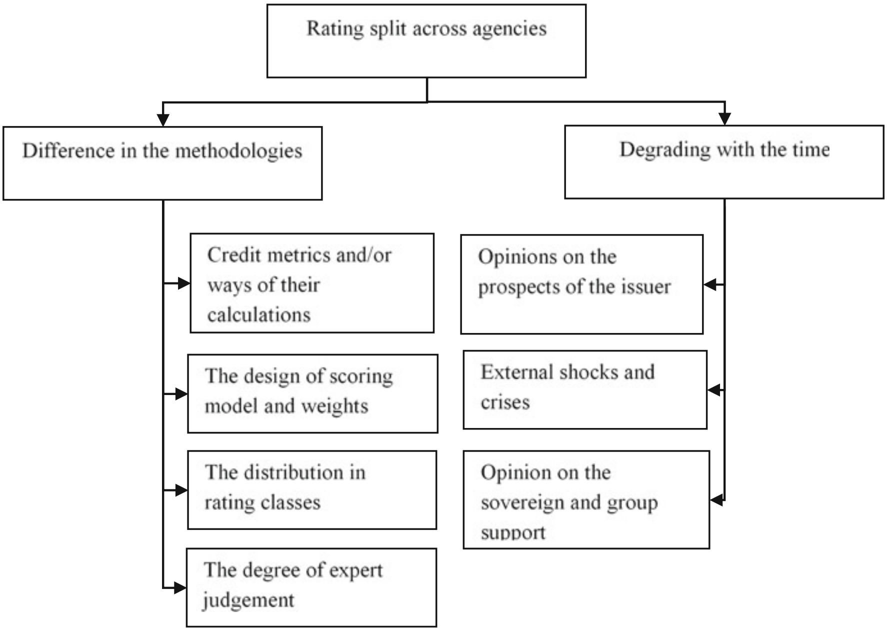


Fig. 5 The differences in methodologies across rating agencies may result in rating split

3 Credit Ratings in Emerging Markets. The Comparison Problem

Emerging markets have specific features that investors usually consider judging credit ratings of the entities from this market. IRA adjusts their methodologies to include additional risk factors specific to emerging economies (Kim and Wu 2018). These risk factors include the following:

Emerging markets are more volatile. Entities in emerging markets experience large fluctuations in business and operating conditions, which affect their creditworthiness. This volatility is underpinned by smaller market size, less market maturity and fluctuations in macroeconomic factors, domestic exchange rate, and interest rates.

Sovereign issues constrain ratings. Sovereign distress is associated with disruption to the economic, business, and financial environment, resulting in a decline in the credit quality of entities in the affected countries compared to entities from countries without sovereign distress. To account for this effect, IRA limits the ratings of all entities to sovereign. However, corporations with a stronger sovereign creditworthiness and characteristics that make them unlikely to default when the sovereign defaults can get a rating well above their sovereign (see discussion below).

Sanctions and other restrictions have a negative impact on the rating. Sanctioned businesses lose the ability to raise the necessary financing at an affordable price and/or sell products to certain customers. Such companies, in addition to the short-term impact of liquidity gaps, outflow of customers and suppliers, experience a long-term credit crisis due to insufficient investment and reduced business potential. In addition, sanctions have a long-term negative impact on the country's economy, for example, on the decline in investor confidence, the country's export potential, as well as on technical modernization.

Institutional governance and lower transparency constrain the rating. In emerging markets, the flow of information is less consistent, transparency is lower, and legal systems are less predictable and reliable. Entities in emerging markets are also subject to constant changes in the local regulatory regime, which can be unpredictable and often have a negative impact on entities.

Liquidity is weaker in emerging capital or banking markets. Local financial markets often lack depth and are more stressed. They tend to have lower liquidity. While demonstrated access to foreign currency capital markets can be beneficial to corporate liquidity by opening multiple sources of financing, it can also lead to significant volatility in the value of local currency liabilities due to changes in exchange rates. Sanctions will further complicate the liquidity problem.

Ferri and Liu (2003) discovered that the contribution of sovereign risk to firm's rating was high in emerging economies but was negligible in developed countries. Mulder and Perelli (2001) showed that the ratios of investment to GDP and sovereign debt to exports have the largest impact in explaining the change in corporate ratings in emerging markets. Williams et al. (2013) analyses the effect of sovereign rating actions on the credit ratings of banks in emerging markets, using a sample from three global rating agencies across 54 countries for 1999–2009. He found that the sensitivity of bank ratings in emerging markets is affected by the countries' economic and financial freedom; by macroeconomic conditions and bank ownership structure.

Our analysis showed that market volatility and weak liquidity in emerging markets are considered by IRA in their regular analysis of entities' business and financial risk. These considerations include: (1) forward-looking assessment of the adequacy and reliability of cash resources and cash flows in relation to the entity's liabilities; (2) conducting strict stress tests to account for volatility; or (3) assessing the ability of entities to withstand shocks due to negative shifts in local currency and interest rates. They also consider other indicators such as the depth of the domestic

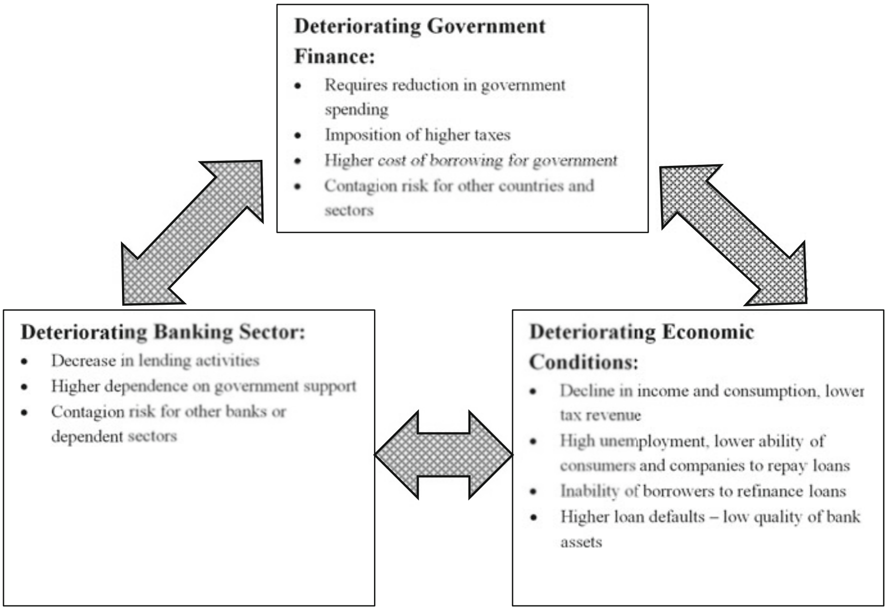


Fig. 6 Credit linkages between sovereigns, financial institutions and the corporates

capital market; the strength of its banking system; and the quality of the company's accounting and corporate governance practices, which can also influence the risk of default. The agencies also take into account the following considerations: (1) market data and information may be difficult to access; (2) political and policy environments may not be transparent; (3) disclosure standards are sometimes not ideal; (4) governance and transparency practices may not meet developed market standards.

For rating agencies, the biggest challenge in assessing ratings in emerging economies is to properly understand the impact that a sovereign's broader credit profile may have on other lenders residing in this market as a result of credit linkages (Fig. 6).

All entities in the same sovereign environment either are subject to the transfer of shocks in a given market, between sectors or through the domestic banking system. The entities will be subject to varying degrees of protective action that the sovereign may take. The sovereign problems often spread over the whole economy reducing economic activity. To account for these systemic risks, IRA limit the corporate ratings at the sovereign level.

In some cases, the corporate may be assigned a higher rating than its sovereign. They include (1) entities with a credit profile that is fundamentally stronger than that of the sovereign; (2) entities which are fenced off from local macroeconomic and financial problems; (3) the entities with the overwhelming share of income, assets, and capital obtained from sources outside of the sovereign; (4) multinational entities which can diversify risks across markets; and (5) entities receiving external support independent of the sovereign environment.

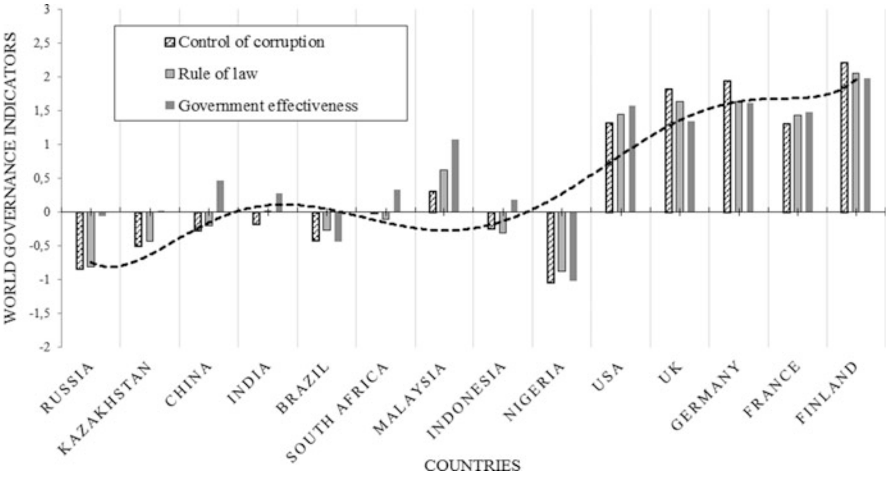


Fig. 7 Institutional governance indicators may signal elevated credit threats for corporates. Source: World Bank Governance Indicators for 2018

For example, Russia's Lukoil (Baa2 with a stable outlook) is rated by Moody's one notch higher than the Russian Federation (Baa3 with a stable outlook). The agency explains these by the robust business profile, the strong financial metrics of the company, its solid liquidity position and the significant share of revenue in foreign currency. These give the company insulation from local market stress and helps cushion the effects of foreign currency debt revaluations on their leverage metrics.[6]

Another constraint for the rating of companies in emerging markets is the weaker governance and the higher political risk than in developed markets. To the measure broader institutional, political and regulatory environment IRA often use the World Bank worldwide governance indicators. For example, the Government Effectiveness Index, the Rule of Law Index, and the Control of Corruption Index show material differences across emerging markets (Fig. 7). The higher the index, the better the institutional environment in the country.

For example, Russia and Kazakhstan are markets where questionable corporate governance and a weak institutional environment negatively influence ratings, while the underlying business and financial profiles of companies are strong. Management practices usually come under the spotlight of agencies in Latin America and Commonwealth of Independent States (CIS). In these regions, corporate governance practices have ranged widely, from market leaders who follow established "best practices" to small family-owned and unlisted firms that often exhibit questionable governance practices. In the latter, while experienced and professional management

[6]Moody's takes rating actions on 16 Russian non-financial corporates; https://www.moodys.com/research/Moodys-takes-rating-actions-on-16-Russian-non-financial-corporates%2D%2DPR_394481.

Table 6 The rating scale of national agencies in emerging markets

Agency	Rating scale
ACRA (Russia)	From AAA(RU) to D(RU)
Expert RA (Russia)	From ruAAA to ruD
CRISIL (India)	From CRISIL AAA to CRISIL D
ICRA (India)	From [ICRA]AAA to [ICRA]D
GSR rating (South Africa)	From AAA(ZA) to D(ZA)

may be employed, such structures raise concerns about the protection of bondholders and the ability of the founding families to exercise control at the expense of other stakeholders.

In contrast to IRA, NRA in emerging markets usually assign the national scale ratings. This is often underpinned by the requirements of local legislation (examples are Russia or China). National scale ratings are the opinion of the relative creditworthiness of issue or issuer relative to benchmarks in national financial markets. The commonly used benchmark is the obligations of national government or their derivatives. In the national scale, the obligations of the national government are assigned the highest rating (e.g. AAA). National scale ratings do not incorporate international comparative risk factors; do not allow comparison between countries and instruments worldwide; do not address certain sovereign risks (such as the possible imposition of foreign exchange controls); and assess the creditworthiness of entities only in the domestic currency. Therefore, ratings of NRA cannot be compared with ratings provided on an international scale by IRA.

National credit rating agencies for rating scale use alpha-numerical symbols adopted from Fitch Ratings and Standard and Poor's from AAA to D. However, the symbols come with the addition of a prefix or suffix to identify the country for which the national rating scale applies. An example of the rating scales of several national rating agencies is presented in Table 6.

Many NRA, in addition to national rating scales, assign international scale ratings. For this purpose, they also adopted rating symbols from IRA (from AAA to D). Usually, no prefix is attached to the symbols. The assigning of ratings on the international scale should remove the constraints specific for national scale ratings. However, the research shows that ratings assigned by IRA and NRA are not comparable. Firstly, the rating standards of NRA can be well below those of IRA, which can result in the inflation of ratings. NRA have not yet established the reputation of organizations free of conflicts of interest, not compromising client confidentiality or having efficient compliance and internal control procedures. Secondly, the methodologies of IRA and NRA may rely on a different set of rating factors (Table 7). Lastly, the sovereign ratings assigned by IRA and NRA may differ by several notches. This creates different anchors for assessing the relative creditworthiness of entities and different rating ceilings (see examples in Table 8).

These differences were noted in various researches. Livingston (Livingston et al. 2008) in his study of Chinese bond ratings found, firstly, that bond investors differentiate ratings from different agencies based on their perceived reputation.

Table 7 Rating factors in methodologies of IRA and NRA are not fully matched (example for non-financial companies in the mining sector)

Group of rating factors	Lianhe ratings global (China)	Moody's
Scale	Not factored	Revenue
Business profile	Market position, competitiveness, diversity, operating efficiency	Products, markets, competitive position, pricing trends, cost efficiency, location of operations, technologies, market demand, susceptibility to environmental, regulatory and political risk
Profitability	Not specified	Margin of earnings before interest and taxes (EBIT)
Leverage	Gross debt over EBITDA (earnings before interest expenses, depreciation and amortization) Debt over capitalization (Total debt + equity)	Gross debt over EBITDA (cash flow from operations after dividends paid) over gross debt Debt over capitalization
Debt interest coverage	EBITDA over interest	EBIT over interest
Liquidity	Current ratio, quick ratio, cash ratio, absolute liquidity ratio	Forecasted liquidity as (cash and equivalents + cash flow from operation + long-term committed available credit line+ equity inflow) over (capital expenditures, dividends and debt repayment)
Financial policy	Not factored	Factored

Retrieved from: Moody's website (https://www.moodys.com/research/Mining%2D%2DPBC_1089739), Lianhe Ratings Global website (https://lhratingsglobal.com/rating-methodology-3/), last accessed January 2020 & General Corporate Rating Criteria (2018)

Table 8 Sovereign ratings of NRA and IRA often are not equal and may be inflated

Country	Sovereign rating assigned by ACRA (Russia)	Sovereign rating assigned by Fitch ratings	Difference in ratings (in notches)
Russia	A−	BBB	+2
Bulgaria	A−	BBB	+2
Hungary	BBB	BBB	0
Czech Republic	AA	AA−	+1
Kazakhstan	BBB+	BBB	+1
Romania	BBB	BBB-	+1
Slovak Republic	A+	A+	0

Retrieved from: Russia's ACRA (www.acra-ratings.ru), Fitch Ratings (www.fitchratings.ru), last accessed January 2020.

Secondly, rating standards vary significantly among different Chinese NRA. Thirdly, while informative, Chinese bond ratings, even when assigned on international scale, are not comparable to the ratings by IRA, although they use similar alpha-numerical symbols. Fourthly, Chinese rating scales are very coarse. A notch difference in ratings results in, on average, a difference of 58 basis points in yields. Thus, a one-notch difference in Chinese ratings is likely equivalent to a one-letter difference in international ratings. In a similar study Jiang and Packer (2019) examine issuer ratings on Chinese firms assigned by both domestic and international agencies on an international scale and find ratings by domestic credit ratings agencies are about 6–7 notches higher than those by international credit ratings agencies.

Therefore, the ratings of various rating agencies, especially NRA and IRA cannot be compared by simple matching. Special techniques of rating scale comparison must be applied (see further discussion in paper 4). However, these techniques could be ineffective for NRA given the possibility of their low rating standards NRA.

4 The Regulation of Rating Activities in European and BRICS Countries

Excessive reliance on credit ratings by investors and regulators can have negative consequences. There were numerous cases when debt securities had investment grade ratings a couple of days before default (Nataf et al. 2018). These cases resulted in the tightening of rating activity regulation in developed markets. In USA, the main initiatives included Sarbanes-Oxley act of 2002; Credit Rating Agency Reform Act of 2006 and Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010. The latter (1) eliminated the references to the credit ratings in regulations; (2) required Securities and Exchange Commission (SEC) to oversight rating activities and take measures to reduce the conflict of interest while fostering corporate governance; and (3) set disclosure requirements (Nataf et al. 2018).

In EU, the main responses for rating crises were Credit Rating Agency regulations of 2009, 2011, and 2013. The European Securities and Market Authorities (ESMA) was established with the status of the only supervisory arm in Europe for the registrar and supervision of rating agency. The regulation (1) set the requirements for the independence of rating agencies boards; (2) prohibits rating agencies from consulting services and services to related parties; (3) set the disclosure requirements for rating agencies including the disclosure of methodologies and quantitative assumptions. The regulation aimed for a reduction in the “mechanical” reliance on external ratings. Lastly, the regulation called for the recognition of structured finance ratings only if the issuer appointed a minimum of two independent rating agencies (Nataf et al. 2018).

Emerging markets inherited these regulatory changes in their regulatory practices. Since 2012 in BRICS countries rating regulation has been tightened to follow the standards of International organization of security commission (IOSCO).[7] It sets (1) the structure and duties of CRA board of directors, internal controls, and outsourcing arrangements; (2) independence and conflicts of interest; (3) the quality and integrity of credit ratings; (4) rating presentations; and (5) disclosure and record keeping.

However, the business environments for NRA varies from country to country (Bhogal 2017). In Brazil and South Africa the regulation does not give preferences to NRA over IRA and provides minimal usage of PCR in the regulation. For example, banks are not required to rely on PCR for prudential and operational purposes while the government gradually remove references to PCR in regulations. As a result, the market share of NRA in these countries are small. They serve the economy segments which are not the primary targets for IRA (e.g. small and medium enterprises or leveraged financial transactions). In Brazil there are four main NRA registered with the country's Securities and Exchange Commission: SR Rating, Austin Rating, Liberum Ratings and Argus. In South Africa the major share of domestic rating market is taken by national rating agency Global Credit Rating Co. Ltd. that applies the methodologies very similar to those of S and P (Bhogal 2017).

Unlike these countries, in India credit ratings are more widely recognized for regulatory purposes. For example, local pension funds can only invest in debt securities that have two ratings. Moreover, Securities Exchange Board of India (SEBI) stipulated that ratings are compulsory on all public issues of debentures with maturity exceeding 18 months. SEBI has also made ratings mandatory for acceptance of public deposits by Collective Investment Schemes. Additionally, capital protection-oriented funds, IPO grading, bank loans (for Basel II/III capital adequacy calculation) are require mandatory CRs before issuance.[8]

NRA in India are regulated by SEBI. It covers registration, obligations, disclosure, conflict of interest, accountability and enforcement. The credit rating agencies in India can be registered by SEBI if it has a minimum net-worth of 50 million rupees (around $675,000), and adequate infrastructure, professionals, and employees to carry out the activity of issuing credit ratings. The agencies must publicly disclose their rating criteria and rating processes, standardize press releases for rating actions, disclose ratings both in the case of non-acceptance by issuers and non-cooperation by issuers and disclose a delay in the periodic review of ratings. The regulation is constantly tightened following the requirement of US and EU regulations as well as following several domestic cases of missed defaults. The recent case is the loan default of Infrastructure Leasing and Financial Services Ltd.[9] In response, SEBI now requires companies to mandatory disclose to the agencies the details on delayed loan repayments and facts of possible defaults. The largest domestic Indian NRA are

[7] https://www.iosco.org/library/pubdocs/pdf/IOSCOPD482.pdf.

[8] Securities and Exchange Board of India (Credit Rating Agencies) Regulations. http://www.sebi.gov.in.

[9] https://economictimes.indiatimes.com/markets/stocks/news/sebi-floats-tighter-norms-for-defaults-disclosure-with-rating-agencies/articleshow/71299007.cms.

controlled by CRA: ICRA (Moody's), CRISIL (S&P), and India Ratings and Research (Fitch Ratings). The largest one is CRISIL, which has 60% of the domestic credit rating market share. The list of largest Indian agencies includes: (1) Credit Analysis and Research Limited (CARE) promoted by Industrial Development Bank of India and providing full scope of rating and auxiliary services; (2) Small and Medium Enterprise Rating Agency of India (SMERA) which functions exclusively for micro, small and medium enterprises; and (3) Credit Rating Agency (ONICRA) which provides credit ratings, conducts risk assessment and provides analytical solutions to individuals, corporates and small and medium enterprises. ONICRA also provides IPO grading services (Sunitha and Sanjeev 2018).

In contrast to the centralized regulation in India, Chinese regulation is decentralized and more complex (Jing 2015). Many types of bonds are traded in separately regulated distinct domestic markets: the over-the-counter market, the exchange-based market, and the inter-bank market. Even though the People's Bank of China (PBC) is mainly responsible for governing the credit rating system, NRA must be accredited in separate markets by separate authorities in order to rate the bonds trading in these markets (Livingston et al. 2018). As of end-2019 there were eight major NRA in bond market: joint domestic–foreign ventures; one engaged in technical cooperation with a foreign enterprise, and the remaining five are domestically funded agencies. The largest NRA in China: Dagong Global Credit Rating, China Chengxin International Credit Rating Co, Lianhe Rating Global, Golden Credit Rating International Co, Pengyuan Credit Rating Co and Shanghai Brilliance Investors Service.

Chinese authorities use domestic CR heavily for regulatory purposes. Examples include (1) establishing minimum rating thresholds for bonds eligible for public offering; (2) calibrating capital requirement of commercial banks; (3) investment guidance for insurance funds, money market funds, and lombard operations (Livingston et al. 2018). These thresholds create incentives for inflating the ratings (as we discussed in 3.3). For example, corporate bonds to be issued to general (public) investors in the exchange bond market are required to have an initial new issue rating of AAA on the domestic rating scale from a NRA.[10] To be eligible for the competitive bidding or centralized auction method for issuing and trading in the main exchanges for qualified investors the bonds must be AA or above. Bonds issued by non-residents are required to have their bonds rated AA or above on the national scale by at least two credit rating agencies (one of them must be NRA). In light of such requirements, bond issuers demand NRA to become a "rubber stamp" needed to meet the prerequisites of listings. The other mentioned problem of Chinese NRA regulation is that the rating decisions are based upon as cited "limited or even bad information" (Jing 2015). Moreover, PBC last year found failures to protect against conflict of interests (five NRA are at least partly state owned), insufficient

[10]ASEAN+3 bond market guide. Exchange bond market in the People's Republic of China. Asian Development Bank, October 2019. URL: https://asianbondsonline.adb.org/documents/abmf_prc_bond_market_guide_2019.pdf (last accessed in January 2020).

quality controls on ratings and failures to update them promptly in response to important developments.[11] These problems intensified recently as the Chinese economy slowed down. Fitch reported that in 2019 bond defaults in China were at an all-time high with 45 issuers defaulting with a combined face value of $17 billion.[12] The July 2017 decision of PBC to allow IRA to rate onshore bonds may increase foreign investor confidence in assigned ratings and improve quality of ratings (the first rating from IRA was assigned in July 2019). Additionally, in September 2018, PBC and the China Securities Regulatory Commission announced the promotion of the gradual unification of rating qualifications in different markets in order to promote the unified regulation of the credit rating industry. Yet, there is no research so far which indicate that these changes have alleviated the problems described above.

The Russian rating industry began to develop in the mid-1990s, when IRA established subsidiaries in the country and began to assign ratings (Moisev 2009). However, due to the reduction in domestic debt issues after the financial crisis of 2009 and limited interest from clients (mainly banks), most NRA were unable to continue working (Jeeyoung 2016). The regulatory landscape has significantly changed since 2015 when a new law regarding rating agencies was passed. According to the law, only NRA or subsidiaries of IRA registered with the Bank of Russia (CBR) may assign ratings for debt issued onshore. These ratings must be graded on a national scale. The agencies are not allowed to withdraw the ratings of onshore debt for any reason including the decision of foreign regulatory arms (e.g. sanctions).

In accordance to the law, CBR became the sole regulatory body of rating activity in the country. It maintains a register of rating agencies, sets methods for calculating the amount of equity of credit rating agencies, performs audit and monitoring of rating activities of NRA. The law also sets certain conditions and prohibitions which (1) prevents conflicts of interest between NRA and issuers; (2) ensures independence and objectivity of rating analysts; and (3) establishes efficient corporate governance. For example, the law stipulates restrictions on financial analysts' equity holdings and restricts the coverage period of the same issuer to 1 year. The rating decisions must be taken by the rating committees consisting of five analysts. Corporate governance requirements suggest that the board of large NRA must contain at least two independent directors (Sasso 2016).

The law sets the minimum amount of equity capital (50 million rubles (around $700,000 as of end-2019) and prohibits financial organization owning an equity stake in exceed of 20%. If NRA violates the law, the regulator can replace the personnel of management bodies and internal control and remove NRA from the register. These changes resulted in a revival of domestic NRA. As of end-2019 four NRA are registered with CBR but only ratings of two of them (ACRA and Export

[11] https://www.ft.com/content/e6ea3c7c-55f8-11e9-91f9-b6515a54c5b1 (last accessed in January 2020).

[12] https://www.fitchratings.com/site/pr/10105201 (last accessed in January 2020).

RA) are accepted for regulatory purposes, the other two NRA and NKR are in start-up phase. Conversely, IRA have no incentives to register their subsidiaries with CBR. They continue to rate only debt issued by Russian companies overseas (Sasso 2016).

The law also imposes disclosure obligations on NRA regarding the methodologies, models and key assumptions used in the rating process. NRA must establish an independent body responsible for the development and approval of the methodologies. The methodologies must be rigorous, systematic, continuous, and subject to validation based on historical experience. In addition, NRA must monitor and review the impact of changes in macroeconomic or financial market conditions on credit ratings. If there are changes in the methodologies, NRA must report to CBR how it will affect the ratings already assigned. Additional disclosure requirements include periodical reporting to CBR of the rating performance, default statistics, and list of agency's largest clients with revenue more than 5% (Sasso 2016).

National scale ratings are used by Moscow Stock Exchange as a criterion of including debt into the certain quotation lists. The ratings are also applied for regulatory purposes. For example, the financial institution must have the rating of at least "A-" to keep federal funds on deposits. The issued debt securities must have the rating of at least "A+" to be accepted by CBR as part of REPO deals. The state pension fund can only accept debt securities with ratings of above "A." The minimal rating thresholds are also part of the regulatory restrictions for certain investment activities of banks, pension funds, dealers, and insurance companies.

5 Conclusion

In this section, we analyzed the literature about credit rating systems and classified the existing rating systems and explained its application. We also performed detailed analysis of methodologies of international and domestic rating agencies to (1) discover key rating factors and compare them across the agencies; (2) analyze causes for rating split of various agencies; (3) identify key adjustments to methodologies which rating agencies made to consider additional risk factors specific to emerging economies. Lastly, we made a detailed analysis of rating regulations and norms in BRICS countries and how these regulations affected the quality of ratings.

In the field of the regulation of rating activities, emerging countries follow the regulatory trends set in Europe and the USA. The common features of the regulation are: (1) establishing requirements for the structure, corporate governance, methodologies, and analytical personnel; (2) demanding the registration of rating agencies with the local regulator; (3) periodical monitoring of rating activities by the regulator. However, the quality and depth of regulation depends significantly of the maturity of rating industry of the particular countries. The rating industry in Brazil, South Africa, and partially India is quite mature. These result in regulation, which clearly establish the industry participants, promote the competition while setting best practices of corporate governance, and restrict the usage of ratings in regulation.

Conversely, rating industries in China and Russia are relatively young. The rating activity regulation in these countries is still untested. The practices of using ratings in regulation are still in place. Improvements are necessary in such fields as (1) promoting competition in the industry and removing barriers for IRA; (2) fostering strong corporate governance practices aimed at conflict of interest reduction; (3) reducing the regulatory franchise of rating agencies to improve objectivity and prevent the inflation of ratings.

References

Al-Sakka, R., & Ap Gwilym, O. (2012). The causes and extent of split sovereign credit ratings in emerging markets. *Emerging Markets Finance and Trade, 48*(1), 4–24.

Bannier, C. E., Behr, P., & Guttler, A. (2010). Rating opaque borrowers: Why are unsolicited ratings lower? *Review of Finance, 14*, 263–294.

Bannier, C. E., & Hirsch, C. W. (2010). The economic function of credit rating agencies—What does the watchlist tell us? *Journal of Banking and Finance, 34*(12), 3037–3049.

Bhogal, P. (2017). Rethinking the relevance of existing credit rating agencies to BRICS. ORF Occasional Paper, 113.

Biglaiser, G., De Rouen, K., Jr., & Archer, C. C. (2011). Politics, early warning systems, and credit rating agencies. *Foreign Policy Analysis, 7*(1), 67–87.

Boguslauskas, V., Mileris, R., & Adlyte, R. (2011). New internal rating approach for credit risk assessment. *Technological and Economic Development of Economy, 17*(2), 369–371.

Crosbie, P., & Bohn, J. (2003). *Modeling default risk* (pp. 1–31). New York: Moody's KMV Company.

Curtis, A., Lewis-Western, M. F., & Toynbee, S. (2015). Historical cost management and the use of DuPont analysis by market participants. *Review of Accounting Studies, 20*(3), 1210–1245.

Davletshina, L. M., Nabieva, L. G., Ustinov, A. E., Muhamadiyarova, R. I., & Mingazova, A. I. (2018). Modeling of the mechanism of the main macroeconomic categories' interrelations of national economies. *Journal of Advanced Research in Dynamical and Control Systems, 10*(7), 1860–1868.

DeYoung, R. (1998). Management quality and X-inefficiency in National Banks. *Journal of Financial Services Research, 13*(1), 5–22.

Ederington, L. (1986). Why split ratings occur? *Financial Management, 15*(1), 37–47.

Ferri, G., & Liu, L. G. (2003). How do global credit-rating agencies rate firms from developing countries. *Asian Economic Papers, 2*(3), 30–56.

Flynn, S., & Ghent, A. (2018). Competition and credit ratings after the fall. *Management Science, 64*(4), 1672–1692.

General Corporate Rating Criteria. (2018). *Lianhe ratings global*. https://lhratingsglobal.com/rating-methodology-3/.

Hilscher, J., & Wilson, M. (2006). Credit ratings and credit risk: Is one measure enough? *Management Science, 63*(10), 3414–3437.

Jansen, J., & Fabozzi, F. J. (2017). CDS implied credit ratings. *Journal of Fixed Income, 26*(4), 25–52.

Jeeyoung, H. (2016). Analysis of regulation of credit rating agencies in Russia and policy recommendations. *Business in Law, 5*, 41–48.

Jiang, X., & Packer, F. (2019). Credit ratings of Chinese firms by domestic and global agencies: Assessing the determinants and impact. *Journal of Banking and Finance, 105*(C), 178–193.

Jing, B. (2015). Regulating Chinese credit rating agencies: Progress and challenges. *Journal of Money Laundering Control, 18*(1), 66–80.

Karminsky, A., Hainsworth, R., & Solodkov, V. (2013). Arm's length method for comparing rating scales. *Eurasian Economic Review, 3*, 113–135.
Karminsky, A., & Polozov, A. (2016). *Handbook of ratings in economy, sport and society* (p. 356). Cham: Springer.
Kim, S.-J., & Wu, E. (2018). Sovereign credit ratings, capital flows and financial sector development in emerging markets. *World Scientific Studies in International Economics, 64*, 431–466.
Livingston, M., Naranjo, A., & Zhou, L. (2008). Split bond ratings and ratings migration. *Journal of Banking and Finance, 32*(8), 1613–1624.
Livingston, M., Poon, W. P. H., & Zhou, L. (2018). A study of the Chinese bond market and credit rating industry. *Journal of Banking and Finance, 87*, 216–232.
Mining Rating Methodology. (2018). *Moody's investors service. Rating methodology*. https://www.moodys.com/research/Mining%2D%2DPBC_1089739.
Moisev, S. (2009). The regulation framework of credit rating agencies activity in the national market. *Voprosy Ekonomiki, 2*, 39–50.
Morgan, D. (2002). Rating banks: Risk and uncertainty in an opaque industry. *The American Economic Review, 92*(4), 240–265.
Mulder, C., & Perelli, R. (2001). *Foreign currency credit ratings for emerging markets*. Washington, DC: IMF.
Nataf, O., De Moor, L., & Vanpee, R. (2018). Was regulation (EC) #1060/2009 on credit rating agencies effective. *Journal of Banking Regulation, 19*(4), 299–316.
Sasso, P. (2016). A critical analysis of the recent Russian regulation on credit rating agencies. *Russian Law Journal, 4*(2), 62–82.
Sunitha, G., & Sanjeev, R. (2018). An analytical study on impact of credit rating agencies in India's development. *International Journal of Pure and Applied Mathematics, 119*(15), 1281–1294.
Topp, R., & Perl, R. (2010). Through-the-cycle ratings versus point-in-time ratings and implications of the mapping between both rating types. *Financial Markets, Institutions & Instruments., 19*, 156–165.
Williams, G., Al-Sakka, R., & Gwilym, O. (2013). The impact of sovereign rating actions on bank ratings in emerging markets. *Journal of Banking and Finance, 37*(2), 563–577.

Aggregation of Rating Systems for Emerging Financial Markets

Sergei Grishunin, Natalia Dyachkova, and Alexander Karminsky

Abstract This paper examines the issues of the aggregation and comparison of the credit ratings of various economic agents for risk management purposes in a commercial bank. The empirical results of the study make it possible to increase the assessment of credit risks based on the constructed system of aggregating credit ratings for industrial companies and commercial banks. The work also confirms the relationship between the level of assigned credit ratings and the various phases of the credit cycle. The dynamics at the macroeconomic level shows that the credit ratings of various economic agents change in different directions and are out of sync with time correlation of credit cycles in various phases. The main scientific result of the study is an aggregate-based approach for credit risk evaluation of various economic agents and to develop the quantitative methods for assessing the relationship between the level of credit ratings and the credit cycle.

Keywords Aggregation of rating systems · Credit ratings · Rating agencies · Commercial banks · Industrial companies · Credit risk · Credit gap · Credit cycles

JEL G17 · G21 · G24 · G32 · E51

S. Grishunin (✉)
National Research University Higher School of Economics, Faculty of Economic Science, ZAO Deloitte & Touche CIS (Russia), Moscow, Russia
e-mail: sgrishunin@hse.ru

N. Dyachkova (✉) · A. Karminsky
National Research University Higher School of Economics, Faculty of Economic Science, Moscow, Russia
e-mail: ndyachkova@hse.ru

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_4

1 Rating Aggregation Systems and Its Opportunities

The development of the rating services market over the past decade has made it possible to use credit ratings in various areas of the economy and finance. It is already difficult to imagine the modern financial market without the everyday use of credit ratings. In the world of developed financial relations, credit ratings play the role of indicators of the financial condition of potential borrowers. That is why the use of credit ratings in models of aggregation and forecasting of financial risks have strong contribution to the risk management systems. The main conclusion regarding the comparison and transition of credit rating estimates is that different indicators show different levels of credit risk. Financial patterns and ratings are based on those factors that have a significant impact on the financial condition and solvency of a company or bank.

It is believed that rating agencies improve their standards in a timely manner and criticism of the methodologies of most rating agencies is not relevant. In this paper, these changes are reliably confirmed by reports, which disclose indicators of the dynamics of the average level of credit risk assessment and company defaults over a period of time (for example, 1 year or 5 years) and compared with previous periods of observations. However, over time, the rating agencies themselves admit that they have not adhered to the best valuation methods and are reviewing their own methodologies. Rating agencies are supposed to follow the principle of self-regulation when crisis phenomena have already affected the financial market and the market power of rating agencies has been decreased.

The credit cycle for 19 countries is presented, calculated using the credit gap using the two-way Hodrick–Prescott filter (HP-filter, two-way-sided). The credit cycle allows to track the possibilities of expanding and reducing access to credit in the economy over time. This, in turn, indirectly affects the business cycle of various companies since access to credit affects the company’s ability to invest. According to empirical observations, over time, the effectiveness of credit-oriented sectors that have a fixed income from their own activities, including large corporations focused on the constant maintenance of their credit ratings and, if possible, investment grade ratings of their securities, is directly related to the credit cycle.

Over time, at various phases of the credit cycle, ratings assigned by rating agencies undergo significant changes. An analysis of data over a 10-year period of time for both developed and developing countries allows us to identify several legitimate reactions of credit ratings received from international rating agencies (IRAs). Their changes are out of sync during the credit cycle and adjust lately to the macroeconomic changes.

Rating agencies are often criticized for being “too slow to update” assigned ratings as a result of a review of methodologies or according to a possible review. This creates a downside migration of credit ratings from one class to another. In response to criticism, rating agencies often emphasize that when ratings are declining, they follow the principles of compromise and choose between the stability of the credit rating and its accuracy.

Thus, based on the grouped data from the sample, including the credit ratings and financial indicators of companies, point estimates can be obtained for the entire sample. Based on this information, we identity financial patterns of credit ratings for companies that have similar quantitative estimates as a result of clustering and median values.

It should be noted that in the context of the short-term forecasting of financial risks and tracking the level of their changes, it is important to take into account the uniformity of objects in the selected groups. A group of clustering entities has similar characteristics and the entities are homogeneous in various subgroups.

An important task of risk management is the allocation of such groups of entities in different time periods. Today's risk management tasks allow us to say that the allocation of common or similar indicators for different objects indicates exposure to a similar level of financial risks. Highlighting common patterns allows us to create accurate calculations about the level of development and exposure to credit risks, regardless of their specifics.

The technique of applying the approach based on the IRB approach (Langohr and Langohr 2008) significantly simplifies the occurrence of a possible situation in which the lender and the borrower are the same entity. For this, they will be unable to meet their obligations. In this aspect, commercial banks are given a basis for developing these approaches to assessing and assigning a credit rating to third counterparties or borrowing from other banks. This provides a comprehensive assessment of their debts according to changes in their credit history for both large customers—corporations and companies in the real sector, and the same is for entrepreneurs, small private firms, and individuals.

The main point of the idea of introducing this type of credit risk assessment, based on the credit rating system, is that credit risk assessment methods, as well as the risk weighting methodology from Basel requirements and forming the final assessment of the borrower's credit risks, were initially and quite successfully applied by rating agencies over the long term. The other side of the issue is the degree of integration and the readiness of most credit organizations to accept the Basel recommendations and turn them into a qualitatively flexible tool for continuous use.

With a positive change in the credit rating, the borrower has the right to apply for additional sources of financing and investments from the side of the commercial bank. In the opposite situation, when a deterioration and lowering of the credit rating is noticeable, the bank will be able to track this by mean values of the rating changes. That creates the situation when the issued funds and late payments will be returned only partially, or they will turn into a category of default and bad loans (non-performing loans).

The rating agencies have new opportunities in the provision of rating and consulting services, including their practice of providing and calculating functions with weight coefficients determining and identifying risk components (Nkomo and Chiwanza 2014). Also, the practice of determining defaults by rating agencies is based on a range of methodologies and 60 years historical observations, they have at their disposal a whole statistical database and empirical observations from various sources. Rating agencies universally compile various types of risks, including the determination of the probability of default (PD), they consider the following additional indicators—specific weight of losses due to default (LGD), maturity (M), their possible losses (EL), unexpected losses (UL), and the exposure at default (EAD).

In addition, in assigning credit ratings by various agencies, it is important to define asset classes; their adopted approaches serve as introductory and best practices for developing risk factors. In the presentation of each of the asset classes, it begins with a consideration of risk components and their relevant factors, they are first based on the function of risk weights for corporate obligations, the distribution of their losses is determined, and another important parameter for considering risk components and tools is to reduce them (Schüler 2018).

This practice allows creating and identifying standards of legal certainty and formulating rules for the recognition and application of risk mitigation tools. Including the flexibility of their use and the improvement of practice, this is especially true for use in commercial banks (Karminsky and Polozov 2016). The mechanisms of this approach allow financial institutions to create their own systems for a quick response to and the instrumental mitigation of emerging risks, this improves the market conditions of market participants.

The recommendations of Basel II and III discuss and justify the adoption by banks of an approach based on the system of creating internal ratings and determining the requirements of risk weighting factors. Such transitional events were taken from the practice of rating services and are the basis for these approaches. Banks have introduced an internal system using IRB, this is an approach that allows bank management to consider the Basel requirements and form the capital base of coverage and the formation of additional reserves for unexpected losses.

One of the tasks of analyzing the credit quality of banks and entities is to set a credit rating. Moreover, for the same company, ratings of several agencies may exist in parallel, which do not always coincide. The difference between the ratings of international and national agencies is especially noticeable. Moreover, not all companies have an international rating or are rated. In this regard, the question arises of setting an internal rating based on information from credit ratings of different rating agencies, and based on additional information—financial statements, market indicators, and benchmark for similar companies.

The aim of the work is a qualitative and econometric analysis of the relationship between the methodologies of various rating agencies, financial and non-financial characteristics of issuers, credit ratings of issuers of both industrial companies and commercial banks, and macroeconomic indicators of a change in the phase of the credit cycle at country level. To this goal, the presented study examined microeconomic and macroeconomic factors that have a significant direct and indirect impact on the assignment and changes of credit rating of various economic entities. The result of the study is up to exposure and build up a system of aggregated credit ratings of various issuers to increase the determination of their credit risks.

2 The Comparison of External Rating Scales for Internal Use

One issuer may be assigned several ratings from a variety of rating agencies, and these ratings may be different, and, in some cases, this difference may be significant. One of the main and widely used methods by both foreign authors and Russian researchers is the mapping method, which allows them to compare various estimates received by the issuer from different rating agencies, and which consists of the following algorithm.

At the first stage, a conditional base rating scale of one of the leading rating agencies is selected. It lists all the other rating scales of the agencies under consideration. Subsequently, the calculated distances between the rating categories will be reduced to the one selected group as the base scale and the aggregation of credit ratings of issuers of one group will be given.

The second stage consists of the following: during the formation of the database of rating estimates and their issuers, a procedure is carried out for the quantitative translation of the symbols of the rating scale of each agency into numerical values.

In this way, numerical estimates of various scales are obtained. According to this procedure, digitized historical ratings are obtained for the selected time periods. The third stage of the procedure for comparing ratings consists of selecting pairs of ratings from the collected primary statistics for homogeneous groups of issuers. Further, it is necessary to determine the number of adjustments and overlapping ratings that are encountered by different issuers.

Data showing the number of different ratings found by the same issuer allows the formation of pairs of observations. Pairs of credit ratings show how possible it is to apply an integrated measure of comparing rating scales and minimizing the distance between rating scales of different agencies.

At the fourth stage of the correlation of rating scales, a method for identifying ratings is used. The most common way of such identification is to linearly correlate and minimize the distance between the rating pairs in one scale and bring them to the selected base scale through the synchronization of several scales into one to create a one-dimensioned rating space.

After that, the procedure for determining the minimum distance between rating categories is completed, and the function of transforming the scales of the observations to the selected base scale is obtained. The result of such a comparison will be obtaining estimates and their correlation with each other in one scale.

As a comparison measure in aggregating and correlating rating scales, the method of finding proximity measures (1) was used:

$$z_k(x) = \sum_{i=1}^{p} \alpha_i \exp\left(\gamma \left\| x_i - x_j \right\|^2\right) + \beta_0 \tag{1}$$

where α_i and β are the main evaluation vectors for selecting the coefficients of the correlation of rating scales; γ is the estimation parameter of aggregated comparison; z_k are the estimation results of credit ratings aggregation.

Further aggregation of rating estimates occurs due to their quantitative correlation with each other using the method of differences for each gradation for all scales of rating agencies (2):

$$\begin{cases} P(Y_m = 1) = F(c_1 - Y_m\beta); \\ P(Y_m = 2) = F(c_2 - Y_m\beta) - F(c_1 - Y_m\beta); \\ \dots\dots\dots\dots\dots\dots \\ P(Y_m = k-1) = F(c_{k-1} - Y_m\beta) - F(c_{k-2} - Y_m\beta) \\ P(Y_m = k) = 1 - F(c_{k-1} - Y_m\beta), \end{cases}; \tag{2}$$

where Y_m is the quantitative translation of the rating symbol into a numerical value; F (c_1-Y_m β) is the correlation function F and the aggregation of each pair of credit ratings in accordance with the values of the estimates of the vector β.

According to the methodology, the assignment of credit ratings by agencies is based on an analysis of six components: (1) business profile and external environment; (2) the size of the company and (3) its profitability; (4) debt and debt coverage; (5) financial policy; and (6) liquidity. Components (2)–(6) relate to estimates of the company's internal environment. From these 6 components, only 3 are determined by the agency on the basis of financial statements, namely: size, profitability, debt and debt coverage. It provides a list of selected explanatory variables for constructing a scale matching model (see the previous section).

Using the multiple mapping method, a linear model was constructed for comparing rating scales of the largest international rating agencies for industrial companies. Based on the coefficients obtained as a result of evaluating the comparison model for all rating agencies considered for the period from 2000 to 2016, a rating scale matching scheme was constructed in accordance with the approach described previously.

The total number of credit rating observations from the IRA pair ratings for industrial companies: 5172 and 1590 for commercial banks. For Russian rating agencies and data on Russia, the number of collected and pairwise matching credit ratings is much smaller—for industrial companies it was possible to find 348 pairs and 166 pairs for commercial banks, including pairs of credit ratings from Russian rating agencies.

The next steps are the processing of the data sources. Firstly, all financial reporting statements are translated into a single currency. Second step is (1) to bring to a common basis the various accounting practices for business transactions between GAAP and IFRS and (2) the reflection of certain business transactions (for example, operating expenses) not in form, but in their economic nature. For this, Moody's methodology for processing financial reporting items was used (Moody's

Table 1 Regression coefficients for international rating agencies for industrial companies

Rating scale	α	β
Fitch	1.041* (0.936)	– 0.098*** (0.055)
Standard & Poor's	1.062 (0.875)	– 0.145*** (0.528)
Moody's	1.012** (1.143)	0.188** (0.009)
Ratings' pairs comparison	5172	

Coefficients are significant at the significance level of *** - 1%, ** - 5%, * - 10%.

2016). As a result, financial and non-financial indicators were obtained, which were included in the model (Bisenov et al. 2019).

The estimated coefficients of multiple mapping functions for industrial companies from 19 countries[1] are shown in Table 1. Based on the results, a number of conclusions can be drawn. According to national scales, Fitch turned out to be the most conservative agency in the investment zone, however, when approaching the speculative zone, ratings of international agencies on national scales begin to converge more and more. The differences are primarily in the speculative zone (in the rating area B (rus)).

According to international scales, Standard & Poor's are more conservative than Moody's and Fitch in the investment zone, but the differences between the scales also begin to decrease in the speculative rating category. Scales of Standard & Poor's and Fitch in the speculative zone are almost the same. The differences between Moody's and Standard & Poor's are strongest in the speculative zone, in the CCC + rating area.

The rating scale of Russian rating agencies is comparable to the national scale of Fitch in the investment zone. In the speculative zone, the Russian agencies are more conservative in comparison with Fitch and have lower standards in comparison with Standard & Poor's on national scales. This shift in the national scale of Standard & Poor's can be explained by a smaller number of observations in the speculative zone.

Among the pairs of ratings Moody's and Fitch, there were no observations in the highest and three lowest grades. Regarding observations by Moody's and Standard & Poor's, 0.7% of pairs are rated highest. Most of the observations are concentrated in the gradation of BBB-/Baa3 ratings, their share was 34%. The ratings of Moody's and Standard & Poor's are characterized by an excess of the number of observations in grades A/A2 and BB-/Ba3 over A−/A3 and BB/Ba2, respectively, which, in turn, is not observed for pairs of observations for the Moody's and Fitch. The model constructed and described above estimate a model similar to the model for Russian industrial companies, but only three scales of international rating agencies were included. A comparison of the scales is shown in Table 2.

[1]Countries included in the sample: Australia, Brazil, Great Britain, Germany, India, Spain, Italy, Canada, China, Mexico, Netherlands, Portugal, Russia, USA, Finland, France, Sweden, South Africa, Japan (in total 19 countries)

Table 2 Multiple mapping model coefficients for commercial banks, including data for Russia

Rating scale	2000–2016, commercial banks for 19 countries		2000–2016, Russian commercial banks	
	α	β	α	β
Moody's	0.038** (0.009)	2.232 (2.089)	0.062*** (0.087)	0.836* (0,898)
Standard & Poor's	0.071** (0.023)	0.074 (0.051)	0.379* (0.433)	0.145** (0.121)
Fitch	0.804* (0.166)	0.514 (0.544)	0.668** (0.241)	0.177** (0.101)
Ratings' pairs comparison	2390		890	
R^2	0.815		0.622	

Coefficients are significant at the significance level of *** - 1%, ** - 5%, * - 10%.

Despite the fact that the largest international rating agencies strive to equalize their scales, they do not always coincide. In the investment zone the scales of agencies are close to each other, but in the speculative Moody's is the most conservative rating agency and assigns the lowest ratings to industrial companies. The scales of Fitch and Standard & Poor's differ in similar valuation methodologies; however, Standard & Poor's has the lowest standards in evaluating industrial companies. These results do not contradict the findings on the comparison of scales for Russian industrial companies regarding the conservatism or loyalty of an agency; however, for Russian companies, the differences in the rating scales of agencies are more significant.

Descriptive statistics for the collected credit rating database of three international rating agencies are presented below. Moody's, whose scale was chosen as the base for displaying the two rating scales of two other international rating agencies, has a left-side shift in the distribution of gradations. The largest number of assigned assessments falls on the mean values of the speculative class B1, B2, B3 and as a percentage 28% of all assigned credit ratings.

For the rating agency Standard & Poor's, there is a situation with the distribution of gradations of credit ratings assigned to commercial banks, the distribution shift in form is abnormal and has two peaks. The largest number of ratings is divided between two classes, investment and speculative, while their share ratios by grades are as follows: the investment class is characterized by the concentration of all ratings in the range from A-, BBB+, BBB- and is 29%, relative to the speculative class—the prevailing number of grades represented by B+, B, B-, CCC+ and expressed in 27% of the total percentage of the distribution of gradations.

The third rating agency—Fitch has a right-hand distribution of gradations, and it is almost close to the normal distribution, most of its ratings are concentrated in the investment class and include the following grades AA, AA-, A+, A, A-, BBB+, BBB, BBB-. All rating agencies have combined a percentage of the entire distribution of gradations of 60% (see Fig. 1).

Figure 1 shows the distribution (in%) of credit ratings to the base scale. The collected data is shown that the largest number of shifts in assigned credit ratings is

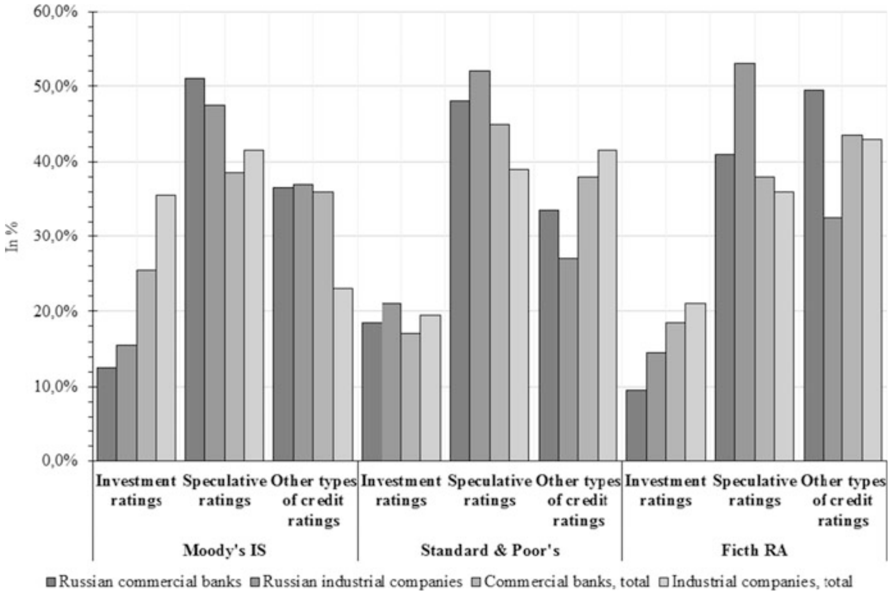


Fig. 1 Distribution of ratings for the period 2000–2016

observed for Moody's, while most of the ratings of this agency are concentrated in completely different classes, both speculative and investment. Regarding the assigned ratings of Standard & Poor's agency, the number of ratings is evenly distributed in different classes. The third rating agency is Fitch R.S. and it mainly assigns its credit ratings in the investment range and tens to extent than the other two international rating agencies.

Conversely, a temporary difference for the three rating agencies is observed. Since 2005 IRAs often lowered credit ratings for many banks, and vice versa, upraised credit ratings for industrial companies. The differences in rating points for companies and banks are significant and are based on the financial indicators.

The aggregated results of comparing and evaluating econometric rating models based on the multiple mapping and clustering methods show that the trend of rating distributions over different time periods relatively to banks is quite high and is average for companies. But since 2002, all the IRA often downgraded the credit ratings of many banks and raised them among industrial companies.

Table 3 shows the results of evaluating a logarithmic model based on the multiple mapping method. The estimation coefficients and the predicted strength of various models for different industrial companies and for different periods of time are relatively high, but for the observed time period there was a significant decrease, especially for Russian industrial companies and commercial banks (see Tables 2 and 3).

The proposed analysis using the multiple mapping method for the period of 2000–2016 gave the least predictive power in comparison; this can be seen from

Table 3 The results of the comparison of ratings of the IRA agencies for industrial companies. Source: based on Dyachkova et al. (2019)

Rating scale	2000–2016 industrial companies for 19 countries		2000–2016 Russian industrial companies	
	α	β	α	β
Moody's	0.355* (0.144)	1.202 (0.908)	0.241*** (0.022)	2.221 (1.344)
Standard & Poor's	0.316** (0.072)	0.546* (0.355)	1.017 (0.903)	−0.044** (0.017)
Fitch	0.349 (0.204)	0.554* (0.658)	1.009* (0.005)	−0.239** (0.360)
Ratings' pairs comparison	2560		1340	
R^2	0.802		0.634	

The coefficient is significant at the significance level of *** - 1%, ** - 5%, * - 10%

the number of correctly predicted ratings. Comparison of rating scales on data on Russian companies and banks highlighted the following results:

1. The differences between Moody's and Standard & Poor's are greatest in the speculative zone, especially in the CCC + rating area.
2. In the speculative zone, the Russian rating agencies are more conservative in comparison with Fitch and more loyal in comparison with Standard & Poor's on national scales.
3. IRA and Russian rating agencies are more conservative to industrial companies in the investment zone, the scales of these agencies are comparable to the national scale of Moody's agency.

It should be noted that the ratings of Russian industrial companies assigned by three international rating agencies on their scales fall into the main zone of transition between investment and speculative classes. The main grades for percentages are low investment grade BB+, BB, BB- and BBB+, BBB, BBB- and their shares are 47% and 36%, respectively. This opinion of international rating agencies is due to a high assessment of country and political risks. In addition, the given gradation tends to a normal distribution, this is due to the fact that in one country there is a general level of risks and its industry adjustments.

3 Clustering Data Analysis for Improving Forecasting Power of Credit Rating Models

A number of works showed that the estimates obtained using structural clustering methods are superior to subjective assessments of the level of credit risks, which significantly different from the use of simple linear models (Distinguin et al. 2013; Grishunin et al. 2018). The approach based on patterns allows us to solve problems, including both obtaining objective estimates in time and improving statistical methods based on these methods and modeling with a higher level of forecasting accuracy. Conversely, the method of patterns is simple enough to exist as a way to determine the quantitative differences of clustering groups of objects within one large unstructured sample.

The methodology for calculating patterns allows the use of various financial and non-financial indicators. The use of accounting and financial information on the object of observation and the frequency of the publication of financial statements of the company, news, etc., allows for the analysis using temporary data. Quantitative information derived from the financial profile of the company is the basis for tracking changes in the financial patterns of the company (Van Laere & Duan 2012). The financial patterns of a company or bank show quality information based on market changes in its financial performance and its sustainability. Their use in financial analysis models allows the tracking of exactly what actions are happening within the company. The changes that take place are reflected in the company's patterns, which make it possible to assess the influence of various factors on changes in the level of credit risks over the past periods. Financial patterns are beginning to play an increasingly important role in the ability to predict changes in credit ratings, since they allow for the consideration of many factors affecting credit ratings and the company's profile.

The purpose of constructing financial patterns for various economic agents is to identify and quantify whether the credit ratings become higher over time, or vice versa, decrease. In order to fulfill the designated goal, this paper will use narrow definitions related to the context of statistical modeling.

If rating agencies, according to their own methodologies, use various indicators to evaluate the company and improve methods, then the credit ratings assigned from 1985 to 2016 to industrial companies and commercial banks will decrease over time, as new methodologies have more inflexible and tights standards for ratings.

The opposite can be assumed that companies or banks with similar characteristics may receive a low rating for objective reasons. Therefore, the construction of financial patterns is considered to give a real picture of how, over time, various economic agents:

(a) could maintain the level of the previously assigned rating or,
(b) change (decrease/increase).

Higher ratings will mean weakening rating standards, while lower ratings will mean their tightening. The method of constructing financial patterns allows the identification of the presence of nonlinear relationships for the allocation and evaluation of homogeneous groups connected by one parameter or pattern (Aleskerov et al. 2013). In this paper, it is stated and shown how a similar statistical evaluation of various economic agents having the same patterns was carried out, based on the selected financial patterns for groups of various economic agents (industrial companies and commercial banks) using the aggregation of ratings constructed on previous comparison of the rating scales.

In the process of constructing financial patterns based on credit ratings, it is showed that:

Table 4 Descriptive statistics for selected variables for BRICS companies

Variable name	Mean values	Standard deviation	Minimum	Maximum
Company's assets, in mln.	19.23	15.16	1.29	757.06
Total equity	4.91	3.19	0.23	21.27
EBITDA/interest expenses	2.26	1.91	1.57	11.30
Net revenue/interest expenses	24.58	10.38	1.66	147.06
Debt/book capitalization	0.52	0.38	0.19	2.97
RCF/debt	47.08	19.61	3.61	219.19
CFS/revenue profit	1.81	1.32	2.92	1.96
Long-term debt, in %	18.32	64.45	17.11	38.12
Operational margin, in %	4.98	2.24	2.29	9.37
Current ratio	6.93	6.18	1.81	9.71
CAPEX/depreciation expenses	0.05	0.15	0.28	0.32
GDP growth, in %	1.86	1.35	1.63	1.92

Industry dummy as categorical variable and financial crisis dummy, Sovereign rating country ceiling were included in the multi-logistic regression model

1. The modeling process and classification of credit ratings is a complex task that requires the preliminary selection of variables. This significantly improves the accuracy of the classification of credit ratings; a preliminary selection of significant variables was made in the previous section.
2. The methodologies of rating agencies vary depending on the rating objects. Best results are achieved by analyzing the industry classification of companies.
3. To obtain estimates of the financial patterns of credit ratings for industrial companies, algorithms were used which are based on previously obtained estimates from the construction of multi-logistic regression models (Table 4).

To conduct a computational experiment and build financial patterns, the credit ratings of two rating agencies, Moody's and Fitch rating agencies, are used. The annual data on credit ratings and various economic agents were uploaded from the financial database of the Bloomberg terminal, from 1985 until the end of 2016. The collected database contains annual and semi-annual observations of changes and revisions of credit ratings of both industrial companies and commercial banks. The general sample is constructed for 16 years, from 2000 to 2016. The credit ratings were converted to an ordinal numerical scale and their ratios with changes of credit ratings are given in Sect. 4, along with Table 5 comparing the rating scales of various agencies.

In the total sample, observations with default ratings were excluded from the sample. The largest number of observations includes speculative rating grades and its share is about 20% of the total sample—566 observations had a level from CCC +, CCC, CCC- to CC. These observations were not excluded from the sample due to their statistical objectivity and formed a subset of speculative rating estimations.

The total number of credit ratings collected by industrial companies and commercial banks was 7336 observations, including notches (upgrades and downgrades

Table 5 Clustering-based rating assessment results using financial patterns for industrial companies and commercial banks

Credit rating symbol	Aa3-Aaa	A1	A2	A3	Baa1	Baa2	Baa3
The quantitative value of the rating	1	2	3	4	5	6	7
Result, in %	1.25%	7%	5%	5.5%	13%	17%	25%
Number of initially assigned ratings	17	129	113	182	255	341	711
Credit rating symbol	Ba1-Ba2		Ba3-B2		Ca-B3		
The quantitative value of the rating	8		9		10		
Result, in %	11.5%		20%		4.8%		
Number of initially assigned ratings	203		549		129		

of credit ratings). In the final sample, 6732 observations remained, 604 observations were excluded (about 10% of general sample).

To analyze the trends and changes in the ratings, firstly the multi-logistic model was evaluated, where credit ratings were a dependent variable on company performance, and dummy variables were added to the model for the crisis year 2008 and the year of the methodology review.

Previous estimations of the coefficients were obtained from a comparison of rating scales; they also showed their stability over time, which allows them to be used in the short-term future and for the method of financial patterns to accurately measure the number of credit ratings deviations.

For each observation, the residual estimation was calculated, and, the evaluation procedure showed that the actual rating is the forecast rating. Then the balance for each period of time was aggregated over 16 years. The average balance of a certain year shows how many grades from the median and mean values of the actual rating of a company or commercial bank differ from their "hypothetical" forecast rating.

The specification of the model for identifying similar groups refers to which financial and non-financial indicators from the methodologies of rating agencies showed the greatest significance (at the level of 10%, 5%, or 1%). In this proposition, indicators were taken that had already shown their significance in the comparison of various rating scales.

The financial patterns for industrial companies are shown in Tables 3 and 4 and contain data on annual downgrades of credit ratings. As a result, the increase in the level of credit risks of these agents shows the results that quantitatively the mean coefficient of lowering credit ratings for different groups of industrial companies is by year: +0.91 in 2002, + 1.12% for 2003, +1.43% for 2004, + 1.37% for 2005, and + 1.48 between 2006 and 2007. These numbers are comparable to an 11% decrease in the number of investment grades credit ratings that occurred after the 2008 financial crisis. Data are based on a report by Moody's. According to the figures, starting from 2008 to 2009, the downgrading trend in credit ratings continued. It is noteworthy that in 2002 the peak ratio of lowering and raising credit ratings among industrial companies reached the highest level and amounted to +4.02% and this is the highest point of all issued credit ratings for the 16 years in the general sample.

The main parameters that had a special impact on the level of credit risks were influenced by financial margin indicators, return on assets (ROA), and the level of debt to equity (financial debt-to-equity ratio).

Based on the analysis of significant coefficients and their signs with explanatory variables, the following conclusions can be drawn:

1. The more stable the situation in the home country, the higher the rating of the industrial company will be (real and forecasting rating).
2. The country's competitiveness level is essential for determining the ranking of industrial companies, but this influence is ambiguous in the model.
3. The size of the company by assets, achieving economies of scale, the ability to produce products at the lowest cost (through the variable of profitability and EBITDA), and the ability to pay obligations on the horizon of 12 months (through current liquidity ratio) positively affect the rating of an industrial company.

Since rating agencies have different methodologies for industrial companies, and each rating agency identifies certain indicators, considering them more significant (see Sects. 2 and 4), the sectoral assessment of each company occurs in different ways. This is one of the reasons for the large number of discrepancies in ratings among industrial companies from various industries.

Table 5 provides descriptive statistics of clustering-based ratings assessment. The mean values of the most significant variables are gross margin, long-term debt, and the total debt-to-assets ratio, which are 31%, 33%, and 34%, respectively. In accordance with these coefficients, it shows that mean values for the selected variables come from various rating categories BBB, C, and AA. As can be seen from Table 4, there are several interesting trends regarding the median values of the variables for industrial companies: the gross margin and the size of long-term debt are steadily increasing throughout the observation period. It is noteworthy that the average value of the total debt-to-assets ratio has improved dramatically since 2002. Relatively to net profit and company assets, these figures increase after 2000. These figures reflect the fact that according to the new methodologies of IRA on the introduction of the new requirements after 2002–2005, according to which rating agencies indicated that a company must have indicators that are 5–10 times higher than indicators before 2002 in order to obtain an investment grade rating.

When dividing the sample into developed and developing countries, one should especially note the group of BRICS countries. The motivation for ranking research and comparison of industrial companies from BRICS is based on: (1) the growing systemic importance of BRICS; (2) an increase in the share of industrial production in these countries; (3) increasing investor interest in BRICS. The data sources are Moody's analytics, 2013, and World Bank reports.

World Bank statistics show that the annual growth of GDP in the BRICS in 2008–2017 has amounted to 5.4%, which was several times higher than in developed countries (according to the World Bank, this growth was only 0.8%) or in other developing (1.1%) countries. Over the past decade, the share of BRICS in global GDP has increased to 30% from 21.9%.

According to World Bank forecasts, further GDP growth is expected in the BRICS countries in 2018–2022 and its rate will grow by 4.7% annually, and this indicator will be higher than in developed countries by 1.5%. According to estimates of the Bloomberg agency, the investments made in BRICS in 2017 amounted to 33.2% of FDI versus 18% in the developed countries of the world and it is forecasted that the level of investments in these countries will remain at the level of 30% until 2030.

For Russian industrial companies, the variables of debt (RCF debt) and liquidity (current ratio) at 1% and 5% significance level were especially important. The estimates also show that Russian industrial companies are higher in creditworthiness than Indian industrial companies and companies from South Africa.

However, when comparing financial patterns, Russian industrial companies have a worse financial picture than Chinese industrial companies. Estimates of Russian industrial companies on the level of debt are comparable with the results of Brazilian companies; however, the production capacities of Russia and Brazil are different. In this aspect, special attention should be paid to macroeconomic factors, including GDP indicators—as the growth of GDP in Russia is lower than in Brazil.

As a result, the econometric modeling and linear specification of a multi-logistic model (see Sect. 4) should take into account the inclusion of control and country dummy variables. The empirical estimates show that the level of the methodologies of rating agencies is similar for industrial companies from BRICS countries.

The rating agency methodologies are based on the assumption that the credit ratings of commercial banks contain the same information regardless of the class of assets to which they are assigned. For example, in regulatory documents that define tight boundaries while restricting investment grade ratings, there are no distinctions made between a bank with a rating of Baa3 or a bank with a rating of Baa2. But this discrepancy in ratings may indicate differences in the risk profiles of the underlying assets owned by the bank and the level of their credit quality.

For an accurate assessment, it should be noted that financial and state regulators relying on credit ratings are likely to receive inaccurate information in this case. This happens under the condition that rating agencies are not guided by relevant information and credit ratings are lowered without considering a revision of the profile of assets for commercial banks.

4 The Relationship Between Credit Ratings and Country's Credit Cycles

Various analysts state that business and economic cycles are no longer applicable to the contemporary world (Aktan et al. 2019; Arteta et al. 2017; Barrell et al. 2017). The main reasoning behind the statement is that the manipulations of rates by Central Banks of developed and developing countries is mostly aimed to encourage people to use borrowed funds, depending on the state of the economy. Therefore, the main attention in this section is paid to the theory and practice of credit cycles.

The basic idea behind the credit cycle theory is the following: the more loans are provided, the higher the speed of development in the real sector will be. The development in the real sector will lead to improvement in the economy as a whole (Loffler 2004; Mian et al. 2013). However, the economic growth and, especially the speed of changes, causes overheating in the financial and capital sectors of the real economy. In the overheated economy, even more credit is provided, so the risks are rapidly increasing. In such cases, the probability of default of a borrower is rising; therefore, a crisis is likely to appear.

Therefore, the crucial part of the whole credit cycle theory is to find the most convenient and reliable ways to determine the point, where steady improvement and a healthy growing economy become an overheated one, which is likely to result in a crisis. The idea of implementation of a counter-cyclical buffer (proposed by Basel Committee and its requirements of Basel II&III) is relatively new and is implemented worldwide. Correct measure of the size of the buffer directly depends on an accurate determination of the credit cycle indicators (Altman & Rijken 2004; Kiff et al. 2013; Repullo & Saurina 2011).

The Basel Committee in December 2010 published new rules, stating that a capital buffer should have counter-cyclical nature. The amount of the buffer is calculated according to the real state of economy, more precisely, and according to the various phases of the credit cycle.

If the theory is true, the probability of a crisis arising may be controlled by the regulation of the amount of loans provided by commercial banks and other financial institutions, and also by government regulation of rates, as, for example, by regulating the interbank lending rate. Conversely, the regulation may concern the amount of reserves held by banks (Ryan 2012). The ways are aimed to reduce the risks that arise with the lending process, especially for the effects of a borrower's default.

The proposition of the Basel Committee is that the amount of money reserved through the process of money creation for a counter-cyclical buffer should increase in good times, when there is an upward trend, and be lower in amount of reserves which are made during economic downturns. Such an approach will provide an opportunity to accumulate a reasonable calculating buffer that will be used as a predictor before financial crises arise.

The components of a financial structure, especially the network of payments, highly depend on financial intermediaries such as commercial banks. Therefore, the efficiency, financial transparency, and strategy of monetary policy define such important parameters as the amount of money borrowed relative to the volume of liquid assets in the economy (which is considered as one of the main dimensions of the stability of financial structure defined by) (Amato and Furfine 2003). The credit risks of various financial relations based on the function are provided not by regular banks, but Central Banks and governments.

Claessens et al. 2011 found that the instability in an economy is mainly caused by growth of contractual payment commitments, relative to both money holdings and specified money flows. The main indicators of financial stability are supposed to be

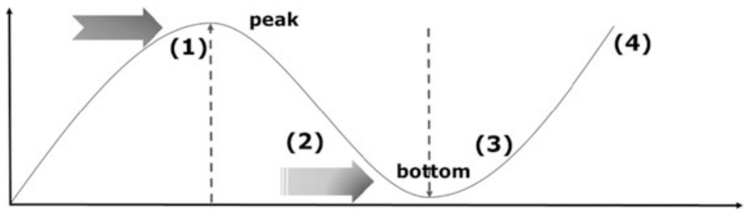


Fig. 2 The credit cycle and its four phases. Source: based on Dyachkova et al. (2019)

asset prices, which are increasing during good times. The higher the ratio of liquid assets, the more stable the economy.

There are two ways to consider the credit cycles: from the side of the demand for credit (the borrower side), and of the supply of credit (the lender side). This research is made with regard to the demand side. There are determined factors that influence the amount that borrowers need.

The main idea of the research is that the liquid assets can be set as factors of production and as collaterals for loans. Therefore, the demand for loans from borrowers is highly connected to the changes in the prices of the assets that are used as collaterals (Bordalo et al. 2018). At the same time, the demand for loans influences the prices. As with changes in the demand for loans, changes in the net worth of the companies/borrowers will be distinguished.

There are only a few specifications of model, including different assumptions about borrower's conditions and their opportunities. Here, the focus will be paid to the conclusions that are applicable to our research. According to most papers calculating and predicting credit cycles, the length of credit cycles is on average around 10 years. Moreover, shocks arise when prices for land and property (assets that are assumed to be the basis of production, and at the same time, the most secure collateral) are at their peak value (see Fig. 2).

Therefore, the major causative agent of shocks is the net worth of borrowers, or in other words, the value of assets and liabilities of borrowers, as well as the value of collateral provided in the real sector of the economy. One of the most significant questions of the whole topic of credit cycles is how to estimate it. The Basel Committee on Banking Supervision (BSBC) proposes to use credit-to-GDP gap indicator as the representation of a credit cycle. According to BIS (2010), financial and business cycles are tightly connected. However, financial cycles are longer and more turbulent than the business and credit cycles.

This happens partly because of the relation between the cycles, the credit-to-GDP ratio starts to rise in advance, and before the actual peak point is achieved. According to the findings of BSBC, the credit-to-GDP ratio tends to rise smoothly, well above trend before the most serious episodes of financial cycles happen (Everaert and Zeng 2015). This fact provides an opportunity to use the parameter to predict a financial crisis and to have enough time to undertake appropriate preventive actions. As the whole story of credit cycles nowadays is built around the necessity to provide a tool

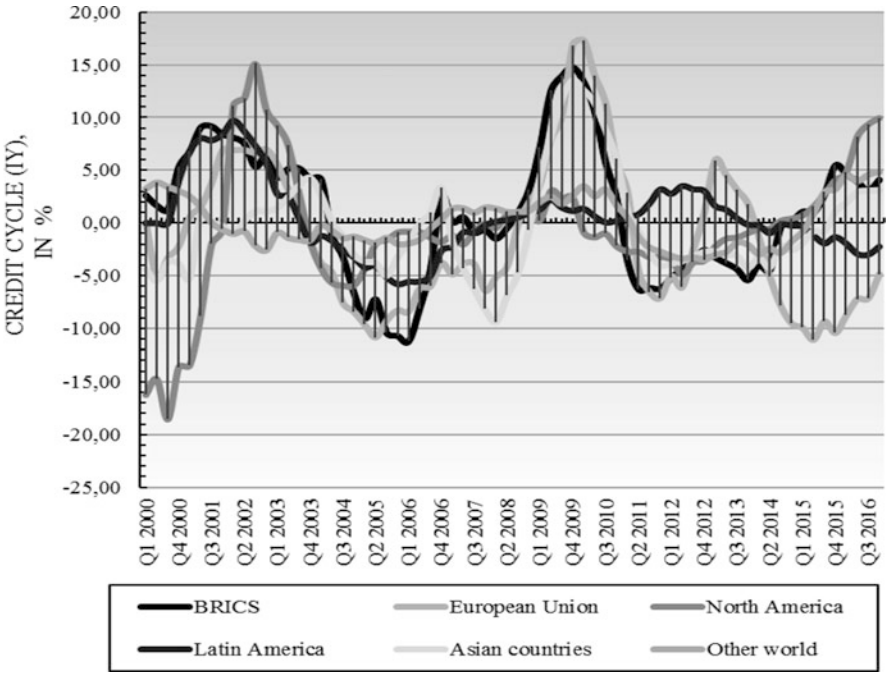


Fig. 3 The credit gap (changes in the loans-to-GDP ratio) in total for 19 countries, using a two-way side Hodrick–Prescott filter (quarterly, in p.p.) Source: Authors' calculations based on IMF data

for ensuring the safety and stability of the financial sector in individual countries and the global community, predicting feature becomes essential.

The vast majority of research that uses credit-to-GDP ratio as the dependent variable tries to implement a technique to remove non-stationarity in the data. For the purpose of this research, a number of de-trending instrument were applied. One of the most important is the Hodrick–Prescott filter, one-sided and two-sided.

In a recent paper conducted within the framework of the Central Bank of Russia (Ponomarenko et al. 2018), emphasis is placed on the application of Hodrick–Prescott filter to figure out the credit gap. Credit gap is calculated according to Giese et al. (2014), as the deviation of credit-to-GDP ratio from its long-term trend. The filter removes long-term trends from the time series. The results are in Figs. 3 and 4 and show that the method is valid in the framework of data used in this research for the quarterly data of developing economies.

The next thing to be determined is the approach to the essence of independent variables. Each paper checks the validity of different groups of parameters, trying to find the most reliable, significant, and applicable in a real-life model.

There are these common ways to split the independent variables into groups. The variables that determine supply and demand are compared to the explanatory power at 1 & 5 % of significance in credit cycle modelling; for more *comparing* practice, the empirical example from research made by (Amato and Furfine 2003) whose

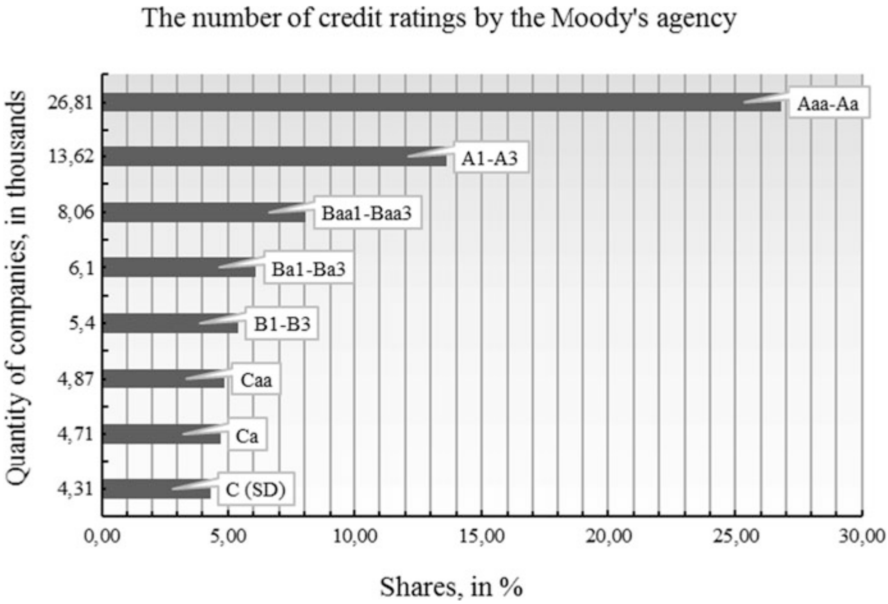


Fig. 4 The number of credit ratings issued by the international rating agency Moody's before the beginning of the credit recession. Source: Authors' calculations based on Bloomberg data

estimation was based only on the demand side. In our research, we took microeconomic and macroeconomic variables and to the relevant empirical evidence, it was obtained the same estimation effects as from research (Everaert and Zeng 2015) whose estimation was based on a model construction using both micro and macro data

However, some research goes further and add:

- Bank variables (bank lending conditions, ROA, ROE);
- Financial market variables (price of oil, etc.);
- Global variables concerning markets, connected to the banking sector (CPI, etc.).

For the purposes of our further model estimation, the most convenient way to discuss the main outcomes of a number of papers is to hold the most relevant independent variables in semantic groups. In this research, the main attention is paid to the demand side of the topic: banking and control variables (such as Central Banks and governments). The variables were split into three large groups.

The model specification is as follows (3):

$$Y = \beta_{it} + \beta_{i,t+1}\left[\sum_{m=1}^{M}\mathrm{MACRO}_{m,it} + \sum_{k=1}^{K}\mathrm{COUNTRY}_{k,it} + \sum_{n=1}^{N}\mathrm{BANKING}_{n,it}\right], \tag{3}$$

where MACRO—macroeconomic factors such as CPI, GDP growth; COUNTRY—most significant variables for emerging countries; BANKING—cluster of special variables for each country-banking sector.

A multi-logistic regression was performed (see Table 6) determining the periods of significant increase or decrease of the variables, which are significant for illustration and description purposes.

To find the dependent variable, we are guided by the propositions of Basel Committee. The dependent variable is the change of credit-to-GDP ratio (the credit gap). All calculations were conducted stating to the recommendations of Basel Committee (Basel Committee on Banking Supervision 2010). The dependent variable was evaluated in two steps:

- First step is the calculation of the credit-to-GDP ratio. Formula (4) shows the calculation:

$$(\text{Banking Loans}/\text{GDP}) * 100\% \tag{4}$$

where banking loans is the total amount of funds provided to the private sector by banks.

Under the condition of the perfect information, the all-statistical figures include all types of banking loans provided by commercial banks to private households and firms from emerging economies. After this, the second step of our iterations of the collected data was the calculation of the credit gap with trend:

$$\text{G} = \text{R} - \text{Trend} \tag{5}$$

where trend component is the long-term fluctuations of the credit-to-GDP ratio; G is calculated as the average of the historical values of banking loans (R).

- The Hodrick–Prescott filter was used to improve the reliability of the figures as it assigns higher weights for our observations.

The empirical analysis made it possible to distinguish the various stages of the credit cycle and to assess the changes in credit ratings, both at the beginning of the credit cycle (at the stage of recovery) and at its end (at the stage of the credit recession). This analysis showed that the onset of the crisis and the downgrading of credit ratings are out of sync: at the beginning of the crisis, credit ratings are relatively high, and only when the crisis starts and the real sectors of economy go into recession, so the ratings begin to fall. This is typical for both developed and developing countries.

The credit gap estimations clearly identify the end of one phase of the credit cycle and the beginning phase of the next. Empirical estimations show that with growth of more than 2.9% of the credit gap over one quarter increases the credit ratings changes. The volume of issued bank loans begins to decrease, and changes in the dynamics of production and inflation rate are often observed simultaneously (see Table 5). In addition, the turning points of the credit cycle can be predicted using the logit models of multiple and ordered choice (for illustration, see Figs. 3 and 4).

Table 6 Empirical results

Variable name	Model specifications: Multinomial logistic, type 1	Multinomial logistic, type 2	Multinomial logistic, type 3	Multinomial logistic, type 4	Multinomial logistic, type 5 (+/with dummy for Russia)	Multinomial logistic, type 6 (+/with dummy for Russia)
Constanta	−0,446** (0,133)	−1672** (0,271)		−0.393*** (0.023)	−0.187*** (0.382)	−0.023 (0.059)
GDP growth		−2377* (1641)	−7948* (5911)	6.935* (0.490)	0.704 (0.411)	0.008* (0.000)
Company's assets	−1627** (0,731)	−1013* (1228)	−0,621 (0,951)	−7.008* (0.121)		−0.007 (0.086)
Inflation rate, IY	−0,185* (0,540)	−3222*** (1187)	−0,118** (0,577)	−0.015 (0.056)	−0.086 (0.331)	1.152* (0.836)
Price index, CPI	−1627** (1348)	0,159* (0,179)	−1137* (0,731)	0.033** (0.120)	−0.155*** (0.317)	−0.000*** (0.002)
Imports			−8007** (4647)	0.006*** (0.008)	0.038** (0.021)	0.261 (1.391)
Exports		−6427** (9577)	7557** (4631)	27.199 (5.129)	0.032^{c*} (0.015)	−0.008*** (0.720)
The lending channel, Tnarrowm	8038** (2.948)	2207** (1119)	−1007^{c*} (7108)	0.000*** (0.725)	0.171** (0.042)	0.042 (0.001)
Money supply	1447** (8.878)	3,17** (2145)		2.602 (0.279)	0.034 (0.088)	−0.118*** (0.119)
Unemployment rate, Stir	0,003*** (0,126)	0,072*** (0,018)	0,036*** (0,099)	90.355* (9.481)		0.050*** (0.025)
Loans to the companies of real sector of the economy, Tloans			0,110* (0,677)		0.126*** (0.209)	
The growth volume of debt securities, stocks		0,396** (0,745)	1758** (1218)	−0.001* (0.004)	0.251*** (0.020)	0.110 (0.881)

(continued)

Table 6 (continued)

Variable name	Model specifications					
	Multinomial logistic, type 1	Multinomial logistic, type 2	Multinomial logistic, type 3	Multinomial logistic, type 4	Multinomial logistic, type 5 (+/with dummy for Russia)	Multinomial logistic, type 6 (+/with dummy for Russia)
Dummy variable for currency revaluation, Xrusd					0.004 (0.011)	0.414*** (0.127)
Russia, Dummy country					0.319 (0.538)	−0.018** (0.316)
Number of observations	*1682*	*1767*	*2282*	*1218*	*1288*	*1322*
R^2	***0.618***	***0.623***	***0.811***	***0.755***	***0.727***	***0.731***

***Significance level at 5%
**Significance level at 1%
*Significance level at 10%
Number of observations (italic) - no significance is needed to add, because it's a total sample;
R^2 (r-squared) is an econometric parameter, which shows the percentage of trully omitted & statistical power, no significance is needed to add

If the theory is true, the probability of a crisis arising may be controlled by the regulation of the amount of loans provided by banks and other financial institutions and also by government regulation of rates. Conversely, the regulation may concern the amount of reserves held by banks. The calculations are aimed to reduce the risks that arise with the lending process, especially the effects of the default of the borrowers.

The selection of macroeconomic and financial variables made it possible to form an optimal set of variables. When constructing a multiple ordered choice model, the distributions of the coefficient estimates and, most importantly, the absolute probability and the factor estimate for the long term were obtained. In the given estimations of the models (for example, see Table 5), all explanatory variables that were significant at the level of 1%, 5%, and 10% are highlighted (see Table 6). If the sign of the coefficient estimation contradicts the economic sense, the composition of the explanatory variables in the models changed and the models were reevaluated.

According to the statistical analysis, the question of the endogeneity of the factors under consideration arose (for example, the credit gap, the lending channel, and the unemployment rate can influence each other). To minimize possible problems associated with endogenous parameters, M2 and M3 estimation models were applied. This approach allowed us to take into account the two-stage assessment of the categorical variable in the models, in which the endogenous variables X_it were explained by those variables that are not dependent on Z_it, taking into account the lags for one period ahead for X_it the dependent variable. The empirical analysis made it possible to look through the different phases of the credit cycle and to calculate the changes in credit ratings.

The results of the analysis of the relationship between the credit cycle, identified by means of a credit gap and the dynamics of credit ratings, based on data from both Russian and foreign companies, are presented in Figs. 3 and 4. Within the framework of joint trajectories of credit ratings and credit activity, the following phases are:

Phase 1. At the beginning of the credit cycle, an increase in the volume of lending from the minimum levels against the backdrop of low ratings.

Phase 2. Credit activity continues to increase, the level of credit ratings is also growing.

Phase 3. Credit activity is no longer growing, but credit ratings are still arising (peak point).

Phase 4. Credit activity is beginning to decline: the market is on the verge of a credit recession. Credit ratings reach their maximum values.

Phase 5. As a result of the credit recession, there is a parallel decrease in the volume of loans issued and the level of credit ratings.

In general, the statistical results of modeling credit ratings using various econometric models indicate the high significance of factors such as GDP dynamics, employment, and the credit gap for rating information. Some specifications are also characterized by the importance of the credit gap. Judging by the “fixed effects” of models, no pronounced fundamental differences in the laws governing the information between developed and developing markets have been revealed. A comparison of the results of evaluating the model of the M4 type with and without

Russia shows that the role of country specificity may be significant (for example, xrusd variable from Table 6).

5 An Application-Based Approach on Credit Ratings Aggregation for Risk Management Purposes in Commercial Banks

There is a convergence of the scales of international agencies but an increase in discrepancies between Russian and international agencies. The results described in previous section show that rating methodologies are similar for different sectors. The possibilities of using aggregated rating estimates and their minimization based on the analysis of patterns through the general level of credit ratings as a quantitative indicator of risks for various objects are evaluated (see Sect. 4).

The study presents the results of a multi-logistic model. The obtained estimates of the coefficients from the model show how different groups, both in terms of assets and factors, have a significant impact. Therefore, for each group, it is easy to interpret the economic significance of the explanatory variables from the credit rating level.

The non-obvious results of the clustering of various economic agents include the revision of rating agency methodologies after 2002. The financial patterns and the results of the multi-logistic model show that a significant difference in the evolution of rating standards for industrial companies and commercial banks was a sharp decrease in credit ratings from investment to speculative level. The introduced categorical variables for changing the methodologies of rating agencies are statistically significant at the level of 1%. These observations show that rating agencies have tightened the standards for assigning investment grade ratings, and there has been a review of previously assigned high ratings in the direction of their lowering and moving to a speculative level.

In the process of assessing the relationship between different credit ratings, quantitative instrumental variables of the credit cycle were introduced, which made it possible to predict the size of possible changes in credit ratings in the long term based on changes in the stages of the credit cycle.

The empirical analysis made it possible to identify the various stages of the credit cycle and evaluate changes in credit ratings, both at the beginning of the credit cycle (at the recovery stage) and at its end (at the stage of credit recession).

This analysis showed that the onset of the crisis and downgrades of credit ratings are out of sync: at the onset of the crisis, credit ratings are relatively high, and only when the stage of credit narrowing, consisting of a successive change in the phases of narrowing and recession, begins, does the crisis go into the active stage, and then most ratings begin to be adjusted by agencies.

Based on the resulting credit gap estimates, it is possible to clearly identify the end of one credit cycle and the beginning of the next. Empirical estimates of the logit and multi-logistic models show that with the growth of the credit gap by more than 2.9% over the course of one quarter point the volume of bank loans begins to decrease. And, at the same time there are changes in the dynamics of other macroeconomic indicators, for example, in GDP growth and inflation.

The constructed models, taking into account the different structure of the credit cycles, are based on tracking changes in rating assessments over a relatively large time horizon. Additionally, issues of interconnection with changes in the credit gap and macroeconomic factors were considered. Obviously, credit ratings are not only subject to cyclical changes within the credit cycle, but are also late in relation to the cycle: at the beginning of a credit recession, the ratings are kept high, and, they can continue to grow, but in the conditions of the beginning of the recession phase, credit ratings are gradually declining. It should also be noted that two macroeconomic factors have a strong influence on the level of credit ratings, as well as on the level of the credit gap: these are the growth rate of GDP and the lending channel, which is a universal mechanism of monetary policy.

6 Conclusion

Most commercial banks build their risk accounting policies and models for evaluating internal ratings based on historical and accumulated data and measure their effectiveness on the basis of how much the resulting assessment predicts the probability of default of the borrower both at the current time and during a certain cycle. The estimates obtained in this way, oriented at a specific moment in time, are used by banks to assign a borrower an internal rating that reflects the current or future assessment of credit risks and the financial condition of the borrower.

Thus, the internal rating requires revision as the borrower's state changes over time, which is also reflected in its relationship, due to changes in the credit cycle. This emphasizes the importance of rating changes, as many of them are focused on the long term.

The relationship between credit cycles and changes in credit ratings is an important issue requiring increased attention. The issue of the mutual influence of changes in ratings and the dynamics of credit cycles is open to discussion and there is no single point of view in the empirical work on this issue. The relationship of ratings and the credit cycle can be considered from two sides. Credit cycles affect the change in the level of ratings of issuers. However, the reverse trend is also true, that with changes in ratings, parallel events occur in the financial market.

The stages of the credit cycle are associated both with the growth of loans issued and with the size of the debt securities market. The stage of the recession and the level of credit ratings are not equally adjusted over time: credit ratings during the recession are highly estimated, and only when the stage of credit recession begins, so the credit ratings begin to reflect the market situation. A similar response mechanism of rating agencies is the same for both developed and developing countries.

References

Aktan, B., Çelik, Ş., Abdulla, Y., & Alshakhoori, N. (2019). The impact of credit ratings on capital structure/ISRA. *International Journal of Islamic Finance, 11*(2), 226–245. https://doi.org/10.1108/IJIF-03-2018-0028.

Aleskerov, F. T., Belousova, V. Y. U., Egorova, L. G., & Mirkin, B. G. (2013). Analiz patternov v statike i dinamike, chast' 2: primery primeneniya k analizu social'no-ekonomicheskih processov. *Biznes-informatika, 4*(26), 2013r.

Altman, E., & Rijken, H. (2004). How rating agencies achieve rating stability. *The Journal of Banking and Finance, 28*, 2679–2714. https://doi.org/10.1016/j.jbankfin.2004.06.006.

Amato, D., & Furfine, C. (2003). *Are credit ratings procyclical?*. BIS Working Papers No. 129. doi: https://doi.org/10.1016/j.jbankfin.2004.06.005.

Arteta, C., Kose, M., & Ohnsorge, F. (2017). *The coming interest rate tightening cycle: Smooth sailing or stormy waters?, policy research note 2*. Washington, DC: World Bank. https://doi.org/10.2139/ssrn.2672683.

Barrell, R., Karim, D., & Macchiarelli, C. (2017). *Towards an understanding of credit cycles: Do all credit booms cause crises?/working paper series*. Rimini, Italy: Rimini Centre for Economic Analysis.

Basel III. (2010). *Guidance for national authorities operating the countercyclical capital buffer*. https://www.bis.org/publ/bcbs187.pdf.

Bisenov, M., Grishunin, S., Dyatchkova, N., & Karminsky, A. (2019). The comparison of statistical methods for modelling of credit ratings of industrial companies from BRICS countries. *Eurasian Economic Review, 9*, 1–16. https://doi.org/10.1007/s40822-019-00130-4.

Bordalo, P., Gennaioli, N., & Shleifer, A. (2018). Diagnostic expectations and credit cycles. *Journal of Finance, 73*(1), 199–227. https://doi.org/10.3386/w22266.

Claessens, S., Kose, M., & Terrones, M. (2011). *How do business and financial cycles interact?. IMF working papers, No. WP/11/88*. Washington, DC: International Monetary Fund. https://doi.org/10.1016/j.jinteco.2011.11.008.

Distinguin, I., Hasan, I., & Tarazi, A. (2013). Predicting rating changes for banks: How accurate are accounting and stock market indicators? *Annals of Finance, 9*(3), 471–500. https://doi.org/10.2139/ssrn.2098964.

Dyachkova, N., Karminsky, A., & Kareva, Y. (2019). The determinants of credit cycle and its forecast. *CBI Proceedings, 1*(2019), 320–329. https://doi.org/10.1109/CBI.2019.00043.

Everaert, G., & Zeng, L. (2015). *Does supply or demand drive the credit cycle?. Evidence from Central, Eastern, and Southeastern Europe, IMF Working Papers 15/15*. Washington, DC: International Monetary Fund.

Giese, J., Andersen, H., Bush, O., Castro, C., Farag, M., & Kapadia, S. (2014). The credit-to-GDP gap and complementary indicators for macroprudential policy: Evidence from the UK. *International Journal of Finance & Economics, 19*(1), 25–47. https://doi.org/10.1002/ijfe.1489.

Grishunin, S., Dyatchkova, N., & Karminsky, A. (2018). Credit ratings patterns for BRICS industrial companies. *Procedia Computer Science, 634*, 17–25. https://doi.org/10.1016/j.procs.2018.10.212.

Karminsky, A., & Polozov, A. (2016). *Handbook of ratings: Approaches to ratings in the economy, sports, and society*. Cham, Switzerland: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-39261-5.

Kiff, J., Kisser, M., & Schumacher, L. (2013). *Rating through-the-cycle: What does the concept imply for rating stability and accuracy?, IMF Working Paper*. Washington, DC: International Monetary Fund. https://doi.org/10.2139/ssrn.3127545.

Langohr, H., & Langohr, P. (2008). The Rating agencies and their credit ratings. *Wiley Finance, 524*.

Loffler, G. (2004). An anatomy of rating through the cycle. *Journal of Banking and Finance, 28*, 695–720. https://doi.org/10.2139/ssrn.275842.

Mian, R., Kamalesh, R., & Amir, S. (2013). *Household balance sheets, consumption, and the economic slump*. Chicago Booth Research Paper No. 13–42. doi: https://doi.org/10.2139/ssrn.1961211.

Moody's rating methodology for banks. (2016). https://www.moodys.com/researchdocumentcontentpage.aspx?docid=PBC_186998

Nkomo, J., & Chiwanza, W. (2014). Predicting a bank's credit rating change using financial data and public financial news. *Interdisciplinary journal of contemporary research in business, 6*.

Ponomarenko, A., Deryugina, E., & Rozhkova, A. (2018). *Kogda ocenki kreditnyh razryvov yavlyayutsya dostovernymi?*. Seriya dokladov ob ekonomicheskih issledovaniyah Banka Rossii No 34, Iyul' 2018.

Repullo, R., & Saurina, J. (2011). *The countercyclical capital buffer of Basel III: a critical assessment*. CEPR Discussion Papers 8304, C.E.P.R. Discussion Papers. https://ssrn.com/abstract=1794894.

Ryan, J. (2012). *The negative impact of credit rating agencies and proposals for better regulation*. Stiftung Wissenschaft und Politik (SWP) Working paper. Research Division EU Integration, German Institute for International and Security Affairs. https://www.swp-berlin.org/.

Schüler, S. (2018). *De-trending and financial cycle facts across G7 countries: mind a spurious medium term*. Working Paper Series ECB, No 2138, March 2018. https://ssrn.com/abstract=3152820.

Van Laere, E., & Duan, J.-C. (2012). A public good approach to credit ratings – from concept to reality. *Journal of Banking & Finance, 36*, 3239–3247. https://doi.org/10.1016/j.jbankfin.2012.03.012.

Part III
Estimating and Modeling Credit and Market Risks in Banking

Bank Credit Risk Modeling in Emerging Capital Markets

Alexander Karminsky and Alexei Morgunov

Abstract Models for assessing the probability of default play an important role in the risk management systems of commercial banks, as they allow assessing the creditworthiness of various counterparties and transactions. Many Russian banks are trying to switch to an advanced approach based on internal ratings (IRB-approach) for evaluating regulatory capital. The main goals that banks pursue when switching to an advanced approach are: stability of credit risk assessment for the ability to carry out strategic planning; the validity of the credit risk assessment to simplify interaction with the regulator and external and internal audit; potential reduction of regulatory capital due to the high quality of the forecast capabilities of the developed models, which leads to a reduction in the regulatory capital of banks. To use internal rating models in the calculation of regulatory capital banks serve the petitions on them to the regulator, on basis of which external validation of the models is carried out and a decision about the possibility of using models for regulatory purposes is made. The main event of credit risk, the default event is determined by banks in the framework of credit policy, is consistent with the Central Bank and is predicted using models for assessing the probability of default. The PD models are the most popular in banking practice due to the fact that according to regulatory requirements, they are developed on the horizon of 1 year, and the minimum amount of statistical data for such models must be at least 5 years. The risk segments are identified using both economic and statistical evaluation criteria based on the banks available empirical data for each group of borrowers to build separate models (Allen, Financial risk management: a practioner's guide to managing market and credit risk. Wiley, Hoboken, NJ, 288 p, 2003; Lobanov and Chugunov, Encyclopedia of financial risk management, 4th edn, Alpina Business books, 932 p, 2009; Rogov, Risk management, Finance and statistics, Moscow, 120 p, 2001). This paper will describe the specifics of developing models for low-default risk segments (bank assets), both low-default and high-default risk segments (corporate borrowers), and

A. Karminsky · A. Morgunov (✉)
National Research University Higher School of Economics, Moscow, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_5

high-default risk segments, including taking into account the availability of a small amount of static data (residential real estate lending and project finance segments).

Keywords Credit risk · Logistic regression · Decision trees · Bayesian approaches

JEL Classification G01 · G28 · G32

1 Building Models for Low-Default Borrowers Using Banks as an Example

In current banking practice, the portfolios of bank assets of the largest Russian banks are low-defaulted due to the fact that most often such lending is carried out with the best and most reliable borrowers in the market, and there are no default statistics for such banks, and the development of a model based on external license revocations or bank failures will not be representative of existing bank portfolios. In this case, the most common approach is based on approximating the frequency of default of external ratings (shadow rating approach). The essence of this approach is to develop a linear regression model of one of the following three types (Karminsky 2015):

$$\ln\left(\frac{\mathrm{PD}}{1-\mathrm{PD}}\right) = \vec{a} \times \vec{x} + b, \tag{1}$$

$$\ln\ \mathrm{PD} = \vec{a} \times \vec{x} + b, \tag{2}$$

$$\mathrm{PD} = \vec{a} \times \vec{x} + b, \tag{3}$$

where PD—average probability of default of the borrower's external ratings; $\vec{a}$—vector row of regression coefficients for normalized risk factors; b—intercept; $\vec{x}$—vector column of normalized values of risk factors that affect the occurrence of a default event.

The weights of risk factors in models are defined as the ratio of modules of values of the corresponding regression coefficients to the sum of modules of regression coefficients, i.e. $\frac{|a_i|}{\sum_{i=1}^{N}|a_i|}$.

The correspondence between the ratings of S&P, Moody's, and Fitch and the annual probability of default is shown in Table 1, obtained from the annual reports of S&P, Moody's, and Fitch, which contain information on the default frequencies of external ratings (these are the probability of default for the economic cycle—TTC, which are used for regulatory purposes) (Mills 2003). The average probability of default of a borrower can be defined as the average probability of default for all external ratings available to it (S&P, Moody's, Fitch) (Pomazanov 2017).

Thus, the developed model in this case is a mapping of the probability of default of borrowers on the average assessment of the probability of default of external

Table 1 Correspondence between the ratings of S&P, Moody's, Fitch, and annual probability of default

Rating S&P, Fitch	Rating Moody's	Scale_num	PD (migration matrix)
AAA	Aaa	1	0.00%
AA+	Aa1	2	0.00%
AA	Aa2	3	0.02%
AA−	Aa3	4	0.03%
A+	A1	5	0.05%
A	A2	6	0.06%
A−	A3	7	0.06%
BBB+	Baa1	8	0.10%
BBB	Baa2	9	0.16%
BBB−	Baa3	10	0.24%
BB+	Ba1	11	0.32%
BB	Ba2	12	0.53%
BB−	Ba3	13	0.95%
B+	B1	14	2.01%
B	B2	15	3.41%
B−	B3	16	6.75%
CCC+	Caa1	17	26.89%
CCC	Caa2	18	26.89%
CCC−	Caa3	19	26.89%
CC+	Ca/C	20	26.89%
CC	Ca/C	21	26.89%
CC−	Ca/C	22	26.89%
C+	Ca/C	23	26.89%
C	Ca/C	24	26.89%
C−	Ca/C	25	26.89%
RD	Ca/C	26	100%

rating agencies Moody's, S&P, Fitch. In such an assessment, additional analysis should be carried out on the compliance of the internal bank definition of default with the definition of default of external rating agencies (Allen 2003; Lobanov and Chugunov 2009; Rogov 2001). In general, the main default criteria for both Moody's, S&P, Fitch, and most Russian banks are the facts of the borrower's insolvency (nonperforming loans for more than 90 days, forced restructuring, bankruptcy), so mapping to information from external rating agencies seems reasonable. The risk indicators (they can be either continuous or discrete) participate in the procedures of logistic transformation (continuous variables) and WOE transformation (qualitative variables) before being used into the regression (Karminsky et al. 2015). The essence of logistics and WOE transformations is to reduce the impact of outliers, the formulas for their implementation will be given below. It is also possible to potentially perform WOE transformation for continuous risk factors, but this often leads to overfitting of the obtained intervals, i.e. the logistic transformation is generally more stable than the WOE transformation.

Logistics transformation in the case when the growth of the factor reduces the level of credit risk of borrowers from an economic point of view is carried out according to formula (4), and in the case when the growth of the factor increases the level of credit risk of borrowers from an economic point of view, formula (5) is used:

$$\mathrm{Ratio}_{\mathrm{tr}} = \begin{cases} \dfrac{1}{1 + \exp\{-\mathrm{Slope} \times (\mathrm{Ratio} - \mathrm{Midpoint})\}} \\ 1 - \dfrac{1}{1 + \exp\{-\mathrm{Slope} \times (\mathrm{Ratio} - \mathrm{Midpoint})\}}, \end{cases} \quad (4)\ (5)$$

where $\mathrm{Ratio}_{\mathrm{tr}}$—transformed value of the risk factor; Ratio—the value of the risk factor; Slope—transformation coefficient for the risk factor; Midpoint—$(\mathrm{Ratio}_{5\%} + \mathrm{Ratio}_{95\%})/2$, where $\mathrm{Ratio}_{5\%}$ and $\mathrm{Ratio}_{95\%}$—the risk factor percentiles are 5% and 95%.

The values of the slope transformation coefficients are found from the following normalization condition (6):

$$\frac{1}{1 + \exp\{-\mathrm{Slope} \times (\mathrm{Ratio}_{95\%} - \mathrm{Midpoint})\}} = 0.95. \quad (6)$$

For qualitative (discrete) risk factors, the woe transformation is performed using the following formula for comparability of discrete values (groups of factors) by default level (Siddiqi 2006):

$$\mathrm{WOE}_i = \ln\left(\frac{(1 - \mathrm{avgPD}_i)/(1 - \mathrm{avgPD}_{\mathrm{all}})}{(\mathrm{avgPD}_i/\mathrm{avgPD}_{\mathrm{all}})}\right), \quad (7)$$

where WOE_i—value of the WOE indicator for the factor group with the sequence number i; avgPD_i—the average probability of default for the average PD of S&P, Moody's, and Fitch borrowers for the factor group with ordinal number i; $\mathrm{avgPD}_{\mathrm{all}}$—the average probability of default for the average PD of S&P, Moody's, and Fitch borrowers in the entire sample.

For the transformed coefficients, in order to bring the risk factors to a single scale in standard deviations, normalization was performed using the following formula:

$$\mathrm{Ratio}_{\mathrm{Norm}} = \frac{\mathrm{Ratio}_{\mathrm{tr}} - \mathrm{Mean}}{\mathrm{Std}}, \quad (8)$$

where $\mathrm{Ratio}_{\mathrm{Norm}}$—normalized value of the risk factor; $\mathrm{Ratio}_{\mathrm{tr}}$—transformed value of the risk factor; Mean—average value of the transformed risk factor; Std—standard deviation of the transformed risk factor.

Then the converted risk factors are substituted into formulas (1), (2), or (3) and the regression coefficients of various model variants are estimated using the least squares method. In other words, all possible models are sorted out and the best one is selected based on various statistical and expert criteria. As statistical criteria, the coefficient of determination R^2 or adjusted R^2 or the Somers'D rank correlation coefficient can be used, calculated using the following formula:

$$\mathrm{SD} = \frac{N_C - N_D}{N_A - N_B}, \tag{9}$$

where SD—value of indicator Somers'D; N_C—the number of consistent pairs between the normalized risk factor values and the probability of non-default of external ratings in the sample (1—the probability of default of external ratings); N_D—the number of inconsistent pairs between the normalized values of the risk factor and the probability of non-default of external ratings in the sample; N_A—the total number of permutations in the sample (for a dimension selection N: $N_A = \frac{N\times(N-1)}{2}$); N_B—the total number of permutations of repeated values of the probability of non-default of external ratings in the sample ($N_B = \sum_{i}^{Q} \frac{t_i\times(1-t_i)}{2}$), where t_i is the number of duplicate ith value of the probability of non-default of external ratings in the sample, and Q is the total number of duplicate ith value of the probability of non-default of external ratings in the sample.

The predictive and discriminative ability of models is considered weak if the values of the coefficients R^2 and *Somers'D* (*SD*) are <40% and strong if the values of these coefficients are more than 60%.

To reduce the number of variables in the model iteration, one-factor analysis results are often used, excluding statistically insignificant variables. When selecting models, its stability is evaluated on a separate out-of-time sample and in the cross-validation procedure on average on out-of-time samples.

The following four main groups of indicators can be used as groups of variables to iterate through bank models (Karminsky and Kostrov 2013):

1. Capital adequacy indicators;
2. Indicators that characterize the quality of bank's assets;
3. Indicators that characterize the quality of management (business activity of banks);
4. Liquidity indicators of the banks.

Examples of group 1 ratios (capital adequacy):

- Sources of own funds/Total liabilities;
- Sources of own funds/Borrowed funds;
- Sources of own funds/Assets generating direct income;
- Authorized capital/Sources of own funds;
- Sources of own funds/Deposits of individuals.

Examples of group 2 coefficients (bank asset quality):

- Assets generating direct income/Total assets
- Risk protection coefficient (retained earnings of previous years (uncovered losses of previous years) + Unused profit (loss) for the reporting period + Reserve Fund)/Assets generating direct income;
- Level of assets with increased risk (Other loans + Loans and other deposited funds with overdue payments + Investments in Finance leases and acquired rights of claim + Investments in securities + |Accounts receivable − Accounts payable|)/Total assets.
- Overdue debt/loans, deposits, and other placed funds.
- Accounts receivable/(Total assets − Assets generating direct income).

Examples of group 3 coefficients (quality of management):

- Loans and other deposited funds/Total assets.
- Loans and other deposited funds/Borrowed funds.

Examples of group 4 coefficients (liquidity indicators):

- Liquidity ratio of the «first stage reserve» (Cash currency and payment documents + In the Bank of Russia)/(Interbank loans (deposits) received (borrowed) + Loans (deposits) received from the Bank of Russia + Funds of clients who are not credit institutions), where In the Bank of Russia = On the organized securities market + Savings accounts of credit organizations in the issuance of shares + Funds of authorized banks deposited with the Bank of Russia + Funds on account with the Bank of Russia + Accounts for other operations with the Bank of Russia.
- Liquidity ratio of the «second stage reserve» (Cash currency and payment documents + In the Bank of Russia + Debt obligations of the Russian Federation + Debt obligations of subjects of the Russian Federation and local governments + Debt obligations of foreign States + Debt obligations of the Bank of Russia)/(Interbank loans (deposits) received (borrowed) + Loans (deposits) received from the Bank of Russia + Funds of clients who are not credit institutions).
- Cash/Total assets.
- Cash/Borrowed funds.
- Balance ratio of the bank's active and passive policies (Cash + Mandatory reserves + Interbank loans (deposits) provided (deposited) + Until demand + Demand loans and promissory notes at sight + Financial assets at fair value + Net investments in HTM-securities + Net investments in available-for-sale securities + Accounts receivable)/(Funds on correspondent bank accounts + Interbank loans (deposits) received (attracted) + Funds in legal entities' accounts of the individuals (non-credit organizations) + Deposits and other borrowed funds on demand + Accounts payable).

By iterating through models with functional dependencies (1), (2), and (3), you can develop a model that maps the probability of default on the probability of default of clients with external ratings that cover a particular bank's portfolio. In addition, it should be noted that often the approach based on external ratings gives an excessively conservative PD forecast and does not take into account internal statistics of customer observations. For this purpose, many banks perform additional calibration of the models based on Bayesian methods, taking into account real bank statistics of observations. This approach consists of obtaining a posteriori probability of default for borrowers with available default statistics using the closest possible loan portfolio (CPP), which has a priori probability of default on external ratings and is based on the Bayes formula of conditional probability density (Surzhko 2017; Pugachev 2002):

$$f(x|t=T,f=F)=\frac{P(t=T,f=F|x)\times f(x)}{\int_0^1 P(t=T,f=F|z)\times f(z)\times dz}, \tag{10}$$

where x—a random variable that characterizes the a priori probability of default; t—a random variable that characterizes the a posteriori number of defaulted borrowers with available statistics; f—a random variable that characterizes the a posteriori number of dissatisfied borrowers with available statistics; T,F—the number (historical number) of defaulted and non-defaulted borrowers, respectively, with available default statistics for the risk segment.

Assuming that the a posteriori number of defaults in the loan portfolio is distributed by the binomial distribution Bin(T, F), and the a posteriori probability of default is distributed by the beta distribution Beta(a,b), we get

$$f(x|t=T,f=F)=\frac{C_T^{F+T}\times x^T\times(1-x)^F\times\frac{x^{a-1}\times(1-x)^{b-1}}{\int_0^1 z^{a-1}\times(1-z)^{b-1}\times dz}}{\int_0^1 C_T^{F+T}\times y^T\times(1-y)^F\times\frac{y^{a-1}\times(1-y)^{b-1}}{\int_0^1 z^{a-1}\times(1-z)^{b-1}\times dz}\times dy}, \tag{11}$$

where C_T^{F+T}—number of combinations without repetitions; t—parameters of the beta distribution that characterizes the a priori probability of default x.

After making the transformations in formula (10), we get

$$f(x|t=T,f=F)=\frac{x^{T+a-1}\times(1-x)^{F+b-1}}{\int_0^1 y^{T+a-1}\times(1-y)^{F+b-1}\times dy}\tilde{\text{Beta}}(T+a,F+b). \tag{12}$$

Thus, the a posteriori distribution in the presence of T default borrowers and F non-default borrowers is approximated by the beta distribution Beta($T+a$, $F+b$). From the properties of the beta distribution, it follows that PD_{TTC} (PD for the economic cycle, used for evaluating regulatory capital) is determined by the formula:

$$\text{PD}_{\text{TTC}}=\frac{T+a}{T+F+a+b}\tilde{\text{Beta}}\,(T+a,F+b). \tag{13}$$

The coefficients *a* and *b* are determined based on approximating the historical probability distribution of default clients with external ratings by beta distribution by maximizing the maximum likelihood function.

The transition from PD accounting for Bayesian adjustment to PD with this adjustment (a posteriori PD) can be performed using the following formula:

$$\mathrm{PD}_{\mathrm{TTC}_{\mathrm{post}}} = N(N^{-1}(\mathrm{PD}_{\mathrm{TTC}_{\mathrm{pr}}}) + b), \qquad (14)$$

where $\mathrm{PD}_{\mathrm{TTC}_{\mathrm{pr}}}$—a priori probability of default (obtained using a linear regression model); $\mathrm{PD}_{\mathrm{TTC}_{\mathrm{post}}}$—a posteriori probability of default (obtained with Bayesian adjustment); *N*—probability function of the standard normal distribution; N^{-1}—quantile function (inverse probability function) of the standard normal distribution; *b*—intercept.

The intercept of calibration is selected for the entire historical loan portfolio using formula (14), taking into account the need to obtain the average $\mathrm{PD}_{\mathrm{TTC}}$ value determined by formula (13).

2 The Construction of Models for Corporate Borrowers

At the moment, Russian banks are focusing considerable attention on lending to the largest borrowers. This is largely due to the fact that in Russia there is instability in the economy with regular recessionary and crisis phenomena (falling GDP, inflation). This is due to the dependence of the Russian economy on raw materials and energy prices, which are quite volatile. The most stable borrowers with minimal credit risk are the largest borrowers. In the largest Russian banks, borrowers with annual revenue or average annual assets of more than 20–30 billion rubles are considered as the largest Russian borrowers. The number of defaults for such clients is insignificant and it is impossible to build standard statistical models for assessing the probability of default (logistic regression, classification trees, and other classical algorithms) in this case. At the same time, the limit of indebtedness of such clients in the largest banks reaches significant amounts. The largest Russian banks prefer to work now with the largest clients and only slightly try to develop the direction of lending to small and medium-sized businesses.

For this reason, to develop rating models for the largest borrowers, an approach for low-default portfolios is used, similar to the shadow rating approach for banks. At the same time, it is necessary to tell about the specifics of allocating the risk segment of the largest (low-default borrowers). There are two main approaches for identifying the largest borrowers segment:

- Based on the identification of the threshold for borrowers, above which no defaults were recorded in the bank, and the development of the model based on external ratings or, more rarely, on expert ranking;

- Based on the identification of a threshold for borrowers that provides optimal coverage of the resulting portfolio of borrowers in terms of external ratings (at this threshold, a small number of clients with external ratings out of all possible clients with external ratings do not fall into the segment, and at the same time, there should also be a minimum number of clients in the segment without external ratings, it is based on maximizing the F1 measure) and in this case, the development of the model is based on external ratings.

The expert ranking approach can also be applied to individual sub-segments of the corporate portfolio if there is insufficient volume of external ratings and default statistics. It consists in the fact that each client is ranked by business departments, receiving an expert rating from 1 to R (1—the worst, R—the best) according to strictly defined criteria. It is better to conduct the ranking for the same client employees in the business department and the underwriting department. Then the model reproduces expert ratings and allows you to get a score for the client, which allows you to rank it in terms of the level of creditworthiness. Calibration of such portfolios is most often based on available external ratings, which are not enough to develop a separate model, but enough to map the resulting ranking score on the external ratings. The algorithm for developing an expert ranking model is based on the construction of ordinal logistic regression models (ordered choice models) and is shown below. It allows you to get cumulative probabilities of being in expert ratings with ordinal numbers 1; 1,2; 1,2,3; 1,2,3,. . .,$R-1$ provided the same values of regression coefficients for risk factors for each cumulative probability using a logistic functional relationship (Ayvazyan 1989; McCullagh and Nelder 1990):

$$P_{1,j} = \frac{1}{1 + \exp\left\{-\left(\vec{a} \times \vec{x}^T + b_j\right)\right\}}, \tag{15}$$

where $P_{1,\ j}$—cumulative probability of finding a borrower in expert ratings with ordinal numbers 1,2,. . ., j; j—ordinal number of the corresponding expert rating ($j = 1,\ldots, R-1$); $\vec{x}^T$—a column vector of the normalized values of risk factors on the expert rating of the borrower; $\vec{a}$—vector row of regression coefficients for normalized risk factors; b_j—the regression coefficient is an intercept when evaluating the cumulative probability of finding a borrower in expert ratings with ordinal numbers 1,2, . . .,j, at the same time for any j и j + 1: $b_{j+1} > b_j$.

The vector coefficients $\vec{a}$ and free regression terms b_j are based on the maximization of the logarithmic likelihood function (16):

$$\mathrm{Log}L = \sum_{i=1}^{N}\left[Y_{i1} \times ln\left(P_{i,1,1}\right) + \sum_{j=2}^{R}\left[Y_{ij} \times ln\left(P_{i,1,j} - P_{i,1,j-1}\right)\right]\right], \tag{16}$$

where Y_{ij}—binary variable from the set {0;1} that records the fact that the ith borrower is in the expert rating with the ordinal number j; $P_{i,\ 1,\ j}$—cumulative probability of finding a borrower with an ordinal number in expert ratings with ordinal numbers 1,2, . . ., j, obtained using the logistics function (15), at the same time $P_{i,\ 1,\ R} = 1$; N—number of borrowers in the sample.

The overlap of the condition for identical values of regression coefficients for risk factors is due to the need to obtain the parameter Score $= \vec{a} \times \vec{x}^T$, which allows you to rank borrowers in terms of creditworthiness. As the value of the Score parameter increases, the probability of being a borrower in expert ratings with a higher creditworthiness (with a larger sequential number) increases.

For corporate borrowers with sufficient default statistics, the most commonly used models are binary logistic regression, an interpreted classification tree (CART algorithm), or an ensemble of interpreted decision trees (usually 2–4 trees).

The most popular approach is based on logistic regression, which is used to predict the event of default/non-default of the borrower {0;1}. The functional dependency PD for logistic regression looks like this (Pomazanov 2017):

$$\mathrm{PD} = \frac{1}{1 + \exp\left\{-\left(\vec{a} \times \vec{x}^T + b\right)\right\}}, \tag{17}$$

where $\vec{x}^T$—vector column of normalized values of risk factors that affect the occurrence of a default event for the borrower; $\vec{a}$—vector row of regression coefficients for normalized risk factors; b—intercept.

The coefficients of the vector $\vec{a}$ and intercept b are based on the maximization of the logarithmic likelihood function (18):

$$\mathrm{Log}L = \sum_{k=1}^{N} \left[Y_i \times \ln\left(\mathrm{PD}_i\right) + (1 - Y_i) \times \ln\left(1 - \mathrm{PD}_i\right)\right], \tag{18}$$

where Y_i—a binary variable from the set {0;1} that records the fact that the borrower has not defaulted in 1 year horizon; PD_i—probability of default for a borrower with an ordinal number obtained using the logistics function.

The transformations of risk factors for a binary variable that records the presence or absence of a default event are similar to those that were described in banks (logistics and WOE transformation and normalization of risk factors), but there is a difference in the implementation of WOE transformation:

$$\mathrm{WOE}_i = \ln\left(\frac{\left(N_{\mathrm{good}_i}/N_{\mathrm{good}_{\mathrm{all}}}\right)}{\left(N_{\mathrm{bad}_i}/N_{\mathrm{bad}_{\mathrm{all}}}\right)}\right), \tag{19}$$

where WOE_i—value of the WOE indicator for the factor group with the sequence number i; N_{good_i}—the number of non-default borrowers in the factor group with the sequential number i; $N_{\mathrm{good}_{\mathrm{all}}}$—total number of non-defaulted borrowers; N_{bad_i}—the number of defaulted borrowers in the factor group with the sequence number i; $N_{\text{bad}_{\text{all}}}$—total number of defaulted borrowers.

The CART (Classification and Regression Tree) algorithm is designed for building a binary decision tree. At each step of building the tree, the rule generated in the node divides the specified set of examples into two parts: the part where the rule is executed (the right subtree) and the part where the rule is not executed (the left subtree) (Breiman 1984, 2001). The CART method is used for continuous and discrete variables. This method iterates through all possible branching options for each node and selects the variable for which the evaluation function gives the best indicator. The estimation function used by the CART algorithm is based on the intuitive idea of reducing the uncertainty (heterogeneity) in the node and is based on the Gini impurity index:

$$\text{Gini}(T) = 1 - \sum\nolimits_{i=1}^{n} \text{p}_i^{\,2}, \tag{20}$$

where p_i is the probability (relative frequency) of class i in T.

If the set T is split into two parts $T1$ and $T2$ with the number of examples in each $N1$ and $N2$, respectively, then the split quality indicator is equal to:

$$\text{Gini}_{\text{split}}(T) = \frac{N1}{N} \times \text{Gini}(T1) + \frac{N2}{N} \times \text{Gini}(T2). \tag{21}$$

The best partition is the one for which $\text{Gini}_{\text{split}}(T)$ is minimal. The choice of the best tree is determined using the definition of such a level of its depth, after increasing which the predictive ability of the tree in cross-validation begins to decrease or slightly increases, not in proportion to the complication of the algorithm. When building a tree, its individual branches may not be interpreted.

For such cases, the tree is manually pruned with minimal loss of accuracy and the absence of non-interpreted branches (most often due to a small number of observations). An example of developing a classification tree for assessing the probability of borrowers defaulting on financial statements is shown in the Fig. 1. An ensemble of decision trees constructed using the CART algorithm from random subsamples of the general population can also be used as a model-approach. This improves the stability of the developed models.

To obtain an estimate of the probability of default on the forecast horizon of 1 year, models based on default statistics are calibrated for the economic cycle using formula (22) (Zhevaga and Morgunov 2015):

$$\text{PD}_{1\ \text{year}} = \frac{1}{1 + \exp\{\alpha \times \text{Score} + \beta\}}, \tag{22}$$

where $\vec{a}$ & b—the regression coefficients of the model; α & β—coefficients determined when calibrating the rating model based on the central tendency (TTC concept).

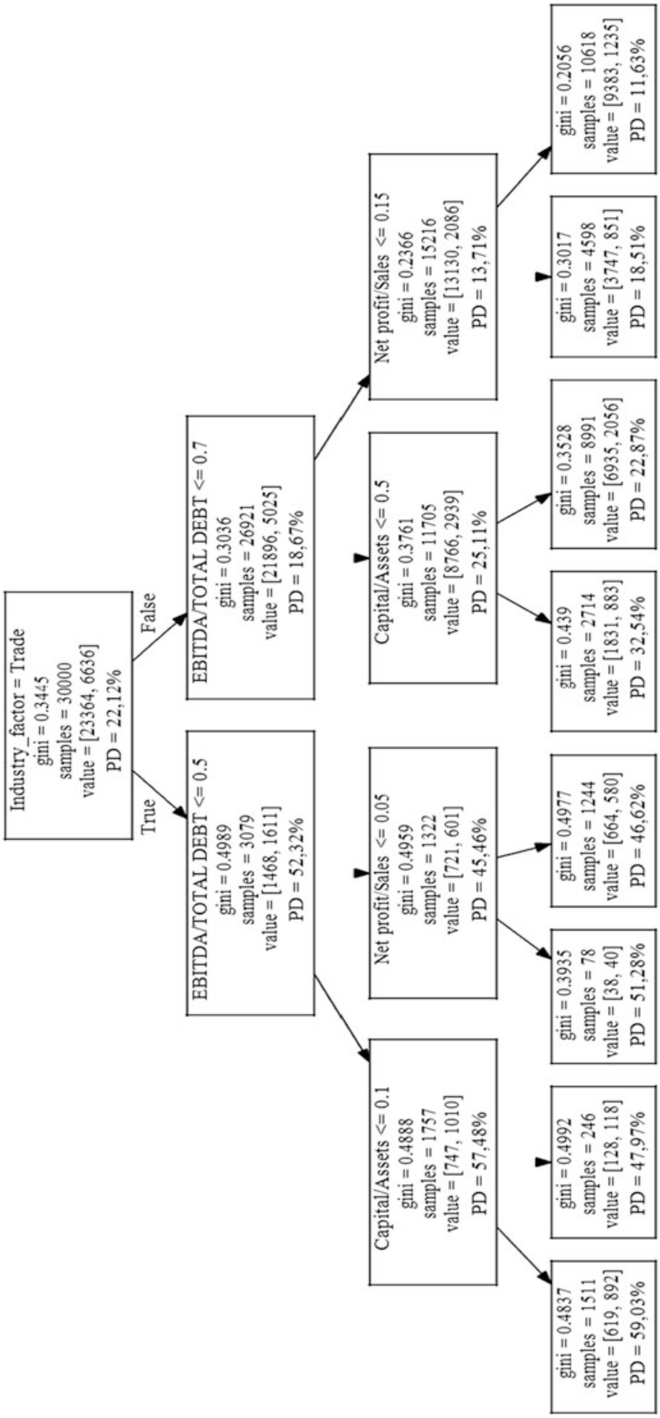


Fig. 1 Example of using the CART classification tree

It should be noted that in the development of corporate models at the moment most often use a modular structure because there is a desire to take into account as much information as possible in the model within individual modules to improve the stability of the model. The modules of corporate model can be divided into the following main groups:

1. The module of financial factors (the groups of factors to develop the module will be listed below);
2. The module of quality factors (Business reputation, market positioning, business ethics, and completeness of information, whether or not the borrower has suppliers/contractors).
3. The module of credit history (external and internal credit history of the borrower in short and long-term time periods).
4. The transaction data (information on internal transactions of the borrower, which allows identifying additional relevant information from the borrower, in particular, the presence of debt on the accounts of card files, that is, the arrest of accounts, which potentially indicates the problems of the borrower in the near future).
5. The information on lawsuits (the amount of claims as a defendant, whether the borrower has bankruptcy claims).
6. The module of borrower's environment influence, which takes into account the characteristics of counterparties and individuals related to the borrower, both by legal (consolidated group of borrowers) and economic criteria (a large number of transactions between clients, connectivity for individual projects with borrowers).
7. The module of warning signals—risk factors that characterize rare negative events that at the same time have a negative impact on the creditworthiness of borrowers.

These modules are aggregated and allow you to get the resulting scoring score, which is converted to the probability of default for the model based on the calibration results.

Also it is important that the benchmarks of risk-factors of corporate borrowers in different sectors are different due to different levels of the turnover in industries. It is necessary to take into account this specifics to develop all credit risk models for the portfolios of corporate borrowers with different industries.

The following Table 2 provides a list of financial risk factors and groups that can be used in developing corporate models. In practice, most often there are groups of liquidity, debt load, financial leverage, borrower turnover, profitability, debt structure, and scale of operations.

Table 2 List of indicators for evaluation for the module of financial factors

Group of factors	Factor name
Liquidity	Cash/Current assets
	(Cash + Short-term financial investments)/Current assets
	(Cash + Short-term financial investments)/Short-term liabilities
	Current assets/Current liabilities
	(Cash + Short-term financial investments + Accounts receivable + Taxes)/Short-term liabilities
	Non-current assets/Non-current liabilities
	Working capital/Assets
Debt load	EBITDA/Interest Expense
	EBITDA/Total debt
	Revenue/Total debt
	Cost/Revenue
	Income from operating activities/(Cost of sales + Commercial expenses + Management expenses)
Leverage	Capital/Assets
	Equity/Current liabilities
	Equity/Long-term liabilities
	Capital/Long-term borrowings + Short-term borrowings
	Equity/Net debt
	Capital/Long-term borrowings
	Capital/Short-term borrowings
Turnover	Mid-annual value of accounts receivable/Revenue
	Mid-annual value of inventory/Revenues
	Mid-annual value of accounts payable/Revenue
	Accounts receivable turnover + Inventory turnover
	Mid-annual value of assets/Revenue
	Mid-annual value of non-current assets/Revenue
Profitability	Gross profit/Revenue
	Gross profit/Assets
	Operating income/Revenue
	Income before taxes/Revenue
	Net profit/Revenue
	EBITDA/Revenue
	Net profit/Mid-annual value of assets
	Net profit/Mid-annual value of capital
	(Commercial expenses + Management expenses)/Revenue
Debt structure	Short-term debt/Short-term debt + Long-term debt
Size	Natural logarithm of assets
	Natural logarithm of capital
	Natural logarithm of revenue

3 Specifics of Developing Models for Clients of Residential Real Estate Lending Segments

Construction occupies a significant part of the Russian GDP structure. In recent years, the share of GDP has decreased from 5.9% in 2014 to 5.4% in 2018. If we look at absolute indicators (the amount of funds and the number of housing units entered), they are also declining. So, in record 2015, the figure of housing commissioning volume was 85.35 million square meters. Summary data for 2015–2019 is shown in Table 3.

Thus, currently, construction, and residential real estate to a greater extent, is stagnating. There are the following reasons for this:

- Construction is sensitive to the economic situation in the country. GDP grows faster during periods of growth and falls more strongly during the times of economic instability.
- If the infrastructure of the projects is mainly financed by the State, then the construction of residential real estate primarily depends on the dynamics of the population's disposable income, which decreased by 9.4% over the same period.
- The third factor is changes in the legislation of the Russian Federation—the ban on shared-equity construction from 2018 and the mandatory transition to escrow accounts.

The conclusion from this is an increase in risks for construction companies, especially those engaged in housing construction, that reflected in an increase in the number of bankruptcies of companies. The number of bankruptcies of housing construction companies during the last years are shown in the Table 4.

In this situation, banks that issue loans for housing construction need to have high-precision models for assessing borrowers. The specifics of developing models for the construction of residential real estate are that many housing companies have separate projects, the success of which directly depends on their creditworthiness. It should also be noted that with a small number of projects, the company's financial statements may not allow us to predict exactly the risks of borrowers of residential real estate construction, especially before such borrowers enter the operational phase when making a decision on their lending. Thus, the role of using in models of risk indicators that characterize individual projects of such companies is increasing (for

Table 3 Volume of housing commissioning in Russia in 2015–2018

Year	2015	2016	2017	2018
Housing commissioning volume, million sq. m.	85.35	80.2	78.6	75.66

Table 4 Number of bankruptcies of housing construction companies (turnover from 50 million rubles to 3 billion rubles)

Year	2013	2014	2015	2016	2017	2018
The number bankruptcies	14	21	24	16	82	88

Table 5 Risk factors of project companies based on financial statements

№	Factor name	Description
1	Profitability of sales, %	Gross profit/revenue
2	EBITDA margin, EBITDA, %	EBITDA/revenue
3	Current liquidity ratio, %	Current assets/Current liabilities
4	Quick liquidity ratio, %	Describes the company's ability to repay short-term liabilities using the sale of liquid assets
5	Absolute liquidity ratio, %	Most liquid assets/short-term liabilities
6	Coefficient of the provision of own working capital, %	Own circulating assets/current assets
7	Total debt to EBITDA ratio, %	Total debt/EBITDA
8	Repayment period of accounts receivable, days	–
9	The period of repayment of accounts payable, days	–
10	Gross margin, %	Gross profit/revenue
11	Total debt to equity ratio, %	–
12	Revenue	–
13	Total assets	–
14	Return on assets ROA, %	Net profit/assets

example, residual value-weighted indicators for 5 major projects of the project company) and the models include both the module based on reporting of the project companies and the module based on financial indicators of major projects (including qualitative, expert risk factors that characterize the construction risks of individual projects). When building a module based on the financial statements of project companies you should pay attention to the following risk factors in the Table 5 (Benninga 2008; Jorion 2007; Joseph 2013; Coleshaw 1989).

The risk factors of the projects module include the following factors (Table 6) (Lynch 2010; Esty 2003; Fight 2006; Finnerty 2013; Davis 2003).

The most significant risk-factors for housing construction in practice are the LTV, The ratio of the market and book value of the project, The fact of balloon payment of the project, because the creditworthiness of project companies is directly affected by the liquidity and quality of collateral of their projects. Approaches to implementing PD estimation models in this area are similar to the main interpreted approaches used in developing models based on default statistics: logistic regression, classification trees, interpreted ensembles of classification trees, and model calibration is performed using formula (21) based on the resulting final score for the models. In this case, the target variable is used as the fact of default of the project company, the module variables based on the project company's financial statements are used only after the company enters the operational phase, and only the project module is used until the company enters the operational phase.

Table 6 Risk factors of the projects module

№	Factor name	Description
1	IRR	The interest discounted rate at which the net cash flow from operating activities, including income from participation in the capital of third parties, is equal to investment costs of the project
2	Weighted DSCR (debt service coverage ratio)	It characterizes the quality of debt service for the project, that is, the adequacy of funds to repay liabilities: $\text{DSCR}_{\text{weighted}} = \sum_{t=1}^{H} t \times \frac{\text{CFADS}_t}{\text{PR}_t + \text{IP}_t + \text{LP}_t}$ where CFADS (cash flow available for debt service) is the cash flow for servicing borrowed funds; PR (principal repayment) payments in part of the principal amount of loans and loans; IP (interests payments) interest payments on borrowed funds; LP (lease payments) lease payments; t number of the payment period (total H payments) relative to the start of the project
3	LLCR (loan life coverage ratio)	It characterizes the company's ability to pay off project debts at the expense of future cash flows: $\text{LLCR} = \frac{\sum_{t=1}^{H} \frac{\text{CF}_t}{(1+i)^t} + \text{DR}}{\text{Debt}}$ where CF is the project cash flow from operating activities; i project interest rate (or WACC); DR provisions for repayment of project obligations; Debt outstanding balance as of the current date
4	LTV	The ratio of the loan amount to the market (or estimated) value of the collateral of the project
5	The ratio of the market and book value of the project	The market value of the project is determined at the current date based on the method of analogues, the book value is equal to the original cost less depreciation
6	Percentage of beneficiaries own participation	Share own participation of the beneficiaries in project financing
7	The payback period of the project	The period of time required for the income generated by the investment to cover the cost of the investment of the project
8	The term of the project	The term of project in years
9	The fact of balloon payment of the project	Payment at the end of the project implementation period
10	The industry of the project	

4 The Specificity of the Development of Models for Investment Projects

The specifics of developing models for investment projects differ from the development of models for housing projects in that the financial reporting indicators of project companies do not work for such transactions. Only weighted indicators of the project module are good for PD prediction, the most significant of which for investment projects are IRR, Weighted DSCR (Debt Service Coverage Ratio), the payback period of the project, the project term, and the share of own participation of the beneficiaries. Statistical approaches to the implementation of portfolio models are similar to those applied to residential real estate models. Additionally, if there is a small amount of statistical data of defaults, an expert ranking approach can be used.

Taking into account the specifics of investment projects, simulation (individual) PD models are often developed for SPV companies with a single project. The definition of default used in simulation models in practice is often taken to be different from the classic one and represents the implementation of at least one of the following events (for the most part—the event №2), due to the fact that investment projects in practice most often pay off the main part of the debt at the end of the term (the project has balloon payment):

1. Default of at least one of the project companies (borrowers) carrying out the project, that is, the presence of at least one company participating in the project, one of the following signs:
   - The project company was declared insolvent (bankrupt);
   - The project company is persistently insolvent, that is, it does not fulfill its obligations to creditors for more than 90 calendar days.
2. The fact of simultaneous implementation of the following two events:
   - Reducing the debt service coverage ratio (DSCR) below 1;
   - Reduction of the principal repayment and servicing ratio (LCR) below 1.

The above definition of default is used in many foreign and Russian credit organizations and is related to the experience of work of the credit organizations with the investment projects.

The simulation model generates a scenario distribution of the project's cash flow based on a number of risk factors. The complexity of the simulation model is determined by the method of selecting risk factors and the method of determining scenarios.

The selection of risk factors for the simulation model can be performed as follows:

- Risk factors are selected by the user;
- Risk factors are selected from a pre-defined set of factors;
- Risk factors are selected from a pre-defined set of factors for each type of project.

You can define scenarios for a simulation model as follows:

- Average values, spreads, and correlation coefficients are set by the model user;
- The average values are set by the model user, and the variance and correlation coefficients are estimated based on empirical data;
- Average values, spreads, and correlation coefficients are estimated using macroeconomic indicators (for example, the GDP index, consumer price index, and others).

Building a simulation model involves three main stages:

- Input of source data;
- Making statistical simulations of the macroeconomical and market scenarios, calculating the values of financial covenants of projects in these scenarios and fixing the facts of default of projects in case of violation of the financial covenants;
- Getting output data and determining the final score.

The source data of the simulation model can be external (exogenous) and internal (endogenous). Internal data—the parameters contained in the model itself that do not depend on a specific project: sector volatility, forward rates, and exchange rate volatility. External data—the project parameters entered in the model and set by the model user.

For the covenant simulations, such as DSCR and LLCR, the following parameters are used:

- Cash flow and scenarios for its development;
- Forward and interest rates;
- The parameters of the deal.

When simulating data on the cash flow of an investment project, the Monte Carlo method is used, which allows you to get a set of iterations (scenarios for the development of the situation) based on a random number generator and mathematical expectations and standard deviations of the cash flow of the investment project. Scenarios for forward and interest rates are also stimulated by the Monte Carlo method based on a random number generator and stochastic process parameters—mathematical expectations and standard deviations of interest/forward rates and the exchange rate. Transaction parameters include incoming data for each tranche for each element that affects the amount and timing of debt coverage in the event of default. Based on the data obtained during scenario simulation, you can calculate the number of implementations of default events and, accordingly, the probability of default (PD).

5 Conclusion

The materials discussed in this paper show the difference in approaches to the development of separate rating models for different risk segments in certain areas of lending.

It should be noted that each bank has a specific loan portfolio related to the volume of available statistics (including default statistics). That is why the approaches used to assess the probability of default for different risk segments in different banks may differ.

The most important ideas and results of this paper:

- For the first time in Russian banking practice, the approaches to assessing the probability of default for various risk segments of lending, depending on the available data, were systematized;
- An approach to assessing the probability of low-default portfolios based on both external ratings and internal statistics is proposed with calibration, using the probit-specification, taking into account Bayesian methods, in continuation of the idea of the article (Surzhko 2017);
- The article demonstrates the possibility of using decision trees (CART algorithm and tree ensembles) in relation to corporate borrowers (in other issues decision trees algorithms are used exclusively for the big data).

In addition to developing PD models, an important stage of working with models is the monitoring stage, i.e. their periodic validation to assess the possibility of using models on actual data using statistical tests. Periodic model validation should take place at least once a year and cover all the main stages of model development:

- The impact of data quality on the model's performance;
- Evaluating the discriminatory and predictive ability of models (including the quality of model calibration);
- Assessment of the discriminatory and predictive ability of models of individual risk factors of models.

In addition, it is necessary to conduct regular risk audits of the models used in banks, covering:

- Assessment of the independence and adequacy of the rating process;
- Quality of filling in information by business and underwriting departments of the bank;
- Adequacy of the results of periodic validation;
- Independent making of recommendations for updating or fully updating models, if necessary, based on the results of their own, alternative researches within the same risk segments.

Only the well-coordinated interaction of the development and validation teams when building models, as well as the independent opinion of the internal audit, allows the bank to develop an independent and best-quality concept when working with models and as part of the rating process. It is also necessary to fully engage in the process of working with models of business and underwriting departments in order to take into account in the models the risk factors that characterize the specifics of individual risk segments that are identified in the lending process of the customers in practice.

References

Allen, S. (2003). *Financial risk management: A practioner's guide to managing market and credit risk* (288 p.). Hoboken, NJ: Wiley.

Ayvazyan, S. A. (1989). *Applied statistics: Classification and reducing of dimension* (607 p.). Moscow: Finance and statistics.

Benninga, S. (2008). *Financial modelling* (3rd ed., 1168 p.) The MIT Press.

Breiman, L. (1984). *Classification and regression trees*. Monterey, CA: Wadsworth & Brooks/Cole Advanced Books & Software.

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.

Coleshaw, J. (1989). *Credit analysis* (240 p.). Woodhead-Faulkner.

Davis, H. A. (2003). *Project finance: Practical case studies* (2nd ed., 237 p.). London: Euromoney Books, Nestor House, Playhouse Yard.

Esty, B. (2003). *Modern project finance: A Casebook* (544 p.). Wiley.

Fight, A. (2006). *Introduction to project finance* (205 p.). Elsevier.

Finnerty, J. (2013). *Project financing: Asset-based financial engineering* (3rd ed., 560 p.). Wiley.

Jorion, P. (2007). *Financial risk manager instruction manual* (4th ed., 736 p.). New York: Wiley.

Joseph, C. (2013). *Advanced credit risk analysis and management* (448 p.). Wiley.

Karminsky, A. M. (2015). *Credit ratings and their modeling* (304 p.). National research University Higher school of Economics, Moscow, Publishing house of the Higher school of Economics.

Karminsky, A. M., & Kostrov, A. V. (2013). Modeling the probability of default of Russian banks: Extended opportunities. *Journal of the New Economic Association, 17*(1), 64–86.

Karminsky, A. M., Morgunov, A. V., & Bogdanov, P. M. (2015). Evaluation of the probability of default of project finance transactions. *Journal of the New Economic Association, 26*(2), 99–122.

Lobanov, A. A., & Chugunov, A. V. (2009). *Encyclopedia of financial risk management* (4th ed., 932 p.). Alpina Business books.

Lynch, P. (2010). *Financial modelling for project finance* (2nd ed., 212 p.). Euromoney Trading.

McCullagh, P., & Nelder, J. A. (1990). *Generalized linear models* (511 p.). New York: Chapman & Hall.

Mills, T. C. (2003). *Modelling trends and cycles in economic time series* (180 p.). New York: Palgrave McMillan.

Pomazanov, M. V. (2017). *Credit risk management in the Bank. Internal rating approach (IRR). Practical guide* (266 p.). YURAYT.

Pugachev, V. S. (2002). *Probability theory and mathematical statistics* (496 p.). FIZMATLIT.

Rogov, M. A. (2001). *Risk management* (120 p.). Moscow: Finance and statistics.

Siddiqi, N. (2006). *Credit risk scorecards developing and implementing intelligent credit scoring* (196 p.). Hoboken, NJ: Wiley.

Surzhko, D. A. (2017). Bayesian approach to PD calibration and stress-testing in low default portfolios. *Journal of Applied Finance & Banking, 7*(2), 1–6.

Zhevaga, A. A., & Morgunov, A. V. (2015). Use of summary macroeconomic indicators for calibration of internal rating models in banks. *Money and Credit, 8*, 39–46.

Loss Given Default Estimations in Emerging Capital Markets

Mikhail Pomazanov

Abstract This paper proposes an approach to decompose the RR/LGD model development process with two stages, specifically, for the RR/LGD rating model, and to calibrate the model using a linear form that minimizes residual risk. The residual risk in the recovery of defaulted debts is determined by the high uncertainty of the recovery level according to its average expected level. Such residual risk should be considered in the capital requirements for unexpected losses in the loan portfolio. This paper considers a simple residual risk model defined by one parameter. By developing an optimal RR/LGD model, it is proposed to use a residual risk metric. This metric gives the final formula for calibrating the LGD model, which is proposed for the linear model. Residual risk parameters are calculated for RR/LGD models for several open data sources for developed and developing markets. An implied method for updating the RR/LGD model is constructed with a correction for incomplete recovery through the recovery curve, which is built on the training sets. Based on the recovery curve, a recovery indicator is proposed which is useful for monitoring and collecting payments. The given recommendations are important for validating the parameters of RR/LGD model.

Keywords Credit risk · Residual risk · IFRS 9 standards · Unexpected losses · Loss given default · Recovery rate · Recovery curve · Capital requirements

JEL C58 · G17 · G28 · G32

M. Pomazanov (✉)
National Research University Higher School of Economics, Promsvyazbank, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_6

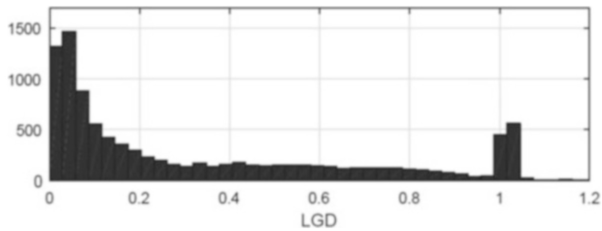


Fig. 1 Typical frequency distribution of the level of losses after LGD model

1 Introduction

LGD—Loss given default is one of the most important credit risk assessment parameters. Along with PD—Probability of default and EAD—Exposure at default, LGD contributes as a key parameter in calculating regulatory requirements, as well as economic capital requirements, as part of an approach based on internal IRB ratings (International Convergence 2006). The purpose of the LGD assessment is to accurately and efficiently quantify the level of recovery risk inherited as part of the default risk. The incentive to build LGD valuation models is the possibility of obtaining permission from the regulator to use the bank's approach based on internal ratings to calculate reserves and requirements for economic capital. The inverse of LGD is the RR (Recovery Rate), RR = 1 – LGD, so the RR simulation is identical to LGD. Recovery from default RR or its inverse value LGD = 1 – RR in practice demonstrates random dynamics and has a typical frequency profile, shown in Fig. 1. Many empirical studies have noted bimodality with a higher concentration of observations at zero and close to one and a higher LGD during periods of economic recession. This is evidenced by the results of a number of empirical works on mortgage lending (Araten et al. 2004; Karminsky et al. 2016) and corporate lending, including corporate bond market (Qi and Zhao 2011; Dermine and de Carvalho 2006; Schuermann 2004; Felsovalyi and Hurt 1998). Therefore, to calculate unexpected losses, it is necessary to take into account the volatility of LGD in addition to its expected estimate. The dispersion of LGD, reinforced by bimodality of distribution, contributes to unexpected losses, which are the basic component of residual credit risk.[1]

The typical model of LGD dispersion is not difficult to determine with the commonly used relation (Gordy and Lutkebohmert 2013):

[1]According to the definition given, for example, by the Bank of Russia (see Bank of Russia Ordinance No. 3624-U, dated April 15, 2015, "On Requirements for the Risk and Capital Management System of a Credit Organization and Banking Group"), residual risk is the risk remaining after the Bank's actions to reduce inherent risk. Suppose a bank takes measures (that is, requires collateral) to recover debt after default, based on which it statistically fairly expects a recovery share of RR = 1-LGD. And, let's say, on a statistically significant portfolio, this share of recovery will take place. However, due to the dispersion of LGD and the granularity of the default part of the portfolio, deviations from the expected value will be observed, including towards losses. This gives unexpected losses related to residual risk.

$$D(\mathrm{LGD}_i) = \gamma \cdot E(\mathrm{LGD}_i) \cdot (1 - E(\mathrm{LGD}_i)), \tag{1}$$

where $D(\cdot)$ is the variance (squared standard deviation), $E(\cdot)$ is the mathematical expectation, $i = 1\ldots N$ is the index of a model-homogeneous population for LGD,[2] γ is a RR/LGD dispersion parameter theoretically belonging to the interval of [0,1], its mean value $\gamma = 0.25$ is proposed, for example, in the CreditMetrics approach (CreditMetrics 1997). Assuming that, within the framework of the TAC, the LGD model corresponds to the average statistical observations of reconstructions, i.e. relatively medium, it does not overestimate or underestimate the calculations, we put $E(\mathrm{LGD}_i) = \mathrm{LGD}_i$. In practice, the parameter γ can be statistically refined at the stage of validation of the internal LGD model, for example, by the formula:

$$\gamma = \frac{\sum\limits_{d \in D} \left(\widehat{\mathrm{LGD}}_d - \mathrm{LGD}_d \right)^2}{\sum\limits_{d \in D} \mathrm{LGD}_d \cdot (1 - \mathrm{LGD}_d)}, \tag{2}$$

where $\mathrm{LGD_d}$ is the model estimate of the one default to the LGD before default, $\widehat{\mathrm{LGD}}_\mathrm{d}$ is the observed loss after the completion of the default debt recovery process.

The study (Antonova 2012) presents the result of the LGD assessment of Russian default issuers according to the information-analytical agency Cbonds. During the observation period from December 31, 2002 to December 31, 2011, 124 Russian corporate issuers made a real default on ruble corporate bonds that were traded on the MICEX. A real default is understood as failure to fulfill an obligation by the issuer before the expiration of the grace period. Based on the calculation method chosen by the author, RR: RR = 1-LGD were calculated for defaults of corporate bonds issued by Russian issuers in 59 cases, which formed a statistical sample. The overall outcome of the assessment was the average rate RR = 48.8% (LGD = 51.2%) with a standard deviation of σRR = σLGD = 29.2%. For the case of an LGD-insensitive assessment model, formula (2) takes a simple form:

$$\gamma = \frac{n-1}{n} \cdot \frac{\sigma \mathrm{RR}^2}{(1 - \mathrm{RR}) \cdot \mathrm{RR}} = 0.34. \tag{3}$$

The numerical estimate of γ is based on the result of the evaluation of LGD model as the average LGD, without constructing a refinement model. This estimate given by issuers can be considered a conservative estimation of uncertainty parameter γ of the LGD for the Russian bond market. It is useful to estimate the statistical error of

[2] A model-homogeneous population should be understood, for example, such industry segments of borrowers as "Banks", "Individuals, consumer loans", "Mass segment of small business", "Large corporate business" including credited to a particular bank, etc. It is reasonable to classify LGD segments of credit assets by business model or financial instrument. For each segment, various parameters γ are possible.

Table 1 Parameters γ for various industry segments of the default bonds of Russia

Industry	Average, in %	Standard deviation, in %	Number of observations	γ
Light Industry	19.4	10	4	0.05
Heavy Industry	63.3	25	11	0.24
Trade	48.5	29	15	0.31
Construction	57.2	27	6	0.25
Agriculture and food processing	50.6	30	18	0.34
Other services	24.4	28.2	5	0.34
Total	48.8	29.2	59	0.34

Table 2 Parameters γ for various industry segments of US default bonds

Industry	Average recovery, in %	Standard deviation, in %	Number of observations	γ
Real estate	41.97	16.05	71	0.10
Transportation	38.17	18.85	70	0.15
Electricity	48.03	22.67	39	0.20
Oil&Gas	44.37	23.68	21	0.22
Manufacturing	38.93	28.55	573	0.34
Service&Leisure	38.65	30.37	190	0.39
Retail	33.4	34.19	33	0.51
Media&Communications	34.7	34.56	163	0.52
Total	38.68	28.22	1160	0.34

the parameter γ, since, when developing the LGD model, statistics are often not enough. The estimate of $\sigma\gamma$ is as follows:

$$\frac{\sigma\gamma}{\gamma} \cong \frac{1}{\sqrt{n}}\left(\sqrt{2} + \sigma\mathrm{LGD}\frac{|2\mathrm{LGD} - 1|}{\mathrm{LGD}(1 - \mathrm{LGD})}\right). \tag{4}$$

Formula (4) gives the standard deviation of the statistical error γ, provided that the model LGD is equal to the average. The statistical error (estimation of the standard deviation of the error) for the above sample of 59 issuers was $\sigma\gamma = 0.06$.

The study of (Antonova 2012) indicators of average RR and standard deviations for several industry segments was also evaluated separately. The results of the evaluation of individual parameters γ are presented in Table 1.

The work of (Jankowitscha et al. 2014) presents the calculation of recovery levels for defaulted US bonds for the period July 2002 to October 2010, as well as standard deviations. A similar calculation of γ for non-financial sector companies is shown in Table 2 by industry and in general.

Figure 2 shows the ranges of γ taking into account standard deviations due to statistical error. It can be seen from Fig. 2 that, taking into account the statistical error

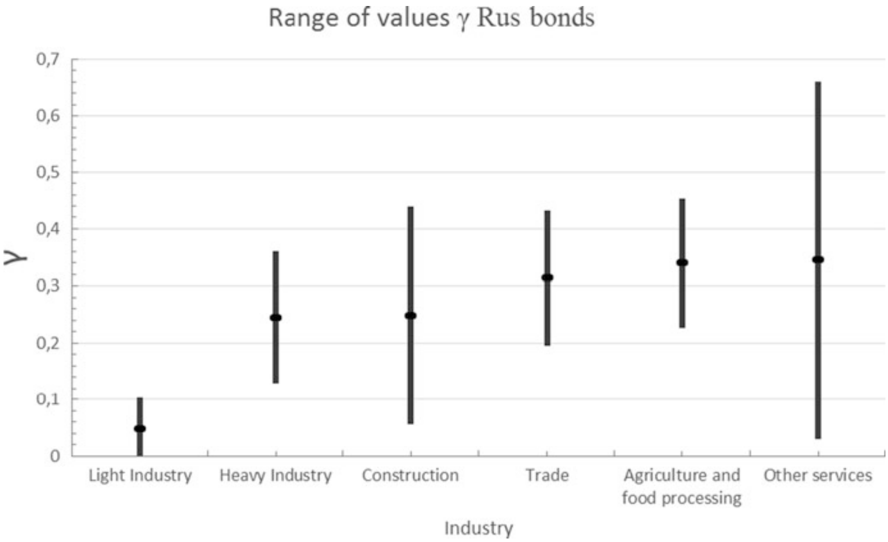


Fig. 2 Ranges γ for different industry segments of Russia, taking into account standard deviations due to statistical error

for different industry segments, the ranges of possible values of γ substantially intersect.

An exception is only for the light industry. But in this segment there are very few measurements and, perhaps, this is just an extreme result, which is usually discarded in statistical measurements (see Fig. 2). Comparing the results of recoveries of default bonds of the US and Russia obtained at the same observation periods, it is obvious that the average recovery level in the US was 10% lower than the Russian ones, however, the average volatility parameter γ practically coincided with the Russian one at the level $\gamma = 0.34$.

Figure 3 shows the ranges of γ according to the standard deviations due to statistical error.

However, a clear stratification of the values of γ by industry segments is revealed, in particular, the real estate differs in the minimum level of the volatility parameter, $\gamma = 0.1$, the sectors Retail and Media & Communications, $\gamma = 0.5$, have the maximum. The inclusion of statistical error, obviously, rejects the hypothesis of independence of γ, in particular, from the industry segment.

Therefore, it makes sense when building the LGD model to a model for the volatility parameter γ, too. With a lack of observations, it is possible to assume that $\gamma =$ const for all measurements within a model-homogeneous population, but this will fix the model error.

In the next part of the work, it is necessary to answer these questions: how to take into account the results of recoveries of default borrowers, if the provided the recovery process is incomplete? How to use statistically implemented recovery dynamics to build recovery indices for early defaults? What functionality should

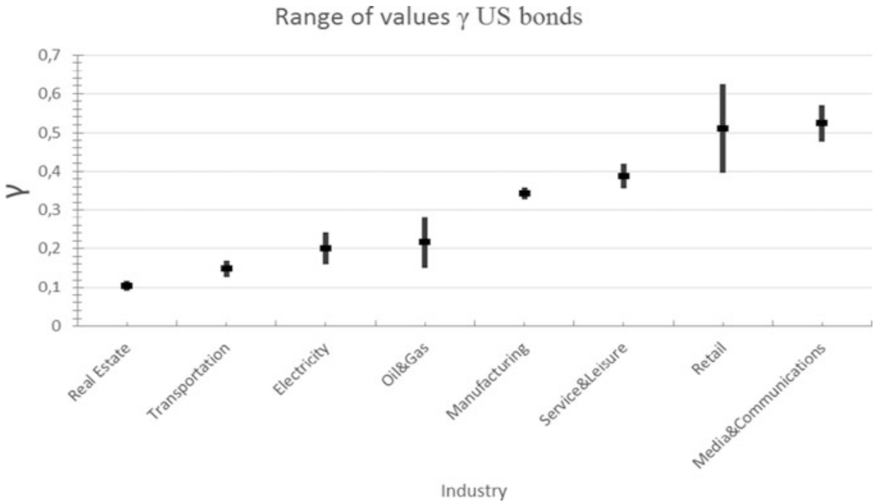


Fig. 3 Ranges γ for different US industry segments, taking into account standard deviations due to statistical error

Table 3 LGD assessment method

	Default count averaging	Exposure weighted averaging
Default weighted averaging	$\text{LGD} = \frac{\sum_{y=1}^{m}\sum_{i=1}^{n_y} LR_{i,y}}{\sum_{y=1}^{m} n_y}$ (5)	$\text{LGD} = \frac{\sum_{y=1}^{m}\sum_{i=1}^{n} EAD_{i,y} * LR_{i,y}}{\sum_{y=1}^{m}\sum_{i=1}^{n_y} EAD_{i,y}}$ (6)
Time weighted averaging	$\text{LGD} = \frac{\sum_{i=1}^{n_y}\left(\frac{\sum_{i=1}^{n_y} LR_{i,y}}{n_y}\right)}{m}$ (7)	$\text{LGD} = \frac{\sum_{i=1}^{n_y}\left(\frac{\sum_{i=1}^{n_y} EAD_{i,y} * LR_{i,y}}{\sum_{i=1}^{n_y} EAD_{i,y}}\right)}{m}$ (8)

be optimized to build an LGD model while minimizing residual risk? How does residual risk affect economic capital requirements? What is the model? A simple, but optimal from the point of view of residual risk, LGD model will be proposed, based on a positively discriminatory rating of LGD.

2 Recovery Curve

The start of identifying the types of RR (LGD) that can be considered as measures of LGD. In the extensive literature on LGD, for example (Vujnović et al. 2016), four are represented (Table 3).

Where i is the observation of default, y is the year of default, n_y is the number of defaults in each year, m is the years of observation, LR is the loss coefficient or LGD for each observation.

For practical purposes, it suffices to contrast on two approaches for calculating RR.

A. Simple recovery index (medium/median or frequency):

$$\mathrm{RR}_{\mathrm{avg}} = \frac{1}{n}\sum_{i=1}^{n}\frac{R_i}{E_i}, \tag{9}$$

where R_i is the amount of funds received to repay the debt of borrower i, discounted to the default date (both direct and indirect recovery are taken into account), E_i is the exposure to default (EAD) of borrower i. EAD—the amount of the main debt, accrued interest, fines, and other charges to the reporting period before default. After the moment of default, fines, interest, and other accruals after default are not included in the EAD exposure, the off-balance part is not included, but the amounts issued after default are included. The net credit exposure is the adjusted (reduced) credit exposure for the amount of the discounted financial collateral. The simple recovery index (RR) is not oriented to amounts; it shows the average share of recovery among defaulting borrowers.

B. Weighted Average Recovery Index

$$RRw = \frac{\sum R_i}{\sum E_i}. \tag{10}$$

The weighted index is sensitive to the defaulted amounts (to losses). Thus, the indicators RRavg and RRw will differ if the share of recovery depends on the amount in default. If large loans recover heavier than small ones, then a simple recovery index exceeds a weighted one and vice versa. The recovery amount is calculated based on recovery payments discounted to the default date.

$$R = \sum_{t=0}^{\infty}\frac{P_t - C_t}{(1+q)^t}, \tag{11}$$

where P_t—recovery payments at time t from the date of default, C_t—costs of bank recovery costs $\frac{1}{(1+q)^t}$ ——discount factor with the rate q, the sign "∞" means that theoretically wait for the a completed collection can indefinite (in practice, of course, the wait is limited and will be seen later). The repayment history for the sample of default loans (at least $\widehat{\tau}$) is presented in Table 4. The sample is taken for a sufficiently wide period of "observing" $\widehat{\tau}$ >3–5 years. Those. on the interval of $[t-\widehat{\tau}, t]$, where t is the current moment of observation of defaults (reporting date—90 days). The list of repayment history parameters:

1. ID (number) of the borrower;
2. Exposure in default (EAD, taking into account possible loans issued after default,

Table 4 Parameters of repayment history

ID	EAD	Discount rate, q in %	Default date	Recovery period after default (year)						
				1	2	3	...	S	...	P
1	E1	10	01.05.2008	R11	R12	R13	...	R1S	...	R1P
2	E2	9	01.08.2008	R21	R22	...	...	...	R2...	
...				...	...	...	...	R...S		
k	Ek	11	01.08.2011	RK1	...	...	...			
...				...	...	...				
...				...	...	...				
...				...	...					
...				...	R..2					
N	EN	6	01.01.2020	RN1						

discounted by default date);

3. Discount rate (q, in practice, the average rate for the lending period is often used in a model-uniform sample of all loans);
4. Date of default, (month of default);
5. Repayment payments discounted with the rate (q) on the maturity date, counted from the date of default (exposure period after default).

For the ease of calculation, repayments are sorted in descending order of exposure after default. The applied formulas for calculating the recovery curve are selected from two possible formats:

1. Simple format (medium/frequency)

$$\mathrm{RR_{Avg}}(\tau) = \frac{1}{\mathrm{n}(\tau)} \sum_{\mathrm{i}:\exists \mathrm{V_i}(\tau)} \frac{\sum\limits_{\mathrm{s}\le\tau} \mathrm{V_i(s)}}{\mathrm{E_i}}, \tag{12}$$

where $n(\tau)$ is the number of default loans that “survived” until the payment of $V_i(\tau)$ in the period τ, i.e. only those loans i are taken into account for which there may be a payment $V\ V_i(\tau),\ i:\ \exists\ V_i(\tau)$ (obviously, if $\tau = 0$, $n(0) =$ all default loans in the database). $V_i(s)$ is discounted payments in the period s from the moment of default (discount), E_i is amount in the default.

Moreover, the square of the standard deviation (the square of the error $RR_{\mathrm{Avg}}(\tau)$) is substantially heterogeneous due to the different dimension $n(\tau)$ for each period τ. $\delta RR^2(\tau)$ is calculated by the formula:

$$\delta {RR_{\mathrm{Avg}}}^2(\tau) = \frac{1}{n(\tau)^2} \sum_{i:\exists V_i(\tau)} \left(\frac{\sum\limits_{s\le\tau} V_i(s)}{E_i} - RR_{\mathrm{Avg}}(\tau) \right)^2. \tag{13}$$

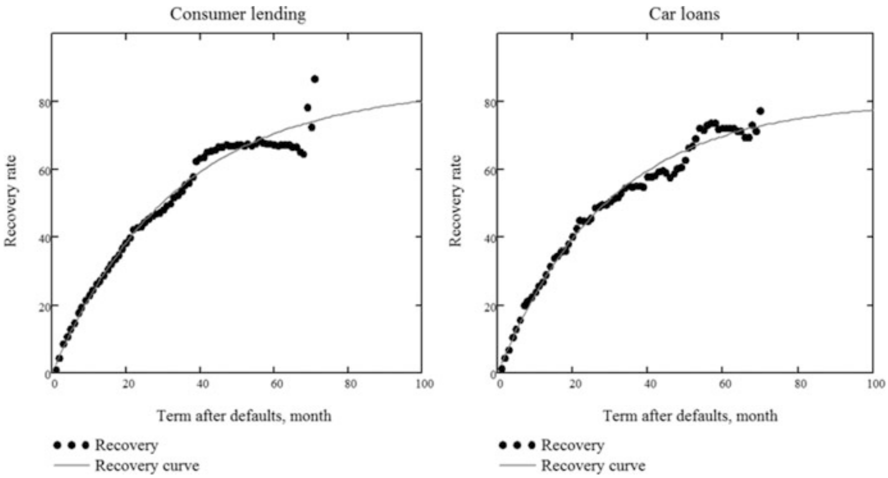


Fig. 4 Examples of constructing recovery curves

2. Weighted average format (taking into account default amounts):

$$\mathrm{RR}_w(\tau) = \frac{\sum\limits_{i:\exists V_i(\tau)} \sum_{s \le \tau} V_i(s)}{\sum\limits_{i:\exists V_i(\tau)} E_i}. \tag{14}$$

The square of the standard deviation can be estimated by the formula:

$$\delta RR_w{}^2(\tau) = \frac{HHI_\tau}{n(\tau)} \sum_{i:\exists V_i(\tau)} \left(\frac{\sum\limits_{s \le \tau} V_i(s)}{E_i} - RR_{Avg}(\tau) \right)^2, \tag{15}$$

where is the Herfindahl–Hirschman index is calculated as:

$$HHI_\tau = \frac{\sum\limits_{i:\exists V_i(\tau)} E_i{}^2}{\left(\sum\limits_{i:\exists V_i(\tau)} E_i \right)^2}. \tag{16}$$

An example of recovery curves is shown in Fig. 4.

The practice implication shows that the curve $RR(\tau)$ can be approximated with high accuracy by a function of the form:

Table 5 Statistical parameters of recovery curves

Product	Recovery period	Total size	R_∞	T, months	R-sq.	Error ΔR
Consumer lending	2011–2016	1309	83.8%	32.8	97.6%	17.6%
Car loans	2011–2016	228	80.0%	29.4	98.5%	12.3%

$$\rho_\tau(R_\infty, T) = R_\infty \cdot \left(1 - e^{-\frac{\tau}{T}}\right), \tag{17}$$

where T is the average recovery time.

The maturity curve limit $RR(\infty)$ is the recovery forecast for a non-default company, and $LGD(0) = 100\ \% \ - RR(\infty)$, term T is the average recovery period. In the work of (Benjelloun 2019) proposed a method for modeling LGD/RR through a random process, averaging of which gives dynamics close to the behavior of Fig. 4. To approximate $RR(\tau)$ of curve (17), the weighted least squares method is used (see, for example, Strutz 2016), in which the residual is calculated in the Euclidean metric with weights $\frac{1}{\delta RR^2(\tau)}$ and is minimized by the parameters $\boldsymbol{R}_\infty$ (limit recovery) and $\boldsymbol{T}$ ((average recovery period):

$$L(\mathrm{RR}_\infty, T) = \sum_\tau \frac{1}{\delta RR^2(\tau)} \cdot (\mathrm{RR}(\tau) - \rho_\tau(R_\infty, T))^2 \to \min{}_{R_\infty, T}. \tag{18}$$

In this case, the error δR_∞ of the estimate R_∞ is estimated using linearized regression (18) at the optimal point $\boldsymbol{R}_\infty, \boldsymbol{T}$. The detailed formula for estimating δR_∞ is given in Appendix 1.

The output is a calculation of the "slow" values of $\boldsymbol{R}_\infty^\Omega$ and $\boldsymbol{T}^\Omega$ in the current long-term "viewing window" for interval $[t - \Omega, t]$. For example, for the data in Fig. 4 values of recovery parameters were calculated (see Table 5).

Numerous empirical calculations show a high level of fit of the recovery curve using the parametric formula (17), for example, for retail products and consumer lending R-sq. = 97–99%.

3 Recovery Indicators

For a company that has an exposure in default with a period of τ and a certain negative account balance, the loss forecast will be estimated using the conditional LGD (τ):

$$\mathrm{LGD}(\tau) = \frac{1 - R_\infty}{1 - RR(\tau)}, \tag{19}$$

or, using the parametric formula (17):

$$\mathrm{LGD}(\tau) = \frac{1 - R_\infty}{1 - R_\infty \cdot (1 - e^{-\frac{\tau}{T}})}. \tag{20}$$

Therefore, based on the current estimations, at the time $\tau > 0$, the recovery value RR^i_τ, we can construct an unbiased estimate of recovery "for infinity" as:

$$RR^i_\infty(\tau) = RR^i_\tau + \left(1 - RR^i_\tau\right)(1 - LGD(\tau)), \text{i.e.}$$

$$RR^i_\infty(\tau) = \begin{cases} RR^i_\tau + \begin{cases} \left(1 - RR^i_\tau\right) \dfrac{R_\infty \cdot e^{-\frac{\tau}{T}}}{1 - R_\infty \cdot (1 - e^{-\frac{\tau}{T}})}, & \text{recovery process is not completed} \\ 0, & \text{recovery process ended.} \end{cases} & ,\tau > 0 \\ 0, & \tau \le 0 \end{cases}. \tag{21}$$

Obviously, for large waiting times τ after default, the correction to RR^i_τ, estimated by the second term in (6), tends to zero and $RR^i_\infty(\infty) = R^i_\infty$, which goes to the statistical base model LGD/RR.

Evaluation (21) should be used as a model estimate of the expected recovery of the debt of borrower in the case when the period after default has not passed, sufficient so that the issue of debt recovery is considered closed. Then it makes sense to determine the recovery indicator for the entire model-homogeneous segment of the population. Recovery indicator determines the forecast of recovery on loans that defaulted on a given "short" indicative moving horizon $[t - \omega, t]$. A simple (or a medium) recovery indicator is constructed as:

$$1. RR^\omega_{Avg}(t) = \frac{1}{N^\omega(t)} \cdot \sum_{i=1}^{N^\omega(t)} RR^i_\infty(t - t_i), \tag{22}$$

2. And, a weighted average indicator, taking into account the amounts of E_i at the time default, t_i, is constructed as:

$$RR^\omega_w(t) = \frac{\sum_{i=1}^{N^\omega(t)} RR^i_\infty(t - t_i) \cdot E_i}{\sum_{i=1}^{N^\omega(t)} E_i}, \tag{23}$$

where $N^\omega(t)$ is the number of borrowers defaulted on a given "short" indicative interval $[t - \omega, t]$.

The recovery indicator is of a great practical importance for monitoring the process of collecting defaulted debts, the strategy for securing loans, segmenting credit policy, etc. If the average recovery indicator exceeds the weighted average, then this means small loans (below average) are more easily repaid than large ones and vice versa.

4 Residual Risk at Loss Given Default Models

The question of residual risk LGD is associated with at least two risk drivers of unexpected losses, which can be underestimated when calculating the requirements for the own economic capital of the loan portfolio. The first driver is macroeconomic, this is a possible correlation of the default rate (i.e. PD) of the loan portfolio and the average LGD, associated with crisis phenomena in the economy, as well as the correlation of the average LGD with other macroeconomic factors. The second driver is local, it is associated with the LGD uncertainty (volatility), for which a "typical" model (1) with parameter γ has been selected. Historical data on the correspondence between the level of default and the level of recovery after default on the corporate bond market in America and Europe (Moody's data) gives the following dependence for the historical period 1982–2016 (Fig. 5).

According to historical data, the credit risk assessment methodology recommends applying a stress correction to the unperturbed value of losses after default LGD in the form $\mathrm{LGD}_{\mathrm{stress}} = \mathrm{LGD}_0 + (1 - \mathrm{LGD}_0)(1 - e^{-17.6 \cdot \mathrm{EDR}})$, where EDR is the

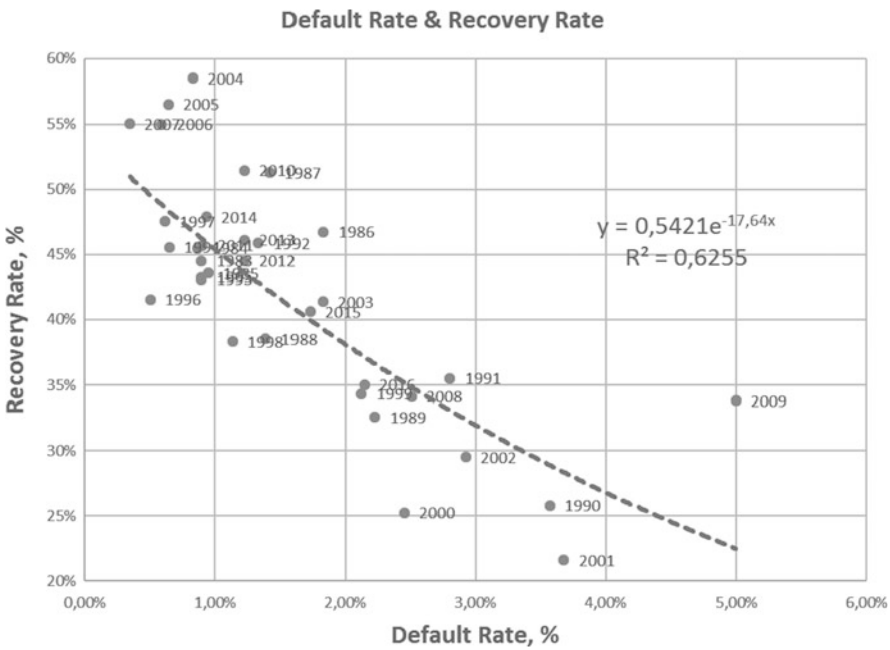


Fig. 5 Historical relationship between the default rate and the recovery rate for the period 1982–2016 according to US corporate bonds and EU (data Moody's 2017)

expected default rate (central tendency), and LGD_0 is the unperturbed LGD value in the stable period. For Moodys data, the $LGD_0 \cong 50\%$.

The correlation problem between PD and LGD (or RR) is one of the key issues in assessing credit risk. For example, a study of (Allen and Saunders 2005) demonstrates calculations according to which the interaction of PD and LGD increases expected losses and capital requirements by up to 30%. However, portfolio credit risk assessment models are often based on the assumption that LGD is fixed and independent of PD. The authors Miu and Ozdemir (2006). note that if PD and LGD correlations are ignored in the model, the LGD should be increased on average by 6% (from 35% to 41%) to compensate for the correlation effect of PD and LGD. At the same time, the results of study of (Ermolova and Penikas 2017) do not allow us to state that there is a relationship between these components of credit risk for the Russian corporate bond market. A generalization of risk metrics that takes into account the dependence of LGD on PD within the framework of the proposed approach can be represented as the dependence of LGD on a random, normally distributed variable, implying that the parameter γ is a constant. In this case, it is recommended to use one of the LGD models (PD (Y)) presented in (Frye and Jacobs 2012) but it should be borne in mind that the basic requirements for the economic capital of an infinitely granular portfolio within the framework of the adjusted one-factor model will differ from the calculation formula recommended by the Basel Committee. Within the framework of approach (1) simulating the dispersion of LGD, the simplest, continuous version of modeling the distribution of losses after default is possible—these are losses Loss = L × EAD with probability pL and losses (Loss = 0) with probability (1 – pL). The parameters L and pL can be determined from the following conditions:

$$\begin{cases} E(\mathrm{Loss}) = \mathrm{LGD} \cdot \mathrm{EAD} \\ D(\mathrm{Loss}) = \gamma \cdot \mathrm{LGD} \cdot (1 - \mathrm{LGD}) \cdot \mathrm{EAD}^2 \end{cases}; \tag{24}$$

These conditions give a unique solution for L and pL:

$$\begin{cases} \mathrm{L} = \gamma + (1 - \gamma) \cdot \mathrm{LGD} \\ \mathrm{pL} = \dfrac{\mathrm{LGD}}{\gamma + (1 - \gamma) \cdot \mathrm{LGD}} \end{cases}. \tag{25}$$

Then, the metrics in which the adjusted PD and EAD can be determined will be set in the form:

$$E_\gamma = \mathrm{EAD} \times (\gamma + (1 - \gamma) \times \mathrm{LGD}), \tag{26}$$

$$\mathrm{PD}_\gamma = \mathrm{PD} \cdot \frac{\mathrm{LGD}}{\gamma + (1 - \gamma) \times \mathrm{LGD}}.$$

The boundary values A: $\gamma = 0$ (the lack of LGD uncertainty) and B: $\gamma = 1$ (maximum LGD uncertainty) will mean, for case A:, $PD_0 = PD$, $E_0 = \mathrm{EAD} \cdot \mathrm{LGD}$; for case B: $PD_1 = PD \cdot LGD$, $E_1 = EAD$.

Obviously, case B implies a greater exposure to default and the capital requirement should be higher for it, despite the fact that the probability of losses will decrease. This issue was investigated in (Witzany 2009). The authors used the one-factor approach to calculating capital recommended by the Basel Committee, taking into account the LGD parameter, first introduced in (Vasicek 1987). Based on the extreme scenarios presented above, it was possible to evaluate VAR (Value at Risk) LGD as the difference between the capital requirement in case B and A. The difference turned out to be positive and monotonous with respect to the model parameters, including expected level of LGD.

In the current approach, we will act similarly in the paradigm of the recommended Basel-2 approach to assessing the requirements for economic capital, created on the basis of the Vasicek formula, under these conditions:

$$\mathrm{UL}_{\gamma} = E_{\gamma} \cdot \left(N\left(\frac{N^{-1}\left(PD_{\gamma}\right) + \sqrt{R} \cdot N^{-1}(0.999)}{\sqrt{1-R}} \right) - PD_{\gamma} \right), \tag{27}$$

where UL is for the estimate of unexpected losses at the recommended reliability level of 0.999 (can be changed), $N(\cdot)$ and $N^{-1}(\cdot)$ are the standard normal and inverse distributions, respectively, R is the correlation parameter, E_{γ}, PD_{γ} from equation (7). The UL_0 is the standard recommended form for evaluating the capital of the Basel-2 Advanced Approach. Define $ULGD_{\gamma}$ as a contribution to equity in relation to EAD:

$$\mathrm{ULGD}_{\gamma} = \frac{\mathrm{UL}_{\gamma} - \mathrm{UL}_0}{\mathrm{EAD}}, \tag{28}$$

which will be responsible for the influence of the dispersion parameter γ of LGD on capital requirements (i.e., unexpected losses).

$$\begin{aligned} \mathrm{ULGD}_{\gamma} = (\gamma + (1-\gamma) \cdot \mathrm{LGD}) \cdot \mathrm{N}\left(\frac{\mathrm{N}^{-1}\left(\mathrm{PD}_{\gamma}\right) + \sqrt{\mathrm{R}} \cdot \mathrm{N}^{-1}(0.999)}{\sqrt{1-\mathrm{R}}} \right) \\ - \mathrm{LGD} \cdot \mathrm{N}\left(\frac{\mathrm{N}^{-1}(\mathrm{PD}) + \sqrt{\mathrm{R}} \cdot \mathrm{N}^{-1}(0.999)}{\sqrt{1-\mathrm{R}}} \right). \end{aligned} \tag{29}$$

Obviously, $\mathrm{ULGD}_0 = 0$. Figure 6 shows graphs of ULGD_{γ} behavior over the entire range of values $\gamma \in [0, 1]$.

Figure 6 shows that, the values of the correlation R, reliability 0.999 and PD, the capital requirements monotonously increase with increasing uncertainty coefficient γ. Figure 7 shows the surfaces $\frac{d}{d\gamma}\mathrm{ULGD}_{\gamma}$ at the extreme points $\gamma = 0$ (upper surface) and $\gamma = 1$ (lower surface). In the entire "working" range PD, LGD $\in [0, 1]$, the surfaces are located above the zero plane.

The study shows that the parameter γ is monotonic with respect to unexpected losses and its growth leads to an increase in the additional capital requirement due to

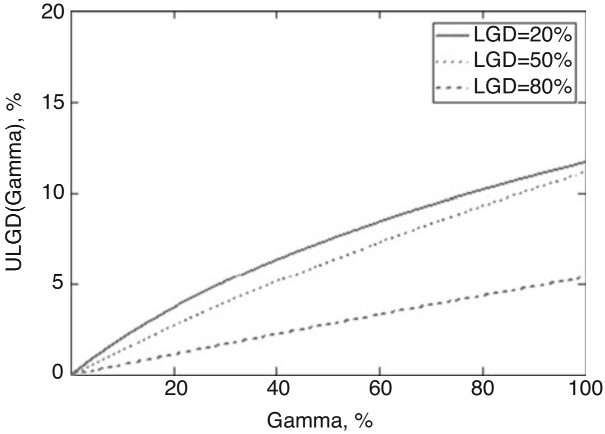


Fig. 6 A graph of the dependence of the additional ULGD requirement for capital on γ (in %) for PD = 10%, correlation $R = 0.2$, and significance level at 99.9%

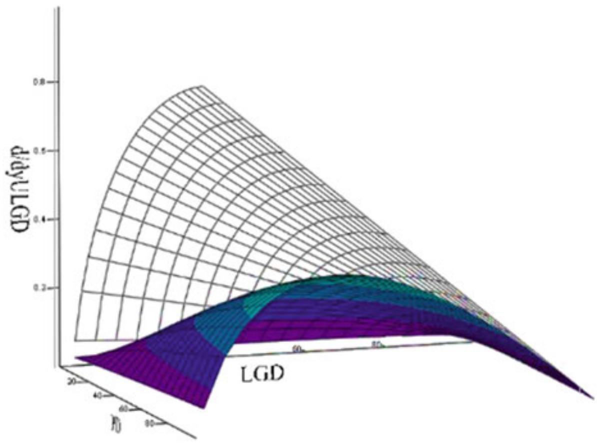


Fig. 7 The surfaces of the derivatives $\frac{\mathrm{d}}{\mathrm{d}\gamma}\mathrm{ULGD}_\gamma$ for the correlation value $K = 0.2$ and the reliability 0.999. Lower for $\gamma = 1$, upper for $\gamma = 0$ over the area space of PD, LGD $\in [0, 1]$

the dispersion of LGD. Therefore, when developing the LGD model, it is reasonable to minimize the uncertainty parameter γ.

The largest contribution to capital will be at $\gamma = 1$ and the probability of default PD = 1:

$$\mathrm{ULGD}_{\max}(\mathrm{LGD}) = N\left(\frac{N^{-1}(\mathrm{LGD}) + \sqrt{R}\cdot N^{-1}(0.999)}{\sqrt{1-R}}\right) - \mathrm{LGD}. \tag{30}$$

Figure 8 shows a graph of $\mathrm{ULGD}_{\max}(\mathrm{LGD})$ and the correlation $R = 0.2$.

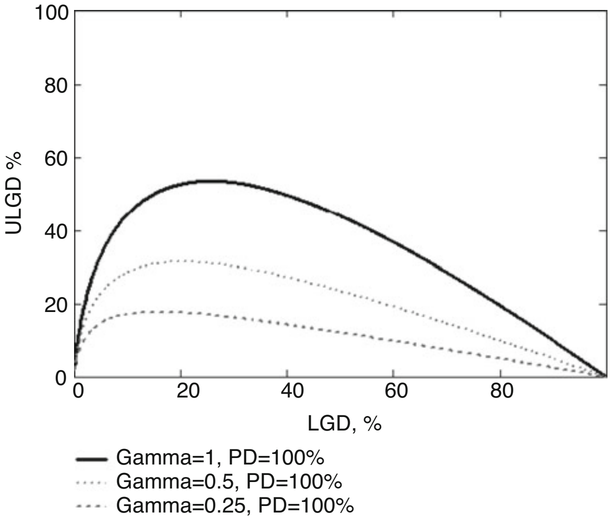


Fig. 8 The graph of the contribution to capital due to the dispersion of LGD for the values PD = 100%, maximum $\gamma = 1$ (black), $\gamma = 0.5$ (light gray), $\gamma = 0.25$ (dark gray)

The maximum function $\mathrm{ULGD}_{\max}(LGD)$ achieved when:

$$\mathrm{LGD}^* = N\left(\frac{\sqrt{(1-R)\left(N^{-1}(0.999)^2 - \ln(1-R)\right)} - N^{-1}(0.999)}{\sqrt{R}}\right). \qquad (31)$$

For correlation parameters $R = 0.2$ and significance level (0.999) $\mathrm{LGD}^* = 25.5\%$. Obviously, the shift of the shift down of unexpected losses is towards LGD $<50\%$. This indicates increased responsibility for the model in the event of a model error in the direction of lowering LGD (increasing RR).

5 Optimal Loss Given Default Model from the Point of Residual Risk

Let introduce θ as the dimension LGD[3] (or RR) rating of an indifferent internal structure. The linear model $\widehat{R}_\theta$ of the recovery level RR relative to the rating θ can be estimated as:

[3]LGD rating means any specially developed function that depends on the risk-dominant parameters of LGD/RR, which correlates with the implemented LGD/RR.

$$\widehat{R}_\theta = \widehat{R} + \mu \cdot \frac{\theta - \widehat{\theta}}{\delta\theta} \cdot \delta R, \tag{32}$$

where $\widehat{R}$[4] is the mean value of n realized recoveries of level R, in other words, $\widehat{R} = \frac{1}{n}\sum_\theta R$, δR is the standard deviation of R, measured by a biased estimation as $\delta R^2 = \frac{1}{n}\sum_\theta \left(R - \widehat{R}\right)^2$.

Equally, $\widehat{\theta}$ is defined as the average value of θ over the entire set of reconstruction implementations on which the model is built, $\widehat{\theta} = \frac{1}{n}\sum_\theta \theta$, $\delta\theta^2 = \frac{1}{n}\sum_\theta \left(\theta - \widehat{\theta}\right)^2$. The most important parameter sought for model (32) is μ—multiplicator, which should depend on the risk-determinism of the LGD rating and minimize the LGD dispersion coefficient indicated by the γ RR/LGD dispersion parameter. The observed recovery of R will be determined by the random variable ε and the model $\widehat{R}_\theta$ in the form $R = \widehat{R}_\theta + \varepsilon$, where the variance ε is modeled, according to (32), by the relation as:

$$D\varepsilon = \gamma \cdot \widehat{R}_\theta \left(1 - \widehat{R}_\theta\right). \tag{33}$$

In this case, the mathematical expectation $M\varepsilon = 0$ by the definition of the model. Further, at the input of the model, it is necessary to determine the correlation ρ between the implemented restorations R and the LGD rating indicated by θ, the estimate of which will be given by the equation:

$$\rho = \frac{1}{N}\sum_\theta \frac{\left(R - \widehat{R}\right)\left(\theta - \widehat{\theta}\right)}{\delta R \cdot \delta\theta}. \tag{34}$$

The more complex, non-linear LGD model in practice makes little sense. It will not provide a significant increase in the estimation accuracy due to the high volatility of LGD due to the two-mode distribution of Fig. 1. The proposed linear LGD model does not automatically guarantee natural restrictions on the simulated recovery level $\widehat{R}_\theta \in [0, 1]$ such as, the popular logistic representation of the type $\widehat{R}_\theta = \frac{1}{1+e^{A\theta+B}}$, but practice shows (see Sect. 6) that the LGD model cannot be created so powerful that the results of its forecast differ by multiples.

For example, if we turn to the recommendations on LGD of the Basel Committee [Basel II 2006], then the recommendations of the minimum LGD vary in the range of 35–45%. Below these values, LGD can be formally evaluated only if there is financial security, which, in fact, should adjust the exposure to default EAD, and not LGD. If this is not done, then LGD uncertainty model is formally destroyed, since financial security is a 100% realizable recovery.

[4]The mean is in the sense of RR_{avg} according to the app. A.2.

Below we will show the range of parameters $\widehat{R}, \rho$ for which the linear model does not go beyond the limits of natural restrictions. Passing to estimates of the observed quantities, it can be equated as[5]:

$$\begin{aligned} n \cdot MSE &= \sum_{\theta}\left(R-\widehat{R}_{\theta}\right)^{2}=\sum_{\theta}\varepsilon^{2}=\sum_{\theta}\mathrm{D}\varepsilon=\gamma\cdot\sum_{\theta}\widehat{\mathrm{R}}_{\theta}\left(1-\widehat{\mathrm{R}}_{\theta}\right) \\ &= \gamma\cdot\left(\sum_{\theta}\widehat{R}\left(1-\widehat{R}\right)-\sum_{\theta}\left(\frac{\theta-\widehat{\theta}}{\delta\theta}\right)^{2}\delta R^{2}\cdot\mu^{2}\right) \\ &= \gamma\cdot n\cdot\left(\widehat{R}\left(1-\widehat{R}\right)-\delta R^{2}\cdot\mu^{2}\right). \end{aligned} \tag{35}$$

Otherwise, it can be written as:

$$\begin{aligned} \mathrm{n}\cdot\mathrm{MSE} &= \sum_{\theta}\left(R-\widehat{R}_{\theta}\right)^{2}=\sum_{\theta}\left(R-\widehat{R}-\mu\cdot\frac{\theta-\widehat{\theta}}{\delta\theta}\cdot\delta R\right)^{2} \\ &= \sum_{\theta}\left(R-\widehat{R}\right)^{2}-2\mu\cdot\delta R^{2}\sum_{\theta}\frac{\left(R-\widehat{R}\right)\left(\theta-\widehat{\theta}\right)}{\delta R\cdot\delta\theta}+\sum_{\theta}\left(\frac{\theta-\widehat{\theta}}{\delta\theta}\right)^{2}\delta R^{2}\cdot\mu^{2} \\ &= n\cdot\delta R^{2}\cdot\left(1-2\mu\cdot\rho+\mu^{2}\right). \end{aligned} \tag{36}$$

Equating the expressions obtained above, the dependence γ (μ) is described as:

$$\gamma(\mu)=\gamma_{0}\cdot\frac{1-2\mu\cdot\rho+\mu^{2}}{1-\gamma_{0}\cdot\mu^{2}}, \tag{37}$$

where $\frac{\delta R^{2}}{\widehat{R}\left(1-\widehat{R}\right)}=\gamma_{0}$ is denoted is the value of the parameter γ for the case that is not sensitive to the LGD estimation model considered in Sect. 2.

To find the solution for the optimal value of μ, the problem can be solved with:

$$\mu^{*}=\operatorname{argMin}_{\mu}\gamma(\mu), \tag{38}$$

where the optimal point for solution is $\gamma^{*}=\gamma(\mu^{*})$.

Problem (38) is solved by the standard method of finding the minimum of a function using the first derivative optimum condition $\gamma'(\mu^{*})=0$. Without bothering the reader with standard mathematical calculations, one can write out the solution to (38):

[5]MSE—Mean Square Error.

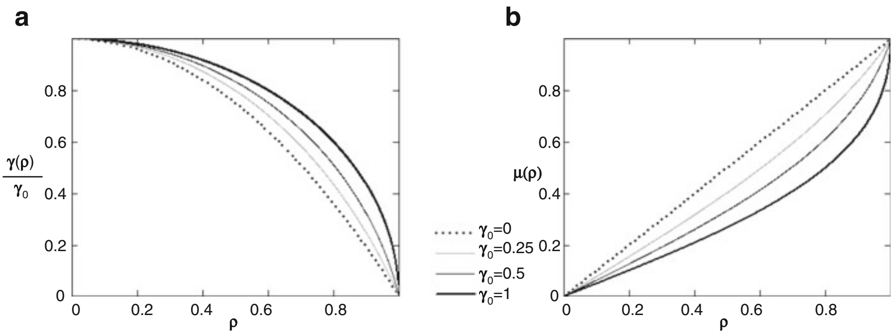


Fig. 9 The dependences of $\frac{\gamma(\rho)}{\gamma_0}$ LGD dispersion parameter (**a**) and μ (ρ) – multiplicator of model (**b**) from correlation ρ upon solution (33)

$$\mu^* = \rho \cdot \frac{2}{1 + \gamma_0 + \sqrt{(1+\gamma_0)^2 - 4\gamma_0\rho^2}},$$

$$\gamma^* = \gamma_0 \cdot \left[1 - \frac{2\rho^2}{1 + \gamma_0 + \sqrt{(1+\gamma_0)^2 - 4\gamma_0\rho^2}}\right], \tag{39}$$

$$\mathrm{MSE}^* = \delta R^2 \cdot \left[1 - 4\rho^2 \cdot \frac{\gamma_0 + \sqrt{(1+\gamma_0)^2 - 4\gamma_0\rho^2}}{\left(1 + \gamma_0 + \sqrt{(1+\gamma_0)^2 - 4\gamma_0\rho^2}\right)^2}\right].$$

For $\rho = 0$ (in the case when the LGD rating does not work properly), an obvious solution is obtained $\mu^* = 0$, $\gamma^* = \gamma_0$, $MSE^* = \delta R^2$.

Figure 9 shows the graphs of solutions (39) in the full range of non-negative correlation of the LGD rating with real measurements for different levels of LGD dispersion.

It can be seen from Fig. 9 that the effect of minimizing the dispersion of LGD becomes most significant as the risk-determinism of the LGD rating increases. However, for the optimal parameter μ of the LGD model, the effect appears immediately and μ becomes less than ρ as soon as the LGD volatility appears. The boundary parameters for the proposed linear model (32) are calculated from the condition: $0 \le \widehat{R}_\theta \le 1$. Assume, without loss of generality, that the rating θ is normally distributed over the interval [0; 1],[6] when $\widehat{\theta} = \frac{1}{2}$, $\delta\theta = \frac{1}{\sqrt{12}}$.

[6]A normal distribution of the random parameter ξ can be described using the substitution for $F(\xi)$, where F is the distribution function of ξ.

According to the model: $\delta R = \sqrt{\gamma_0 \cdot \widehat{R}\left(1 - \widehat{R}\right)}$, then the boundary values of recovery will be

$$\widehat{R}_\theta^{\pm} = \widehat{R} \pm \mu \cdot \sqrt{3 \cdot \gamma_0 \cdot \widehat{R}\left(1 - \widehat{R}\right)}. \tag{40}$$

It means that: $\mu_{\max} = \frac{\min\left(\widehat{R}, 1-\widehat{R}\right)}{\sqrt{3 \cdot \gamma_0 \cdot \widehat{R}\left(1-\widehat{R}\right)}}$.

Avoiding the analysis of the full variety of the three-dimensional parameter region $\widehat{R}, \gamma_0, \rho$, in which the restriction $0 \le \widehat{R}_\theta \le 1$ is satisfied, we will calculate μ_{max} for typical LGD parameters according to the recovery of US corporate bonds (see Sect. 2). For them, $\gamma_0 = 0.34$, $\widehat{R} = 38.7\%$ $\mu_{\max} = 0.79$, which corresponds to very high risk-determinism indices of the LGD model with a correlation $\rho > 0.8$, which is not achieved by any models.

In the practically significant range of possible models of LGD ratings and not "extreme" practical levels of average recovery $\widehat{R}$ (that is, not close to 0 and 1), the linear LGD model (32) will not give out a range of predictive recoveries $\widehat{R}_\theta$ beyond the limits of [0,1]. In practice, when constructing the LGD model, it is recommended to convert the LGD rating to a range of uniformly distributed values, evaluate μ^* (39) and check constraint (39).

In the next section, we will consider several public models for the LGD rating and their authors' assessments show the applicability of the approach described.

6 Practical Drivers of Loss Given Default Models

The level of recovery of the borrower after default is very specific and depends on many factors. In the literature (see, for example, (Grunert and Weber 2009) four categories of factors for corporate borrowers are defined (see Fig. 10), which correspond to:

- for the borrower, the company of the borrower, incl. creditworthiness (rating) above all;
- for macroeconomics, incl. default rate;
- for the condition of the loan, incl. collateral in the first place;
- for business relations of the borrower, incl. their intensity.

Factors are divided into quantitative and qualitative groups, involving expert assessment. A set of factors forms a long-list from which factors are selected that correlate with the level of implemented LGD results.

To build models for various asset classes, data sources, and measurement methods, which are classified in Table 6.

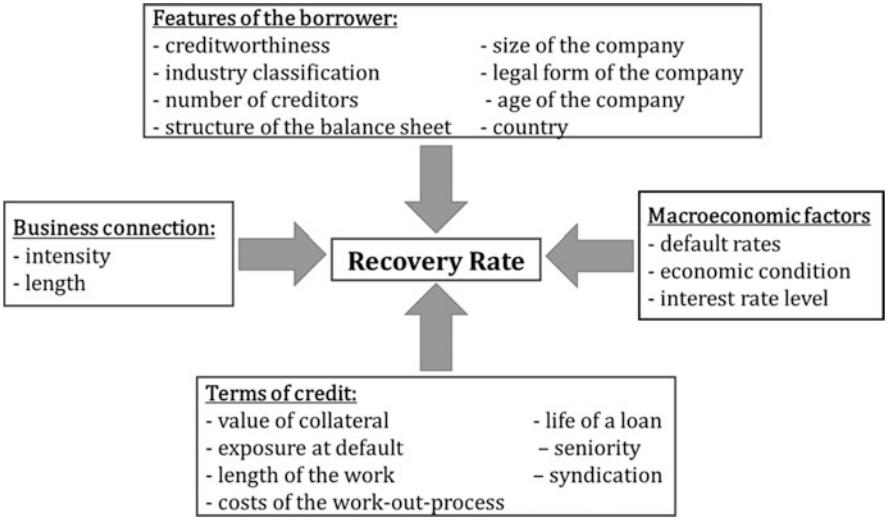


Fig. 10 Drivers for RR/LGD

Table 6 Classification of evaluation methods LGD

Source	Measure	Methods	Exposure
Market values	Price differences	Market LGD	Large corporate, sovereigns, banks
	Credit spreads	Implied market LGD	Large corporate, market LGD sovereigns, banks
Recovery and cost experience	Discounted cash flows	Workout LGD	Retail, SMEs, large corporate
	Historical losses and estimated PD	Implied historical LGD	Retail

Various linear and non-linear algorithms are used to train the LGD classification model. In the literature, (Loterman et al. 2012; Qi and Yang 2009; Bonini and Caivano 2014), a range of methods are analyzed:

– Ordinary Least Squares (OLS);
– Ridge Regression (RiR);
– Robust Regression (RoR);
– Ordinary Least Squares with Beta transformation (B-OLS);
– Beta Regression (BR);
– Ordinary Least Squares with Box-Cox transformation (BC-OLS);
– Regression trees (RT);
– Multivariate Adaptive Regression Splines (MARS);
– Least Squares Support Vector Machines (LSSVM);
– Artificial Neural Networks (ANN);
– Linear regression + non-linear regression (OLS+);
– Logistic regression + (non)linear regression (LOG+).

Table 7 Source data

Dataset	Type	Total size
BANK1	Personal loans	47,853
BANK2	Mortgage loans	119,211
BANK3	Mortgage loans	3351
BANK4	Revolving credit	7889
BANK5	Mortgage loans	4097
BANK6	Corporate loans	4276

Source: Loterman et al. (2012)

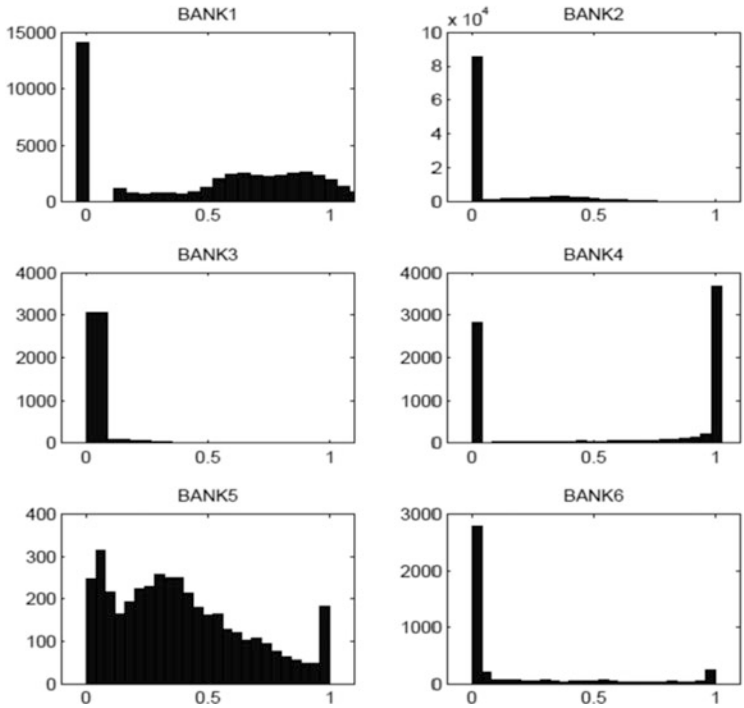


Fig. 11 Density of LGD distribution by Loterman G

Nevertheless, even on impressive empirical data (Table 7), with tens of thousands of measurements for corporate and consumer portfolios of banks, it was found that the obtained models have limited predictive characteristics regardless of which method is used, although non-linear methods give higher characteristics than traditional linear methods. The banks analyzed by the author have unique LGD distributions, which are shown in Fig. 11.

Table 8 shows the result of measuring the linear Pearson correlation predicted and implemented by LGD for different banks. Table 8 shows that significant differences in the results obtained by different methods are observed only for Bank N 3, and for the data of this Bank, even the best models show a weak result. In general, one can

Table 8 The result of measuring the linear Pearsons' correlation predicted and implemented for different LGD methods for different banks

Pearson's R (Cohen et al. 2002) measures the degree of linear relationship between predictions and observations.

Technique	BANK1	BANK2	BANK3	BANK4	BANK5	BANK6
OLS	0.311	0.485	0.117	0.664	0.474	0.350
B-OLS	0.295	0.477	0.077	0.651	0.507	0.305
BR	0.260	0.464	0.157	0.653	0.456	0.321
BC-OLS	0.240	0.472	0.137	0.573	0.501	0.286
RiR	0.306	0.492	0.146	0.666	0.478	0.354
RoR	0.306	0.477	0.173	0.653	0.454	0.349
RT	0.300	0.582	0.387	0.692	0.506	0.339
MARS	0.321	0.558	0.502	0.692	0.567	0.362
LSSVM	0.347	0.569	0.453	0.702	0.579	0.396
ANN	0.360	0.603	0.378	0.705	0.596	0.362
LOG+OLS	0.326	0.484	0.076	0.668	0.498	0.348
LOG+B-OLS	0.317	0.529	0.121	0.665	0.512	0.323
LOG+BR	0.280	0.453	0.074	0.668	0.457	0.335
LOG+BC-OLS	0.213	0.463	0.167	0.666	0.510	0.310
LOG+RiR	0.329	0.539	0.132	0.676	0.492	0.341
LOG+RoR	0.326	0.535	0.151	0.673	0.474	0.339
LOG+RT	0.330	0.555	0.455	0.666	0.500	0.335
LOG+MARS	0.332	0.553	0.488	0.675	0.569	0.329
LOG+LSSVM	0.340	0.559	0.415	0.677	0.580	0.365
LOG+ANN	0.350	0.559	0.538	0.670	0.585	0.369
OLS+RT	0.338	0.579	0.258	0.678	0.536	0.362
OLS+MARS	0.339	0.562	0.502	0.692	0.577	0.363
OLS+LSSVM	0.371	0.567	0.465	0.700	0.576	0.349
OLS+ANN	0.372	0.601	0.261	0.705	0.557	0.350
<r>	**0.32**	**0.53**	**0.28**	**0.67**	**0.52**	**0.34**
dr	**0.04**	**0.05**	**0.17**	**0.03**	**0.05**	**0.02**

Source: Cohen et al. (2002)

notice that the linear OLS model gives an average level result, for corporate bank N6 even above the average.

The study (Seidler et al. 2017) presented the LGD model, trained in the Czech consumer lending market. The aim of the study was to show that lag macrovariables involved in the delayed model are still strong risk factors. As a result, the authors agreed on a meaningful set of factors presented in Table 9.

The following informative LGD model is presented in (Košak and Poljšak 2010). The model has been trained in the rapidly developing small and medium business borrowing market (SME) of Eastern Europe. Table 10 shows the risk-dominant variables that were identified by the authors as defining the LGD model.

Table 12 also presents calculations of model parameters (32) for Košak and Poljšak (2010). The authors used a limited number (124 observations), which

Table 9 Variables included in the LGD model

Explanatory variable logit LDG	Macroeconomic variables, current values	Macroeconomic variables, lagged and lead values
Client- specific factors	Real GDP growth (y-o-y)	Real GDP growth (y-o-y) (t-1)
Exposure at default	Real GDP growth (y-o-y)	Real GDP growth (y-o-y) (t-2)
Relationship with bank	Real Consumption Growth (y-o-y)	Real investment growth (y-o-y) (t-2)
Age	Real Investment Growth (y-o-y)	Unemployment rate (t-8)
Children	Real Pribor3m	Real wage growth (y-o-y) (t-3)
Phone	Inflation rate (y-o-y)	Real wage growth (y-o-y) (t-4)
Employment	Property prices (y-o-y)	Real wage growth (y-o-y) (t-5)
Education	Default rate	
Female	Retail loan growth (y-o-y)	

Source: Seidler et al. (2017)

Table 10 Variables included in the LGD model

Collateral type	Industry	Period	Rating of the borrower before default	EAD
Assignment of receivables Financial collateral Personal guarantee Physical collateral Real Estate collateral Unsecured	Manufacturing Real Service Trade	Long-term loan Short-term loan	Last rating C Last rating D Last rating E.	Large Medium Small

Source: Košak and Poljšak (2010)

gives rise to a tangible statistical error in determining the parameters characterizing the uncertainty. For the parameter γ_0 and γ^* according to formula, the statistical error is at the level of 10%. A third example of the LGD model is proposed to consider a model prepared by linear regression based on 10 years of historical development of real data on corporate and retail loans from a group of European commercial banks under the control of the ECB [Bonini and Caivano 2016]. 26,000 cases were processed, including 7500 large and medium corporate defaults. The result is a recovery level model presented in Table 11.

Table 12 shows the calculations of the parameter γ_0 of the "LGD dispersion" without taking into account the LGD model, the optimal γ^* from the point of view of residual risk after applying model (8), the optimal sensitivity parameter μ^*, and also the range $\widehat{R}_\theta^{\pm}$ of possible values for the model RR as it applied in (8). The correlation ρ between the implemented LGD and the model was estimated by the formula $\rho = \sqrt{R_{\text{squared}}}$. The calculations were carried out for three sources in which the parameters of the models are indicated.

Table 12 shows that the model recovery level (8) does not go beyond the range (0.1). Judging by the relation ${}^{\gamma^*}\!/_{\gamma_0}$ and Fig. 6, the models presented in Table 12 can

Table 11 Model RR (recovery rate)

Variables	Grouping	Coefficient	*p*-value	Variable weight
Macro-geographical area	Intercept	0.1001	<.0001	
	Center	0.2145	<.0001	13.87%
	North East	0.1113		
	Sud & Island	0.0788		
	North West	0		
Exposure at Default	EAD	0.1567	<.0001	10.13%
Portfolio segmentation	Medium – Large Corporate	0.594	0.0033	38.40%
	Small Business (Retail)	0.377	0.0022	
	Individuals (Retail)	0	<.0001	
Type of product	Mortgages	0.1876	<.0001	12.13%
	Other products	0		
Presence of personal guarantee	Absence	0.1134	<.0001	7.33%
	Presence	0		
Presence of mortgages	Absence	0.1609	<.0001	10.40%
	Presence	0		
Type of recovery process	Out of court	0.1189	<.0001	7.69%
	In court	0.0533		
	No information	0		

Source: Bonini and Caivano (2016).

provide a 10–25% reduction in the residual risk of LGD relative to how if LGD were assessed in the zero-approximation by the average LGD.

Summing up the results of a sample study of the results of RR /LGD modeling performed by different authors on different statistical recovery databases, we can draw the following conclusions:

1. It is impossible to unequivocally give preference to a particular method that is optimal in terms of modeling accuracy. In many cases, for example, see Table 7, an increase in the complexity and accuracy of the methods does not lead to a noticeable improvement in the results of the RR/LGD model and, on the other hand, often to a deterioration;
2. The set of risk-dominant parameters of the RR/LGD model can vary significantly when analyzing the statistical bases of different banks and different economies or different model-homogeneous populations;
3. The average recovery parameters and their dispersion can fluctuate significantly with a narrowing of model-homogeneous populations, including lending segments including in different banks. The maximum accuracy achieved on certain optimal models is also significantly heterogeneous.

The general results of the maximum achieved accuracy of LGD modeling, measured in various metrics, such as the correlation of the realized and model LGD, show a rather modest result. Very rarely a correlation greater than 0.6 is achieved, the average achieved on the best models is about 0.45.

Table 12 Calculations of the parameter γ_0 for LGD dispersion without taking into account the LGD model, optimal γ^* in terms of residual risk after solution of problem (33), which are presented (39), optimal value multiplicator μ^*, and also the boundary values of recovery $\widehat{R}_\theta^{\pm}$, which are described in (40)

Source	Seidler (2017)	Košak and Poljšak (2010)	Bonini and Caivano (2016)
LGD model	GLM[a]	GLM	OLS
Type of asset	Retail, 2003q1-2010q2, 18 698 obs.	SME, 2002 – 2005, 124 obs.	Individuals (Retail), Small size Corporate (Retail), Medium—Large size Corporate, 2002q4-2012q4, 26 000 obs.
Mean value of realized recoveries $\widehat{R}$	0.42	0.73	0.51
Standard deviation of recoveries δR	0.40	0.35	0.46
Pseudo R-squared	0.152 (Adjusted)	0.363 (Nagelkerke)	0.31 (Adjusted)
Starting value dispersion parameter γ_0	0.657	0.622	0.847
Optimal value dispersion parameter γ^*	0.594	0.468	0.692
Optimal value multiplicator of model (9) μ^*	0.245	0.421	0.329
Boundary values of recovery $\widehat{R}_\theta^{\pm}$	0.25–0.59	0.48–0.98	0.25–0.77

[a]Generalized linear model/GLM

All this convincingly argues the practical expediency of using simple methods, such as (9), for which the optimal sensitivity setting is possible to minimize residual risk. The construction of the model is based on the maximum Pearson correlation. The results of other models can be compared with the results of model (9) to identify their effectiveness.

7 Conclusion

In this study, it is proposed an approach to divide the RR/LGD model development process into two stages, namely: the RR/LGD rating model and calibrate the latter using a linear form that minimizes residual risk. The RR/LGD rating model is constructed in such a way as to ensure the maximum Pearson correlation with the implemented RR/LGD on the training statistical sample. In preparing the RR statistical base, correction (4) for the incomplete recovery process for part of the sample is taken into account. To do this, the recovery curve parameters (4) should be estimated using the method (5) on the historical recovery base (see Table 4). At the same time, recovery payments, net of costs, must be cleared of non-payments and discounted at the time of default. Financial support should be included in the EAD model. The RR/LGD rating model is based on risk-dominant factors, examples of which are presented in Sect. 6. In the process of setting the optimum, from the point of view of correlation, RR/LGD rating model, it should be normalized so that the distribution of ratings is statistically (with an acceptable error) uniform.

At the next step, the optimal sensitivity parameter μ is calculated by formula (12) with allowance for the parameter γ_0 of the LGD dispersion and the correlation parameter ρ. When calculating these parameters, the correction for the incomplete recovery process should be taken into account. Including for the recovery sample $ID = 1...N$ according to Table 4:

$$\gamma_0 = \frac{\sum_{ID} (R_{ID} - R(\tau_{ID}))^2}{\sum_{ID} R(\tau_{ID}) \cdot (1 - R(\tau_{ID}))},$$

$$\rho = \frac{\sum_{ID} (R_{ID} - R(\tau_{ID})) \cdot \left(\theta_{ID} - \widehat{\theta}\right)}{\sqrt{\sum_{ID} (R_{ID} - R(\tau_{ID}))^2 \cdot \sum_{ID} \left(\theta_{ID} - \widehat{\theta}\right)^2}}, \qquad (41)$$

where R_{ID} is the share of the implemented borrower recovery ID, $R(\tau_{ID})$ is the recovery function (4) if recovery is not completed, or $R(\tau_{ID}) = R_\infty$ if it is completed by the time τ_{ID} after default, $\theta_{ID}, \widehat{\theta}$ is rating borrower's RR/LGD ID and average rating respectively.

The verification of the model is determined by formula (9). The validity of the model within the limits of the model RR restriction should be verified by formula (13). The value of the final adjustment and calibration of the LGD model can be estimated as a percentage of the EAD of economic capital savings on residual risk through the difference $\mathrm{ULGD}_{\gamma_0} - \mathrm{ULGD}_{\gamma^*}$ according to formula (8). For example, a capital saving of 1% EAD is tangible and comparable to the countercyclical capital premium (buffer) introduced by Basel—III (maximum 2.5% from Basel III, 2011). In addition, it is necessary to take into account the forecast/adjustment of the expected average RR (parameter $\widehat{R} = R_\infty$ in formula (9), taking into account the macroeconomic scenario and forecast. A reliable LGD driver, according to Moody's (see Fig. 5), is the central trend of PD.

To check and validate the already built "M" of RR/LGD model, it is necessary to compare it with the reference model (9), built on the data of the "M" model being tested. To do this, calculate the correlation ρ of the implemented LGD- construction with LGD_M, taking into account the possible incompleteness of recovery (all values for LGD_M are recommended to be consistent to a normal distribution). The second step will be the direct calculation of γ_M by the formula (2) for "M." Obviously, the average value of the realized LGD model should follows the rule, i.e. $\widehat{R} \cong R_\infty \pm \delta R_\infty$ (4), where δR_∞ is the error of the R_∞ estimate in problem (5), estimated (A2) in Appendix 1. One of the concepts of the recovery calculation format, or simple frequency, should be adhered to be weighted by means. It is generally accepted to adhere to the "simple" format, and balance on EAD should be taken into account in the LGD model, which depends on EAD. After calculating the optimal γ^* reference model (9) using formula (11), the obtained parameters of the LGD dispersion should be compared. If $\gamma_M > \gamma^* + \sigma\gamma$, where σγ is the statistical error (3), then the "M" model is not optimal and can be improved.

The next step is to check whether the LGD_M values goes beyond the lower limit of constraints (13). The values of LGD_M significantly (outside the statistical error) lower than the lower limit of the constraints $1 - R^+$ (13) are not permissible, since the conservative principle should be violated. In this case, the power of the "M" model is not enough to assign significantly lower values to the LGD model level. This can lead to a significant model risk, transformed into credit risk with the significant volumes for individual loans.

The Estimation Procedure of the Calculated Standard Error for the Average Marginal Share of Repayment

The solution of problem (5) gives the optimal values of the recovery period T and the limiting recovery R_∞. The error of the values depends on the quality statistics of the approximation of the cumulative recovery of the recovery curve (4). The linear problem of the parameter estimation question $\theta = \{R, T\}$ for the non-linear regression problem $(\tau) = \rho_\tau(\theta) + \delta_\tau \cdot \varepsilon_\tau$, near the optimal solution θ of problem (5) is given a linear regression relation for the error $\Delta\theta = \boldsymbol{\theta} - \theta$ in the standardized form:

$$\frac{RR(\tau) - \rho_\tau(\theta)}{\delta_\tau} = \frac{\partial_\theta \rho_\tau}{\delta_\tau}\ \Delta\theta + \varepsilon_\tau, \tag{42}$$

where $\partial_\theta \rho_\tau$ is composed by the $n \times 2$ partial derivatives matrix $\left[\frac{\partial}{\partial R}\rho_\tau(R, T), \frac{\partial}{\partial T}\rho_\tau(R, T)\right]$, ε_τ assumed to be normal uncorrelated random variable with unknown variance for each recovery period τ , of which there are n. Apparently, for an optimal solution in the sense of equation (5) for θ, the solution of problem (A1) for $\Delta\theta$ will be obvious $\Delta\theta = 0$. However, the error $\Delta\theta$ will be expressed through the covariance matrix according to the well-known formula (see, for example, Strutz 2016):

$$cov(\Delta\theta) = \left(\left[\frac{\partial_\theta \rho_\tau}{\delta_\tau} \right]^T \times \left[\frac{\partial_\theta \rho_\tau}{\delta_\tau} \right] \right)^{-1} \cdot \frac{RSS}{n-2}, \tag{43}$$

where for (A1):

$$RSS = \sum\nolimits_\tau \frac{1}{\delta_\tau^{\ 2}} (RR(\tau) - \rho_\tau(\theta))^2.$$

Denoting the partial derivatives as:

$$\begin{aligned} \rho_\tau &= R \cdot \left(1 - e^{-\frac{\tau}{T}}\right); \\ \partial_R \rho_\tau &= 1 - e^{-\frac{\tau}{T}}; \\ \partial_T \rho_\tau &= -Re^{-\frac{\tau}{T}} \frac{\tau}{T^2}, \end{aligned} \tag{44}$$

and according for the estimation error R, the only the upper diagonal element of the matrix $cov(\Delta\theta)$, it is needed to obtain

$$\delta R^2 = \frac{1}{n-2} \cdot \frac{-\sum_\tau \frac{\partial_R \rho_\tau \cdot \partial_T \rho_\tau}{\delta_\tau^{\ 2}} \cdot \sum_\tau \frac{(RR(\tau) - \rho_\tau)^2}{\delta_\tau^{\ 2}}}{\sum_\tau \frac{\partial_R \rho_\tau^{\ 2}}{\delta_\tau^{\ 2}} \cdot \sum_\tau \frac{\partial_T \rho_\tau^{\ 2}}{\delta_\tau^{\ 2}} - \left(\sum_\tau \frac{\partial_R \rho_\tau \cdot \partial_T \rho_\tau}{\delta_\tau^{\ 2}} \right)^2}. \tag{45}$$

To estimate the error R_∞ as the measure for the standard deviation δR_∞, it is necessary in formula (45) to substitute the solution of problem (5) as R—the limiting recovery R_∞, the time for recovery T, and $\delta_\tau^{\ 2} = \delta RR_{\mathrm{Avg}}^{\ 2}(\tau)$ or $\delta RR_w^{\ 2}(\tau)$, these replacements depend on the calculation of the recovery curve.

References

Allen, L. Saunders, A. (2005). *A Survey of Cyclical Effects in Credit Risk Measurement Models*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=315561

Antonova, B. N. (2012). Ocenka stavki vosstanovleniya po rossijskim korporativnym obligaciyam. *Journal of Corporate Finance Research, 6*(4), 130–143. [in Russian].

Araten, M., Jacobs, M., & Varshney, P. (2004). Measuring LGD on commercial loans: An 18-year internal study. *RMA Journal, 86*(8), 96–103.

Basel II. (2006). *International convergence of capital measurement and capital standards: a revised framework - comprehensive version*. Part 2. pp. 295. https://www.bis.org/publ/bcbs128b.pdf

Basel III. (2011). *A global regulatory framework for more resilient banks and banking systems*. Part 1, pp. 139. https://www.bis.org/publ/bcbs189.pdf

Benjelloun, M. (2019). *Stochastic modelling of the loss given default (LGD) for non-defaulted assets*. This work was supported by the Global Research & Analytics Dept. of Chappuis Halder

& Co. https://www.chappuishalder.com/wp-ontent/uploads/2019/03/GRA_White_Paper_LGD_stochastic.pdf

Bonini, S., & Caivano, G. (2014). Development of a LGD model basel2 compliant: A case study. In *Mathematical and Statistical Methods for Actuarial Sciences and Finance, 2014* (pp. 45–48). Springer. https://doi.org/10.1007/978-3-319-05014-0_10. ISBN: 3319050141.

Bonini, S., & Caivano, G. (2016). *Econometric approach for Basel II Loss Given Default Estimation: From discount rate to final multivariate model*. White paper. https://www.semanticscholar.org/paper/Econometric-approach-for-Basel-II-Loss-Given-%3A-from/fc954a9ac25c2fd9f2de67b9f89b53aa36742e6c

Cohen, M., Nelson, R., & Walsh, P. (2002). Links and impacts: The influence of public research on industrial R&D. *Management Science, 48*(1), 1–23.

CreditMetrics. (1997). *Technical Document*. JP Morgan.

Dermine, J., & de Carvalho, C. N. (2006). Bank loan losses-given-default: A case study. *Journal of Banking and Finance, 30*(4), 1219–1243.

Ermolova, M. D., & Penikas, H. I. (2017). PD-LGD correlation study: Evidence from the Russian corporate bond market. *Model Assisted Statistics and Applications, 12*(4), 335–358.

Felsovalyi, A., & Hurt, L. (1998). Measuring loss on Latin American defaulted bank loans: A 27-year study of 27 countries. *Journal of Lending and Credit Risk Management, 80*, 41–46.

Frye, J., & Jacobs, M., Jr. (2012). Credit loss and systematic loss given default. *The Journal of Credit Risk, 8*(1), 109–140.

Gordy, M. B., & Lutkebohmert, E. (2013). Granularity adjustment for regulatory capital assessment. *International Journal ofCentral Banking, 9*(3), 38–77.

Grunert, J., & Weber, M. (2009). Recovery rates of commercial lending: Empirical evidence for German companies. *Journal of Banking & Finance, Elsevier, 33*(3), 505–513. https://doi.org/10.1016/j.jbankfin.2008.09.002.

Jankowitscha, R., Naglerb, F., & Subrahmanyam, M. G. (2014). The determinants of recovery rates in the US corporate bond market. *Journal of Financial Economics., 114*(1), 155–177. https://doi.org/10.1016/j.jfineco.2014.06.001.

Karminsky, A. M., Lozinskaia, A. M., & Ozhegov, E. M. (2016). Estimation methods of creditor's loss in residential mortgage lending. *HSE Economic Journal, 20*(1), 9–51. [in Russian].

Košak, M., & Poljšak, J (2010). *Loss given default determinants in a commercial bank lending: an emerging market case study*. Zbornik radova Ekonomskog Fakulteta u Rijeci: časopis za ekonomsku teoriju i praksu. - Rijeka, ISSN 0353-3689, ZDB-ID 12830379. - 28.2010, 1, p. 61–88.

Loterman, G., Brown, I., Martens, D., Mues, C., & Baesens, B. (2012). Benchmarking regression algorithms for loss given default modeling. *International Journal of Forecasting, 28*, 161–170.

Miu, P., & Ozdemir, B. (2006). Basel requirement of downturn LGD: Modeling and estimating PD & LGD correlations. *Journal of Credit Risk, 2*(2), 43–68.

Moody's Corporation. (2017). Annual Default Study: Corporate Default and Recovery Rates, 1920–2016 - Excel data.

Qi, M., & Yang, X. (2009). Loss given default of high loan-to-value residential Mortgages. *Journal of Banking and Finance, 33*(5), 788–799.

Qi, M., & Zhao, X. (2011). Comparison of modeling methods for loss given default. *Journal of Banking and Finance, 35*(11), 2842–2855.

Schuermann, T. (2004). *What do we know about loss given default?* (February 2004). Wharton Financial Institutions Center Working Paper No. 04-01. Available at SSRN: https://doi.org/10.2139/ssrn.525702

Seidler, J., Konečný, T., Belyaeva, A., Belyaev, K. (2017). *The time dimension of the links between loss given default and the macroeconomy*. Working Paper Series 2037, European Central Bank. https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp2037.en.pdf

Strutz, T. (2016). *Data Fitting and Uncertainty (A practical introduction to weighted least squares and beyond)*. Springer Vieweg. ISBN 978-3-658-11455-8., paper 3.

Vasicek, O. (1987). *Probability of Loss on a Loan Portfolio*. [on-line], San Francisco, KMV, c1987, [cit. 27th March, 2009]. http://www.moodyskmv.com/research/files/wp/Probability_of_Loss

Vujnović, M., Nikolić, N., & Vujnović, A. (2016). Validation of loss given default for corporate. *Journal of Applied Engineering Science, 14*(4), 465–476. https://doi.org/10.5937/jaes14-11752.

Witzany, J. (2009). Unexpected recovery risk and LGD discount rate determination. *European Financial and Accounting Journal*, ISSN 1805-4846, University of Economics, Faculty of Finance and Accounting Prague, *4*(1), 61–84. https://doi.org/10.18267/j.efaj.63.

Comparing Bankruptcy Prediction Models in Emerging Markets

Roman Burekhin

Abstract This paper presents an overview of the main models for predicting bankruptcies of companies and considers the classification of existing approaches. Examples of using algorithms such as logistic models, classification trees, random forests, and artificial neural networks are highlighted. Particular attention is paid to comparing traditional and advanced (based on ML) algorithms. The main development trends of this class of models are considered in Russia, China, and in developed markets of the USA and Europe. This paper forms the basis for the practical use of such models in solving risk management problems.

Keywords Bankruptcy · Machine learning models · Deep learning models · Parametric models of prediction of bankruptcy · Imbalance data

JEL G01 · G11 · G17 · G32 · G33

1 Introduction

The ability of investors or potential lenders to correctly assess the credit risks of companies is a problem that has historically attracted the attention of financial experts. To achieve this goal, different methods of assessing credit risks are used, the purpose of which is to effectively predict the onset of an unfavorable situation at the enterprise. Typically, these methods represent traditional models (logistic models, multiple discriminant analysis models), characterized by a relatively simple mathematical apparatus and simple qualitative interpretation. Nevertheless, these methods are quite static and do not consider subtle economic or behavioral factors; the predictive ability of these models decreases with the non-linear nature of the relationships between the indicators.

R. Burekhin (✉)
National Research University Higher School of Economics, Moscow, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_7

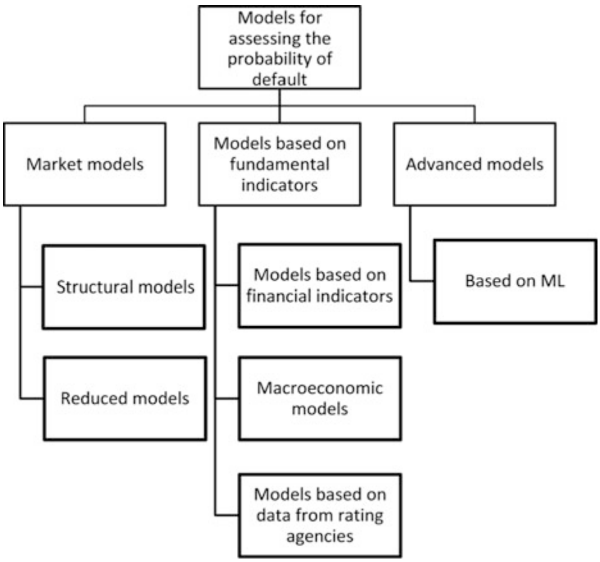
To conduct an effective credit policy, new methods must be flexible and adaptable to the changing realities of market economies. Therefore, there is currently an interest in new, advanced models built on the basis of artificial intelligence: classification forests, random forests, gradient boosting, artificial neural networks, etc. Today even a minimal improvement in accuracy is a significant achievement, leading to the increased financial stability of the company. This paper provides an overview of the main approaches to the prediction of bankruptcies and discusses the advantages and disadvantages of these methods (Sect. 2). Section 3 provides the examples of the use of these techniques in the Russian market. In Sect. 4, the main research trends in the prediction of bankruptcy are considered.

2 An Overview of Default Probability Models

Currently, many models have been developed and tested to assess the credit risk of borrowers. The classification of existing models is extremely important for the selection, implementation, and adaptation of the most appropriate model for assessing credit risk. The choice of approach depends on the nature and quality of the data, the mathematical apparatus available, the planning horizon, the research objectives pursued, and the availability of IT infrastructure in the organization.

Totmyanina (2011) provides an overview of the fundamental models for assessing the probability of default. The author considers the advantages and disadvantages, prerequisites, and the classification of bankruptcy forecasting models (Fig. 1). She distinguishes the following types of models for assessing default probabilities: market models (structural models, reduced form models); models

Fig. 1 Classification of default probability models



based on fundamental indicators (models based on financial indicators, macroeconomic models, models based on data from rating agencies); advanced models (based on ML algorithms).

2.1 Market Models

Market models make up a large block of default forecasting models. They are based on market information, primarily on the value and various characteristics of the issuer's securities.

The founder of the structural approach is Merton (1974). This approach assumes that equity is a European call option on the assets of the company, and the default of the company occurs when the value of the assets, which are subject to the simplest diffusion process, falls to a level determined by the constant amount of debt. However, this approach has several limitations, the main one being the assumption that all the company's assets are traded on the market and their market value is uniquely determined. In reality, this does not happen. However, the Merton model, based on various assumptions, led to the emergence of a number of models which attempt to ease restrictions used. Black and Cox (1976) expanded Merton's model, allowing default to maturity. Taking into account that zero-coupon bonds are a special case of models used for coupon bonds, Longstaff and Schwartz (1995) include in the model the possibility of the default of the company before the maturity date. The empirical evidence for the use of structural models is mixed. Eom et al. (2004) conducted an empirical study to compare the effectiveness of various structural models. Their analysis is based on cross-sectional data of US corporate bonds. They conclude that none of the models considered confirmed the observed data.

Jarrow (2009) argues that a structural model is preferable to internal (corporate) risk management. A reduced model is preferable when assessing credit risks. Huang et al. (2009) uses a structural model to predict defaults of Taiwanese construction companies. The authors note the difficulty of collecting the necessary information and the dominant role of market factors in building insolvency forecasting.

Structural models are based on the premise that economic agents are well informed about the value of assets and liabilities of a company. In reality, this is not always the case. Structural models rely on information about changes in the value of the company and its modeling, while reduced models miss the problem of determining the value of a company and directly model the probability of default and the scale of default as a random process. Unlike structural models, reduced models consider default as an unexpected event and associate it primarily with prices, bond yields, and not with the value of the firm's assets. It is assumed that this approach uses only available market information. One of the limitations of this approach is the assumption that the default processes within the same rating class are the same when it is empirically determined that bond credit spreads can vary significantly within the same rating group. Another key limitation of the reduced models is that they ignore the fundamental indicators of the company's functioning, such as the value of assets, financial leverage, or level of profitability. In these models, it is problematic to link the probability of default and the recovery rate with the fundamental characteristics of the bond and its issuer, which makes the models more difficult to interpret from an economic point of view.

Trujillo-Ponce et al. (2014) determined which model, based on financial statements or market information, better assesses credit risk. As a proxy for credit risk, the authors use 2186 CDS spreads on the European market from 2002 to 2009. Models based on accounting information are criticized because of the historical nature of the information used as input and because they do not consider the volatility of the value of the company during the analyzed period. However, the proponents of this approach argue that capital market inefficiencies can lead to more significant errors in predicting credit risks. The authors emphasize the inconsistency of the results of previous works. Trujillo-Ponce et al. (2014) consider various credit risk proxies: the default or non-default of a company, credit ratings, corporate bond spreads, and CDS spreads. They give preference to CDS spreads because these reflect information about credit risks continuously, reflect the market's perception of a possible default, and not the opinion of a rating agency, reflect information about the risk of default, and not about the return of the principal amount; CDS spreads are less affected by taxes and liquidity, unlike bond spreads calculated using an "unknown" risk-free rate, CDS spreads themselves reflect credit risk.

The authors compare three models in which the dependent variable is represented by the natural logarithm of CDS spreads. The first model includes variables derived from financial statements. The second model includes factors based on market information (based on the structural approach). The third model includes both financial and market regressors. They did not find significant differences in the predictive ability of the approaches, concluding that the two types of data are complementary, and the complex model shows the best result. The predictive power of the models increases during periods of macroeconomic uncertainty (e.g. financial crisis).

Market models are often too complex or market dependent. Their application requires access to a large amount of data (knowledge of the market value of share capital, debt obligations, spreads of bond yields, etc.). Despite the widespread use of market models by Western companies, their use in the Russian market is difficult due to the small number of listed securities.

2.2 Models Based on Fundamental Indicators

Totmyanina (2011) identifies three groups of models, based on fundamental indicators, depending on the nature of the indicators underlying them:

1. based on macroeconomic indicators;
2. based on indicators of financial and accounting statements;
3. based on indicators of external rating agencies.

A feature of models based on *macroeconomic indicators* is the idea that the probability of default is cyclical and increases during an economic recession. The macroeconomic indicators used in these include: GDP, inflation, national currency, and unemployment rate. That allows us to give a long-term estimate of the probability of default. A classic example of such models is the Wilson model (1997a, b), which was the basis for the development of CreditPortfolio View, designed to assess credit risk and developed by the consulting group McKinsey & Co.

Financial ratios derived from *financial statements* are an important source for constructing default forecasting models. Beaver (1966) and Altman (1968) were the first to use financial ratios to analyze and predict the probability of default; their work was continued by Abidali and Harris (1995). These works focus on the application of multiple discriminant analysis in determining the probability of defaults in the corporate sector using financial ratios. Ohlson (1980) was one of the first to successfully use logistic analysis to predict company insolvency.

Illustrative is the work of Ping Tserng et al. (2014), which is devoted to assessing the probability of defaults of construction companies based on financial ratios using the logit model. The authors conduct multivariate and univariate analysis. The logistics model included 21 ratios, divided into five groups (liquidity, financial leverage, turnover, profitability, market factors). The final sample consisted of 87 US companies, 29 of which defaulted between 1970 and 2006. Forecasting horizons of 1, 2, and 3 years were considered. The results show that the addition of market variables (the ratio of the market value to the book value) increases the accuracy of default forecasting, especially when the forecast horizon is within one year. Of the models considered, the best was one that included the following factors: ROA, financial leverage, total assets turnover, current liquidity ratio, and the ratio of the market value to the book value. The AUC of this model is 0.7918 and 0.7951 when forecasting for one and two years, respectively. The greatest predictive ability was shown by ROA.

The class of models based on data from *rating agencies* is widespread. The rating contains important information with an average market efficiency if it provides the market with non-public confidential information. An important argument in the favor of this thesis is that rating agencies have long-term relationships with various issuers and investors. Discussions with senior management, the telephone and personal contacts of analysts with issuers provide valuable and reliable information about the internal affairs of companies, which is not always available to external users. Rating agencies learn about planned issues, strategic plans, reserves, future dividend policies, and anticipated corporate actions. They analyze financial statements, assess risks, and extract more accurate information about the company's profit and loss. It is also more preferable for a company to disclose information to rating agencies than to the public or the media, as rating agencies are required to maintain confidentiality under the terms of the rating assignment agreement.

To determine the probability of default, a *cohort approach* is used, based on which *transition matrices* are constructed, which estimate the frequency of credit ratings changing for a given sample of companies. In this case, the probability of default can be obtained on the basis of the analysis of historical data as the ratio of the number of firms that made the transition to the default rating to the total number of observations. This information is periodically published by the largest world rating agencies.

2.3 Advanced Models

Discriminant, logistic analysis is a popular traditional tool for predicting bankruptcies, but it has a number of drawbacks associated with its low predictive power and the presence of restrictions on its use. Therefore, nonparametric methods have become widespread.

Frydman et al. (1985) were among the first to use *classification trees* to predict company bankruptcies. They found that their classification trees outperform discriminant analysis. It was also noted that with the complication of the model (including more factors), the accuracy of the model deteriorated due to overfitting. However, this success did not increase the frequency use of decision trees in this area.

Further development of the use of classification trees is the use of algorithms based on bootstrap approaches. *Random forest* is a ML algorithm that represents a combination of using classification trees.

Based on financial reporting data, Behr and Weinblat (2017) use random forest models to predict the defaults of companies from seven European countries (Finland, France, Germany, Italy, Portugal, Spain, and the United Kingdom) and to identify specific signs of defaults of companies in various countries. The authors note that the source data cannot be used as input due to the low share of insolvent companies (the problem of imbalance). The most common method of dealing with data imbalance is undersampling or oversampling. Since undersampling in the work would lead to the loss of more than 96% of observations, the authors use the oversampling approach, paying attention to the fact that the calculation process is complicated, since such a model requires about two million objects. The authors note that the model is highly dependent on internal parameters (the number of trees; the number of parameters used to construct one tree; the maximum number of layers in a tree; the minimum number of objects in a descendant node or the parent node), and their determination is based on the cross-validation procedure.

Behr and Weinblat (2017) note the advantages of random forests such as high accuracy and resistance to emissions. In addition to high forecasting accuracy (AUC is in the range from 0.6903 to 0.8530), random forests made it possible to identify country-specific factors that have the greatest impact on a company's insolvency. It is determined that the use of a general model which does not take into account country specifics leads to a decrease in the effectiveness of the model. It was found that the greatest impact on a company's insolvency is provided by the ratios: Debt ratio, ROA, ROS, Net Debt to EBITDA Ratio.

The idea of *neural networks* is based on how the human brain analyzes data. Currently, this algorithm is used in various tasks, for example, pattern recognition, classification, and time series forecasting. Neural networks have a built-in ability to adapt their weights to changes in the environment.

Odom and Sharda (1990) were among the first to use a neural network to predict bankruptcy. They built a neural network with several hidden layers and used the financial ratios from the Altman model as input. The share of correctly classified companies was about 80%.

Tam and Kiang (1992) were among the first to compare the predictive power of the logit model, the k nearest neighbors method, classification trees, and neural networks. They conclude that the neural network is superior to all the other methods.

The main disadvantage of neural network modeling is the fact that a neural network acts as a "black box," i.e. the result is not interpretable. Altman et al. (1994) conduct a comparative analysis of neural networks and linear discriminant analysis. They conclude that neural networks show high accuracy in determining the solvency category of a company. Nevertheless, this model is inferior to the predictive quality of the traditional logit model. The authors note the disadvantages of neural networks: the problem of overfitting, training time, and the non-interpretability of the model parameters. Altman (2018) remains skeptical whether practitioners will accept "black box" methods for assessing credit risk of counterparties.

We can conclude that at present there are many works proving the possibility of using advanced methods for predicting the insolvency of companies. These algorithms often show higher efficiency, even though they are characterized by significant time and physical costs. The next section discusses the examples of the successful application of various bankruptcy forecasting methods in Russian practice.

3 Russian Modeling Experience

Despite the importance of the task of predicting the bankruptcy of counterparties using more advanced methods, there are not so many Russian works in this area. Works devoted to comparing the accuracy of traditional and non-traditional models in predicting bank defaults are more likely the exception. In many Russian studies that use non-traditional methods, special attention is not paid to the training of the algorithm. In this case, default algorithm parameters are often used, which may not be the most optimal.

Karminsky et al. (2012) consider the features of modeling the probability of a bank default in the context of Russian reality using a logistic model. Based on Russian banking statistics, macroeconomic and institutional data for 1998–2011, a number of default probability models for the Russian banking sector were constructed. The logistic model, combined with the CAMELS approach in selecting the best explanatory variables, demonstrated high predictive power when testing outside the sample: more than 60% of defaults that occurred in 2010–2011 were correctly predicted. The authors conduct a comparative analysis of traditional models with neural networks. According to the results of testing on the test set of the neural network, 42% of defaults were predicted, which is a low indicator compared to logistic models.

Bogdanova et al. (2013), based on the data of financial indicators of public reporting, conducted an analysis of the solvency of Russian enterprises in the manufacturing industries. The authors compare neural network models with well-known traditional models. In this study, the best model had one inner layer consisting of four neurons, providing a forecast accuracy of 85.1%. Researchers conclude that neural network models are superior to logit models in accurately identifying potential defaults.

Demeshev and Tikhonova (2014) compare approaches to modeling the critical financial situation of Russian SME in various industries using financial and non-financial indicators from 2011 to 2012. The authors consider four industries: manufacturing, real estate, wholesale and retail, construction. A feature of the work is the amount of data (almost 1 million observations), the number of statistical methods: logit and probit models, linear discriminant analysis, quadratic discriminant analysis, discriminant distribution mixture analysis, the classification tree method, and random forest algorithm. The greatest predictive power was shown by the random forest algorithm, regardless of industry, and type of sample (balanced or unbalanced). They concluded that non-linear algorithms show the best results. The most significant non-financial factors were industry, federal district, and the age of the enterprise. The size of the enterprise and its organizational legal form had a weak impact on defaults.

Despite the advantages of non-linear models, works using traditional binary choice models to predict the probability of defaults prevail in Russian practice. Rybalka (2017) uses logistic regressions to test the hypothesis of the influence of corporate structure (such as characteristics of the general director, board of directors, ownership structure) on the predictive power of the models. He confirms his hypothesis and notes the convenience of using traditional models to solve similar tasks.

Kostrov (2016) compares statistical classification methods for predicting bankruptcies of Russian banks. The author notes that only a small number of Russian banks have an international rating, however, the relevance of the forecast for revoking a bank's license at that time was especially high (60–80 financial institutions went bankrupt annually). The author described a linear discriminant analysis, a naive Bayesian classifier, logistic regression, decision trees, a neural network when forecasting the bankruptcy of Russian banks over a 6-month horizon. In this case, to combat imbalance, the author uses the oversample method with the duplication of observations of the bankruptcy type m times, where m takes the following values $\{1, 5, 10, 25, 50, 100\}$. As a measure of the quality of the models, the author used the arithmetic mean of the proportion of outcomes of the True Positive Rate (TPR) and the proportion of outcomes of the True Negative Rate (TNR). The author concludes that cases of bankruptcy of a bank with negative capital are predictable on the forecast horizon of six months. The naive Bayesian classifier was the best model; logistic regression was next. The use of neural network modeling and the decision tree method showed poor results. The author used the default neural network with one hidden layer and ten neurons. In our view, the process of learning and the search for the optimal architecture of the neural network could improve the predictive accuracy of the models (which is also characteristic of decision trees).

Karminsky and Burekhin (2019) compare the ability of traditional and advanced models to predict the bankruptcy of Russian construction companies on a one-year horizon. They consider logistic models and their modifications using the WOE metric, classification trees, random forests, artificial neural networks. Particular attention is paid to the features of ML models, the problem of data imbalance, the analysis of the influence of non-financial factors on the predictive ability of models. The authors used financial and non-financial indicators from 2011 to 2017. AUC was used as a metric for the quality of the models. The authors focus on identifying companies which were in danger of bankruptcy, including companies for which the legal bankruptcy procedure had been launched and companies that have liquidated voluntarily.

It is concluded that the algorithms show acceptable quality for use in bankruptcy forecasting. Artificial neural networks were found to outperform other methods, while logistic regression models combined with WOE adjustments closely follow them. It was found that the effectiveness of the method of overcoming data imbalances depends on the type of models used. For logistic regressions, artificial neural networks, and classification trees, oversampling showed higher quality. However, using oversampling in the random forest method leads to overfitting. Therefore, for random forests undersampling is more efficient. A significant effect of the imbalance of the training set on the predictive ability of the model was not revealed. The significant effect of non-financial indicators on the likelihood of bankruptcy was also not confirmed.

4 The Main Trends in Forecasting Bankruptcies

In the last decade, most studies have focused on improving and comparing existing models. A broad review was conducted by Kumar and Ravi (2007) who reviewed 128 scientific papers from 1968 to 2005. They note that most methods (discriminant analysis, logit analysis, classification trees, etc.) can be used to predict bankruptcies and give satisfactory results. However, the neural network algorithm has the greatest accuracy. At the time of writing, the authors noted that there is a tendency for algorithms based on one method to lose popularity, while ensemble or hybrid models are becoming more popular and show better performance. A striking example is provided by Xiao et al. (2012), where the prognostic ability of logistic regression, support vector machine (SVM), and neural networks are combined. The results of three separate models were combined into an "ensemble model" and weighted. They conclude that the combined method was superior to the predictions of the three methods individually. They also note that the lack of generally accepted procedures when building hybrid models are serious barriers to the use of these techniques.

Qu et al. (2019) review bankruptcy forecasting models using ML and DL models. They note the interest of researchers in the use of DL not only in problems related to pattern recognition, voice, NLP, but also in financial fields, including in solving problems of forecasting defaults. They consider the work of Hosaka (2019), as a successful application of a convolutional neural network in predicting the bankruptcy of Japanese companies. Mai et al. (2019) is an example of the use of NLP and neural networks in assessing the credit risk of US public companies. Mai et al. (2019) note the significant contribution of textual information (such as financial reports, expert opinions, and media reports) in improving the accuracy of the models. This textual information can become a new driver for the development of predictive models. The authors also note a tendency to obtain interpretable results from the black box while maintaining the high accuracy of these models.

In Russia, there is also a tendency towards more complicated predictive models using ML algorithms. However, there are few such studies, which may be due to the lack of similar models in business processes, insufficient management awareness of the possibilities of such algorithms, and the high cost of developing and implementing such models. There is also a clear interest in the development of more diversified models. Despite the clear superiority of non-linear algorithms in accuracy over traditional models, Russian researchers continue to use them because of their simple interpretation, the ease of construction, and the ability to answer questions of interest to the researcher.

References

Abidali, A. F., & Harris, F. C. (1995). A methodology for predicting company failure in the construction industry. *Construction Management and Economics, 13*(3), 189–196.

Altman, E. (2018). A fifty-year retrospective on credit risk models, the Altman Z-score family of models and their applications to financial markets and managerial strategies. *The Journal of Credit Risk, 14*, 1–34.

Altman, E. I. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*(4), 589–609.

Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking & Finance, 18*(3), 505–529.

Beaver, W. H. (1966). Financial ratios as predictors of failure. *Journal of Accounting Research, 4*, 71–111.

Behr, A., & Weinblat, J. (2017). Default patterns in seven EU countries: A random forest approach. *International Journal of the Economics of Business, 24*(2), 181–222.

Black, F., & Cox, J. (1976). Valuing corporate securities: some effects of bond indenture provisions. *Journal of Finance, 31*, 351–367.

Bogdanova, T. K., Shevgunov, T. Y., & Uvarova, O. M. (2013). Using neural networks for predicting solvency of Russian companies on manufacturing industries. *Business Informatics, 2*, 40–48.

Demeshev, B. B., & Tikhonova, A. S. (2014). Bankruptcy forecasting of Russian companies: Cross-industry comparison [Electronic resource]. Working paper WP2/2014/04 (Series Quantitative analysis in economics).

Eom, Y. H., Helwege, J., & Huang, J. Z. (2004). Structural models of corporate bond pricing: An empirical analysis. *Review of Financial Studies, 17*(2), 499–544.

Frydman, H., Altman, E. I., & Kao, D. L. (1985). Introduction recursive partitioning for financial classification: the case of financial distress. *Journal of Finance, 40*(1), 269–291.

Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications, 117*, 287–299.

Huang, Y. (2009). Prediction of contractor default probability using structural models of credit risk: an empirical investigation. *Construction Management and Economics, 27*(6), 581–596.

Jarrow, R. A. (2009). Credit risk models: Annual review of financial economics. *Journal of Contemporary Accounting & Economics, 6*(1), 34–45.

Karminsky, A. M., & Burekhin, R. N. (2019). Comparative analysis of methods for forecasting bankruptcies of Russian construction companies. *Business Informatics, 13*(3), 52–66.

Karminsky, A.M., Kostrov, A.V., & Murzenkov, T.N. (2012). Approaches to evaluating the default probabilities of Russian banks with econometric methods. Working paper WP7/2012/04 (Series Mathematical methods of decision analysis in economics, business and policies).

Kostrov, A. V. (2016). Comparison of statistical classification methods to predict Russian banks failures. *Management of Financial Risks, 47*(3), 162–180.

Kumar, P. R., & Ravi, V. (2007). Bankruptcy prediction in banks and firms via statistical and intelligent techniques: A review. *European Journal of Operational Research, 180*(1), 1–28.

Longstaff, F., & Schwartz, E. (1995). A simple approach to valuing risky fixed and floating rate debt. *Journal of Finance, 50*(3), 789–819.

Mai, F., Tian, S., Lee, C., et al. (2019). Deep learning models for bankruptcy prediction using textual disclosures. *European Journal of Operational Research, 274*(2), 743–758.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance, 29*(2), 449–470.

Odom, M. D., & Sharda, R. (1990). A neural network model for bankruptcy prediction. *Neural Networks in Finance and Investing, 2*, 163–168.

Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*, 109–131.

Qu, Y., Quan, P., Lei, M., et al. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science, 162*, 895–899.

Rybalka, A. I. (2017). Modeling the probability of default in the construction sector: Factors of corporate governance. *Journal of Corporate Finance Research, 3*(13), 79–99.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science, 38*(7), 926–947.

Totmyanina, K. M. (2011). A review of default probability models. *Financial Risk Management, 1* (25), 12–24.

Trujillo-Ponce, A., Samaniego-Medina, R., & Cardone-Riportella, C. (2014). Examining what best explains corporate credit risk: accounting-based versus market-based models. *Journal of Business Economics and Management, 15*(2), 253–276.

Tserng, P., Chen, P.-C., Huang, W.-H., Lei, M. C., & Tran, Q. H. (2014). Prediction of default probability for construction firms using the logit model. *Journal of Civil Engineering and Management, 20*(2), 247–255.

Wilson, T. (1997a). Portfolio credit risk: part I. *Risk, 9*, 111–117.

Wilson, T. (1997b). Portfolio credit risk: part II. *Risk, 10*, 56–61.

Xiao, Z., Yang, X., Pang, Y., & Dang, X. (2012). The prediction for listed companies' financial distress by using multiple prediction methods with rough set and Dempster-Shafer evidence theory. *Knowledge-Based Systems, 26*, 196–206.

Measures and Assessment of ALM Risks in Banks: Case of Russia

Ekaterina Seryakova

Abstract This chapter focuses on the assessment and management of ALM risks: liquidity risk and interest-rate risk. The first part is devoted to liquidity risk: various types of liquidity risk, its sources, measures, and the principles of liquidity risk management, as well as scenarios for stress testing of liquidity risk. The second part focuses on the concept and types of interest-rate risk, the methods of evaluation (metrics) and approaches to its management. In the conclusion, current challenges in assessing and managing ALM risks are presented.

Keywords ALM risks · Stress testing · Interest-rate risk · Liquidity risk · Management

JEL G21 · G32

1 Introduction

Liquidity risks in banks are divided into two categories: market liquidity risk and liquidity funding risk. Market liquidity risk occurs due to a slump in the price of any financial instrument which a bank possesses (as an asset). Liquidity funding risk arises due to a mismatch in the terms of assets and liabilities. This paragraph will mainly concentrate on funding liquidity risk which can be divided into three types: physical liquidity risk; risk of regulatory liquidity; structural liquidity risk.

Physical liquidity risk occurs due to the incapability of a bank to fulfill its obligations in any currency due to a deficit of cash or non-cash money in this currency. The risk of regulatory liquidity occurs when a bank violates the regulatory requirements for liquidity ratios. The risk of structural liquidity is explained with existing disbalances on both the asset and liabilities side of balance sheets. For

E. Seryakova (✉)
MGIMO, Gazprombank, Moscow, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_8

instance, the concentration risk of non-stable deposits or imbalances of deposits in different currencies are examples of structural liquidity risk. Sources of liquidity risk can be classified into external (shocks in market interest rates, client panic) or internal (the default of a client and the realization of credit risk). Liquidity funding risk can be realized through several channels:

1. Slump in liquidity buffer. The price of high-liquidity assets portfolio slumps due to an increase in market interest rates. High-liquidity asset portfolios include cash, nostro accounts with other banks, accounts with the Central Bank and high-liquidity bonds (with rating at least BBB- according to S&P rating scale). The liquidity buffer represents sources of funding available in a stressed period which consists of confirmed funding sources adjusted by a surplus or deficit of cash. Elements of liquidity buffers are repurchase agreement operations with Central Bank and funding collateralized by credit portfolios recognized with a discount.
2. An increase in off-balance operations which are driven by a higher part of loan drawing in stress periods when borrowers anticipate further growth in interest rates. For instance, during crises, the average loan drawing could increase from 50% up to 90%.
3. The realization of a credit risk of a heavy borrower.
4. Client outflows which occur due to a lack of confidence in the banking sector. Such behavior could provoke a chain of bankruptcies in the banking sector and lead to the realization of systemic liquidity risk.

There are two main methods of liquidity risk estimation:

1. Cashflow forecasting. Cashflow forecasting amid normal market conditions is based on behavior balance models (models of prepayment, models of renegotiation) and suggests measures in case of liquidity risk aggravation.
2. Stress testing. Stress testing is necessary for defining an adequate volume of the liquidity buffer and elaborating financial resilience restoration plan in case of crisis realization.

2 Measures of Liquidity Risk and Principles of Liquidity Management

After the 2007–2009 financial crisis, liquidity risk became a key banking risk. The standards of Basel III, which appeared in 2010, implemented new requirements for liquidity risk measurement. In particular, such metrics as LCR (Liquidity coverage ratio) and NSFR (Net Stable Funding Ratio) were introduced to measure bank capabilities to resist stress within 1 month and within more than 1 year, respectively. LCR is a ratio of high-liquidity assets to net cash outflow for 1 month (BIS, BCBS, 2013. Liquidity coverage ratio and liquidity risk monitoring tools). Minimum requirements for LCR are set by Central banks. The minimum value of LCR starting from 01.01.2019 in Russia is 100%.

Managerial LCR (MLCR) is calculated as.

$$\mathrm{MLCR} = \frac{\text{Available high} - \text{liquidity assets}}{\text{Net cash outflow}}, \tag{1}$$

where *Available high-liquidity assets* = cash + nostro accounts with other banks + account with Central Bank + high-liquidity bonds (with rating at least BBB- according to S&P rating scale);

$$\mathit{OutflowNet\ cash\ outflow} = \mathit{Outlow} - \mathit{min}\,(0,75 \cdot \mathit{Inflow}). \tag{2}$$

where *Outflow* is cash outflow (deposit outflow, utilization of credit lines); *Inflow is* cash inflow (e.g., credit redemption).

NSFR defines the volume of stable resources necessary for funding long-term assets amid stress. NSFR can be represented as the ratio of available stable funding to required stable funding. The minimum value of NSFR is 100% starting from 01.01.2018. Banks regularly conduct *gap-analysis* which represents the difference between all cash inflows and outflows. There are three scenarios which define the breakdown of cashflows into time buckets:

1. Gap_Plan CCYi defines the liquidity gap, calculated by currencies and time buckets according to the planned operations of a bank within a period of operational plan of a bank (usually 3 months).
2. Gap_Stress CCYi defines the liquidity gap, calculated by currencies and time buckets considered in stress periods.
3. Gap contractual_CCYi defines the liquidity gap, calculated by currencies and time buckets according to the contract maturity of instruments.

The solvency horizon of a bank which is usually called the survival horizon is a period within which the solvency of a bank is provided by a liquidity buffer in stress periods.

3 Principles of Liquidity Risk Management

Important *principles of liquidity risk management* are the following (BIS, BCBS (2000). Principles for sound risk-management and supervision):

1. The management of liquidity risk is conducted in accordance with risk appetite which is constrained by liquidity risk measures: constraints on the liquidity contract gap and constraints on bank-calculated MLCR and NSFR.
2. The management of banking balance. In a normal market situation, liquid assets are planned first and then their funding is provided.
3. The diversification of resources by clients, sources, instruments, and terms.
4. The costs of liquidity risk management are allocated by business-departments by means of transfer pricing.
5. The principle of "three lines of defence" (see Table 1).

Table 1 Principle of "three lines of defence"

First line of defence	Treasury	Forward-looking approach to risk of liquidity management, setup of limits
Second line of defence	Risk management	Control of limits on risk of liquidity measures
Third line of defence	Internal audit	Independent validation of models, procedures, and processes of the risk of liquidity management in Treasury and risk management

The elaboration of scenarios for stress-testing risk of liquidity is considered to be the most sophisticated in banking risk-practices as it requires assumptions for both assets and liabilities (BIS, BCBS (2018) Stress-testing principles). *Assumptions for assets* can be:

- a rise in overdraft drawings;
- an increase in the probability of defaults for some borrowers;
- a reduction of cashflows from interbank operations due to the default of one of another bank;
- the default of one or several heavy borrowers (concentrated risk realization).

Assumptions for liabilities can be:

- the reduction of refinancing and the reduction of attracting long-term deposits;
- the early termination of the agreement on current accounts with minimum balances;
- the reduction of cash "on demand" below a stable level;
- the outflow of funds of the largest creditor within one quarter;
- the early demand of deposits;
- the increase in collateral for margin transactions;
- the inaccessibility of a bank to capital markets.

Scenario analysis by product when setting or reviewing limits and approving new types of operations may include a set of risk factors corresponding to certain types of risk of the instrument or product, operational risk factors, and other risk factors. If modeling technologies allow a bank to take into account time factors in the results of stress testing, dynamic stress testing is used. Dynamic stress testing is applied to take into account the severity of risk losses when forming or adjusting the strategy of a bank. Dynamic stress testing can include (individually or in combination) the following elements:

- the dynamic Stress scenario, with a certain duration, gradual deployment, reaching peak values, and then reducing the intensity, etc.
- the deferred reaction of financial indicators, which is especially relevant when assessing the impact of changes in the macro-parameters on the stress-tested indicators of a bank's performance.

4 Interest-Rate Risk

4.1 Types of Interest-Rate Risk

Interest-rate risk management is a set of actions and procedures that manage and control a bank's interest-rate risk arising from its assets and liabilities, including their effect on the balance sheet and income statement. Interest-rate risk is the risk of losses due to adverse movements of market interest rates. A bank manages its assets and liabilities regularly and measures, manages, and monitors its interest-rate risk on a stand-alone and consolidated basis. Interest-rate risk limits are set for all interest-rate risk metrics and comply with risk-appetite of a bank for interest-rate risk. A bank recognizes the importance of asset-liability management (ALM) as part of the effective management of its balance sheet and income statement. Before granting substantial new loans, purchasing bonds or making any other type of investment, the impact of the transaction on a bank's interest-rate risk profile and liquidity situation is assessed. The main interest-rate sensitive assets are corporate loans, bonds, term deposits, nostro accounts with other banks and corresponding accounts of the Central Bank, client account overdrafts and financial derivatives. The main interest-rate sensitive liabilities are received term funds (deposits), client accounts, accounts from other banks, and financial derivatives. Table 2 contains risk mitigation actions and risk remediation actions in respect of interest-rate risk.

According to "Interest-rate risk in the banking book" (IRRBB) types of interest-rate risk:

- *Repricing risk:* risk which occurs due to maturity term mismatches or repricing term mismatches. The examples illustrating this type of risk:
  - 1-year assets are funded with 3-month deposits;
  - a loan with floating rate (6-month LIBOR +0,2% spread) is funded with a 3-month deposit.

Table 2 Risk mitigation actions and risk remediation actions in respect of interest-rate risk

Risk mitigation actions	Risk remediation actions
• Generally avoid, minimize or hedge open interest-rate risks. • Manage the balance sheet in a term-congruent manner. • Ensure that the approved limits are sufficient for the business plan or strategy. • Pre-check the available limits before entering into new transactions. • Establish netting (ISDA) and credit support (CSA) agreements and sufficient limits with hedge counterparties to be prepared for risk transfers. • Show early warning indicators in the limit utilization report	• Hedge excessive interest-rate risk through risk transfer with hedge counterparties. • Interest-rate risk management through transfer pricing policy. • Temporarily or permanently review limits via the authorized approval body

- *Yield curve shift risk:* unfavorable parallel or non-parallel shifts of the market yield curve, leading to a Net Interest Income (NII) slump and the aggravation of sensitivity of Net Present Value (NPV).
- *Basis risk:*
  - risk which occurs due to the different pace of loan and deposit rates changes on the condition that loans and deposits are of the same term and with fixed rates; this difference in pace is explained by the different sensitivity of loans and deposits to changes in market rates: for example, risk of losses due to adverse changes in the spreads between the rates of borrowing and placement for one term in one currency (Mosprime 3 M and RUONIA) or for one term in different currencies (Mosprime 3 M and EURIBOR 3 M);
  - risk which occurs due to the different bases of interest floating rates (loan has a rate equal to RUONIA +2% spread and deposit has Mosprime Overnight +1% spread);
  - risk which arises due to the fact that loans could be attached to fixed rates and deposits – to floating rates and vice versa;
- *Optionality risk:* risk of losses due to behavioral models (prepayment and repricing) applied to financial instruments subject to interest-rate risk;
- *Risk of funding spread change:* risk of losses due to changes in the spread between the cost of borrowing resources by a bank in the financial market and the credit spread of this bank, which depends on the market interest curve. Banks in Russia are exposed to this type of risk on positions in foreign currencies.

4.2 Metrics of Interest-Rate Risk

Sensitivity of interest-rate risk (ΔNII):

$$\Delta NII_{CCY,H}(I, \Delta R_{CCY}) = X_{ccy} * \Delta R_{CCY} * \sum\nolimits_{k=0}^{n} CF_{k,CCY}(I, \Delta R_{CCY}) * (H - t_k), \quad (3)$$

where X_{ccy} is the currency exchange rate for instruments in non-national currencies;

$CF_{k,\ CCY}(I, \Delta R_{CCY})$ are cash flows (positive for assets, negative for liabilities) for instrument I in currency CCY; cash flows do not include interest payments;

t_k is the term of maturity of $CF_{k,\ CCY}(I,\ \Delta R_{CCY}$or the term of interest-rate repricing on financial instrument I;

ΔR is the market interest-rate parallel shift.

Interest-rate risk gap in CCY currency for financial instrument I for period T is computed with a breakdown into time buckets for $CF_{k,\ CCY}(I, \Delta R_{CCY})$. Such gaps calculated by term buckets are called marginal. The consequent summing of marginal gaps can give a cumulative gap for each bucket:

$$GAP_{CCY,T}(I) = X_{CCY} \sum_k (CF_{k,CCY}(I) |\, t_k \leq T). \tag{4}$$

Sensitivity of Net Present Value (ΔNPV) in CCY currency for financial instrument *I* is calculated as

$$\Delta NPV_{CCY}(I, \Delta R_{CCY}) = \sum_k \left(\frac{CF_{k,CCY}(I, \Delta R_{CCY})}{\exp([R_{CCY}(t_k) + \Delta R_{CCY}(t_k)] * t_k)} - \frac{CF_{k,CCY}(I, o)}{\exp(R_{CCY}(t_k) * t_k)} \right) X_{CY}. \tag{5}$$

Table 3 contains approaches to evaluating interest-rate risk according to Basel III recommendations.

4.3 Managing Interest-Rate Risk

There are three main objectives for managing interest-rate risk:

1. *Hedging the interest position of the bank*. This is carried out in order to reduce net interest income and to minimize the risk of a parallel shift in the market interest curve. The goal is mainly applied in countries with low interest rates and flat market interest curves.
2. *The transformation of balance term-structure* (placement of short-term liabilities into long-term assets). The goal is to maximize income within predetermined limits. The goal is applied in countries with an increasing market interest curve.
3. *The acceptance of interest-rate risk within the specified limits*. The goal is justified in countries with low market liquidity and insufficiently developed market of derivatives (IRS).

The main factors influencing *the choice of the interest-rate risk-management g*oals are:

1. The shape and slope of the market interest curve;
2. Market development of derivative financial instruments;
3. The share of bank assets in the banking sector;
4. A bank's risk appetite for interest-rate risk (the level of limits on interest-rate risk metrics);
5. The ratio of interest and commission income in the total income of a bank.

Interest-rate risk management in a bank is conducted using the transfer pricing system and is concentrated in the internal audit service of the bank. The internal audit carries out centralized ALM risk management. The main functions of transfer pricing are the redistribution of risks and the determination of the internal cost of resources in a bank. A bank has a special unit—the internal treasury department—which manages interest-rate risk. The essence of management is to transform the

Table 3 Approaches to evaluating interest-rate risk according to Basel III (BIS, BCBS, 2016. IRRBB)

Metrics	Scenarios	Calculations	Portfolios
EV/EVE economic value of equity	Six scenarios by currency for EVE (ΔR): (i) Parallel shift of market curve up; (ii) parallel shift of market curve down; $\Delta R_{parallel,\ c(tk)} = \pm R_{c(tk)} \cdot \alpha_{parallel,}$ $Rc(tk)$ is the parallel shift c is currency tk is the time bucket from o to k; (iii) change of the short, middle or long end of the curve (shift up); (iv) change of the short, middle or long end of the curve (shift down) $\Delta R_{short,c(tk)} = \pm R_{c(tk)} \cdot \alpha_{short} \cdot e^{(-tk/4)}$ $\Delta R_{medium,c(tk)} = \pm R_{c(tk)} \cdot \alpha_{medium} \cdot S_{medium(tk)}$ $\Delta R_{long,c(tk)} = \pm R_{c(tk)} \cdot \alpha_{long} \cdot (1 - e^{-tk/4})$ (v) Anti-clockwise turn of market curve (steepening); $\Delta R_{c,(tk)} = -0{,}65 * \Delta R_{short,c(tk)} + 0{,}9 * \Delta R_{long,c(tk)}$ (vi) clockwise turn of market curve (flattening) $\Delta R_{c,(tk)} = 0{,}8 * \Delta R_{short,c(tk)} - 0{,}6 * \Delta R_{long,c(tk)}$	$EVE_{i,c} = CF_{i,c}(k) * e^{(-Ri,c(tk) * tk)}$ Risk-parameters: • For parallel shift $\alpha = 60\%$; • For short end of the market curve $\alpha = 85\%$; • For middle part of the market curve $\alpha = 55\%$; • For long end of the market curve $\alpha = 40\%$	Banking and trading books
EaR (earnings-at-risk) (ΔNII)	Scenarios by currencies: (i) Parallel shift up of the market curve; (ii) parallel shift down of the market curve	$\Delta NII_{i,c} = \Delta NII_{i,c}^{g} + \Delta NII_{c}^{b}$ i is the scenario; c is the currency; g is the component 1: Change of NII due to scenario i; b is component 2: Change of NII due to basis risk	Banking and trading books

current interest position (interest gap) into the target one, which corresponds to the risk appetite of a bank, namely, the existing limits on the interest gap (Kulik and Vedyakhin 2017). The following tools are used to manage interest-rate risk:

– limits on interest-rate risk metrics (interest gap, sensitivity of interest income, sensitivity of net present value);
– hedging an interest position using interest-rate derivatives (IRS and CIRS);

– changing the transfer curve in different time buckets to stimulate attracting or placing bank units to attract or place funds for necessary terms;
– performing operations:
– to purchase or sell securities in the available-for-sale portfolio of the bank;
– in the money market;
– in the capital market: issuing bonds, issuing subordinated loans;
– in the market for derivative financial instruments: the conclusion of interest-rate transactions (IRS) and currency interest-rate swaps (CIRS).

Models used in calculating interest-rate risk:

– model of the prepayment of loans to individuals: for instance, for mortgage loans, two options are taken into account: prepayment and refinancing at a lower rate in case of decrease in market rates.
– model of the prepayment of corporate loans;
– model for revising interest rates for loans with a quasi-floating interest rate.
– The purpose of using these models is to reduce the effective term of loans.
– -model of the early termination of term deposits due to an increase in market interest rates.

4.4 Challenges in ALM Risks Assessment and Management

It is worth mentioning current challenges in asset-liabilities (ALM) risk assessment and management:

1. The separation of interest-rate risk from other types of risk when elaborating scenarios for integrated stress tests.
2. The elaboration of complex stress scenarios: it is difficult to separate the effects of changes in interest-rate risk, total credit spread (CSRBB), and individual credit spread when calculating interest-rate risk metrics.
3. The issue of attributing an instrument to the banking or the trading book is ambiguous: in international practice and in practice of leading Russian banks, derivatives, and debt instruments of the trading portfolio (re-evaluated daily through profit or loss) are referred to the trading book, while the rest are referred to the banking book.
4. It is challenging to evaluate ΔNII metric for a non-parallel shift of a market curve, i.e. for different shifts of the curve in different time buckets (Bank of Russia Report for public Consultation (2020)).
5. The qualitative judgment of changes in the interest-rate risk gap is not obvious: changes in the shape of the interest-rate risk gap cannot be clearly interpreted as better or worse.
6. It is difficult to conduct dynamic modeling when stress testing both interest-rate risk and liquidity risk. Dynamic modeling involves: (1) changes in market rates on the evaluation horizon more than once; (2) and/or changes in the balance

structure on the evaluation horizon. The main challenge is to develop assumptions for changing the balance sheet structure (scenarios and assumptions for reinvesting contracts), and scenarios for the evolution of market interest rates. For some metrics, a full dynamic analysis is not possible (changes in both market interest rates and balances): for example, the sensitivity of net present value (NPV) is a static measure and can be only used to evaluate sensitivity from a point in the future, taking into account changes only in the balance in this point.
7. The final challenging issue is to select the base indicator for products with a floating rate, for instance, the key rate of Central Banks which serves a base indicator for loans and causes both liquidity and interest-rate risks in the absence of deposits with the key rate as the base indicator.
8. The hedging of interest-rate risk, an underdeveloped IRS market.

5 Conclusion

This paper presents current approaches to evaluation and management of ALM risks today in systemically important banks in Russia. Principles of liquidity risk management and scenario assumptions for liquidity risk stress testing are mentioned. The last but not the least, current challenges in managing and assessing interest rate risk in Russian banks are highlighted.

References

Bank of Russia Report for Public Consultation. (2020). *On the best practices for managing interest rate risk on the banking portfolio in credit institutions*. https://cbr.ru/Content/Document/File/98190/Consultation_Paper_200120.pdf

Basel III. (2013). *The liquidity coverage ratio and liquidity risk monitoring tools, BIS, BCBS, 2013*. https://www.bis.org/publ/bcbs238.pdf

BIS, BCBS. (2000). *Principles for sound liquidity risk management and supervision*. https://www.bis.org/publ/bcbs144.pdf

BIS, BCBS. (2016). *Interest rate risk in the banking book* (*IRRBB*). https://www.bis.org/bcbs/publ/d368.pdf

BIS, BCBS. (2018). *Stress-testing principles*. https://www.bis.org/bcbs/publ/d450.pdf

Kulik, V., & Vedyakhin, A. (2017). *Fundamentals of risk management, textbook for the programs of Sberbank Corporate University* (p. 384). Moscow, Russia: ANO DPO "Sberbank Corporate University.

Forecasting and Backtesting of Market Risks in Emerging Markets

Dean Fantazzini

Abstract Emerging markets often go through periods of financial turbulence and the estimation of market risk measures may be problematic. Online search queries and implied volatility may (or may not) improve the model estimates. In these situations a step-by-step analysis with R and Russian market data is provided. Four classes of models are considered (GARCH, HAR, ARFIMA, and realized-GARCH), and a detailed forecasting and backtesting investigation is performed.

Keywords Forecasting · Value-at-risk · Realized volatility · Google trends · Implied volatility · GARCH · ARFIMA · HAR · Realized-GARCH

JEL Classification C22 · C51 · C53 · G17 · G32

1 Introduction

The Value-at-Risk (VaR) is the most well-known market risk measure and can be defined as the maximum portfolio loss over a determined time horizon at a given confidence level, see Jorion (2007) and the Basel Committee on Banking Supervision (2013, 2016) for more details. VaR is not sub-additive when the portfolio returns are not elliptically distributed, and for this reason, the risk of a portfolio can be larger than the sum of the separate risks of its components, see Artzner et al. (1997) and Artzner et al. (1999). An alternative risk measure which satisfies the property of sub-additivity is the Expected Shortfall (ES), which computes the average of the portfolio losses given a specific probability level, see Acerbi and Tasche (2002). Gneiting (2011) showed that ES does not satisfy a mathematical property called elicitability (while VaR does), and it cannot be directly backtested. In this regard, Emmer et al. (2015) and Kratz et al. (2018) showed that ES is elicitable

D. Fantazzini (✉)
Moscow School of Economics, Moscow, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_9

conditionally on VaR and it can be backtested using a multinomial test of VaR violations at multiple confidence levels.

This paper provides a step-by-step analysis with R and Russian market data to verify whether adding Google search queries and implied volatility (IV) from option prices to several volatility models can improve their estimated market risk measures. This analysis was recently performed by Fantazzini and Shangina (2020) and this paper is a practical complement to that paper, by showing step-by-step how to implement this backtesting exercise using R.

Google search data is a useful indicator of the behavior of the general public and small investors (Da et al. (2011), Goddard et al. (2015), Vlastakis and Markellos (2012), and Vozlyublennaia (2014), Campos et al. (2017), while IV represents a forward-looking estimate of the volatility mainly driven by the expectations of institutional investors and market makers (Mayhew (1995), Martens and Zein (2004), Busch et al. (2011), Bazhenov and Fantazzini (2019).

These two variables are added to four volatility models to forecast VaR at multiple levels for the daily data of the Russian RTS index. The forecasted VaR of these models are then compared using the tests by Kupiec (1995) and Christoffersen (1998), the asymmetric quantile loss (QL) function proposed by González-Rivera et al. (2004), the Model Confidence Set by Hansen et al. (2011), and the multinomial test of VaR violations by Kratz et al. (2018). Moreover, a robustness check to measure the accuracy of VaR forecasts obtained with a multivariate model is also discussed.

The rest of this paper is organized as follows. Section 2 reviews the literature dealing with Google Trends and IV, while the forecasting methods for VaR are briefly discussed in Sect. 3. The empirical exercise with R is reported in Sect. 4, while a robustness check is discussed in Sect. 5. Section 6 concludes.

2 Literature Review

There is a large body of literature that shows that IV delivers better forecasts for volatility than GARCH models, see Christensen and Prabhala (1998), Corredor and Santamaría (2004), Martens and Zein (2004), Busch et al. (2011), Haugom et al. (2014), and the references therein. Nonetheless, there are a few cases when this was not true as shown by Agnolucci (2009) and Birkelund et al. (2015). Moreover, the best results are often achieved when both IV and other market variables are included in the forecasting model, see Taylor and Xu (1997), Pong et al. (2004), and Jeon and Taylor (2013). Instead, the results are not that favorable when IV is used to forecast the future quantiles of the returns' distribution, see Chong (2004), Christoffersen and Mazzotta (2005), Giot (2005), Jeon and Taylor (2013), just to name a few.

Bams et al. (2017) represent the largest backtesting exercise dealing with VaR forecasts, using more than 20 years of daily data from US markets. Their analysis shows that IV based VaR tends to be outperformed by GARCH based VaR, due to the volatility risk premium embedded in IV. In general, Bams et al. (2017) showed that even though IV can be useful for forecasting future volatility, this is not the case for forecasting the returns' distribution quantiles, due to the complex dependence structure between the volatility risk premium and the extreme returns.

Google online queries can be a proxy for investor attention and information demand, see Ginsberg et al. (2009), Choi and Varian (2012), Da et al. (2011), Vlastakis and Markellos (2012), Vozlyublennaia (2014), Goddard et al. (2015), and Fantazzini and Toktamysova (2015). They can help to forecast future volatility, as discussed by Vozlyublennaia (2014), Dimpfl and Jank (2016), Campos et al. (2017), Xu et al. (2019) and Seo et al. (2019). For market risk management, the literature working with Google Trends is almost nonexistent: there are very few with empirical studies limited in scope and time, see Hamid and Heiden (2015), Basistha et al. (2018), and Bazhenov and Fantazzini (2019). Fantazzini and Shangina (2020) was the first work analyzing almost two decades of daily data for an emerging market, using a large scale backtesting analysis similar to the work by Bams et al. (2017): they found that the predictive power of several models did not increase if IV and Google data variables were added, while other models augmented with these variables did not reach numerical convergence. Fantazzini and Shangina (2020) showed that, in the case of Russian future markets, T-GARCH models with IV and Student's t errors are the best choice if robust market risk measures are of concern.

3 Methodology

This section shows how to implement and replicate with R most of the empirical analysis presented in Fantazzini and Shangina (2020), to ultimately verify whether adding IV and Google data to volatility models improves the quality of the forecasted VaR at multiple confidence levels for the Russian RTS index. I provide below a brief review of the theoretical aspects involved, while I refer the interested reader to Fantazzini and Shangina (2020) for more details.

3.1 Measures of Volatility

I use two volatility measures: the realized variance (RV) and IV from options prices (IV). The RV is a nonparametric and consistent estimator of the daily integrated variance, see Meddahi (2002) and Andersen et al. (2001):

$$\mathrm{RV}_{t+1} = \sum_{j=1}^{M} r_{t+j\Delta}^2, \tag{1}$$

where $\triangle = 1/M$ is the time interval of the intraday prices, M is the number of intraday returns, while $r_{t+j\triangle}$ is the intraday return. The weekly RV w at time t is the given by:

$$\mathrm{RV}_t^{(w)} = \frac{1}{5}\left(\mathrm{RV}_t^{(d)} + \mathrm{RV}_{t-1d}^{(d)} + \ldots + \mathrm{RV}_{t-4d}^{(d)}\right), \tag{2}$$

where we considered a weekly time interval of five working days. If the underlying stochastic process for the log-prices contains jumps, then it is possible to show that the RV converges to the sum of the integrated variance and the cumulative squared jumps, see Barndorff-Nielsen and Shephard (2004a, 2006), Andersen et al. (2007). The continuous sample path variation can be estimated non-parametrically using the standardized Realized Bipower Variation measure:

$$\mathrm{BV}_{t+1}(\Delta) = \mu_1^{-2}\sum_{j=2}^{\frac{1}{\Delta}} \left| r_{t+j\Delta}\right|\left|r_{t+(j-1)\Delta}\right| = \mu_1^{-2}\sum_{i=2}^{M}\left|r_{t,i}\right|\left|r_{t,i-1}\right| = C_{t+1}(\Delta), \tag{3}$$

while the jump component can be estimated by $J_{t+1}(\Delta) = \max\left[\mathrm{RV}_{t+1}(\Delta) - \mathrm{BV}_{t+1}(\Delta), 0\right]$, where the non-negativity truncation on the actual empirical jump measurements was proposed by Barndorff-Nielsen and Shephard (2004b) because the difference between RV and BV can become negative with real data. More elaborate methods to compute the jump components were proposed by Huang and Tauchen (2005) and Andersen et al. (2007).

If the financial market under consideration is not open 24 h (like for cryptocurrencies, see Fantazzini (2019), then the RV must be adjusted for the return in the overnight gap from the market close on day t to the market open on day $t + 1$. I scaled up the market-open RV using the unconditional variance estimated with the daily squared returns,

$$\mathrm{RV}_{t+1}^{24H} = \left(\frac{\sum_{t=1}^{T} r_t^2}{\sum_{t=1}^{T}\mathrm{RV}_t^{\mathrm{OPEN}}}\right)\mathrm{RV}_{t+1}^{\mathrm{OPEN}}, \tag{4}$$

where r_t^2 are the daily squared returns computed using the close-to-close daily prices, while $\mathrm{RV}_{t+1}^{\mathrm{OPEN}}$ is the RV computed with intraday data when the RTS future market is open, see Hansen and Lunde (2005), Christoffersen (2012), and Ahoniemi and Lanne (2013).

An implied volatility (IV) index computes the market expectations for future volatility implied by options prices. Differently from Fantazzini and Shangina (2020), I will use only the Russian Volatility Index (RVI) which was introduced on 16 April 2014 and which measures the market expectations for volatility over a

30-day period using prices of the nearby and next RTS Index option series.[1] The RVI formula is reported below:

$$\mathrm{IV} = 100\sqrt{\frac{T_{365}}{T_{30}} * \left| T_1\sigma_1^2 \frac{T_2 - T_{30}}{T_2 - T_1} + T_2\sigma_2^2 \frac{T_{30} - T_1}{T_2 - T_1} \right|}, \tag{5}$$

where T_{30} stands for 30 days expressed as a fraction of a calendar year, T_{365} for 365 days expressed as a fraction of a calendar year, T_1 is the time to expiration of the near-series options expressed as a fraction of a calendar year, T_2 is the time to expiration of the next far-series options expressed as a fraction of a calendar year, σ_1^2 is the variance of the near-series options and σ_2^2 is the variance of the next-series of options.[2]

3.2 Volatility Models

I employ the four models considered by Fantazzini and Shangina (2020): the Threshold-GARCH(1,1) proposed by Glosten et al. (1993) and Zakoian (1994), models the conditional variance as follows:

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1\varepsilon_t^2 + \beta_1\sigma_t^2 + \gamma_1\varepsilon_t^2 I(\varepsilon_t < 0), \tag{6}$$

where $I = 1$ if $\varepsilon_{t-k} < 0$ and the error term takes the leverage effect into account. A specification including the (implied) volatility index and Google Trends as additional regressors is considered,

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1\varepsilon_t^2 + \beta_1\sigma_t^2 + \gamma_1\varepsilon_t^2 I(\varepsilon_t < 0) + \delta IV_t + \psi GT_t, \tag{7}$$

The second model will be the HAR model by Corsi (2009),

$$\mathrm{RV}_{t+1} = \beta_0 + \beta_D \mathrm{RV}_t + \beta_W \mathrm{RV}_{t-5,t} + \beta_M \mathrm{RV}_{t-22,t} + \epsilon_{t+1},$$

where D,W, and M stand for daily, weekly, and monthly values of the realized volatility, respectively. The HAR model augmented with the implied volatility and Google data will also be considered:

[1]Fantazzini and Shangina (2020) used a composite volatility index ranging from January 2006 till April 2019, containing both the new RVI index and the previous RTSVX (Russian Trading System Volatility Index) which was discontinued on 12 December 2016. At the time of writing this paper, the time series for the RTSVX was no more available for free, so that I stuck to the current RVI index to make the analysis fully reproducible also for readers who have no access to the commercial database.

[2]The full description of the RVI methodology: http://fs.moex.com/files/6757

$$\mathrm{RV}_{t+1} = \beta_0 + \beta_D \mathrm{RV}_t + \beta_W \mathrm{RV}_{t-5,t} + \beta_M \mathrm{RV}_{t-22,t} + \delta \mathrm{IV}_t + \psi \mathrm{GT}_t + \epsilon_{t+1} \tag{8}$$

The third model is the *Auto-Regressive Fractional Integrated Moving Average—ARFIMA(p,d,q)*—model proposed by Andersen et al. (2003):

$$\Phi(L)(1-L)^d(\mathrm{RV}_{t+1} - \mu) = \Theta(L)\varepsilon_{t+1}, \tag{9}$$

where L is the lag operator, $\Phi(L) = 1 - \varphi_1 L - \ldots - \varphi_p L^p$, $\Theta(L) = 1 + \theta_1 L + \ldots + \theta_q L^q$ and $(1 - L)^d$ is the fractional differencing operator defined by $(1-L)^d = \sum_{k=0}^{\infty} \times \frac{\Gamma(k-d)L^k}{\Gamma(-d)\Gamma(k+1)}$ and $\Gamma(\bullet)$ is the gamma function. Similarly to the HAR and GARCH models, I will also consider IV and Google Trends as additional regressors:

$$\Phi(L)(1-L)^d(\mathrm{RV}_{t+1} - \mu) = \gamma \mathrm{GT}_t + \alpha \mathrm{IV}_t + \Theta(L)\varepsilon_{t+1}, \tag{10}$$

Finally, I also estimate the realized GARCH with a log-linear specification proposed by Hansen et al. (2012), which jointly models the returns and the realized measures of volatility:

$$\begin{aligned} r_t &= \mu + \sqrt{\sigma_t^2} \cdot z_t, \quad z_t \sim i.i.d.(0,1); \\ \log\ \sigma_t^2 &= \omega + \sum_{i=1}^{q} \gamma_i \log\ \mathrm{RV}_{t-\mathrm{i}} + \sum_{i=1}^{p} \beta_i \log\ \sigma_{t-i}^2; \\ \log\ \mathrm{RV}_t &= \xi + \psi \log\ \sigma_t^2 + \tau_1 z_t + \tau_2\left(z_t^2 - 1\right) + u_t, \quad u_t \sim i.i.d.\left(0, \sigma_u^2\right). \end{aligned} \tag{11}$$

Similarly to previous models, an augmented model with IV and Google Trends as additional regressors is considered.

3.3 Market Risk Measures and Backtesting Methods

VaR can be defined as the maximum market loss of a financial position over a time horizon h at a pre-defined confidence level (1-α), or alternatively, the minimum loss of the worst losses (α) over the time horizon (h). For GARCH and Realized-GARCH models with Student's t errors, the 1-day ahead VaR is given by $\mathrm{VaR}_{t+1,\alpha} = \widehat{\mu}_{t+1} + t_{\alpha,\upsilon}^{-1} \cdot \sqrt{(\upsilon-2)/\upsilon} \cdot \sqrt{\widehat{\sigma}_{t+1}^2}$, where $\widehat{\mu}_{t+1}$ is the 1-day-ahead forecast of the conditional mean, $\widehat{\sigma}_{t+1}^2$ is the 1-day-ahead forecast of the conditional variance, while $t_{\alpha,\upsilon}^{-1}$ is the inverse function of the Student's t distribution with υ degrees of freedom at the probability level α. For HAR and ARFIMA, the 1-day ahead VaR is computed as follows, $\mathrm{VaR}_{t+1,\alpha} = \Phi_\alpha^{-1}\sqrt{\widehat{\mathrm{RV}}_{t+1}}$, where Φ_α^{-1} is the inverse function of a standard normal distribution function at the probability level α, while $\widehat{\mathrm{RV}}_{t+1}$ is the 1-day-ahead forecast for the realized volatility.

The Expected Shortfall (ES) measures the average of the worst losses, where α is a percentile of the returns' distribution, and it is computed as follows: $\mathrm{ES}_\alpha = \frac{1}{\alpha}\int_0^\alpha F_z^{-1}(X)dz = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_z(X)dz$, where F^{-1} is the inverse function of the returns' distribution, that is the VaR. Wimmerstedt (2015) and Emmer et al. (2015) showed that the ES2.5% proposed by the Basel Committee on Banking Supervision (2013, p. 18) can be approximated using the VaR computed at different probability levels as follows:

$$\mathrm{ES}_{2.5\%} \approx \frac{1}{5}\left[\mathrm{VaR}_{2.5\%} + \mathrm{VaR}_{2.0\%} + \mathrm{VaR}_{1.5\%} + \mathrm{VaR}_{1.0\%} + \mathrm{VaR}_{0.5\%}\right]. \qquad (12)$$

The null hypothesis that the average number of VaR violations is equal to α% can be tested using the unconditional coverage test by Kupiec (1995), while the joint null hypothesis that the average number of VaR violations is correct and the violations are independent can be tested using the conditional coverage test by Christoffersen (1998).

The magnitude of the VaR violations can be evaluated by computing the asymmetric quantile loss (QL) function by González-Rivera et al. (2004), $\mathrm{QL}_{t+1,\alpha} = (\alpha - I_{t+1}(\alpha))(r_{t+1} - \mathrm{VaR}_{t+1,\alpha})$, where $I_{t+1}(\alpha) = 1$ if $r_{t+1} < \mathrm{VaR}_{t+1,\alpha}$ and zero otherwise. The losses of the competing models can be compared using the Model Confidence Set (MCS) by Hansen et al. (2011) to select the best VaR forecasting models at a specified confidence level.

Finally, the estimated VaR at different confidence levels can be jointly tested using the multinomial VaR test by Kratz et al. (2018), which implicitly backtests ES using the previous idea by Emmer et al. (2015) to approximate the ES at several VaR levels. A discussion at the textbook level of these backtesting methods and market risk management in general can be found in Fantazzini (2019).

4 Empirical Analysis

As anticipated in the previous section, this analysis will use only free resources to be fully reproducible. I consider the following data:

- *RTS index future*: intraday data sampled every 5 min is downloaded from the website finam.ru. The sample ranges from January 2015 till August 2019. The 5-min squared log-returns are then used to calculate the daily, weekly, and monthly realized variance measures. Daily returns are also computed.
- *RVI (Russian Volatility Index)*: this is the IV of the RTS index future computed from option prices.
- *Google Trends*: this is a standardized index ranging between 0 and 100 which shows the number of search queries for a topic or a keyword over a specific period

and a specific region. Its computation requires dividing the number of searches by the total amount of searches for the same period and region, and the resulting time series is then divided by its highest value and multiplied by 100. The average of the Google Trends data for the query "RTS index," both in English and in Russian is used. It is now time to introduce R to download the data and to perform the VaR backtesting analysis with competing volatility models:

```
# Load the Russian Volatility Index (RVI)
library(rusquant)
getSymbols("SPFB.RVI", from='2015-05-05', to='2019-08-09',
src="Finam", period="day")
RVI<-SPFB.RVI$SPFB.RVI.Close; colnames(RVI)<-"RVI"; rm(SPFB.RVI)
# Load RTS intraday data (max 3 years of 5-min data per single download)
getSymbols("SPFB.RTS", from="2015-01-01", to="2017-12-31",
src="Finam", period="5min"); a1=SPFB.RTS
getSymbols("SPFB.RTS", from="2018-01-01", to='2019-08-09',
src="Finam", period="5min"); a2=SPFB.RTS
dat<-rbind(a1,a2); rm(SPFB.RTS);rm(a1);rm(a2)
# Compute the daily returns and the daily RV for the RTS index
library(xts);library(highfrequency)
closep<-dat[,"SPFB.RTS.Close"]
intraday_squared_returns <- highfrequency::makeReturns(closep)^2
daily_RV <- aggregatets(intraday_squared_returns, on = 'days', k =
1, dropna = T, FUN="sum")
daily_returns <- highfrequency::makeReturns(aggregatets(closep, on =
'days', k = 1, dropna = T))
A<-cbind(daily_returns, daily_RV); colnames(A)<-c
("daily_returns","daily_RV")
rm(intraday_squared_returns); rm(dat)
# Merge the datasets
A <- merge(A,RVI, all=F)
rm(daily_returns); rm(daily_RV); rm(RVI)
# Download first Google monthly data, then daily data and finally
concatenate them
library(gtrendsR)
res_en_all <- gtrends(keyword = c("RTS index"), time = "2015-05-01
2019-07-30")
res_en_all<-xts::xts(x = res_en_all$interest_over_time$hits, order.
by = res_en_all$interest_over_time$date)
res_en_all<-xts::as.xts(aggregate(res_en_all, as.yearmon, mean))
res_ru_all <- gtrends(keyword = c("Индекс РТС"), time ="2015-05-01
2019-07-30")
res_ru_all<-xts::xts(x = res_ru_all$interest_over_time$hits, order.
by = res_ru_all$interest_over_time$date)
res_ru_all<-xts::as.xts(aggregate(res_ru_all, as.yearmon, mean))
len=length(res_ru_all)
startdate<- seq(as.Date("2015-05-01"),length=len+1,by="months")
enddate<- seq(as.Date("2015-05-01"),length=len+1,by="months")-1
  GT<-NULL
for (i in 1:len){
 daily_date<-seq(startdate[i], enddate[i+1], by="days")
 res_en <- gtrends(keyword = c("RTS index"), time = paste(startdate[i],
```

```
enddate[i+1],sep=" "))
 if (is.null(res_en$interest_over_time$hits)==FALSE){
  res_en<-res_en$interest_over_time$hits*(as.numeric(res_en_all
[i])/100)
 }else{
  res_en <-0
 }
 res_ru <- gtrends(keyword = c("Индекс РТС"), time = paste(startdate
[i],enddate[i+1],sep=" "))
 if (is.null(res_ru$interest_over_time$hits)==FALSE){
  res_ru<-res_ru$interest_over_time$hits*(as.numeric(res_ru_all
[i])/100)
 }else{
  res_ru <-0
 }
 res<-(res_en+res_ru)/2; rts<-xts::xts(x = res, order.by = daily_date)
 GT<-rbind(GT,rts)
}
# Substitute zero values in GT with small positive number
GT[GT==0] <- 0.1
# Merge the datasets
A <- merge(A,GT, all=F); rm(GT)
# Adjust the daily Realized Variance for the night market closure
A$daily.RV.adj<-(sum(A$daily_returns^2)/sum(A$daily_RV))*A
$daily_RV
```

Note that the quality of this downloaded dataset is worse than the dataset used by Fantazzini and Shangina (2020) because there are several missing values. Nevertheless, I continue working with these data to allow readers without access to commercial databases to fully reproduce this analysis. After the data download, we estimate the volatility models using a rolling window of 400 observations and then compute the VaR forecasts till the end of the available sample.

The R code below considers only the models which reached numerical convergence, whereas models which failed to converge are discarded. The R scripts *HARRV_forecast_functions.R* and *ARFIMA_LOG_forecast_functions.R* which are loaded below contains functions to estimate and forecast with HAR models and with ARFIMA models using the logarithm of the RV as dependent variable, respectively. Their full contents are reported in Appendix 1.

```
library(rugarch);library(doParallel); library(xts); library
(highfrequency); ncores=detectCores()-1
A[A==0]<-0.0000001 ### Problems with too many zeroes in the data:
substitute small pos. numbers

# 1) =================== GARCH models
=================================================
# Basic T-GARCH(1,1)
v_alpha <- c(0.005, 0.01, 0.015, 0.02, 0.025)
garch.spec = ugarchspec(variance.model = list(model = "gjrGARCH",
garchOrder=c(1,1)),
```

```
            mean.model = list(armaOrder=c(0,0), include.mean = TRUE),
            distribution.model = "std")
ctrl = list(outer.iter = 100, inner.iter = 650, tol = 1e-5)
cl<-makeCluster(ncores)
registerDoParallel(cl)
tgarch11.roll = ugarchroll(spec=garch.spec, data = A$daily_returns,
n.ahead = 1,
             n.start = 400, refit.every = 1, refit.window = "moving",
          solver = "solnp", solver.control = ctrl, calculate.VaR = TRUE,
VaR.alpha = v_alpha,
             keep.coef = FALSE, cluster=cl, window.size = 400)
stopCluster(cl)

# 2) =================== HAR models
====================================================
source('D:/Dean/Papers/Shangina/HARRV_forecast_functions.R')
results.HARRV.LOG <- HARRV.all.1step.forecast.night(dat=closep,
roll.window = 400, type="HARRV",
                   transform="log")
results.HARRVIV.LOG <- HARRV.all.1step.forecast.night(dat=closep,
roll.window = 400, type="HARRV", external =
                   lag(A$RVI), transform="log")
results.HARRVGT.LOG <- HARRV.all.1step.forecast.night(dat=closep,
roll.window = 400, type="HARRV", external =
                   lag(A$GT), transform="log")

# 3) =================== ARFIMA models
=================================================
# Basic ARFIMA (1,1)
arfima.spec<-arfimaspec(mean.model = list(armaOrder =c(1,1), include.
mean=TRUE,arfima=TRUE ))
cl<-makeCluster(ncores); registerDoParallel(cl); n.start=400
arfima.roll = arfimaroll(arfima.spec, data = A$daily.RV.adj, n.ahead = 1,
n.start = n.start,
            window.size = 400, refit.every = 1, refit.window = "moving",
            solver="hybrid",calculate.VaR=FALSE,keep.coef=FALSE,
cluster=cl)
stopCluster(cl)
RV_fore<-arfima.roll@forecast$density$Mu
RV_fore<-ifelse(RV_fore<0,min(RV_fore[RV_fore>0]),RV_fore)
RV_fore<-xts::xts(RV_fore, order.by = arfima.roll@model$index[(n.
start+1):nrow(A$daily_returns)] )
# Compute VaR
v_alpha <- c(0.005, 0.01, 0.015, 0.02, 0.025)
m <- matrix(sqrt(RV_fore),nrow=length(sqrt(RV_fore)),ncol=length
(v_alpha), byrow=FALSE)
m_VaR <- xts::xts( t(t(m) * qnorm(v_alpha)), index(RV_fore))
source('D:/Dean/Papers/Shangina/ARFIMA_LOG_forecast_functions.R')
RV.VaR.fore.IV <-ARFIMA.RV.1step.log.fore(dat.daily.RV=log(A
$daily.RV.adj), external=lag(log(A$RVI)),
                                    windowsize = 400)
RV.VaR.fore.GT <-ARFIMA.RV.1step.log.fore(dat.daily.RV=log(A
$daily.RV.adj), external=lag(log(A$GT)),
```

```
                                          windowsize = 400)
# 4) ====================== Realized-Garch
==========================================
# Basic Realized-Garch(1,1)
rgarch.spec <- ugarchspec(mean.model = list(armaOrder=c(0,0),
include.mean=TRUE),
              variance.model = list(model = 'realGARCH', garchOrder = c
(1, 1)))
cl<-makeCluster(ncores)
registerDoParallel(cl)
rg.roll = ugarchroll(rgarch.spec, data = A$daily_returns, n.ahead = 1,
           n.start = 400, refit.every = 1, refit.window = "moving",
          solver = "hybrid", calculate.VaR = TRUE, VaR.alpha = v_alpha,
          keep.coef = FALSE, cluster=cl, realizedVol = A$daily.RV.adj,
window.size = 400)
stopCluster(cl)
```

We now proceed to merge all VaR forecasts to compute the previously discussed market risk backtests:

```
# ========================== LOAD and MERGE VaR forecasts
=========================
library(MCS);library(rugarch);library(highfrequency)
tgarch_all05<-xts::xts(cbind(tgarch11.roll@forecast$VaR[,1]),
order.by =
           as.Date(rownames(tgarch11.roll@forecast$VaR)) ); colnames
(tgarch_all05)=c("TGARCH")
tgarch_all10<-xts::xts(cbind(tgarch11.roll@forecast$VaR[,2]),
order.by =
           as.Date(rownames(tgarch11.roll@forecast$VaR)) ); colnames
(tgarch_all10)=c("TGARCH")
tgarch_all15<-xts::xts(cbind(tgarch11.roll@forecast$VaR[,3]),
order.by =
           as.Date(rownames(tgarch11.roll@forecast$VaR)) ); colnames
(tgarch_all15)=c("TGARCH")
tgarch_all20<-xts::xts(cbind(tgarch11.roll@forecast$VaR[,4]),
order.by =
           as.Date(rownames(tgarch11.roll@forecast$VaR)) ); colnames
(tgarch_all20)=c("TGARCH")
tgarch_all25<-xts::xts(cbind(tgarch11.roll@forecast$VaR[,5]),
order.by =
           as.Date(rownames(tgarch11.roll@forecast$VaR)) ); colnames
(tgarch_all25)=c("TGARCH")
tgarch_realized<-xts::xts(tgarch11.roll@forecast$VaR[,6],order.by
=
            as.Date(rownames(tgarch11.roll@forecast$VaR))); colnames
(tgarch_realized)=c("realized")

HARRV.all05.LOG<- cbind(results.HARRV.LOG$m_VaR[,1],results.
HARRVIV.LOG$m_VaR[,1],
                    results.HARRVGT.LOG$m_VaR[,1]);
                 colnames(HARRV.all05.LOG)<-c("HARRV.LOG","HARRV_IV.
```

```
LOG","HARRV_GT.LOG")
HARRV.all10.LOG<- cbind(results.HARRV.LOG$m_VaR[,2],results.
HARRVIV.LOG$m_VaR[,2],
                    results.HARRVGT.LOG$m_VaR[,2]);
                 colnames(HARRV.all10.LOG)<-c("HARRV.LOG","HARRV_IV.
LOG","HARRV_GT.LOG")
HARRV.all15.LOG<- cbind(results.HARRV.LOG$m_VaR[,3],results.
HARRVIV.LOG$m_VaR[,3],
                    results.HARRVGT.LOG$m_VaR[,3]);
                 colnames(HARRV.all15.LOG)<-c("HARRV.LOG","HARRV_IV.
LOG","HARRV_GT.LOG")
HARRV.all20.LOG<- cbind(results.HARRV.LOG$m_VaR[,4],results.
HARRVIV.LOG$m_VaR[,4],
                    results.HARRVGT.LOG$m_VaR[,4]);
                 colnames(HARRV.all20.LOG)<-c("HARRV.LOG","HARRV_IV.
LOG","HARRV_GT.LOG")
HARRV.all25.LOG<- cbind(results.HARRV.LOG$m_VaR[,5],results.
HARRVIV.LOG$m_VaR[,5],
                    results.HARRVGT.LOG$m_VaR[,5]);
                 colnames(HARRV.all25.LOG)<-c("HARRV.LOG","HARRV_IV.
LOG","HARRV_GT.LOG")

arfima.all05<- cbind(m_VaR[,1], RV.VaR.fore.IV$m_VaR[,1],RV.VaR.
fore.GT$m_VaR[,1]);
          colnames(arfima.all05)=c("ARFIMA","ARFIMA_IV","ARFIMA_GT")
arfima.all10<- cbind(m_VaR[,2], RV.VaR.fore.IV$m_VaR[,2],RV.VaR.
fore.GT$m_VaR[,2]);
          colnames(arfima.all10)=c("ARFIMA","ARFIMA_IV","ARFIMA_GT")
arfima.all15<- cbind(m_VaR[,3], RV.VaR.fore.IV$m_VaR[,3],RV.VaR.
fore.GT$m_VaR[,3]);
          colnames(arfima.all15)=c("ARFIMA","ARFIMA_IV","ARFIMA_GT")
arfima.all20<- cbind(m_VaR[,4], RV.VaR.fore.IV$m_VaR[,4],RV.VaR.
fore.GT$m_VaR[,4]);
          colnames(arfima.all20)=c("ARFIMA","ARFIMA_IV","ARFIMA_GT")
arfima.all25<- cbind(m_VaR[,5], RV.VaR.fore.IV$m_VaR[,5],RV.VaR.
fore.GT$m_VaR[,5]);
          colnames(arfima.all25)=c("ARFIMA","ARFIMA_IV","ARFIMA_GT")
rg_all05<-xts::xts(cbind(rg.roll@forecast$VaR[,1]), order.by = as.
Date(rownames(rg.roll@forecast$VaR)) );
         colnames(rg_all05)=c("RG")
rg_all10<-xts::xts(cbind(rg.roll@forecast$VaR[,2]), order.by = as.
Date(rownames(rg.roll@forecast$VaR)) );
         colnames(rg_all10)=c("RG")
rg_all15<-xts::xts(cbind(rg.roll@forecast$VaR[,3]), order.by = as.
Date(rownames(rg.roll@forecast$VaR)) );
         colnames(rg_all15)=c("RG")
rg_all20<-xts::xts(cbind(rg.roll@forecast$VaR[,4]), order.by = as.
Date(rownames(rg.roll@forecast$VaR)) );
         colnames(rg_all20)=c("RG")
rg_all25<-xts::xts(cbind(rg.roll@forecast$VaR[,5]), order.by = as.
Date(rownames(rg.roll@forecast$VaR)) );
         colnames(rg_all25)=c("RG")
VaR.all.05 <- merge(tgarch_all05, arfima.all05,rg_all05, HARRV.all05.
```

```
LOG, tgarch_realized, all=F)
VaR.all.10 <- merge(tgarch_all10, arfima.all10,rg_all10, HARRV.all10.
LOG, tgarch_realized, all=F)
VaR.all.15 <- merge(tgarch_all15, arfima.all15,rg_all15, HARRV.all15.
LOG, tgarch_realized, all=F)
VaR.all.20 <- merge(tgarch_all20, arfima.all20,rg_all20, HARRV.all20.
LOG, tgarch_realized, all=F)
VaR.all.25 <- merge(tgarch_all25, arfima.all25,rg_all25, HARRV.all25.
LOG, tgarch_realized, all=F)
```

The following R code computes Kupiec's and Christoffersen's tests for all competing models with $\alpha = 0.5\%$. These two tests are computed using two alternative R functions: *VaRTest* from the rugarch package, and *BacktestVaR* from the GAS package. The latter has better numerical routines for zero violations or too many violations, as is visible below (see Table 1):

```
test_VaR_mat = NULL;test_VaR_mat2 = NULL
for (i in 1: 8){
 test_Var_RG <- rugarch::VaRTest(alpha=0.005,actual=VaR.all.05[,9],
VaR.all.05[,i])
 test_VaR_mat <- rbind(test_VaR_mat, cbind(test_Var_RG$uc.LRp,
test_Var_RG$cc.LRp,
                    100*test_Var_RG$actual.exceed/243))
 test_Var_RG2 <- GAS::BacktestVaR(alpha=0.005,data=VaR.all.05[,9],
VaR=VaR.all.05[,i])
 test_VaR_mat2 <- rbind(test_VaR_mat2, cbind(test_Var_RG2$LRuc[2],
test_Var_RG2$LRcc[2],
                test_Var_RG2$AE*0.5))
}
rownames(test_VaR_mat) = rownames(test_VaR_mat2) <-colnames(VaR.
all.05[,1:8])
colnames(test_VaR_mat) = colnames(test_VaR_mat2) <- c("p-value
UC","p-value CC", "% violations")
test_VaR_mat; test_VaR_mat2
```

The results of the Kupiec's and Christoffersen's tests are similar to those reported by Fantazzini and Shangina (2020): the TGARCH model and the models without additional regressors tend to perform better than the competitors and, importantly, they managed to reach numerical convergence in the very volatile Russian market. The computation of the Kupiec's and Christoffersen's tests for the remaining quantile levels $\alpha_2 = 1\%$, $\alpha_3 = 1.5\%$, $\alpha_4 = 2\%$ and $\alpha_5 = 2.5\%$ is left to the reader as a small exercise.

The next step is to compute the asymmetric MCS by Hansen et al. (2011) with the quantile loss by González-Rivera et al. (2004) to select the best VaR forecasting models at a specified confidence level (see Table 2):

```
# MCS
loss.VaR05 = loss.VaR10 = loss.VaR15 = loss.VaR20 = loss.VaR25 = matrix
(0,nrow = nrow(VaR.all.05)-1,ncol=8);
colnames(loss.VaR05)=colnames(loss.VaR10)=colnames(loss.VaR15)
```

```
=colnames(loss.VaR20)=colnames(loss.VaR25)=colnames(VaR.all.05
[,1:8])
for (i in 1:8){
 loss.VaR05[,i] = LossVaR(VaR.all.05[-1,9], VaR.all.05[-1,i], which =
'asymmetricLoss', type = 'normal', tau=0.005)
 loss.VaR10[,i] = LossVaR(VaR.all.10[-1,9], VaR.all.10[-1,i], which =
'asymmetricLoss', type = 'normal', tau=0.01)
 loss.VaR15[,i] = LossVaR(VaR.all.15[-1,9], VaR.all.15[-1,i], which =
'asymmetricLoss', type = 'normal', tau=0.015)
 loss.VaR20[,i] = LossVaR(VaR.all.20[-1,9], VaR.all.20[-1,i], which =
'asymmetricLoss', type = 'normal', tau=0.02)
 loss.VaR25[,i] = LossVaR(VaR.all.25[-1,9], VaR.all.25[-1,i], which =
'asymmetricLoss', type = 'normal', tau=0.025)
}
cl <- makeCluster(4);clusterEvalQ(cl, library(MCS))
MCS05<-MCSprocedure(loss.VaR05,alpha=0.15,B=5000,cl=cl,ram.
allocation=TRUE,statistic="Tmax",k=NULL)
MCS10<-MCSprocedure(loss.VaR10,alpha=0.15,B=5000,cl=cl,ram.
allocation=TRUE,statistic="Tmax",k=NULL)
MCS15<-MCSprocedure(loss.VaR15,alpha=0.15,B=5000,cl=cl,ram.
allocation=TRUE,statistic="Tmax",k=NULL)
MCS20<-MCSprocedure(loss.VaR20,alpha=0.15,B=5000,cl=cl,ram.
allocation=TRUE,statistic="Tmax",k=NULL)
MCS25<-MCSprocedure(loss.VaR25,alpha=0.15,B=5000,cl=cl,ram.
allocation=TRUE,statistic="Tmax",k=NULL)
stopCluster(cl)

MCS05
```

Table 1 Results of various models

Model type	*p*-value UC	*p*-value CC	% violations
TGARCH	0.51387733	0.7947332	0.8230453
ARFIMA	0.51387733	0.7947332	0.8230453
ARFIMA_IV	NA	NA	NA
ARFIMA_GT	NA	NA	NA
RG	0.51387733	0.7947332	0.8230453
HARRV.Log	0.17188569	0.3787542	1.2345679
HARRV_IV.Log	0.04564615	0.1268776	1.6460905
HARRV_GT.Log	0.17188569	0.3787542	1.2345679
TGARCH	0.51387733	0.7947332	0.8230453
ARFIMA	0.51387733	0.7947332	0.8230453
ARFIMA_IV	0.00000000	0.0000000	37.4485597
ARFIMA_GT	0.17188569	0.3787542	1.2345679
RG	0.51387733	0.7947332	0.8230453
HARRV.Log	0.17188569	0.3787542	1.2345679
HARRV_IV.Log	0.04564615	0.1268776	1.6460905
HARRV_GT.Log	0.17188569	0.3787542	1.2345679

Table 2 Superior set of models

	Rank_M	v_M	MCS_M	Rank_R	v_R	MCS_R	Loss
TGARCH	3	−0.79914	1.0000	1	−0.34961	1.0000	0.00057465
ARFIMA	2	−0.83326	1.0000	3	0.55399	0.9704	0.00058826
ARFIMA_GT	4	0.47721	0.9158	4	0.89691	0.9394	0.00060364
RG	1	−0.93460	1.0000	2	0.34961	0.9834	0.00058567
HARRV_LOG	7	1.17788	0.5276	7	1.59330	0.4710	0.00061794
HARRV_IV_LOG	6	0.90605	0.7394	6	1.29186	0.6946	0.00061178
HARRV_GT_LOG	5	0.59588	0.8894	5	0.95178	0.9260	0.0006040

Statistic: Tmax
Elapsed Time: Time difference of 18.07589 s
Number of eliminated models: 1

The TGARCH model showed the smallest quantile loss, while the ARFIMA model with IV was eliminated. The results for the remaining quantiles are left to the reader.

Finally, the last step of our (baseline) analysis is the computation of the multinomial VaR test by Kratz et al. (2018) to implicitly backtest the ES by approximating it with several VaR levels:

```
# Multinomial test
test_VaR_mat = NULL
for (i in 1: 8){
 test_Var_05 <- VaRTest(alpha=0.005,actual=VaR.all.05[-1,9], VaR.
all.05[-1,i])
 test_Var_10 <- VaRTest(alpha=0.01,actual=VaR.all.10[-1,9], VaR.
all.10[-1,i])
 test_Var_15 <- VaRTest(alpha=0.015,actual=VaR.all.15[-1,9], VaR.
all.15[-1,i])
 test_Var_20 <- VaRTest(alpha=0.02,actual=VaR.all.20[-1,9], VaR.
all.20[-1,i])
 test_Var_25 <- VaRTest(alpha=0.025,actual=VaR.all.25[-1,9], VaR.
all.25[-1,i])
 tv<- c(test_Var_05$actual.exceed, test_Var_10$actual.exceed,
test_Var_15$actual.exceed,
      test_Var_20$actual.exceed,test_Var_25$actual.exceed)
 test_VaR_mat <- rbind(test_VaR_mat, tv)
}
#Number of VaR violations in each cell
rownames(test_VaR_mat)<-colnames(VaR.all.05[,1:8]); test_VaR_mat
                [,1] [,2] [,3] [,4] [,5]
TGARCH            2   3   4   7   9
ARFIMA            2   3   4   6   6
ARFIMA_IV        91  93  96  96  97
ARFIMA_GT         3   4   6   8   8
RG                2   4   6   6   7
HARRV.LOG         3   6   6   8  10
HARRV_IV.LOG      4   6   6   9  10
HARRV_GT.LOG      3   5   6   7   9
test_VaR_multi = NULL
for (i in 1: 8){
 # Compute the number of violations in each cell
 n_cell<-c(test_VaR_mat[i,], 242) - c(0, test_VaR_mat[i,])
 #and test all VaR jointly using the multinomial VaR backtest by Kratz
et al. (2018)
 theo_cell <- c(v_alpha, 1) - c(0, v_alpha)
 aa=XNomial::xmonte(n_cell, theo_cell, detail=2)
 test_VaR_multi <- rbind(test_VaR_multi, aa$pLLR)
}
#P-values of the multinomial test for each forecasting model
rownames(test_VaR_multi)<-colnames(VaR.all.05[,1:8]);
test_VaR_multi
TGARCH         0.88254
ARFIMA         0.79732
ARFIMA_IV      0.00000
```

```
ARFIMA_GT      0.57569
RG              0.73951
HARRV.LOG     0.32676
HARRV_IV.LOG  0.15547
HARRV_GT.LOG  0.88254
```

The results of the multinomial test confirm the previous empirical evidence, where the null hypothesis is strongly rejected only for the ARFIMA model with the IV index. This model showed very unstable numerical estimates which resulted in extremely poor VaR forecasts.

5 A Robustness Check: Forecasting the VaR Using Hierarchical-VAR Models

Similar to Fantazzini and Shangina (2020), we now proceed to check how our results change with a multivariate model able to accommodate a large number of regressors and parameters.

More specifically, we employ the *Hierarchical Vector Autoregression* (HVAR) model estimated with the Least Absolute Shrinkage and Selection Operator (LASSO) proposed by Nicholson et al. (2018). The starting point is the following VAR model,

$$Y_t = \boldsymbol{\nu} + \sum_{l=1}^{22} \Phi^l Y_{t-l} + \boldsymbol{u}_t, \quad \boldsymbol{u}_t \sim \mathrm{WN}(\mathbf{0}, \Sigma_u), \tag{13}$$

where Y_t is a 4×1 vector containing the daily returns, the daily realized volatility, the implied volatility, and the Google data, $\boldsymbol{\nu}$ is an intercept vector, while Φ^l are the usual coefficient matrices.

This model is estimated using the following penalized least squares optimization:

$$\min_{\boldsymbol{\nu}, \Phi} \sum_{t=1}^{T} \left\| Y_t - \boldsymbol{\nu} - \sum_{l=1}^{22} \Phi^l Y_{t-l} \right\|_F^2 + \lambda(\mathcal{P}_Y(\Phi)), \tag{14}$$

where $\|A\|_F$ denotes the Frobenius norm of matrix A (that is, the elementwise 2-norm), $\lambda \geq 0$ is a penalty parameter, while $\mathcal{P}_Y(\Phi)$ is the group penalty structure on the endogenous coefficient matrices. The *elementwise penalty function* which allows every variable in every equation to have its own maximum lag was used in the estimation process (see Nicholson et al. (2018) for more details):

$$\mathcal{P}_Y(\Phi) = \sum\nolimits_{i=1}^{4} \sum\nolimits_{j=1}^{4} \sum\nolimits_{l=1}^{22} \left\| \Phi_{ij}^{l:22} \right\|_2. \tag{15}$$

```
# HVAR model VaR forecast
=======================================================
library(BigVAR); library(doParallel); library(xts)
A= as.xts(read.zoo("A.csv", sep="", header=T))
A.adj = A; A.adj$daily_RV =NULL
num.data=nrow(A.adj)
window_roll=400
n = nrow(A.adj$daily_returns) - window_roll - 1

# Prepare function to create multivariate forecasts
col_for <- function(i) {
 fcst_on<-vector('numeric')
 # Prepare the data
 data<-A.adj[i:(i+window_roll),]
 #Elementwise HVAR for data in log-returns
 try({
  ModelHVAR<-cv.BigVAR( constructModel(as.matrix(data),p=22,
                    struct="HVARELEM",gran=c(25,10), verbose=FALSE,
IC=TRUE) )
  fcst_on[1]<- max( predict(ModelHVAR, n.ahead=1)[4], 0 )
 })
 fcst_on[2]<-i
 return(fcst_on)
}

# Parallel computation setup
no_cores <- detectCores()-1
cl <- makeCluster(no_cores)
clusterExport(cl, varlist <- c("A.adj","window_roll","col_for"))
clusterEvalQ(cl, library(BigVAR))
# Small trial with 3 out-of-sample data
seqa= 1:n #1:n /(n-1):n
sh <- parLapply(cl,seqa, col_for)
stopCluster(cl)

# Organise forecasts
forecasts<-data.frame(matrix(unlist(sh),nrow=length(seqa),
byrow=T)) #length(1:n)
colnames(forecasts)<-c("RV.HVAR", "row")
forecasts<-xts::xts(forecasts, order.by= zoo::index(A
$daily_returns)[seqa+window_roll+1] )

# Compute VaR
RV_fore<-forecasts$RV.HVAR
v_alpha <- c(0.005, 0.01, 0.015, 0.02, 0.025)
m <- matrix(sqrt(RV_fore),nrow=length(sqrt(RV_fore)),ncol=length
(v_alpha), byrow=FALSE)
HVAR_VaR <- xts::xts( t(t(m) * qnorm(v_alpha)), index(RV_fore))
```

```
# Test the VaR forecasts for each quantile using Kupiec (UC) and
Chirstoffersen (CC) VaR tests
test_HVaR_mat = NULL
for (i in 1: length(v_alpha)){
 test_HVar<-VaRTest(alpha=v_alpha[i],actual=as.numeric(tail(A
$daily_returns,242)),VaR=HVAR_VaR[,i])
 test_HVaR_mat <- rbind(test_HVaR_mat, cbind(test_HVar$uc.LRp,
test_HVar$cc.LRp, 100*test_HVar$actual.exceed/242))
}
colnames(test_HVaR_mat)= c("UC pvalue", "CC pvalue", "Actual
exceed.")

test_HVaR_mat
#     UC pvalue CC pvalue  Actual exceed.
[1,] 0.5106661 0.7920842   0.8264463
[2,] 0.7179413 0.9020792   1.2396694
[3,] 0.8473260 0.9175503   1.6528926
[4,] 0.6909964 0.8637100   1.6528926
[5,] 0.9835531 0.8577689   2.4793388

# P-values of the Multinomial VaR test by Kratz et al. (2018) with
α1=0.5%,α2=1%,α3=1.5%,α4=2%, α5=2.5%
n_cell<-c(test_HVaR_mat[,3], 242) - c(0, test_HVaR_mat[,3])
#and test all VaR jointly using the multinomial VaR backtest by Kratz
et al. (2018)
theo_cell <- c(v_alpha, 1) - c(0, v_alpha)
XNomial::xmonte(n_cell, theo_cell, detail=2)

P value (LLR) = 0.71366 +/- 0.00143
1e+05  random trials
Observed:  0.8264463 0.4132231 0.4132231 0 0.8264463 239.5207
Expected Ratio:  0.005 0.005 0.005 0.005 0.005 0.975
```

In contrast to empirical evidence reported by Fantazzini and Shangina (2020), the HVAR model passes all specification tests. This should not come as a surprise, given that the time sample used for this backtesting analysis is very small and it ranges from mid-2017 till mid-2019, which was much less volatile than the (larger) sample used by Fantazzini and Shangina (2020).

6 Conclusions

This work provided a step-by-step analysis with R and Russian market data to partially replicate the analysis performed by Fantazzini and Shangina (2020) to verify whether adding Google search queries and IV from option prices to several volatility models could improve their estimated market risk measures.

Despite the fact that the dataset used in this work was much smaller than the one employed by Fantazzini and Shangina (2020) due to the limitations of freely available resources, the results reported here did not greatly differ from those reported in the original publication: the TGARCH model without regressors was able to pass the Kupiec and Christoffersen's tests for almost all quantiles, and it also reported the lowest asymmetric quantile losses. Moreover, very few models augmented with IV and Google data managed to reach numerical convergence, thus highlighting the importance of choosing a model able to withstand volatile periods and sudden market crashes, which is the typical situation for an emerging market.

It is hoped that this work can be helpful to professionals and students in finance who want to see a detailed application of backtesting techniques for market risk measurement and management, particularly in view of the Basel III agreement that will come into force on January 1, 2022.

Acknowledgments The author gratefully acknowledges financial support from the grant of the Russian Science Foundation no. 20-68-47030.

Appendix 1

I report below the two functions contained in the R scripts *HARRV_forecast_functions.R* and *ARFIMA_LOG_forecast_functions.R*, respectively. Appendix 1:

```
# ================ HAR-RV model with corrections for night returns
=======================
HARRV.all.1step.forecast.night <- function(dat, roll.window = 2000,
type="HARRV", external=NULL,
                              transform=NULL,v_alpha=c
(0.005,0.01,0.015,0.02,0.025)){
 dat_ret <- highfrequency::makeReturns(dat)
 daily_returns <- highfrequency::makeReturns(aggregatets(dat, on =
'days', k = 1, dropna = T));colnames(daily_returns)="daily_returns"
 #btc_harrv <- highfrequency::harModel(data=dat_ret,periods=c
(1,5,22),type=type, h=1,transform=transform)
 btc_harrv <- highfrequency::harModel(data=dat_ret,periods=c
(1,5,22),type=type, h=1,transform=transform,inputType = "returns")
 daily_dat<-xts::as.xts(btc_harrv$model, order.by =btc_harrv$dates)
 zoo::index(daily_dat)<-as.Date(zoo::index(daily_dat))
 # Merge daily returns and daily RV and adjust RV for night returns
 if (is.null(transform)==TRUE){
  correction<-merge(daily_returns, daily_dat$y, all=F)
  daily_dat<-(sum(correction$daily_returns^2)/sum(correction$y))
*daily_dat
 }
 if (transform=="log"){
  correction<-merge(daily_returns, exp(daily_dat$y), all=F)
  correction<-sum(correction$daily_returns^2)/sum(correction$y)
 }
 # Merge possible external data and original daily RV data
```

```
 if (!is.null(external)==TRUE){ daily_dat<- merge(daily_dat,
external, all=F) }
  names_for_eq <- colnames(daily_dat)
 formula_RV<-stats::as.formula( paste(names_for_eq[1], paste0
(names_for_eq[2:length(names_for_eq)], collapse="+"), sep = '~') )
 h=1
  prediction_recursive<-function(series){
  mod <- stats::lm(formula = formula_RV, data = series)
  date_last<-zoo::index(last(series))
  nextOb<-nrow( window(daily_dat, start=index(daily_dat)[1],
end=date_last) ) + 1
  # t+1
  fore_all<-matrix(NA, ncol = ncol(daily_dat)+1, nrow=h)
  if (is.null(transform)==TRUE){
   predicted <- max( stats::predict( mod,newdata=data.frame(daily_dat
[nextOb,] )), 0)
   realized<-zoo::coredata(daily_dat[nextOb,"y"])
  }
  if (transform=="log"){
   predicted <- correction*exp( stats::predict( mod,newdata=data.
frame(daily_dat[nextOb,] )) )
   realized<-correction*exp( zoo::coredata(daily_dat[nextOb,"y"]) )
  }
  dat_pred<-c(realized, predicted)
  names(dat_pred)=c("realized", "predicted")
  return(dat_pred)
 }
  roll.fore<-zoo::rollapply( daily_dat[1:(nrow(daily_dat)-h),],
width=roll.window, FUN=prediction_recursive, by.column=F,
align='right')
 roll.fore<-xts::xts( roll.fore, zoo::index(daily_dat)[(1+h):nrow
(daily_dat)] )
 HARRV.fore=na.omit(roll.fore$predicted)
 # Compute VaR
 m <- matrix(sqrt(HARRV.fore),nrow=length(sqrt(HARRV.fore)),
ncol=length(v_alpha),
       byrow=FALSE)
 m_VaR <- xts::xts( t(t(m) * qnorm(v_alpha)) , index(HARRV.fore))
  results <-list(roll.fore=roll.fore, m_VaR=m_VaR)
 return(results)
}
```

Appendix 2

```
# ======================== ARFIMA model with log dependent variable
========================
ARFIMA.RV.1step.log.fore <- function(dat.daily.RV, windowsize =
```

```
500, external=NULL,
                          v_alpha=c(0.005,0.01,0.015,0.02,0.025)){
 dat.arf<-dat.daily.RV[-1,]
 n.fore= nrow(dat.arf)-windowsize
 m_VaR <- matrix(NA,nrow=length(dat.arf),ncol=length(v_alpha),
byrow=FALSE)
 m_RV <- matrix(NA,nrow=length(dat.arf),ncol=1,byrow=FALSE)
 if (!is.null(external)==TRUE){
  external<-external[-1,]
 }
 for (i in 1:n.fore){
  if (!is.null(external)==TRUE){
   arfima.spec<-rugarch::arfimaspec(mean.model=list(armaOrder=c
(1,1), include.mean=TRUE, arfima=TRUE, external.regressors=as.matrix
(external[i:(i+windowsize-1),]) ) )
  } else {
   arfima.spec<-rugarch::arfimaspec(mean.model=list(armaOrder=c
(1,1), include.mean=TRUE, arfima=TRUE, external.regressors= NULL ) )
  }
  arfima.fit <- rugarch::arfimafit(arfima.spec, data = dat.arf[i:(i
+windowsize-1),], out.sample = 1, solver="hybrid")
  arfima.fcst <- rugarch::arfimaforecast(arfima.fit, n.ahead=1)
  sigma.hat <- sqrt( exp(arfima.fcst@forecast$seriesFor) )
  # Insert VaR and RV
  m_VaR[(i+windowsize),] = sigma.hat*qnorm(v_alpha)
  m_RV[(i+windowsize),] = sigma.hat^2
 }
 m_VaR <- xts::xts( m_VaR , order.by=index(dat.arf))
 m_RV <- xts::xts( m_RV , order.by=index(dat.arf))
  results <-list(m_RV=m_RV, m_VaR=m_VaR)
 return(results)
}
```

References

Acerbi, C., & Tasche, D. (2002). On the coherence of expected shortfall. *Journal of Banking and Finance, 26*(7), 1487–1503.

Agnolucci, P. (2009). Volatility in crude oil futures: A comparison of the predictive ability of GARCH and implied volatility models. *Energy Economics, 31*(2), 316–321.

Ahoniemi, K., & Lanne, M. (2013). Overnight stock returns and realized volatility. *International Journal of Forecasting, 29*(4), 592–604.

Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics, 89*(4), 701–720.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics, 61*(1), 43–76.

Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica, 71*(2), 579–625.

Artzner, P., Delbaen, F., Eber, J. M., & Heath, D. (1997). Thinking coherently. *Risk, 10*(11), 68–71.

Artzner, P., Delbaen, F., Eber, J.-M., & Heath, D. (1999). Coherent measures of risk. *Mathematical Finance, 9*(3), 203–228.

Bams, D., Blanchard, G., & Lehnert, T. (2017). Volatility measures and value-at-risk. *International Journal of Forecasting, 33*(4), 848–863.

Barndorff-Nielsen, O. E., & Shephard, N. (2004a). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics. *Econometrica, 72*(3), 885–925.

Barndorff-Nielsen, O. E., & Shephard, N. (2004b). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics, 2*(1), 1–37.

Barndorff-Nielsen, O. E., & Shephard, N. (2006). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics, 4*(1), 1–30.

Basel Committee on Banking Supervision. (2013, October). *Fundamental review of the trading book: A revised market risk framework*. Consultative Document.

Basel Committee on Banking Supervision. (2016, January). *Minimum capital requirements for market risk*. Consultative Document.

Basistha, A., Kurov, A., & Wolfe, M. (2018). *Volatility forecasting: The role of internet search activity and implied volatility*. West Virginia University working paper.

Bazhenov, T., & Fantazzini, D. (2019). Forecasting realized volatility of Russian stocks using Google trends and implied volatility. *Russian Journal of Industrial Economics, 12*(1), 79–88.

Birkelund, O. H., Haugom, E., Molnár, P., Opdal, M., & Westgaard, S. (2015). A comparison of implied and realized volatility in the Nordic power forward market. *Energy Economics, 48*, 288–294.

Busch, T., Christensen, B. J., & Nielsen, M. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics, 160*(1), 48–57.

Campos, I., Cortazar, G., & Reyes, T. (2017). Modeling and predicting oil VIX: Internet search volume versus traditional variables. *Energy Economics, 66*, 194–204.

Choi, H., & Varian, H. (2012). Predicting the present with Google trends. *Economic Record, 88*, 2–9.

Chong, J. (2004). Value at risk from econometric models and implied from currency options. *Journal of Forecasting, 23*(8), 603–620.

Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realized volatility. *Journal of Financial Economics, 50*(2), 125–150.

Christoffersen, P. (2012). *Elements of financial risk management*. Cambridge, MA: Academic Press.

Christoffersen, P., & Mazzotta, S. (2005). The accuracy of density forecasts from foreign exchange options. *Journal of Financial Econometrics, 3*(4), 578–605.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review, 39*, 841–862.

Corredor, P., & Santamaría, R. (2004). Forecasting volatility in the Spanish option market. *Applied Financial Economics, 14*(1), 1–11.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics, 7*(2), 174–196.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *Journal of Finance, 66*(5), 1461–1499.

Dimpfl, T., & Jank, S. (2016). Can internet search queries help to predict stock market volatility? *European Financial Management, 22*(2), 171–192.

Emmer, S., Kratz, M., & Tasche, D. (2015). What is the best risk measure in practice? *Journal of Risk, 18*, 31–60.

Fantazzini, D. (2019). *Quantitative finance with R and cryptocurrencies*. Amazon KDP.

Fantazzini, D., & Shangina, T. (2020). The importance of being informed: Forecasting market risk measures for the Russian RTS index future using online data and implied volatility over two decades. *Applied Econometrics*, forthcoming.

Fantazzini, D., & Toktamysova, Z. (2015). Forecasting German car sales using Google data and multivariate models. *International Journal of Production Economics, 170*, 97–135.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature, 457*(7232), 1012.

Giot, P. (2005). Implied volatility indexes and daily value at risk models. *The Journal of Derivatives, 12*(4), 54–64.

Glosten, L., Jagannathan, R., & Runke, D. (1993). Relationship between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance, 48*, 1779–1801.

Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association, 106*(494), 746–762.

Goddard, J., Kita, A., & Wang, Q. (2015). Investor attention and FX market volatility. *Journal of International Financial Markets, Institutions and Money, 38*, 79–96.

González-Rivera, G., Lee, T. H., & Mishra, S. (2004). Forecasting volatility: A reality check based on option pricing, utility function, value-at-risk, and predictive likelihood. *International Journal of Forecasting, 20*(4), 629–645.

Hamid, A., & Heiden, M. (2015). Forecasting volatility with empirical similarity and Google trends. *Journal of Economic Behavior and Organization, 117*, 62–81.

Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics, 27*(6), 877–906.

Hansen, P. R., & Lunde, A. (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics, 3*, 525–554.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*(2), 453–497.

Haugom, E., Langeland, H., Molnár, P., & Westgaard, S. (2014). Forecasting volatility of the US oil market. *Journal of Banking and Finance, 47*, 1–14.

Huang, X., & Tauchen, G. (2005). The relative contribution of jumps to total price variance. *Journal of Financial Econometrics, 3*(4), 456–499.

Jeon, J., & Taylor, J. W. (2013). Using CAViaR models with implied volatility for value-at-risk estimation. *Journal of Forecasting, 32*(1), 62–74.

Jorion, P. (2007). *Financial risk manager handbook* (Vol. Vol. 406). Hoboken, NJ: Wiley.

Kratz, M., Lok, Y. H., & McNeil, A. J. (2018). Multinomial VaR Backtests: A simple implicit approach to backtesting expected shortfall. *Journal of Banking and Finance, 88*, 393–407.

Kupiec, P. H. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives, 3*(2), 73–84.

Martens, M., & Zein, J. (2004). Predicting financial volatility: High-frequency time-series forecasts Vis-à-Vis implied volatility. *Journal of Futures Markets, 24*(11), 1005–1028.

Mayhew, S. (1995). Implied volatility. *Financial Analysts Journal, 51*(4), 8–20.

Meddahi, N. (2002). A theoretical comparison between integrated and realized volatility. *Journal of Applied Econometrics, 17*(5), 479–508.

Nicholson, W. B., Wilms, I., Bien, J., & Matteson, D. S. (2018). High dimensional forecasting via interpretable vector autoregression. *arXiv preprint arXiv:1412.5250*.

Pong, S., Shackleton, M. B., Taylor, S. J., & Xu, X. (2004). Forecasting currency volatility: A comparison of implied volatilities and AR (FI) MA models. *Journal of Banking and Finance, 28* (10), 2541–2563.

Seo, M., Lee, S., & Kim, G. (2019). Forecasting the volatility of stock market index using the hybrid models with Google domestic trends. *Fluctuation and Noise Letters, 18*(01), 1950006.

Taylor, S. J., & Xu, X. (1997). The incremental volatility information in one million foreign exchange quotations. *Journal of Empirical Finance, 4*(4), 317–340.

Vlastakis, N., & Markellos, R. N. (2012). Information demand and stock market volatility. *Journal of Banking and Finance, 36*(6), 1808–1821.

Vozlyublennaia, N. (2014). Investor attention, index performance, and return predictability. *Journal of Banking and Finance, 41*, 17–35.
Wimmerstedt, L. (2015). *Backtesting expected shortfall: The design and implementation of different backtests*. Technical report, Swedish Royal Institute of Technology.
Xu, Q., Bo, Z., Jiang, C., & Liu, Y. (2019). Does Google search index really help predicting stock market volatility? Evidence from a modified mixed data sampling model on volatility. *Knowledge-Based Systems, 166*, 170–185.
Zakoian, J. M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control, 18*(5), 931–955.

Integrated Risk Measurement System in Commercial Bank

Alexander Zhevaga and Alexei Morgunov

Abstract Integrated risk management means the comprehensive and effective management all significant risks (affecting the bank's activities) and their interrelation, including building a corporate culture of risk management and integrating risk management into strategic planning. The significant risks have big impact on the financial result of the bank, its capital, and liquidity, business reputation, their consideration is required for the assessment of banking creditworthiness and stability for regulators. In the context of economic crises and sanctions, the role of effective risk management in banks is significantly increasing, as it allows the bank to adequately distribute its capital and reserves and contributes to its stable existence in the face of uncertainty. The most significant risks in banking are credit and liquidity risks. In the banking sector, a significant methodological base has now been accumulated for assessing and managing these types of risks. The purpose of this study is to systematize the approaches to the formation of a risk management system in Russian and world practice, to assess their advantages and disadvantages, and also to formulate a list of recommendations for improving the existing system. Decision-making at management levels takes place in conditions of uncertainty in the external and internal environment, which causes partial or complete uncertainty in the final results of activities. In economics, uncertainty is understood as incompleteness or inaccuracy of information on the conditions of economic activity, including the costs and the results. The causes of uncertainty are three main factors: ignorance, randomness, and competition. In particular, the uncertainty is explained by the fact that the problems are reduced to the tasks of choosing from a certain number of alternatives, while the banks do not have full knowledge of the situation to work out the optimal solution, and do not have the resources to adequately account for all the information available to them. A measure of uncertainty is risk, i.e. the probability of occurrence of events, as a result of which unexpected losses of

A. Zhevaga (✉)
MGIMO, Moscow, Russia

A. Morgunov
National Research University Higher School of Economics, Moscow, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_10

income, property, cash, and other assets are possible. In modern banking risk management systems, procedures for influencing individual risk events or types of risk are increasingly being replaced by the organization of continuous monitoring of the bank's aggregate risk and the management of the value of various businesses of a credit institution adjusted for their inherent risk. This conceptual approach is called Integrated Risk Management (IRM). In the international banking regulation standards, the IRM logic is disclosed by the requirements of Component 2 of the Basel II and Basel III agreements (BKBN 2004, 2010), in Russian practice—Bank of Russia Ordinance No. 3624-U "On requirements for the risk and capital management system credit organization and banking group"(Bank of Russia, *On Requirements for the Risk and Capital Management System of a Credit Institution and a Banking Group*, 2015).

Keywords IRM · Credit risk · Market risk · Alm risk · Liquidity risk · Operational risk · Risk culture

JEL G21 · G24 · G32

1 Problems of IRM Implementation in Russian Banks

The introduction of IRM in Russian commercial banks faces a number of challenges related to the imperfection of corporate and strategic management systems, the lack of processes and technologies for accumulating and verifying risk information, and the insufficient resources to implement large-scale tasks. On the one hand, IRM procedures make it possible to build bank management in the context of individual lines of business and products based on determining the target ratio of their profitability and risks (risk appetite), determined by the shareholders. On the one hand, IRM procedures make it possible to build bank management for the specific business directions and products based on the target ratio of their profitability and risks (risk appetite), determined by the shareholders. The costs associated with the implementation of the IRM system may turn out to be higher than the savings from reducing the risk level if this implementation is formal and does not lead to a change in the risk culture of the bank and the harmonization of the processes of strategic development and risk management. Therefore, the introduction of IRM in bank management practice requires a systematic approach that integrates risk management, strategic and financial planning, performance management, and liquidity management. This integration should be based on the use of unified tools for managing these processes: a unified financial structure, a unified methodology of financial estimates and forecasts, and a unified information space. This section is devoted to the description of the standard of building an IRM bank system that meets these requirements. It reflects the ideas of standardizing the quality of banking, developed by the Russian banking community as part of the activities of the ARB Committee on Banking Quality Standards (BCBS 2004, 2010; Banking quality standards 2014;

Mardanov 2008) including IRM and ICAAP standards (Berger and Mester 1997; Pomorina 2015; Bondarenko and Pomorina 2016), as well as the best international practices of organizing IRM (ISO 2018; COSO 2017; FERMA 2003).

2 The Main Content and Elements of the IRM System

The integrated risk management system can be represented by the following scheme (see Fig. 1).

IRM supports the aggregation various types of risks and their connection with business processes. Risk management is based on large volumes of data and requires a modern, industrial IT infrastructure. Analysis of the accumulated data allows banks to identify risks and assess their materiality. Risk management processes should be launched for each significant risk and closely connected with business processes, while risk management is carried out by means of modeling and quantitative risk assessment. Estimates for certain types of risks are aggregated to assess the cumulative and adjusted impact both on the credit institution as a whole and on the group to which the organization is a member, while resistance to market disasters and specific crises is assessed through stress testing. The integration risks into the assessment of the effectiveness of a credit institution allows us to assess real profitability, risk appetite, and the limits of various levels to link the achievement of business goals with the goals of ensuring stability and sufficient capital to cover losses. A separate role is given to risk reporting, which allows banks to see a slice of quantitative and qualitative information. IRM can be described by sequentially determining the content of its main elements:

- Targeted;
- Resultant;
- Methodological;
- Organizational;
- Informational;
- Technological;
- Resource.

Target elements determine the desired result of the functioning of the system and, therefore, the content of all other elements. In the modern interpretation, the goal of the IRM is to optimize the value of the bank in the long run taking into account risks. Often, the term “risk-return management” is used to denote it. The resulting elements of the system are its final product, transmitted to the external environment (in our case, the bank’s internal product, which is used by all its subjects). The functioning of the IRM system are its results such as:

- Bank’s risk and capital management strategy integrated into the bank’s development strategy;
- Results of the identification and assessment of bank risks;

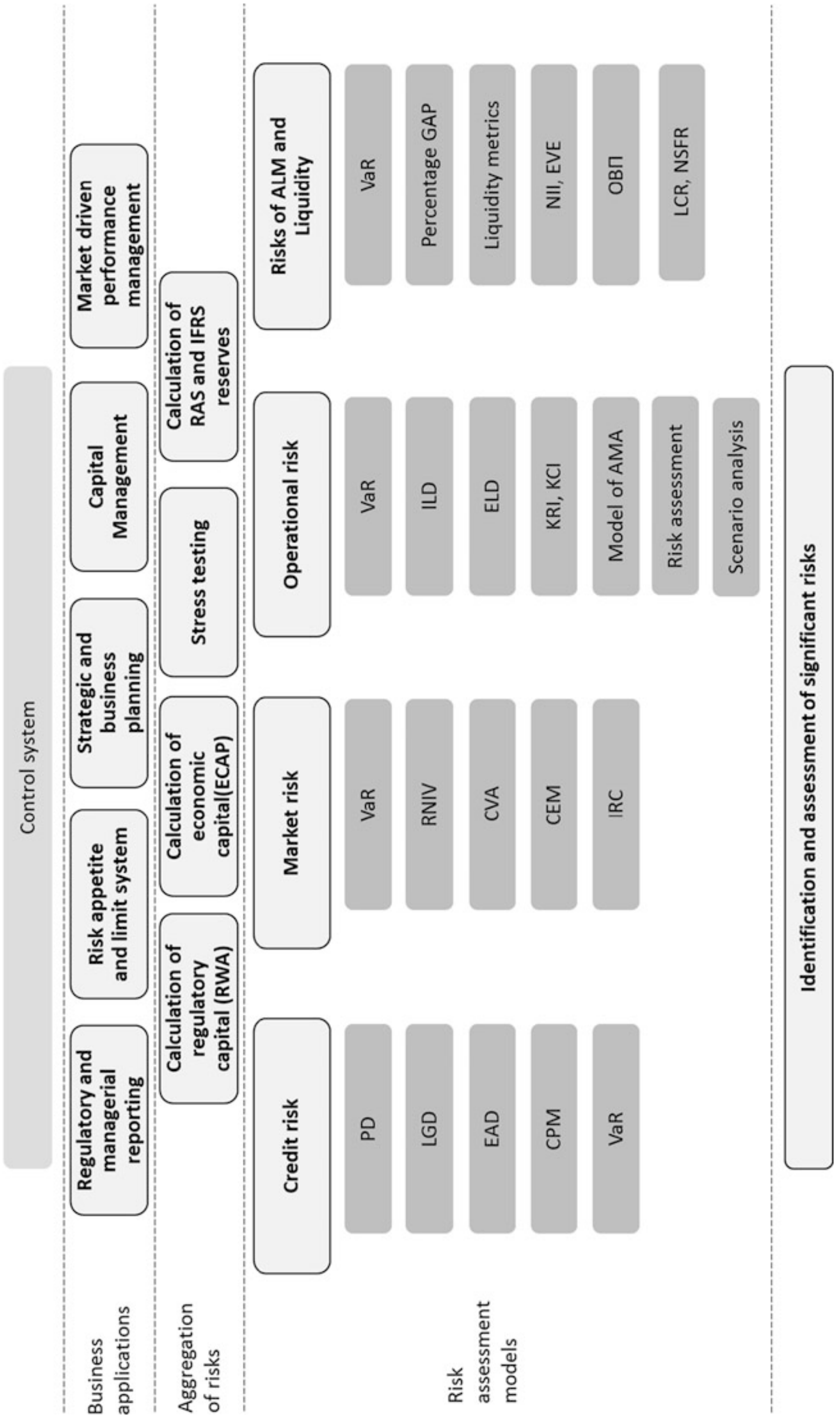


Fig. 1 The integrated risk management system

- A system to limit the level of risk;
- Management decisions regarding measures aimed at maintaining consistency between the target and actual risk levels,
- Reporting, providing control over the level of risks.

Methodological elements determine the standards and methods of risk assessment, as well as approaches to risk management.

Organizational elements determine the subjects of the IRM system, among which the following bodies and subdivisions of the bank traditionally distinguish:

- Corporate governance bodies (Board of Directors and IAS),
- Executive bodies (the Board and its committees),
- Specialized departments responsible for risk management,
- Other divisions of a credit institution, from which risks arise.

The most important element of the IRM system is its information component: data that allows banks to evaluate both realized and potential risks, as well as the technological component, which is a combination of tools for accumulating, analyzing, exchanging, and using risk information.

Resource elements determine the infrastructure of the IIR: human and technological potential that can be used for risk management purposes.

The IRM system contains subsystems for managing certain types of banking risks: credit, market, operational, liquidity risks, etc. It does not simply combine all these subsystems, but forms general principles and approaches for building management systems for the specific types of banking risks. Let us consider in more detail the content of each of the listed elements of the IRM system and their main subsystems.

3 Target Elements of the IRM System

The goals of these systems have recently changed significantly. The focus is no longer on risks, but on their impact on the value of the bank (see Fig. 2).

4 Methodological Elements of the IRM System

To ensure the effective functioning of the IRM system, it is necessary to form a continuously repeating risk management cycle based on a unified methodology.

The main stages of this cycle are the identification of risks, the determination of their quantitative and qualitative assessments, and the actual risk management and control over management effectiveness based on the assessments made.

As part of the identification:

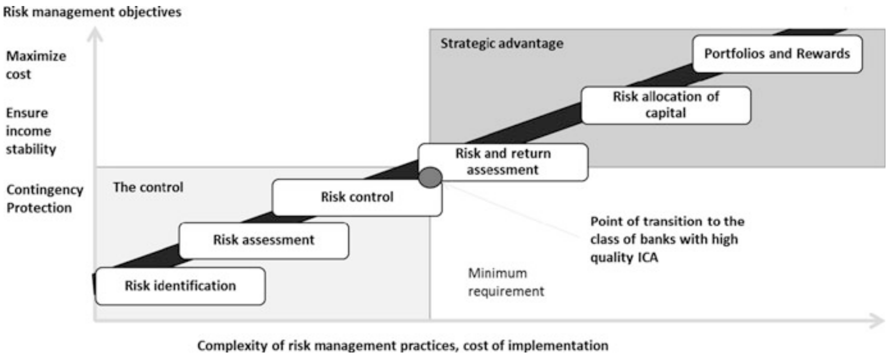


Fig. 2 Target elements of IRM system

– risk information is collected;
– fundamentally measurable and unmeasured risks are identified;
– risks are identified that require special attention and risks that can be neglected.

At the stage of risk assessment, the relationship between the balance sheet structure and risk positions is clarified. The sources of information in this case are the balances of certain types of risks. These balances systematize the company's positions in accordance with the risks associated with them. Using risk balances, an analysis of the sensitivity of the result to specific risk factors is carried out.

At the risk management stage, the following occurs:

– The development of methods, regulations, procedures. Effective risk management involves the professional selection and application of special methods and tools: statistical analysis methods, expert forecasting methods, hierarchy analysis methods, simulation methods, etc.
– Information support for decision makers. Risk decision makers need comprehensive information. Along with information on possible risks, information is also needed on the positive or negative consequences of various risk management measures. The successful implementation of this task involves solving the problem of ensuring the completeness, reliability, efficiency, and visibility of the provision of information.
– The creation of a risk reporting system. The risk reporting system serves to inform the management of the organization and the structural units of systematized data on the identification, analysis, and assessment of risks. Reporting also serves to control and monitor risks and is an important component of bank documentation, which is provided to market regulators, exchange analysts, and other market participants.

Table 1 Risk identification methods

Type of risk	Materiality	Proposed management approach
Concentration risk	Significant	Capitalization
Reputational risk	Significant	Control
Insurance risk	Insignified	Monitoring

4.1 Risk Identification Methods

An integral component of Integrated Risk Management is the procedure for determining the significant risks of a credit institution. This procedure includes:

1. The list to identify risks:
    1.1. Annual process initiation.
    1.2. Formation of a long list, including identification of emerging risks and confirmation of existing ones.
    1.3. Drawing up a work plan for the process as a whole. Here, the main focus is on default risk on loans, market and operational risks, concentration and liquidity risks, business risk, strategic and compliance risks, reputational risk, etc.
2. Materiality assessment:
    2.1. Creation and identification of scripts.
    2.2. Modeling the impact of scenarios.
    2.3. Assessment of materiality of risks and documentation of conclusions.
3. Risk management decision:
    3.1. Selection of significant risk management options.
    3.2. Reporting findings.
    3.3. Linking results to other processes.

Examples of risk identification, their assessment of significance and the development of control measures are given in the following table (Table 1).

4.2 Risk Assessment Methods and Risk Metrics

Risk analysis and assessment can be qualitative and quantitative (Altman and Saunders 1998; Dowd 2002; Allen 2003; Jorion 2007). Qualitative analysis aims to identify factors, areas, and types of risk. Quantitative analysis allows you to evaluate the value of individual risks and the value of the overall risk of the enterprise. To evaluate counterparties, the monitoring of data and publications about the counterparties (partner, client, competitor), assessing independent appraisers, analyzing materials about them or the relevant industry, region, country, etc. is carried out. A distinctive feature of these approaches is the relatively high cost and in relatively low efficiency.

In this regard, remote assessment tools that are based on current and historical data on the subject of assessment and do not imply face-to-face contacts with the analyzed subject seem to be quite important. These methods are significantly less costly, do not require expert opinion, but the cost for this is a potentially higher possibility of estimation and forecast errors. Indeed, the corresponding estimates may not take into account some factors and are probabilistic in nature. Among such tools are remote ratings, including these based on models.

Risk aggregation and the business application of integrated risk management are carried out taking into account the risk appetite of the bank. Risk appetite is the risk limit set by the governing body, within the framework of which the strategy is determined and the budget of the credit organization is formed.

The following are the goals of defining risk appetite:

1. Fulfillment of regulatory requirements for an internal assessment of capital adequacy taking into account risks.
2. Management of a holistic and structured picture of risks, consistent with the expectations of the shareholders of a credit institution.
3. Ensuring the transparency of an acceptable level of risk in business units.
4. Understanding the maximum accepted risks as part of the planning process, ensuring the implementation of a long-term strategy of a credit institution.
5. Involving stakeholders in the risk management process, such as senior management through their direct participation and external participants (shareholders, investors, analysts, etc.) through regular informing.
6. The possibility of cascading risk appetite through mechanisms of communication with lower-level limits and risk control in business processes.

Risk appetite is also a tool for managing the risk profile of a credit institution. For indicators of risk appetite, it is desirable to establish signal limits. Signal levels are necessary for effective management and early response to worsening situations. Examples of such restrictions can be: at the level of the credit institution—the target rating, as well as capital adequacy levels; at the level of the risk types for credit—economic capital of the loan and retail portfolios; at the portfolio level—NPL and EL; at the market level—the VAR limit and stop loss at the operating level—the share of operating expenses in the expenses of the credit institution, at the level ALM—the economic capital of the interest and currency risks of the bank book.

Risk appetite metrics can be divided into three categories:

1. Metrics with covenants S & P/other credit rating agencies. Here, the limits are set depending on the covenant S&P.
2. Regulatory standards. Here, the limits are set depending on the covenants of the Bank of Russia and the results of scenario planning.
3. Additional internal metrics, excluding the external factor. Limits are set based on historical data, current and planned values, and expert judgment.

An important tool for integrated risk management is the risk analysis of profitability, which can be based on economic and regulatory capital. Depending on current needs, a credit institution can use both risk-based performance indicators based on the calculation of economic and regulatory capital.

In the case where a more stringent restriction for a credit institution sets economic capital, the Economic Profit, and RAROC indicators should be maximized. The features of this approach are the following postulates:

1. The increase in profitability per unit of economic capital and ROE, in a situation where economic capital is very scarce.
2. Economic capital is more sensitive to risk (internal models of a credit organization are used, additional forms of risk manifestation, for example, concentration, are taken into account).
3. Capital is allocated to business lines, portfolios, and transactions that create the highest added shareholder value.
4. Transparent rules that allow managing a credit organization on a portfolio basis.

If regulatory capital sets a stricter restriction on a credit institution, the RoRWA indicator should be maximized. The following postulates are important:

1. Increase in profitability per unit of RWA and ROE, in a situation when Regulatory capital is very scarce.
2. Automatic implementation of regulatory restrictions and standards.
3. RWAs are less risk sensitive, especially in the absence of TAC models in a credit institution (additional forms of risk manifestation, for example, concentration, are not taken into account).
4. RoRWA optimization can lead to the formation of ineffective portfolios based on the risk/return criterion.

4.3 Planning and Stress Testing Methods

The process of business planning must be carried out taking into account risk metrics and indicators of risk return of a credit institution.

The main tasks of introducing risk metrics into the business and strategic planning process are improving the quality of planning and increasing the level of risk control. The main objectives of business planning are:

1. Determining the level of risk at the planning stage. As part of this task, it is necessary to assess the level of risks taken for the established goals and strategic objectives of the credit institution and plan measures to control and minimize risks.
2. Maximizing profitability. Maximize the achievable level of profitability by allocating resources to the most effective risk-taking business units.
3. Communication with risk appetite. The actualization of risk appetite is carried out as part of the business planning process, and the consistency of business indicators and risks is ensured.

4. Improving the quality of planning. The most qualitative forecast of the business plan of a credit institution are forecasts of reserves, arrears, losses using forecasts of macroeconomic factors.

An important tool for managing the risks and activities of a credit institution is stress testing.

Stress testing is a tool for assessing the impact of specific exceptional events, such as the financial crisis, the collapse of the securities market, etc., on the activities of a credit institution. The stress testing procedure can be carried out both from top to bottom and from bottom to top. In the first case, an upper-level analysis of the impact of stress scenarios on the performance of a credit institution is carried out. The main tool is a single model for all types of risk, which uses simplified sensitivities and correlations. As a forecasting horizon, a forecast of up to 3–5 years is used on the basis of the portfolio, which takes into account the dynamics over time and the dependence on macro factors.

In the case of applying the bottom-up stress testing procedure, a detailed analysis of the impact of stress scenarios on the performance of a credit institution is carried out. In this case, the main tools are bottom-up individual stress testing models that use regression analyzes, correlation matrices. The planning horizon is narrowed to 1 year based on a static portfolio (as of the date of stress testing).

4.4 Risk Management Methods

Risk management methods can be divided into groups:

- Obtaining additional information;
- Risk distribution;
- Risk insurance;
- Reservation of funds;
- Diversification;
- Measures of active influence (for example, incoming quality control).

Risk management is implemented as a complex process and involves a preliminary and final (post-event, posterior) analysis. This analysis allows to identify risks, to control risk limits, to improve the distribution of risks and their diversification.

The risk management philosophy is based on three main principles:

- Risk management along with the risks of individual transactions requires special attention to structural risks. As practice shows, managing specific risks is not enough to manage the risk of a credit institution, since the most important risks arise not at the level of specific operations, but at the level of the structure of the entire business system.
- Risk management focuses on the allowable loss potential of the entire bank. The potential loss is determined by all categories of risks associated with its activities.

– The level of the maximum allowable loss potential is determined by such factors as:
– The ability of the bank to take risks;
– The probability of losses;
– The need to ensure operations.

5 Organizational Elements of the IRM System

For the successful development of the IRM, it is mandatory to have a risk unit for resolving issues that are not related to a certain type of risk, providing a link between risk management and financial and strategic planning processes. The functional responsibilities of this unit include the calculation of aggregate risk indicators of the entire credit institution, including group relationships and the implementation of macroeconomic stress testing.

The fundamental factors in the formation of such a unit are the involvement of senior management and the clearly defined role of the risk function. In particular, it is mandatory that management participate in determining the risk appetite for the credit institution as a whole and for certain types of risks and business areas, including the existence of regular management risk reporting. A necessary condition for the qualitative formation of such a unit is the introduction of a model of three lines of defense, within which the functions of risk acceptance, risk management and audit are clearly separated. A clear understanding of the risks of its role as a service function aimed at creating benefits for the business, given the functional, organizational, and staff separation from the business.

The greatest role is played by the understanding of the importance of risk management both at the leadership level and at the level of performers (risk culture). The desire to use quantitative indicators in risk analysis, including the practice of making decisions based on risk analysis, becomes undeniable.

Risk culture is the established standards of employee behavior in the organization aimed at identifying and managing risks.

In credit risk management organizations, either formal procedures or informal principles and beliefs often dominate. The most successful organizations develop a risk culture in all areas.

In an ideal credit institution, a risk culture pervades the organization and defines the actions of employees, including risk-prudent business behavior, strengthening the methodological and expert functions of risk management and impact through communication/risk-based compensation.

The following tools for developing a risk culture in the organization are distinguished.

- Work with job seekers when applying for a job. The level of risk culture of employees is assessed at the stage of an interview for vacancies in a credit institution. Recruitment criteria should include risk culture issues. Activities for new employees should include topics promoting the organization's risk culture.
- Development. Training all employees in risk culture and its principles. It is advisable to conduct separate, specialized training in risk culture, depending on the role, function, and status of the employee in the organization.
- activity. Each employee should have access to the materials on risk culture. Within the framework of a credit institution, standards of risk culture in the style of "Do/Do not do" should be in place. It is mandatory to consider incidents in terms of risk culture.
- Rewards and promotions policy. Dependence of career progression incl. must steadily take into account compliance with the rules of risk culture.
- The role of leadership. Behavioral approaches that define elements of risk culture.

High-quality risk management gives a credit institution a competitive advantage, and therefore the identification and assessment of risks is the task of each employee. All employees of a credit organization strive to be professionals in risk management and for this works openly and together. Each employee of a credit institution complies with the rules, and if they are incomplete or imperfect, they openly speak about this and are guided by the interests of the credit institution.

6 Information and Technological Elements of the IRM System

Financial risks, due to their high volatility, are becoming increasingly susceptible to crises. The tools for collecting, maintaining integrity, analysing data and for forecasting use databases of financial reports, transaction results and macroeconomic indicators (Aleskerov et al. 2004).

7 Credit Risk Management System

Credit risk presents the possibility of losses due to the counterparty's failure to fulfill its contractual obligations.

The most typical manifestation of credit risk is default—the counterparty's failure to fulfill the terms of the loan agreement. The category of credit risk primarily includes losses associated with the announcement by the counterparty of default. In addition, losses associated with lowering the borrower's credit rating can also be attributed to credit risk, since this usually leads to a decrease in the market value of its obligations and losses in the form of lost profits due to early repayment of the loan by the borrower.

Credit risk includes credit concentration risks, such as country risk, industry risk, and counterparty risk. Country risk arises when it becomes impossible for the counterparty to fulfill its obligations as a result of government actions (for example, when implementing currency control measures). Country risk is primarily determined by the specifics of the country, state control, macroeconomic regulation and management. Industry risk is associated with specific market situations and relations both within the country and internationally. Counterparty credit risk can be divided into two components: risk to settlements and risk calculations.

The risk before settlements is the possibility of losses due to the counterparty's refusal to fulfill its obligations during the term of the transaction (before settlements). This type of credit risk is typical for long-time intervals: from the moment of the transaction to the settlement.

Settlement risk refers to the possibility of the non-receipt of funds at the time of settlement of the transaction due to default or lack of liquidity or operational failures. In other words, this is a risk that transactions will not be settled on time. This risk is characteristic for relatively short time intervals.

By source of manifestation, credit risk can be divided into two groups:

– External risk (counterparty risk);
– Internal risk (credit product risk).

External risk is due to the solvency or reliability of the counterparty, the likelihood of defaulting and potential losses in the event of default. The composition of the external risk includes:

– Counterparty risk—the risk of the counterparty not meeting its obligations;
– Country risk—the risk that all or most of the counterparties (including authorities) in a given country will not be able to fulfill their financial obligations for any internal reason;
– The risk of restricting the transfer of funds outside the country due to a shortage of foreign exchange reserves;
– -Concentration risk—the risk of an unbalanced distribution of funds between various industries, regions or counterparties.

Internal risk is associated with the specifics of the loan product and the possibility of losses due to the non-performance by the counterparty. The composition of internal risk includes:

– Risk of non-payment of principal and interest;
– Risk of completion of the operation—the risk of the counterparty failing to fulfill its obligations on time or late fulfillment;
– Loan security risk—of losses associated with a decrease in the market value of the loan security, the inability to enter into the right to own collateral, etc.

An important concept in assessing credit risk is a credit event. A credit event refers to a change in the borrower's creditworthiness or the credit quality of a financial instrument, the onset of which is characterized by clearly defined conditions. There are 6 main types of credit events:

1. Bankruptcy of a subject or instrument. This type of credit event may include:
   - Liquidation of the company (with the exception of mergers);
   - Insolvency (insolvency) of the company;
   - -Assignment of claims (cession);
   - Initiating bankruptcy proceedings in court;
   - Appointment of an external debtor's property manager;
   - -Seizure by a third party of the property of the debtor.
2. Early maturity of the obligation, which means a default (other than non-payment of the due amount) for any other similar obligation of the borrower and the entry into force of the reservation on the early maturity of this obligation.
3. Default on the obligation (cross-default), which means the declaration of default (other than non-payment of the due amount) for any other similar obligation of the borrower.
4. Insolvency, which implies non-payment by the borrower of a certain (exceeding the agreed limit) amount on time (after the expiration of the agreed grace period).
5. A moratorium in which the counterparty refuses to make a payment or disputes the legal force of the obligation.
6. Debt restructuring which entailed a unilateral refusal, deferral or change of the debt repayment schedule on less favorable terms for the lender.

   The following facts can also be recognized as a credit event:

7. Downgrade or recall by the rating agency of the borrower's credit rating;
8. Currency inconvertibility caused by state restrictions;
9. Actions of state bodies jeopardizing the legal force of the obligation; war or hostilities that impede the government or the banking system.

The credit risk management system in the bank is formed on the following key principles of formation:

- Independence of decision-making. Organizational independence of risk management departments and direct reporting of the head of these departments to the management of the company.
- Representation in specialized committees. Representation of heads of risk management units on all relevant committees of the bank, which are competent to accept credit risk.
- Systematic credit risk management. Using a systematic approach to risk management of both the loan portfolio as whole and individual transactions with specific borrowers (a group of related borrowers).
- Integration in the lending process. Mandatory availability of an independent risk assessment of all operations bearing credit risk.
- Adequacy of credit risk management methods. Application of an adequate methodology to the scale of operations to identify and quantify credit risk.
- Granting authority to limit risk. The head of the risk management departments has the authority to promptly suspend the limits on counterparties and credit organizations and limits on transactions with securities.

– Unity of approaches to credit risk management. The credit process in the head office, branches, and subsidiaries is based on common approaches, principles, and regulatory documents of the bank.
– Using a delegation of authority system. It includes a balanced combination of centralized and decentralized decision-making in transactions involving the adoption of credit risk.
– Reliability and independent evaluation. Independence and objectivity is ensured by its obligatory coordination with representatives of risk management divisions.

A significant role in the development of a credit institution and in the management of its risks is played by the credit policy, which is the program and direction in the provision of loans to legal entities and individuals. The credit policy is based on a risk-return ratio of operations acceptable for a credit institution.

The main objective of the credit policy is to maximize profits with minimal risk. Based on the possible correlation of these components, as well as available resources, the credit institution determines the current tasks: areas of lending, technology for carrying out credit operations and control in the lending process.

Credit policy should be reviewed depending on changing economic conditions. Credit risk management is carried out as part of an integrated risk analysis, management and control system, which includes a combination of qualitative (expert) and quantitative (statistical) assessment of credit risk. Credit risk assessment is carried out on the basis of individual (examination of individual transactions) and portfolio (assessment of risk concentrations) approaches. Credit risk management is carried out at all stages of the lending process from the moment the client's application for the provision of borrowed funds is examined until the full repayment of the obligations. The main elements of a credit risk management system at the level of individual transactions are:

– An independent comprehensive examination of credit risk;
– Analysis of the forecast cash flow of the borrower;
– Assessment of the business reputation of the counterparty;
– Monitoring the level of accepted credit risk;
– Assessment of the need to include borrowers in the register of counterparties subject to special supervision by the bank;
– System of limits for accepting credit risk.

The main elements of the credit risk management system at the level of the loan portfolio (certain areas of lending) are:

– Minimal level of internal rating, below which operations are not allowed;
– Loan portfolio quality indicators;
– Minimal discount rates used in assessing the effectiveness of projects;
– Minimal collateral discounts used in assessing the adequacy of collateral;
– Standard parameters of lending programs and limits for self-acceptance of credit risks;
– Arguments and restrictions in the field of lending to borrowers of certain sectors or areas of lending.

An important element of the credit risk management system is monitoring, which allows you to identify in advance an increased level of credit risk in the early stages of its occurrence and quickly implement measures to minimize and limit it. The main tools for monitoring:

- system of limits for accepting credit risk;
- Control conditions that must be met before the transaction and additional conditions that must be met within a specified period after the transaction.

Monitoring of the financial position of the counterparty is carried out by credit units with subsequent monitoring of the results. Monitoring includes assessing the financial position of the borrower based on official financial statements, cash flow forecasts, and other information characterizing the current and future solvency of the borrower. Along with the expert opinion included in the file of the borrower, the monitoring results are recorded in the form of an internal rating of the borrower, the category of which characterizes the level of accepted risk. In turn, the internal rating affects the amount of reserves for the transaction and the need to take additional measures to monitor the transaction and minimize the risks taken. In order to limit the bank's operations with counterparties having a dubious business reputation, the counterparty's business reputation is monitored. In addition, the bank may maintain a register of counterparties subject to special monitoring (the so-called Watch List). The criteria for inclusion of counterparties in this registry may be:

- The presence of any negative (financial and non-financial) information received from open or other sources of information that calls into question the ability of the counterparty to timely fulfill its obligations;
- The presence of overdue obligations;
- Restructuring of obligations;
- The occurrence of debt as a result of repayment of debt on a pre-existing asset;
- Loss of part of collateral.
- Default of the counterparty;
- Initiation of bankruptcy proceedings against the debtor;
- Adoption of measures by the third parties regarding the debtor to take over the business or reorganization actions;
- Repeated failure to submit reports and other documentation required by agreements, poor-quality preparation of necessary documents, and similar violations of obligations;
- The seizure or adoption of other restrictive measures in respect of the property of the borrower in favor of third parties;
- Actions to withdraw the borrower's assets without prior approval from the bank;
- Identification of facts of obtaining false or incomplete information at the stage of issuing a loan.

The result of a qualitative assessment of credit risk is the preparation of expert opinions on the acceptability of the requested transaction parameters, the required measures to minimize the accepted credit risks and the compliance of the requested form and the purpose of the transaction to finance the cash flow model. A qualitative assessment of credit risk is usually carried out in the context of the following groups of transactions:

- Current and investment financing;
- Project financing;
- Transactions with financial institutions;
- Transactions with administrations;
- Transactions with individuals;
- Operations in financial markets.

A Qualitative assessment of credit risk allows you to:

- Structure the loan transaction in accordance with the individual characteristics of the borrower's business and the forecast of its cash flow;
- Evaluate the sufficiency and validity of the sources of repayments of obligations available to the borrower;
- Identify risks inherent in the activities of the borrower and develop measures to minimize them;
- Evaluate the appropriateness of the availability and sufficiency of the security accepted for the transaction;
- Establish pricing conditions adequate for the level of accepted credit risk.

The results of a qualitative assessment of credit risk are usually presented in the form of a report by an expert unit, which is mandatory to be included in the materials submitted to the authorized bodies of the bank when considering issues of accepting credit risk. A quantitative assessment of credit risk complements the qualitative one and allows you to get a quantitative expression of the credit risk accepted by the bank for individual transactions and the loan portfolio as a whole. A tool for quantitative assessment of credit risk is the mathematical apparatus, which includes various approaches to modeling risk events, in particular:

- Econometric models allow based on regression analysis (in particular, binary and multiple choice models. These models are used to predict the probability of default and ratings as a function of several independent variables). They allow you to get estimates of the probability of an event (for example, default, with the sufficiency of the available statistics of defaults) and ratings;
- Neural networks—computer algorithms that simulate the work of the human brain through interconnected neurons. The neural networks use the same input data as with the econometric approach, and the relationships between them are highlighted by repeated repetition by trial and error;
- Optimization models based on mathematical programming methods that allow you to minimize lender errors and maximize profits, taking into account various restrictions. Using mathematical programming methods, it is possible to determine, in particular, the optimal parameters of credit products;
- Expert models used to simulate the risk assessment process carried out by an experienced and qualified specialist (models reproducing the work of credit

experts, including ratings of international rating agencies, used for low-default portfolio of borrowers);

– Hybrid models that use statistical estimation and simulation and can be based on cause-effect relationships (for example, if there is insufficient default statistics, new defaults can be modeled and used to build econometric models);
– Simulation models—allow you to determine the risk characteristics of borrowers for individual borrowers and transactions based on a scenario analysis of the borrower's cash flows—generating a scenario distribution of the project's cash flow based on risk factors relevant to the borrower.

An integral part of the quantitative assessment is the classification of the assets of the banking book. The banking book is the assets classified as "corporate," "sovereign," "banking," "retail" or "participation" in accordance with the requirements of Section III of the Basel Agreement. The Bank Book does not include assets that meet the criteria of the trading book (according to the requirements of the Basel Agreement). Classification objective: to determine the classification algorithms for the assets necessary to highlight the individual components of credit risk used in the calculation of expected and unexpected losses. Five classes are distinguished in the Bank Book Assets: "Corporate Assets," "Retail Assets," "Banking Assets," "Sovereign Assets," "Participation." Within these classes of assets, risk segmentation of borrowers is carried out: separate risk segments are distinguished, characterized by a single list of indicators that affect the level of credit risk of these counterparties. In particular, examples of risk segments in corporate assets include: "Largest and largest corporate assets," "Medium and other corporate assets," "Project finance," Income-generating real estate,""Commodity financing,""High-risk commercial real estate." As part of the credit risk management system, all models must undergo validation and internal audit procedures. The purpose of these events is to improve the quality, visibility, and interpretability of the developed models and to reduce model risks arising during the development. To form common standards, banks formulate methods for the development and validation of models, covering the specifics of models developed by banks. Validation and internal audit of models should be carried out at least once a year in order to assess the quality of existing models on relevant data, as well as take into account the conformity of the models used to current business processes and business strategies of the bank. Validation of models and internal audit of business processes in the bank can be divided into "deep" (as part of the development of new models) and "periodic" (as part of the verification of existing models in the bank)

8 Liquidity Risk Management System

Specialists in the field of risk management do not have unity in approaches to determining the liquidity risk of a credit institution. Some believe that the liquidity risk is the risk of losses resulting from the bank's inability to meet its obligations at the expense of the funds at its disposal due to the unbalanced timing and volume of future incoming and outgoing cash flows.

Another group of specialists determines the liquidity risk as the risk of insufficient (or negative) liquidity:

– Lack of assets for timely fulfillment of obligations;
– The impossibility of a quick conversion of financial assets into means of payment without significant losses;
– Losses due to the need for a quick conversion of financial assets;
– Change in net income and market value of shares.
– There is a known classification of liquidity risks in terms of excess or shortage of cash or highly liquid assets:
– Excess liquidity risk—the risk of losses resulting from a decrease in bank profitability due to an imbalance in the timing and volume of future incoming and outgoing cash flows (Cash flow);
– Insufficient liquidity risk—the risk of default due to the lack of cash or other highly liquid assets (this risk seems to be significantly more dangerous for the financial stability of the bank).

For banks, compliance with liquidity at any given time is one of the primary goals, as they live off the trust of customers. Therefore, the exclusion or significant limitation of liquidity risks is the central task of banking risk management.

The tasks of managing short-term liquidity risks of a credit institution include:

– Determination of the net outflow of funds based on historically observable statistical data (statistical analysis and valuation), as well as by analyzing the status of all accounts with the Bank of Russia and cash positions at the beginning and end of the day.

Liquidity Calculation at Risk (LAR)—the expected excess of payments (Net need for financing) for a certain period of time, which is with a given probability (95%—under normal financial load; 99%—with increased load; 99.9%—with maximum load) will not be implemented.

– Optimization of liquidity reserves, which consists in classifying the potential of the assets at the bank's disposal in terms of their ability to turn into liquid assets and contrasting the potential with the risks arising from net cash outflows as a result of external factors.

The classification of liquidity risks typical of a credit institution can be carried out as follows:

– Refinancing risks arise as a result of the transformation of the terms, which is carried out in order to obtain profitability through the formation of a normal interest structure (interest on long-term investments should be greater than on short-term attraction). With repeated refinancing, there is a danger that funds cannot be raised at all to close long-term positions or they will be very expensive.

– The risks of an unplanned extension of the capital binding period lead to the fact that the debt and interest on the debt return more slowly than planned.
– The risk of unexpected withdrawal of deposits from the accounts is the risk that the agreed loan is unexpected, that is, earlier than the scheduled term, is claimed, or deposits are withdrawn before the agreed term. This type of risk is typical for large banking transactions.

Professional liquidity management of a credit institution involves the structuring of measures to guarantee liquidity.

The above measures are aimed primarily at managing the balance sheet structure and are a long-term oriented structural liquidity management. But in order to ensure sustainable solvency, operational liquidity management is necessary, in which the movement of specific means of payment is analyzed.

In liquidity risk management, instrumental and organizational aspects should be distinguished. Naturally, without the systematization of the instrumental component, it is impossible to create an effective risk management process. However, an equally important aspect of the successful functioning of the process and the liquidity risk management system is its organizational and cultural component. The latter includes such essential components as: risk culture, decision-making culture in a conflict situation, that is, in the presence of opposing alternative solutions, as well as risk management methodology and approaches.

9 Market Risk Management System

Market risk is the risk of losses resulting from adverse changes in market risk factors. Market risks are associated with the uncertainty of market fluctuations—price and exchange rate (currency) risks, interest rate risks, liquidity—and sensitivity to these fluctuations of risk-bearing objects (for example, assets). Market risks are sometimes called technical risks in association with technical analysis used to study and forecast prices, rates, volumes, and other indicators related to the market. Not only direct price factors are sources of market risks. For example, the correlation between the returns of various instruments is not a direct price factor, but indirectly affects the price characteristics of a portfolio containing these instruments.

Classification of market risks allows you to clearly structure the problems and affects the analysis of situations and the choice of effective management. The classification of market risks should correspond to the specific goals of each study and be carried out from the perspective of a systematic approach. Based on these principles, we can distinguish the most widely used classification of market risks by market segments:

- Interest rate risk (risk of losses on positions in debt securities and other instruments sensitive to changes in interest rates);
- Currency risk (risk of fluctuations in the value of positions in foreign currencies);

- Stock risk (risk of fluctuations in the value of positions in shares and their derivatives);
- Market risk of derivative financial instruments (risk of a decrease in the value of derivative financial instruments (options, futures contracts and others);
- Commodity risk (the risk of fluctuations in the value of positions under contracts for goods).

Each of the above types of risk is affected to one degree or another by the risk of market liquidity, which is associated with losses that a participant may suffer due to insufficient market liquidity. A measure of market liquidity risk is the realized spread—the difference between the weighted average prices of transactions for a certain period of time, committed at the bid price, and transactions, made at the bid price. Calculating this value is quite problematic.

As the problem is examined, the types of risks associated with a particular aspect of the problem or parameter are often introduced: for example, the risk associated with the possibility of a parallel shift in the interest rate curve; risk associated with changes in financial results due to currency fluctuations and others.

As methods of managing market risk, the approaches most often used are those associated.

The risk limitation system may be as follows:

- VAR—the value of the possible (with probability %) maximum depreciation of the trading portfolio on the horizon of T days;
- DV01—the value of the possible depreciation of the trading portfolio when the rates change by 1 bp (sometimes 1 p.p. is used);
- CS01—the value of possible impairment of the trading portfolio when the credit spread changes by 1 bp (sometimes 1 p.p. is used);
- Stop Loss—the amount of the maximum allowable loss for a financial instrument/portfolio. Upon reaching the specified limit, the position in the financial instrument is closed in whole or in part;
- Max Loss—the amount of the maximum allowable loss on the portfolio. After its achievement, trading in the portfolio is suspended and the question of further plans is reviewed by the management of the company together with the shareholders or the Board of Directors;
- The maximum allowable amount of open positions in financial instruments within the portfolio;

Limitations on the period of holding securities in a portfolio.

The structure of limits for transactions with derivative financial instruments (hereinafter—the derivatives) its own characteristics; the following types of limits are distinguished for such transactions:

- Maximum portfolio volume.
- Restriction on types of underlying assets.
- Limitation on types of derivatives.
- Limitation on the urgency of transactions.
- Restriction on the Greeks.

– Restriction on currency risk.
– Limit on interest rate risk.
– Limitation on the delta hedged position.
– The Stop-loss indicator shows at what negative difference between the position and the delta hedge the trader is obliged to take actions for additional hedging.
– Max Loss—the maximum allowable loss on the portfolio.
– Limit on negative or positive Fair Value.

The main approaches to assessing the cost-based measure of market risk for the value of the possible maximum depreciation of the trading portfolio on the horizon of T days:

– Delta-normal approach (taking into account the log-normality of the distribution of return on assets);
– The method of historical modeling (based on a complete revaluation of the current portfolio at market prices modeled on the basis of historical scenarios, that is, the method is based on the assumption that the behavior of market prices is stationary in the near future);
– Monte Carlo method (based on modeling random processes with given characteristics).

Approaches to assessing market risk, as well as credit, should undergo annual procedures for periodic validation and internal audit.

10 Operational Risk Management System

Operational risk (as defined by the Basel Committee on Banking Supervision) is the risk of losses caused by inadequate or erroneous internal processes, employee actions, systems, or the influence of external events. Includes legal risk, but excludes reputational and strategic risks.

At the same time, according to the definition of the Bank of Russia, operational risk is the risk of losses resulting from unreliability and inadequacies in the internal management procedures of a credit institution, failure of information and other systems or due to the impact of external events on the activities of a credit institution.

Sources of operational risk are people, systems, processes, external influences.

Operational risk management principles:

• Compliance with legislation.
• Anti-corruption policy.
• Protection of information.
• Countering internal fraud.
• Ban on concealing information on facts/threats of loss.
• Risk analysis when creating new/changing existing products.
• Separation of powers, prevention of conflicts of interest.

- Prohibition of transactions on customer accounts in the absence of an appropriate order.
- Acquisition of goods /works/services on a competitive basis.
- Documentation of business processes and control procedures.
- Availability of plans to ensure the continuity and restoration of the credit institution.

An integral part of operational risk is a risk event. A risk event is an event that has occurred due to operational risk, which has caused or is likely to lead to operational losses of the bank and has occurred due to erroneous or faulty banking processes, actions of people and systems, as well as due to external events—realization of risk (threat).

The consequences of risk events can carry both a financial and non-financial component. The former include both the past and possible (including expected/forecasted) financial consequences (except for the lost profit), and the latter—both past and possible (including expected/forecasted) negative consequences of a non-financial nature, including lost profits.

The difference between risk and risk event is presented in the table (Table 2).

The registration of data on events caused by operational risk that entailed or could result in losses in an amount exceeding the established cut-off level and their consequences is carried out by the unit that revealed the risk event or the unit-owner of the risk.

An important component of operational risk is a key indicator of operational risk. A key indicator of operational risk is a quantitative indicator that allows monitoring the level of risk and the effectiveness of control procedures aimed at minimizing risk. For key indicators of operational risk, threshold risk levels are established, which serve to determine the level of risk, the measured indicator.

The main advantages of monitoring the level of risk with the help of key indicators of operational risk:

Table 2 Risk vs Risk event

Risk	Risk event
Risk (threat)—opportunity, uncertainty, perspective	Risk event—specific fact, past event
Risk (threat) may exist but not be realized Risk (threat) can be realized by several risk events of various types	A risk event is the realization of risk (threat) A risk event, in turn, can have consequences: realized, not realized—expected, possible, uncertain.
Risk (threat) may or may not be tied to a specific time and place	A risk event has a specific time and place (even if they are unknown)
Risk minimization—these are measures related to preventing the implementation of relevant events/reducing their adverse effects in the future	Minimizing the consequences of a risky event are measures related to the settlement of the consequences of a specific adverse event

- The ability to effectively localize problem areas (departments, systems, employees) for their further in-depth analysis and organize risk management depending on the dynamics of indicator values;
- The ability to take risk management measures in a proactive mode;
- Absence of discrepancies in assessments of the significance of risk.

Key indicators of operational risk are most effective for use:

- For monitoring and forecasting the level of risk, which is easily measured by quantitative indicators and has an "audit trail";
- To assess the effectiveness of control procedures (at the level of business processes);
- To identify areas of increased attention—"hotspots";
- For use in establishing risk-based KPIs.

At the same time, key indicators of operational risk may not be effective enough, for example, to monitor rare events such as the Black Swan, to measure risks associated with the distribution of powers, including a conflict of interest and to measure specific risks associated with the provision of complex non-mass services.

The life cycle of a key indicator of operational risk is presented in the figure (Fig. 3).

One of the main tasks of the operational risk management system is to ensure the continuity of the credit organization. Ensuring continuity is a set of organizational, technical and software events to minimize bank losses in case of emergency situations. It is based on the action plans in case of emergency situations, including the following sections:

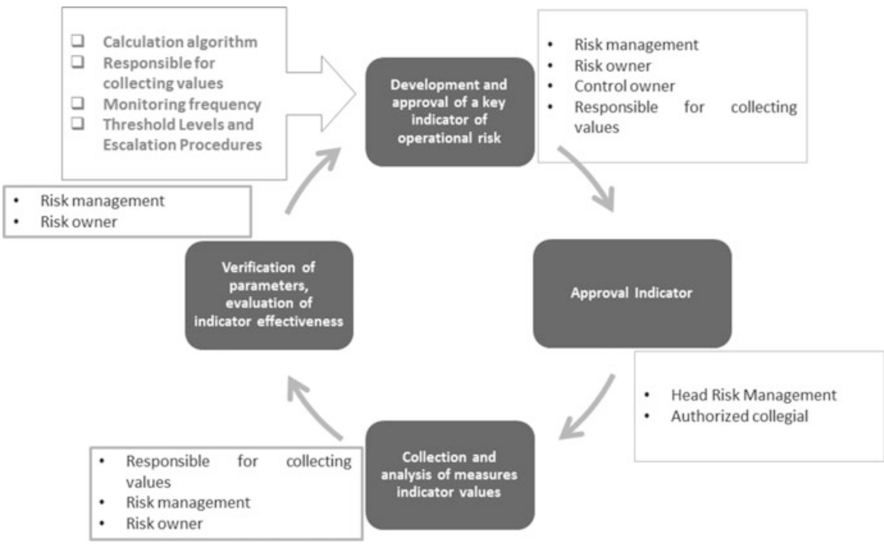


Fig. 3 The life cycle of a key indicator of operational risk

- General provisions;
- Necessary resources;
- Staff and external services;
- Reserves;
- Criteria for identifying a problem situation;
- Procedure for notification of a possible situation;
- Sequence of actions to restore activity;
- Test plan;
- The order of introduction and updating.

Important elements of a system for ensuring business continuity are:

- Identification of threats (risk factors) significant for the continuity of a credit institution.
- Formation of scenarios for the implementation of continuity threats.
- Analysis of the impact of downtime on the business of a credit institution, determination of recovery targets.
- Developing response strategies for the implementation of the scenario.
- Implementing activities to ensure the feasibility of strategies.
- Maintaining the strategy in readiness for execution.
- Monitoring the situation for signs of emergency situations.

11 Conclusion

A significant factor in the sustainable development of a credit institution is risk management as one of the key requirements of corporate governance. This is an integral management system in the face of uncertainty in production and economic situations. Integrated risk management is focused on reducing risks associated with the variability of the external environment and internal conditions of the credit institution. The main emphasis is on forecasting trends and using appropriate forecasts when making planning decisions that take into account the dynamics of markets, as well as on the use of monitoring results and forecasts.

Management of certain basic risks of a credit institution (credit, liquidity, market, operational) does not ensure the stability of its development and financial stability in the long term. Only the integrated risk management system described in this Paper will allow a credit institution to form adaptive strategies that quickly respond to continuous technological changes, the impact of new crisis factors, and the tightening of global competition, as well as to ensure the stability of its development and financial stability in the long term. It is the integrated risk management system that forms such approaches to Bank risk management that allow us to take into account the widest possible range of risks and their interaction, long-term aspects of their impacts and constantly changing forms of their manifestation.

References

Aleskerov, F., Ersel, H., & Yolalan, R. (2004). Multicriterial ranking approach for evaluating bank branch performance. *International Journal of Information Technology and Decision Making, 3* (2), 321–335.

Allen, S. (2003). *Financial risk management: A practitioner's guide to managing market and credit risk*. Hoboken, NJ: John Wiley & Sons, Inc.

Altman, E. I., & Saunders, A. (1998). Credit risk measurements: Developments over the last 20 years. *Journal of Banking and Finance, 21*, 1721–1742.

Bank of Russia. (2015). *On Requirements for the Risk and Capital Management System of a Credit Institution and a Banking Group*. Bank of Russia Directive No. 3624 U - www.cbr.ru, 15.04.2015, pp. 1–28.

Banking quality standards. Key Points and Requirements (Версия 2014). Association of Russian Banks - https://arb.ru/arb/bureaux-and-committees/29634/documents/, 20.02.14, CC. 1-21.

BCBS. (2004). *Basel II: International convergence of capital measurement and capital standards: A Revised Framework*. Basel Committee on Banking Supervision - www.bis.org, 10 June 2004, pp. 1–251.

BCBS. (2010). *Basel III: A global regulatory framework for more resilient banks and banking systems*. Basel Committee on Banking Supervision - www.bis.org, 16 Dec 2010, pp. 1–77

Berger, A., & Mester, L. (1997). Inside the Black Box: What explains differences in the efficiencies of financial institutions? *Journal of Banking and Finance, 21*, 895–947.

Bondarenko, D. V., & Pomorina, M. A. (2016). Quality standard for integrated risk management and organization of internal procedures for assessing capital adequacy in banks. *Money and Credit 1*, CC.61–68.

COSO. (2017). *Enterprise Risk Management*. http://www.coso.org/

Dowd, K. (2002). *Measuring market risk Hoboken*. John Wiley & Sons Ltd.

Enhancing Corporate Governance for banking organizations. Basel Committee on Banking Supervision. BIS Publication. February 2006. Basel. URL: http://www.bis.org/publ/bcbs122.pdf

FERMA. (2003). *Risk management standards.* https://www.aoosk.ru/about/vnutrenniy-kontrol-upravlenie-riskami/standart%20ferma_russia.pdf, pp. 1–16.

International regulatory framework for banks (Basel III). URL:http://www.bis.org/bcbs/basel3.htm?m=3%7C14%7C572

Jorion, P. (2007). *Financial risk manager handbook* (4th ed., p. 736). Wiley, 4.8. Merton, R. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance*, *29*(2).

Mardanov, R. Kh. (2008). On the development of conceptual approaches to standardization of the quality of banking. *Money and Credit*, *2*, C. 8–17.

Pomorina, M.A. (2015). *About the draft quality standard for the organization of IRM and ICAAP in banks*. XI Scientific and Practical Conference "Banks and Financial Organizations: Processes. Standards Quality" http://npk.akforb.ru/upload/doc/2015/Поморина.pdf

The international standard Risk Management – Guide ISO 31000:2018. https://pqm-online.com/assets/files/pubs/translations/std/iso-31000-2018-(rus).pdf

Economic Capital Structure and Banking Financial Risks Aggregation

Marina Pomorina

Abstract Banks must maintain a balance between their own capital and the level of accepted aggregate risk to ensure financial stability. This paradigm is expressed in terms of capital adequacy requirements to both the minimum capital required to cover regulatory risks and the risk capital required to fully cover bank's total risk (economic capital). Therefore, the Basel Committee on Banking Supervision requires banks to implement ICAAP procedures to ensure regular risk assessment and maintain a sufficient level of capital. The Basel Committee on Banking Supervision regularly analyzes the implementation of ICAAP by global systemically important banks (G-SIB). Following the results of the analysis, the Committee has identified a number of relevant development areas: selection of approach to aggregate different material risks, detection and allocation of risk capital taking into consideration the effect of diversification, and setting limits as a function of capital allocation by activities and types of risks. This section offers a solution to the problem. It presents a conceptual approach to determining economic capital structure, which is based on material risk identification and on the determination among them of financial risks, assessed using quantitative methods. We propose a simulation model of the bank's economic capital where the total risk is presented as a composition of the products of the material risk's factors on the P&L elements exposed to these risks. Thus, the elements of the P&L define the weights for the material risk's distributions in the economic capital model. The economic capital model makes it possible to assess the distribution of the bank's total risk at different management levels (products—departments—total bank), disaggregate the available capital by products, business lines, and types of risks and, on this basis, establish limits based on the distribution of capital in accordance with the Pillar-2 requirements of Basel II.

Keywords Internal capital adequacy · Basel II · Basel III · Total risk · IRM system · Allocation of capital · Economic capital structure · Risks aggregation · RORAC · Corporate governance integration · Simulation methods

---

M. Pomorina (✉)
National Research University Higher School of Economics, Sberbank, Moscow, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_11

JEL C53 · D81 · G21 · G32 · G38

1 Internal Capital Adequacy Assessment Processes in the IRM System

The IRM system is aimed at the comprehensive management of all banking risks. One of its most important functions is to determine the level of materiality of the impact of each bank risk on the activities of a bank, to assess the correlations of such impacts and to make decisions about the acceptable level of both individual risks and their totality.

Ultimately, in order to maintain the stability of the bank, it is necessary to ensure that all its risks are covered by its own capital since servicing the borrowed capital (attracted resources) requires regular interest payment for its use. The source of these payments is the interest and commissions that the bank receives for placing the attracted resources in income-generating assets. If the assets depreciate and/or show signs of default, the operating income flow becomes insufficient to meet the obligations on the borrowed capital, which ultimately leads to the bankruptcy of the bank.

This fact has led to the emergence and constant development of regulatory requirements for capital adequacy. These requirements are aimed per se at controlling the maintenance of the required capital level, which provides coverage of all risks accepted by the bank (risk-capital), as well as capital expenditures (cost-capital). Maintaining this balance is the main goal of the bank's IRM system (Fig. 1).

To maintain capital adequacy in the IRM system, banks must establish procedures for assessing the required capital and risk management to ensure that the

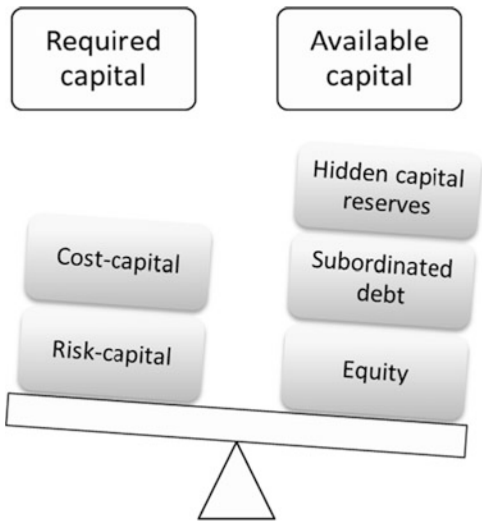


Fig. 1 Proper risk—capital balance

risks taken are limited within the bank's available capital (equity). In modern banking practice, they are called internal capital adequacy assessment procedures (ICAAP).

2 Reasons for the Introduction of the ICAAP Concept in the Basel II and Basel III Recommendations

Since compliance with the principle of covering the risks of a credit institution with its own capital is one of the most important factors for its stability, this issue is the focus of banking regulation systems in all countries. Standards for such regulation for the affiliated countries are developed by the Basel Committee on Banking Supervision (BCBS 1988).

The requirements for assessing the need for risk capital first appeared in 1988 in the Basel I agreement and only concerned capital to cover credit risk (CR). In 1996, a market risk amendment (MR) supplemented the capital adequacy requirements (BCBS 1996). In 1999, the development of Basel II began, which significantly changed the approaches to assessing CR, and also introduced a capital requirement to cover operational risk (OR) (BCBS 2004). Basel III continued these changes taking into account the factor of the crisis of 2008–2010 (BCBS 2010a, b).

However, current practice, including recent financial crises, has shown the weaknesses of the standardized approach to assessing capital adequacy, since many banks with officially adequate capital under Basel I, II, and III in times of crisis could not always meet their obligations to customers and investors.

Therefore, the most significant change in Basel II in terms of risk capital requirements was the introduction of Pillar 2, which formulated *the concept of economic capital* as a more accurate assessment of the overall banking risk compared to regulatory capital. Pillar 2 suggested that regulatory capital requirements (Pillar 1) should be treated as a minimum assessment of risk capital. The bank must now determine the real need for capital based on estimates of economic capital.

Economic capital in Basel II is considered as an assessment of the overall bank's risk based on internal models (BCBS 2004, p. 158). The list of risks that bank must allocate capital to cover was significantly expanded. In addition to CR, MR, and OR capital should be allocated for all types of *material risks*. The functions of identifying material risks are assigned to credit institutions. The list of potential material risks includes the interest rate risk of the banking book (IRRBB), liquidity risk, concentration risks, as well as legal, reputational, and regulatory risks. The list of potential material risks is not closed: credit institutions can expand it during the identification process.

To assess material risks, banks must develop internal models that may differ from regulatory capital assessment models. If regulatory approaches do not provide a sufficiently accurate assessment of the level of the bank's risk under consideration, the regulator may require creating *an internal model* that adequately assesses this risk.

To assess the total capital requirement to cover all material risks, the bank must also determine *the aggregation methods* and establish *procedures for determining the available risk capital* and *its allocation by business lines and types of material risks.*

In the course of operations and risk monitoring, the bank should focus on this distribution of capital and set and control *the limits based on the capital allocation*.

The bank strategy should also be based on the available capital adequacy to cover the overall risk inherent in the strategy. In this sense Basel II requires the integration of risk management processes and strategic management processes.

Thus, Pillar 2 defines the following structure of ICAAP:

- material risk identification;
- material risk measurement (quantification);
- material risk aggregation;
- allocation of capital by material risk types;
- maintaining compliance with the strategy and the available capital allocation.

The global financial crisis of 2008–2009 revealed the weaknesses of the Basel II agreement, which led to its revision and the emergence of Basel III. However, the requirements of Pillar 1 and Pillar 2 were not canceled. An important addition to the concept of risk capital assessment appeared. In accordance with this, banks should switch to risk assessments based on “going concern,” as opposed to previous approaches based on “gone concern.” Accordingly, the risk assessment models developed by banks should be based not only on statistics of historical losses, but also include factors that change the level of risks in the future depending on external and internal fundamental factors: the macro situation, the loan portfolio structure, the client base composition, changes in the profile of banking products, etc.

Pillar 2 defines the following basic principles for the organization of ICAAP:

- *Principle 1*: Banks should have procedures for assessing their overall capital adequacy in relation to the risk profile and a strategy for maintaining the capital level.
- *Principle 2*: Supervisors should review and evaluate banks’ internal capital adequacy assessments and strategies, as well as their ability to monitor and ensure their compliance with regulatory capital ratios. Supervisors should take appropriate supervisory actions if they are not satisfied with the result of this process.
- *Principle 3*: Supervisors can expect banks to operate above the minimum regulatory capital ratios and should be able to require banks to maintain capital in excess of the minimum.
- *Principle 4*: Supervisors must intervene proactively to prevent capital from falling below the minimum level required to support the risk characteristics of a particular bank and must take urgent corrective measures if capital is not maintained at a sufficient level or is not restored to a sufficient level.

The issues of ICAAP methodology and organization received more detailed coverage in such BCBS documents as “Range of practices and issues in economic capital frameworks”(BCBS 2009) and “Principles for effective risk data aggregation and risk reporting” (BCBS 2012). The documents were prepared by the Risk Management and Modelling Group and the Standards Implementation Group. It was assumed that the Financial Stability Board (FSB), in cooperation with task managers, would develop methods for supervising risk aggregation tools, especially for G-SIB. The goal is for the supervisory authorities to be confident that the management reporting fully reflects the level of existing credit institution risk.

The Bank of Russia introduced Pillar 2 requirements in the regulation of Russian credit institutions by issuing Directive No. 3624 U ‘On Requirements for the Risk and Capital Management System of a Credit Institution and a Banking Group’[1] (together with “Requirements for the organization of procedures for managing certain types of risks”).

To meet the requirements of Directive No. 3624-U, banks must ensure the implementation of such ICAAP procedures as

- material risk identification,
- material risk assessment,
- identifying and setting limits on risk appetite,
- risk limit control,
- ICAAP and corporate governance integration.

This chapter is devoted to developing the appropriate methodological approaches to the economic capital assessment that meet the requirements of BCBS and of the regulator for their further implementation ICAAP procedures.

3 Concept and Structure of the Bank’s Economic Capital and Available Risk Capital

The bank’s *economic capital* can be defined as the amount of potential losses of the Bank from all types of risks it accepts, which will not be exceeded with a high level of probability (usually 99.99%).

The structure of economic capital is determined by the types of *material risks accepted by the Bank* and is shown in Fig. 2.

BCBS determined the following economic capital covering principle, which determines the structure of the bank’s risk capital:

[1]Bank of Russia Directive No. 3624 U ‘On Requirements for the Risk and Capital Management System of a Credit Institution and a Banking Group’.

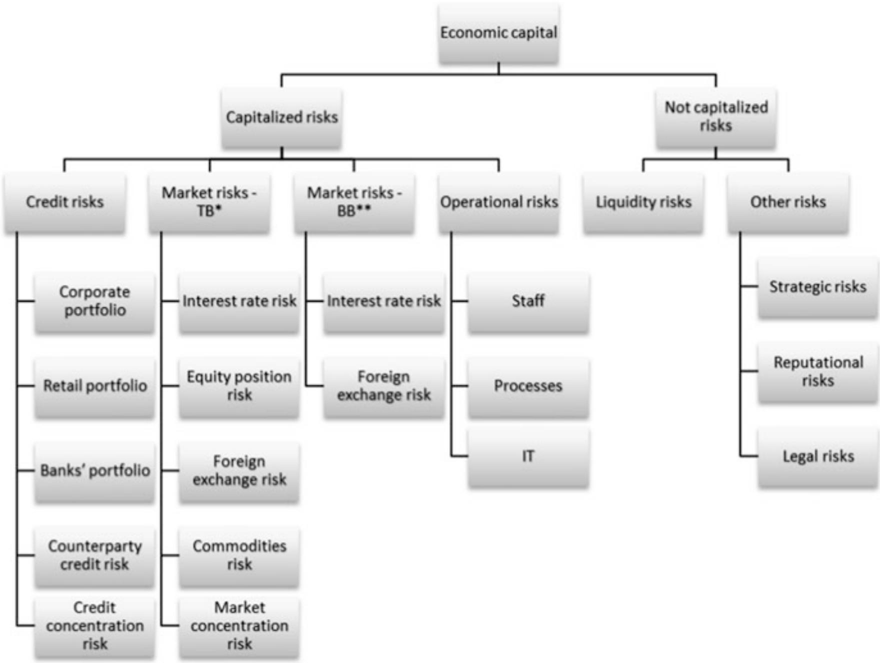


Fig. 2 The Bank's economic capital structure. **TB* trading book, ***BB* banking book

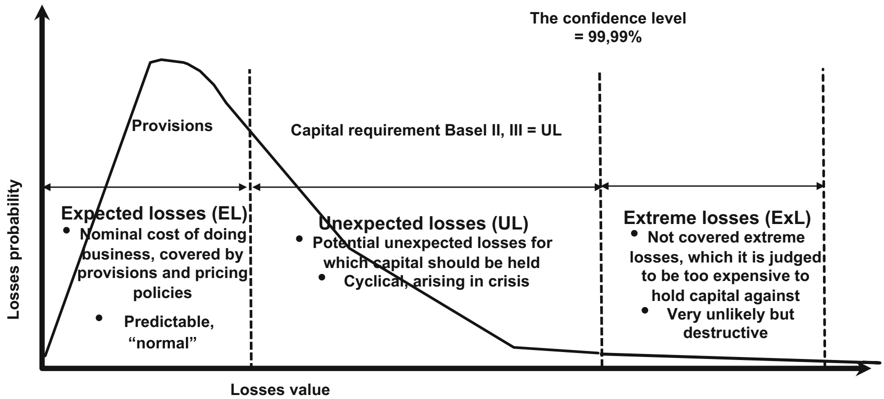


Fig. 3 The principles of banks' economic capital covering as defined by the BCBS

- the expected risk level (EL—Expected losses) should be covered by reserves for losses from the bank's profit;
- the remaining risk—unexpected losses (UL—Unexpected losses) must be covered by the bank's own capital;
- extreme losses are covered by the remaining capital, as well as hidden (quasi-capital) reserves (Fig. 3).

4 Financial and Non-Financial Risks: Specificity of Assessment and Capital Requirements

The basis for assessing risk capital is the value of each individual material banking risk. The BCBS and the Bank of Russia recommend determining an approach to capital allocation based on two alternative principles (Bank of Russia 2015, p. 10, paragraph 4.9.1):

- on quantitative methods based on empirical models for risk allocation assessment. This approach should be applied at least to CR, MR, and OR;
- on capital buffer allocation methods. This approach is applied to risks for which not enough historical data has been accumulated or quantitative models development is impractical due to cost and benefit mismatches.

The question of choosing a method for assessing capital requirements for material risks is closely related to the problem of dividing risks into financial and non-financial. The BCBS and the Russian regulator do not specify any definition for non-financial risks. In paragraph 3.3 of Directive No. 3624-U, the Bank of Russia simply lists the non-financial risk types: legal risk, regulatory risk, strategic risk, and the risk of loss of business reputation (Bank of Russia 2015, pp. 6–7, paragraph 3.3). Some authors believe that CR and MR are financial, while others are non-financial (Bank Saint Petersburg 2013). Others associate non-financial risks with the external environment or stakeholder impacts (Green 2009; Galushkin et al. 2007; Dugin et al. 2019). They refer to non-financial risks such as reputational, regulatory (compliance risk), legal, business risk, strategic risk.

Another group of authors emphasize the peculiarity of implementing non-financial risks, saying that “non-financial risks differ from the main business risks since no one expects any benefits from them”(Orlova 2012).

A fourth group allows for the occurrence of both damage and benefits from the effects of non-financial risks, but they speak about a broader impact of these risks on the companies’ activities, as opposed to financial risks: “In this case, the measure of damage or benefit is not only a direct impact on profit/costs and share price but also the impact on the reputation and human capital development as the main intangible companies’ assets, as well as on the general socio-political situation in the territories of its presence and the country as a whole” (Galushkin et al. 2007).

We are closest to the latter point of view. In our opinion, non-financial risks can be defined as the external and internal impacts which directly or indirectly affect the company’s value, profit, strategy, reputation, and other intangible assets. At the same time, a quantitative financial assessment of the results of such impacts is difficult due to the uncertainty of the model of the direct impact of these risks on the company’s financial results and/or due to the lack of sufficient historical data on risk losses to build a statistical model. Each type of risks or their individual components can be

Table 1 Differences in methods for assessing financial and non-financial risks

Assessment and management methods	Financial risks	Non-financial risks
Evaluation methods		
Historical data	Regular and observable realization	There were no historical realizations or they were extremely rare. The sample size is not sufficient to build a risk distribution assessment
Risk factors	Sufficiently determined for quantitative models building	The risk factors list is not uniquely defined
Assessment methods	Statistical models	Expert assessments
Management methods		
Provisions for losses from current profits	There are underlying assets or other financial indicators to regularly determine the size of the reserve	A reserve is formed for specific events on the basis of expert judgment
Capital coverage	Based on quantitative models individually for each material risk position	Based on expert judgment in the form of a capital buffer
Capital limits	Set	Not set

both regular and observable,[2] and rare, and therefore not available for quantitative statistical forecasting. In the first case, for minimizing the damage of risk realization, the bank builds risk management procedures based on quantitative risk distribution assessments. In the second case, they capture and evaluate the risk impact, including potential risk and direct damage, using qualitative methods, but they do not use statistical models to forecast, because they do not have a sufficient base for their construction and validation. Thus, it is possible to associate bank's financial risks with the risks that can be assessed based on applying quantitative assessment models of risk capital requirements, and non-financial risks with those ones for which it is advisable to apply qualitative assessments and capital buffer allocation methods. However, it should be emphasized that the inability to conduct quantitative statistical assessments does not mean that the risk is not managed and measured. Simply, the control and measurement methods will be different (Table 1).

The choice of methods for quantitative assessments of capital requirements for financial risks and determining the capital buffer size is left to the bank. For risks that are subject to quantitative capital requirement models, the task of aggregating the distributions of individual risks into the distribution of total risk arises at the aggregate risk assessment stage. One of the approaches to solving this problem is suggested in the next section.

[2] i.e., the organization collects data on risk exposure factors and outcomes.

5 A Structural Model for the Economic Capital Assessment

Mathematically, the task of overall risk assessment is reduced to constructing the random value sum distribution of losses from risks exposed to aggregation, the so-called distribution convolution:

$$\xi = \xi_1 + \xi_2 + \ldots \xi_i + \ldots \xi_n \,, \tag{1}$$

where ξ_i is the distribution of losses from the i-th type of risk.

Practices and techniques in risk aggregation are generally less developed than the methodologies that are used in measuring individual risk components. The BCBS highlights such problems in risk aggregation methodology as

- assuming diversification gains across all components;
- estimating the variance-covariance matrix which represents the co-movement between risks;
- lack of relevant data to assess risk interactions;
- lack of unification risk account units (risk measures), risk metrics, confidence levels and time horizons, used for different risk type assessments (BCBS 2009).

From our point of view, the main problem of material risk aggregation is the lack of a universal approach to various kinds of risk assessment. Not only types of risk distributions are distinguished, but also *risk metrics and measures*.

By risk metric we understand the approach to assessing the losses incurred from the risk implementation. In practice, there are various risk type metrics that do not match significantly:

- credit risk in the current approach, IFRS 9 is estimated on the basis of expected losses, measured as the difference in the present value of the contractual and risk-adjusted cash flow based on the actual state of affairs. Thus, the credit risk metric is the difference between the planned cash flow indicator and the expert assessment based on current facts (the so-called plan-fact analysis);
- market risk measurement is based on the dynamic analysis of the financial instruments prices volatility (FI volatility). In this case, the risk is perceived as a change in the FI fair value over time;
- operational risk is measured mainly as incurred losses, sometimes adjusted for other bank benchmark data;
- gap analysis or duration methods are used to assess interest rate risk, which is the net interest income sensitivity to changes in market interest rates, etc.

Obviously, it is pointless to sum up various risk metrics to determine the overall risk. The indicator will not have a clear economic interpretation.

The frequency of various risk metric assessment also varies in financial statements: credit and operational risks are assessed on a monthly basis, market risk on a daily basis, non-financial risks losses can be assessed once a year, etc.

A risk measure refers to the characteristics of the random risk distribution used for risk assessment. They may vary in practice as well. Typically, the measure of risk is VaR, but the standard deviation is often used for market risks. Recently, risk measures such as Shortfall and spectral measures have become widespread.

In addition, different time horizons and levels of confidence are used for the statistical assessment of various risk type measures:

- time horizons for market risks, the intraday and monthly VaR are calculated, for credit risks, the annual VaR is calculated,
- levels of confidence can take values 97.5%, 99%, 99.9%, 99.99%.

At the stage of Pillar 2 Basel II implementation, the BCBS conducted an analysis and systematized the risk aggregation practice in banks (BCBS 2009). It found gaps in the economic capital evaluation and management practice of G-SIB. These gaps related to the application of various risk measures and metrics, the inability of management systems and information systems to implement ICAAP processes.

Following the results, in January 2013 the BCBS formulated the principles of effective data aggregation and risk reporting (BCBS 2013a, b), highlighting 4 main areas of activity:

1. *Comprehensive management and infrastructure*, requiring the bank management to form adequate and effective mechanisms for managing data aggregation processes and preparing risk reports and the practice of risk reporting integrated with other principles and guidelines of the Basel Accords (Principle 1), as well as the development, creation and support of appropriate data architectures and IT infrastructure that fully provide data aggregation and risk reporting capabilities beyond just normal times, but also during periods of stress or crisis (Principle 2).
2. *Risk Data Aggregation Capabilities* that provide.

   - the accuracy and integrity of risk data and conditions for automating their aggregation processes (Principle 3);
   - the completeness of data on all material risks of the bank and its group by business lines, types of assets, industries, regions and other groups (Principle 4);
   - the timely generation of summary and up-to-date risk data (Principle 5);
   - the adaptability of risk data to meet a wide range of requests (Principle 6).

3. The practice of risk reporting, guaranteeing the accuracy, completeness, transparency and usefulness of reporting for its users (Principles 7–9), as well as the establishment by stakeholders of its frequency and confidentiality (Principles 10–11).
4. Regulatory supervision, tools and cooperation that imply compliance with the periodicity of supervisory audits of data aggregation and reporting processes, ensuring the implementation of measures to eliminate identified shortcomings, and international cooperation of supervisory authorities in this area. (Principles 12–14).

Since 2013, BCBS has been analyzing the progress G-SIB in implementing the principles of risk aggregation and the preparation of risk reporting (BCBS 2013a, b, 2015a, b, 2017, 2018). However, in 2018, BCBS's findings highlighted the difficulties of implementation and the need to continue working on improving risk aggregation and risk reporting systems, despite the fact that the project completion date was initially set as 2016. In this regard, the relevant area of scientific and applied research is the further development of aggregating risk methodology.

In international and domestic practice to assess economic capital various methods of simulation modelling and stress testing of bank portfolios are used. Most of them belong either to the class of one-factor models or to the class of models with the same type of risk factors. From our point of view, the best direction of this methodology development is *multifactor model development for assessing total financial risk based on the full modelling method*.

We offer an example of building such a model based on the identifying elements of the bank's profit formation that are exposed to certain material risks, as well as the determination of the factors of these risks, the change of which determines the volatility of the corresponding profit element. This approach, combined with the management accounting methodology, reveals the bank's financial result (profit) structure in terms of business lines, products, customers, and geographical regions, allows us to assess the total impact of material risks to the bank, and to implement risk aggregation/disaggregation procedures, to assess allocation capital to cover risks, and to establish risk limits based on the capital allocation and risk strategies.

The proposed approach will allow the harmonization of the procedures for assessing individual components of economic capital based on the following principles:

1. ensuring consistent disaggregation of the total financial result to positions exposed to various risks types;
2. the unified risk measures used for all material risks based on deviations of the financial result from the planned indicators both at the overall bank and business line levels, as well as individual product lines and customer groups;
3. the use of a universal risk metric—VaR and a universal tool for its assessment—stochastic modelling;
4. integration of ICAAP procedures in the processes of strategic and financial planning by using indicators of the risk appetite/the disposable capital of the bank when selecting planned alternatives.

To solve this problem, we suggest choosing the following *metric and risk measures*:

- as a risk indicator—the absolute change in profit (or one of its components exposed to risk) compared with the target indicator with the opposite sign (SRisk)[3];

[3]The use of the opposite sign will lead to the fact that the reduction in profits will mean and reflect the effects of risk, and the increase—its opposite side—the receipt of economic benefits, which will

- as a horizon for risk forecasting—1 year:
- as a step of risk modelling—1 day;
- as a risk measure—VaR.

To select the overall risk components we use differential function that reflects the influence of qualitative (intensive) and quantitative (extensive) revenue generation factors. The qualitative factor is ROA^t, and the quantitative factor is asset volume—A^t. Profit is the product of these two factors:

$$\mathrm{Profit}^t = ROA^t * A^t. \tag{2}$$

We use this function differential with the negative sign as the overall risk metric at all the analysis levels: total bank operations, business units, products, and clients:

$$\begin{aligned}\text{Total risk}^t &= -\partial ROA^t * A^t - ROA^t * \partial A^t \\ &= -\frac{\partial ROA^t}{ROA^t} * \mathrm{Profit}^t - \frac{\partial A^t}{A^t} * \mathrm{Profit}^t,\end{aligned} \tag{3}$$

$-\partial ROA^t * A^t$ (the first component of the expression (3)) reflects the impact of risk factors determining assets' profitability.

$-ROA^t * \partial A^t$ (the second component (3)) reflects the factors, determining assets volume change.

The second component of *expression* (3) can be interpreted as an indicator of strategic/business risk, as it reflects the bank's ability to expand its business and attract the necessary capital for this purpose (both its own and borrowed). Therefore, we can use this indicator to model and assess business risk:

$$\text{Business risk}^t = -ROA^t * \partial A^t = -\frac{\partial A^t}{A^t} * \mathrm{Profit}^t. \tag{4}$$

The first component of expression (3) reflects the influence of all other risks. In order to highlight the individual aggregate risk components, we associate these components with various elements of a banks' Profit and Loss Statement (P and L). In doing so, we try to find the appropriate type of financial material risk prescribed by ICAAP for each P&L element. Bank profit is a combination of the following elements:

$$\mathrm{Profit}^t = NII^t - ALLL^t + NTI^t + NFXE^t + NFCI^t + OOI^t - O\mathrm{Exp}^t - \mathrm{Tax}^t, \tag{5}$$

where:

NII^t is net interest income, which is the sum of

---

be reflected in the left tail of the risk distribution, which completely coincides with the approaches used in VaR models for assessing market risk.

- interest income of banking book ($IIBB^t$) = average credit interest rate (ICR^t) *average credit portfolio volume (CP^t),
- and interest income of trading book ($IITB^t$) = average interest yield of trading book (CY^t)*average trade portfolio volume (BP^t),
- minus interest expenses ($IExp^t$) = average deposits interest rate (IDR^t)*average deposits portfolio volume (DP^t).

$ALLL^t$ is Allowance for Loan and Lease Losses (or provisions charge for loan impairment) = average provisions rate (PR^t)* CP^t.

NTI^t is net trade income, which is the sum of

- net gain of equity portfolio ($NGEP^t$) = average equity portfolio profitability (EPP^t) *average equity portfolio volume (EP^t),
- and net gain of commodity portfolio ($NGComP^t$) = average commodity portfolio profitability ($ComPP^t$) *average commodity portfolio volume ($ComP^t$).

$NFXE^t$ is net foreign exchange earnings, which is equal to the difference of

- FX gain (FXG^t)–FX loss (FXL^t),

or product of

- average exchange rate change ($FXCh^t$) *average open currency positions (OCP^t),

$NFCI^t$ is net fee and commission income, which is equal to the difference of

- fee and commission income (FCI^t)–fee and commission expenses ($FCExp^t$),

or product of

- net average fee and commission profitability ($NFCP^t$)*A^t.
- OOI^t is other operations income, which is equal to product of.
- net average other operations profitability ($NOOP^t$)*A^t,
- $OExp^t$ is operations expenses, which is equal to product of.
- average operations cost for assets unit (UOC^t)*A^t.
- Tax^t is taxes paid, which is equal to product of.
- average income tax rate for assets unit ($ITaxR^t$)*A^t.

Using expression (5) we can present the first component of expression (3) as:

$$
\begin{aligned}
-\partial ROA^t * A^t = {} & \left(-\partial ICR^t * CP^t + \partial IDR^t * DP^t\right) + \partial ALLL^t * CP^t \\
& + \left(-\partial CY^t * BP^t - \partial EPP^t * EP^t - \partial ComP^t * ComP^t - \partial FXP^t * OCP^t\right) \\
& - \partial NFCP^t * A^t - \partial NOOP^t * A^t - \partial UOC^t * A^t - \partial ITaxR^t * A^t.
\end{aligned} \tag{6}
$$

Now we correlate the components of expression (6) with individual material risk types:

- *credit risk:*

$$CR^t = \partial PR^t * CP^t = \frac{\partial PR^t}{PR^t} * PR^t * CP^t = \frac{\partial PR^t}{PR^t} * ALLL^t; \tag{7}$$

market risk:

$$\begin{aligned} MR^t &= -\partial CY^t * BP^t - \partial EPP^t * EP^t - \partial ComP^t * ComP^t - \partial FXP^t * OCP^t \\ &= IRRTB + ER + ComR + FXR, \end{aligned} \tag{8}$$

where

– $IRRTB^t$ is interest rate risk of trading book,

$$IRRTB^t = -\partial CY^t * BP^t = -\frac{\partial CY^t}{CY^t} * CY^t * BP^t = -\frac{\partial CY^t}{CY^t} * IITB^t; \tag{9}$$

– ER^t is equity risk of trading book,

$$ER^t = -\partial EPP^t * EP^t = -\frac{\partial EPP^t}{EPP^t} * NGEP^t; \tag{10}$$

– $ComR^t$ is commodity risk of trading book,

$$ComR^t = -\partial ComPP^t * ComP^t = -\frac{\partial ComPP^t}{ComPP^t} * NGComP^t; \tag{11}$$

– FXR^t is foreign exchange risk,

$$FXR^t = -\partial FXCh^t * OCP^t = -\frac{\partial FXCh^t}{FXCh^t} * NFXE^t; \tag{12}$$

- *operational risk*

$$OR^t = \partial UOC^t * A^t = \frac{\partial UOC^t}{UOC^t} * UOC^t * A^t = \frac{\partial UOC^t}{UOC^t} * OExp^t; \tag{13}$$

- *interest rate risk of banking book*

$$IRRBB^t = \left(-\partial ICR^t * CP^t + \partial IDR^t * DP^t\right) = -\frac{\partial ICR^t}{ICR^t} * IIBB^t + \frac{\partial IDR^t}{IDR^t} * IExp^t; \quad (14)$$

- *price risk on fees, commissions, and other deals*

$$\mathrm{PriceRisk}^t = (-\partial NFCP^t * A^t - \partial NOOP^t * A^t) = -\frac{\partial NFCP^t}{NFCP^t} * NFCI^t + \frac{\partial NOOP^t}{NOOP^t} * OOI^t; \quad (15)$$

- *tax risk*

$$\mathrm{TaxRisk}^t = (\partial ITaxR^t * A^t) = -\frac{\partial ITaxR^t}{ITaxR^t} * \mathrm{Tax}^t. \quad (16)$$

We presented the overall financial risk of the bank as the sum of the main material risks. Each material risk corresponds to a certain risk factor and an element of P and L which determines the weight for aggregating the risk factor into the overall risk model (Table 2).

The model created (3–16) *unambiguously links the components of the total risk with the elements of the banks' profit formation*. Further within each component, it is possible to separate more granulated risk elements. For example, it is possible to divide credit risk into certain credit portfolios, FX risk for currencies, equity risk for security portfolios and other kinds of securities.

Thus, in general terms, the aggregate risk model (3–16) can be represented as a linear combination of various risk factors:

$$\mathrm{TotalRisk}^t = \sum_{i=1}^{N} \partial RF_i^t * RP_i^t, \quad (17)$$

where N is the number of risk factors in the economic capital model; ∂RF_i^t is the i-th risk factor change at time t; $RP_i^t$is risk position, corresponding with the i-th risk factor.

The result of expression (17) is a random variable[4] of the total risk. To assess the aggregate risk metrics, it is necessary to evaluate its distribution based on the given distributions of individual material risks types.

If all the risk factors presented in expression (17) have a normal distribution, then the ratio can be used to calculate the total risk VaR:

[4]more precisely, a random process.

Table 2 The overall financial risk model components

Material risks	Risk factors	Risk weights
Business risk (BusinessRisk)	Assets volume change (ach)	Profit
Credit risk (CR)	Change of average provision rate—PR	Allowance for loan and lease losses—ALLL
Market risks (MR)		
Interest rate risks of trading book (IRRTB)	Change of average interest yield of trading book—CY	Interest income of trading book—IITB
Equity risk of trading book (ER)	Change of average equity portfolio profitability—EPP	Net gain of equity portfolio—NGEP
Commodity risk of trading book (ComR)	Change of average commodity portfolio profitability—ComPP	Net gain of commodity portfolio—NGComP
Foreign exchange risk (FXR)	Change of average exchange rate change—FXCh	Net foreign exchange earnings—NFXE
Interest risks of banking book (IRRBB)		
Interest rate risks of credit portfolio (IRRBB-CP)	Change of average credit interest rate (ICR)	Interest income of banking book (IIBB)
Interest rate risks of deposit portfolio (IRRBB-DP)	Change of average deposit interest rate (IDR)	Interest expenses (IExp)
Operational and other risks		
Operational risks (OR)	Change of average operational cost for assets unit—UOC	Operational expenses (OExp)
Price risk on fees, commissions and other deals (PriceRisk)	Change of net average fee and commission profitability—NFCP Change of net average other operational profitability—NOOP	Net fee and commission income (INFCI) Other operations income (OOI)
Tax risk (TaxRisk)	Average tax rate—TaxR	Tax paid (tax)

$$VaR_{\text{TotalRisk}} = \sqrt{VaR^T * COR * VaR}, \tag{18}$$

where VaR^T is a vector-line whose coordinates are the VaR values for the i-th risk factor multiplied by the corresponding position: $VaR\left(RF_i^t\right) * RP_i^t, VaR$ is a vector-column with similar coordinates.

COR is a matrix of correlation coefficients between different risk factors ∂RF_i^t 0.

However, if the risk factors have different types of distributions or these distributions are not normal, the task of assessing the total risk is complicated. Here you can use either a simple historical method or parametric methods for calculating the convolution of random variables (for example, the copula method) or the stochastic modelling method.

Due to the complexity of the convolution parametric approximation, one of the most popular methods for assessing total risk is the stochastic modelling method (or Monte Carlo method). However, for its application it is necessary to evaluate not only the individual distributions of risk factors, but also their joint distribution, taking into account the correlation between individual factors.

If the individual risk factor distributions laws are different or cannot be approximated based on known parametric distributions, a discrete approximation can be used:

$$\varphi(y) = \sum_{l_1, l_2, \ldots l_N = 1}^{K} p_{l_1}^1 * p_{l_2}^2 * \ldots l p_{l_N}^N * S(y, l_1, l_2, \ldots l_N), \tag{19}$$

where $\varphi(y)$ is the convolution distribution; K is the interval number of risk distributions; $p_{l_k}^k$ is the probability that the k-th risk value falls into the interval l_k; $S(y, l_1, l_2, \ldots l_N)$ is the intersection area of the hyperplane $x_1 + x_2 + \ldots x_N = y$.with cube $L = \{x_1, x_2, \ldots x_N : m(l_1 - 1) * RP_1 \leq x_1 \leq m(l_1) * RP_1, m(l_2 - 1) * RP_2 \leq x_2 \leq m(l_2) * RP_2, \ldots m(l_N - 1) * RP_N \leq x_N \leq \} m(l_N) * RP_N\}$

$m(l_k)$ is the interval l_k center, RP_k is the risk-weight of the k-th factor.

Note that expression (17) is valid both for a single transaction/a separate financial instrument profit calculation and for profit in the context of products, projects, customers, business lines, and the total financial result of the bank.

The overall risk model is integrated with the financial banking model. Due to this, it is possible to assess the total risk within the financial planning process using stochastic modelling and stress testing. Financial planning and aggregate risk models have the same parameters including interest rates on loans and deposits, reserve rates for the impairment of assets, exchange rates and rates of return on financial market instruments, unit costs for bank processes and products, tax rates, etc.

The proposed approach to economic capital assessment makes it easy to disaggregate it in the context of business areas, products, customer groups based on traditional methods of functional-value analysis (cost-effective value engineering) used in management accounting systems. The data accumulation for this model can also occur within the framework of traditional managerial accounting and budgeting systems (Teplova 2019; Pomorina 2017; Cedric Reed, Hans-Dieter Scheuerman 2012; Karpov 2007). It should be noted that the model (3–16) also allows us to evaluate the diversification effect for covering economic capital at different levels of analysis. Note that in the model, such risk components as business risk, tax risk and price risk on bank fees, commissions, and other deals have appeared. These risks are not traditionally considered in economic capital models, but, nevertheless, their impact on bank profits can be significant.

The operational risk measurement differs from that adopted in regulatory approaches as it shows its impact on the bank's costs. This approach is more consistent, since its manifestations lead to an increase in bank costs, and the occurrence of fines and compensation for losses incurred.

Note that BCBS allows the possibility of using various risk assessment methods in individual models and economic capital models, as identical measures and metrics must be used when aggregating risks.[5]

The aggregated financial risk model advantages include:

1. quantitative accounting of a wide range of financial risks;
2. defining a unified approach to assessing certain material risks based on the bank's financial results;
3. the reflection of the effect of diversification;
4. using non-parametric methods for aggregate financial risk distribution assessment;
5. integration of the economic capital model with the financial planning processes.
6. universality: the applicability of the approach for any financial institution.

6 Forming of ICAAP Procedures Based on the Economic Capital Model and Available Risk Capital. The Bank's Risk Limit System Based on the Distribution of Available Risk Capital

BCBS and the Bank of Russia put forward requirements for ICAAP both for individual material risks and for economic credit institution capital (total risk). These procedures must include:

- methods for assessing and forecasting material risks and the economic capital of a bank;
- capital management procedures based on determining the planned (target) level of available risk capital, current capital requirements, and the allocation of capital by types of material risks and activities;
- a system for monitoring capital adequacy and limits for material risks.
- In order to control the accepted risks, the bank determines the planned (target) risk levels, the target risk structure, and the risk limits system for each material risk based on the business cycle phase, tolerance for risks and strategic and business objectives.

In order to control its capital adequacy, the bank establishes procedures for the capital allocation through a limits system in business line and in material risks, taking into account reserves for non-financial risks and for the new business projects implementation. The limits system must have a multi-level structure. Control over the established limits is carried out by setting signal values. If these limits are exceeded, an anti-crisis measure system must be developed and implemented and,

[5]Basel Committee on Banking Supervision. Range of practices and issues in economic capital frameworks. Part IV.B—March 2009, www.bis.org.

Table 3 Diversification factors for the allocation of capital determining

	Bank economic capital/diversification ratio								
Diversification ratio for aggregated level	DR1 = (85 + 65 + 40)/95 = 2								
EC—aggregated level	EC = 95 billion rub								
Diversification ratio for business areas	DR21= 85/(55 + 40 + 32,5) =1,5			DR22 = 65/ (55 + 40 + 35) =2			DR23 = 40/ (25 + 30 + 5) = 1,5		
EC for business areas	Corporate block—85 billion rub			Retail block—65 billion rub			Development projects—40 billion rub		
Material risks	CR	MR	OR	CR	MR	OR	CR	MR	OR
EC for SR	55	40	32,5	55	40	35	25	30	5

possibly, capital reallocation should be carried out (Bank of Russia 2015, paragraph 4.11–4.14).

The proposed model of economic capital allows the creation of the above-described procedures for managing economic capital, based on the allocation and distribution of available capital by types of material risks and setting limits. We can suggest the following scheme for implementing the regulator's requirements based on the integrated assessment of the economic capital model.

At *the first stage*, the bank's need for risk capital is estimated, using the economic capital model proposed above. Required capital depends on the business structure, development plans, including bank projects, the external economic situation, client base features, business process efficiency, etc.

Consider an example in which the estimate of the bank's economic capital amounted to 95 billion rubles (Table 3).

At *the second stage*, the need for risk capital is settled, which is determined by the economic capital assessment regulated by the bank's financial model formed as part of its strategy development and development plans, and the available risk capital allocated by the Board of Directors.

The procedure must be organized *within the framework of the strategic and business planning process* and may require a number of iterations if the allocated risk capital is not sufficient to cover the risks of the bank's strategy and development. Shareholders can review their risk appetite or require the bank to implement a more conservative strategy.

Suppose that according to the results of the coordination, the Board of Directors allocated 100 billion rubles capital to cover risks.

Risk capital can be allocated both in absolute terms and on the basis of various indicators of risk appetite, for example, on the basis of setting a target level of the RAROC indicator, the possibilities of using which are described in the next section.

At *the third stage*, the allocated risk capital should be distributed between business areas, customer groups, products, and material risks. To do this, based on the model, diversification coefficients must be determined. The function is implemented based on the calculation of economic capital at the selected planning

Table 4 Allocation of risk capital taking into account diversification factors

	Risk capital/diversification ratio								
Aggregated level	RC = 100 billion rub								
Business areas	Corporate block			Retail block			Development projects		
RC for 2 level = 2 × 100 = 200 billion rub	88 billion rub			68 billion rub			44 billion rub		
Material risks	CR	MR	OR	CR	MR	OR	CR	MR	OR
RC for 3 level	88 × 1,5 = 132 billion rub			68 × 2 = 136 billion rub			44 × 1,5 = 66 billion rub		
	57	42	33	58	42	36	27	32	7

levels. In this case, a unified assessment model is used, which is applied to calculate the EC of all business areas, bank projects, and certain types of material risks (see Table 3).

Further, the obtained diversification coefficients can be used to allocate capital at different hierarchical levels of setting risk limits (see Table 4).

At *the fourth stage*, limits must be set based on the allocated risk capital. This can be both direct restrictions on the volume of losses incurred from risks in the context of business areas and material risks, and limits that restrict the volume of operations.

When setting limits for losses incurred, the bank must determine the metrics of losses incurred on loans and financial assets, on operational and other risks. For example, the losses incurred on loans can be estimated as direct costs to write them off, and the difference between the amortized costs of actually received loan flow from its target value.

The latter is more consistent with the current concept of measuring the fair value of financial assets, as reflected in IFRS 9.

When setting limits on the volume of operations, the risk factor VaR estimates obtained in the model can be used. If based on the ratio $VaR_{KRisk} = VaR_{\partial NR} * CP$, the limit on the volume of the loan portfolio can be set in the amount equal to $EC_{CRisk}/VaR_{\partial NR}$.

As a result, on the basis of the described principles, the bank can form a hierarchical system of limits, presented in Fig. 4. Further, if necessary, they are disaggregated by individual products or financial instruments.

7 ICAAP Integration in the Processes of Strategic and Operational Bank Management Using RORAC

For business development, it is important to balance risks and benefits. In this sense, restrictions can be set not only on the maximum amount of risk accepted, but also on the profitability/risk ratio. This approach compares the level of risks with the benefits

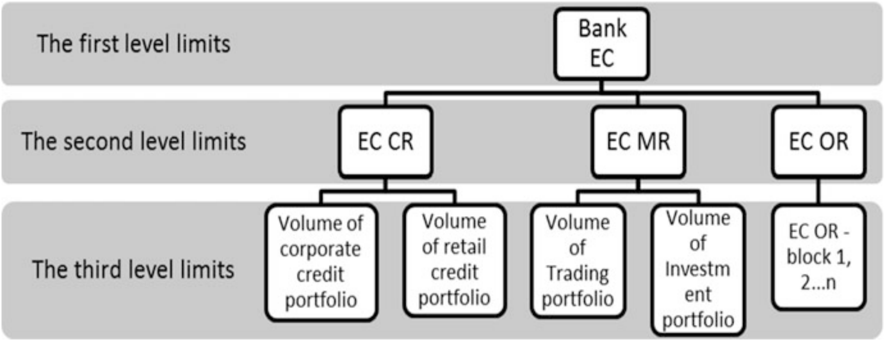


Fig. 4 An example of a system of limits based on the allocation of bank risk capital

received, which makes it easier for shareholders to determine their risk tolerance (risk appetite). Most often, RORAC is used as such a target.

$$RORAC = \frac{Profit - EaR}{VaR}. \tag{20}$$

The economic meaning of RORAC can be interpreted as the ratio of the annual profit received by the bank under normal operating conditions to loss in a crisis situation. Accordingly, if the indicator is 100%, then unforeseen losses will be covered within one year, if 50%—within 2 years, 30%—within 3.3 years, etc.

This indicator can be calculated within the multifactor model of bank economic capital described here in the context of departments, products, groups of customers. The target level can be defined as one common value for the entire bank, and differentiated by levels of analysis.

An example of the RORAC application in the strategic management process is shown in Table 5.

In the presented example, the target for the bank is set to 30% RORAC level. In developing the financial plan for both the main business units and projects, the RORAC goal was observed (line 5 of Table 2). However, in the process of implementing the plan, the profitability level of the Projects turned out to be lower than planned. The target RORAC level has been violated (row 9 of Table 2). In this situation, managerial impacts (system of measures) should be defined to eliminate excess risk. Since in our example the indicator for the bank is generally respected, one of the solutions may be the reallocation of risk capital if the Board of Directors is ready to accept a higher risk for projects that are significant for the future development of the bank.

Thus, the RORAC indicator can be used as one of the key indicators for strategy implementation, as it will allow the identification of high risk points at various management levels and the timely formulation of anti-crisis plans to reduce risks. Similar to other limits for RORAC, a warning and critical level can be defined, and anticipatory actions should be formed when moving to the warning zone.

Table 5 An example of RORAC assessment and use for strategic management

Business areas	Corporate block	Retail block	Development projects	Total
1. Position at risk (in billion rubles)	81,600	34,000	20,400	136,000
2. Profitability (in %) (plan)	**3.25%**	**4.60%**	**6.00%**	4.00%
3. Expected income (in billion rubles)	2652	1564	1224	5440
4. VAR (in billion rubles)	8054	4396	3550	16,000
5. RORAC—plan	**32.93%**	**35.58%**	**34.48%**	34.00%
6. The purpose of RORAC	30.00%	30.00%	30.00%	30.00%
7. Profit (in billion rubles)	2660	1618	890	5168
8. Actual return	**3.26%**	**4.76%**	**4.36%**	3.80%
9. RORAC—fact	**33.03%**	**36.80%**	**25.07%**	32.30%
10. RORAC deviation from the target	3.03%	6.80%	−4.93%	2.30%

8 Conclusion

The multifactor model for assessing economic capital proposed above can be implemented within the traditional bank model, as the parameters of the financial model are also risk factors in the proposed model for assessing economic capital. Thus, this model is naturally integrated into the strategic planning and management system, since, simultaneously with the selection of planned alternatives, their risks will be assessed and economic capital estimates will be calculated at all hierarchical levels of management: departments, products, customers.

The approach based on the use of Income Statement Analysis to identify positions exposed to various types of risk allows you to automatically determine weights for aggregating estimates of economic capital of certain material risks into an assessment of the total risk of a bank.

As shown above, the use of the proposed model fulfills all the requirements of Pillar 2 and Bank of Russia regulations regarding ICAAP procedures.

References

Bank of Russia. (2015). *On requirements for the risk and capital management system of a credit institution and a banking group* (pp. 1–28). Bank of Russia Directive No. 3624 U. Retrieved Apr 15, 2015, from www.cbr.ru.

Bank Saint Petersburg. (2013). *Annual report 2013* (p. 56). http://www.bspb.ru/investors/annual-reports/2013.

BCBS. (1988). *International convergence of capital measurement and capital standards* (pp. 1–30). Basel Committee on Banking Supervision. Retrieved July 15, 1988, from www.bis.org.

BCBS. (1996). *Amendment to the capital accord to incorporate market risks* (pp. 1–56). Basel Committee on Banking Supervision. Retrieved Jan 4, 1996, from www.bis.org.

BCBS. (2004). *Basel II: International convergence of capital measurement and capital standards: A revised framework* (pp. 1–251). Basel Committee on Banking Supervision. Retrieved Jun 10, 2004, from www.bis.org.

BCBS. (2009). *Range of practices and issues in economic capital frameworks* (pp. 1–73). Basel Committee on Banking Supervision. Retrieved Mar 27, 2009, from www.bis.org/publ/bcbs152.pdf.

BCBS. (2010a). *Basel III: A global regulatory framework for more resilient banks and banking systems* (pp. 1–77). Basel Committee on Banking Supervision. Retrieved Dec 16, 2010, from www.bis.org.

BCBS. (2010b). *Basel III: International framework for liquidity risk measurement, standards and monitoring* (pp. 1–53). Basel Committee on Banking Supervision. Retrieved Dec 16, 2010, from www.bis.org.

BCBS. (2012). *Principles for effective risk data aggregation and risk reporting* (pp. 1–24). Basel Committee on Banking Supervision Consultative Document. Retrieved Jun 25, 2012, from www.bis.org/publ/bcbs222.pdf.

BCBS. (2013a). *Principles for effective risk data aggregation and risk reporting* (pp. 1–28). Basel Committee on Banking Supervision Consultative Document. Retrieved Jan 8, 2013, from www.bis.org/publ/bcbs239.pdf.

BCBS. (2013b). *Progress in adopting the principles for effective risk data aggregation and risk reporting* (pp. 1–34). Basel Committee on Banking Supervision Consultative Document. Retrieved Dec 17, 2013, from www.bis.org/publ/bcbs268.pdf.

BCBS. (2015a). *Progress in adopting the principles for effective risk data aggregation and risk reporting* (pp. 1–18). Basel Committee on Banking Supervision Consultative Document. Retrieved Dec 16, 2015, from www.bis.org/publ/d348.htm.

BCBS. (2015b). *Progress in adopting the principles for effective risk data aggregation and risk reporting* (pp. 1–25). Basel Committee on Banking Supervision Consultative Document. Retrieved Jan 23, 2015, from www.bis.org/publ/d308.htm.

BCBS. (2017). *Progress in adopting the principles for effective risk data aggregation and risk reporting* (pp. 1–23). Basel Committee on Banking Supervision Consultative Document. Retrieved Mar 27, 2017, from www.bis.org/publ/d399.htm.

BCBS. (2018). *Progress in adopting the principles for effective risk data aggregation and risk reporting* (pp. 1–26). Basel Committee on Banking Supervision Consultative Document. Retrieved Jun 20, 2018, from www.bis.org/publ/d443.htm.

Cedric Reed, Hans-Dieter Scheuerman. (2012). *Financial Director as an integrator of business*. Moscow: Alpina Publisher.

Dugin, A. D. (2019). In A. D. Dugin & G. I. Penicas (Eds.), *Risk and capital management system development (ICAAP)* (pp. 145–167). Moscow: Urait.

Galushkin, S. V., Zagidullin, J. K., & Flyar, M. G. (2007). Business and society: Corporate integration. Corporation and development. In O. B. Alekseev & O. I. Genisaretsky (Eds.), *Collection of works on the philosophy of corporate development* (Vol. 2, pp. 122–140). Moscow: Europe.

Alexander Green. (2009). *Management of risks: Business strategy development* (pp. 1–3). https://subscribe.ru/archive/economics.school.riskmanagement/200906/25140618.html.

Karpov, A. E. (2007). 100% of practical budgeting. In *Budgeting as a management tool* (pp. 1–8). Moscow: Knorus.

Orlova, O. E. (2012). Non-financial risk management. *Actual Issues of Accounting and Taxation, 10*, 89–97.

Pomorina, M. A. (2017). *Financial management in a commercial bank*. Moscow: Knorus.

Teplova, T. V. (2019). *Effective financial director: textbook*. Moscow: ID Urait.

Part IV
Systemic Risks Modeling and Stress Testing

Exploring the Interplay Between Early Warning Systems' Usefulness and Basel III Regulation

Elena Deryugina, Maria Guseva, and Alexey Ponomarenko

Abstract We analyse the ability of credit gap measures to predict banking crises by estimating the usefulness measure conditionally on policymaker's preferences. The results show that the signals based on the credit gap indicators are most useful when the policymaker's preferences regarding Type I and Type II errors are approximately equal. However, according to the current consensus, the preferences to avoid missing a crisis are higher than issuing a false signal. This means that the usefulness of the credit-gap-based early warning systems is likely to increase once the static Basel III regulative measures are implemented (assuming that their implementation results in lower financial crises' costs).

Keywords Credit gap · Early warning system · Macroprudential policy · Basel III regulation

JEL G01 · G28 · G32

1 Introduction

One of the key goals of central banks is the timely adoption of measures to prevent or mitigate financial crises, as well as to improve the financial stability of the banking system as a whole. In 2010, the Basel Committee on Banking Supervision published an assessment of the long-term economic impact of stronger capital and liquidity requirements introduced by Basel III (BCBS 2010). The Basel Committee's assessment of the long-term economic impact finds that there are clear, net, long-term economic benefits from increasing the minimum capital and liquidity requirements from their current levels in order to increase the safety and soundness of the global banking system. The benefits of higher capital and liquidity requirements accrue

E. Deryugina · M. Guseva · A. Ponomarenko (✉)
Bank of Russia, Moscow, Russia
e-mail: ponomarenkoaa@cbr.ru

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_12

from reducing the probability of a financial crisis and the output losses associated with such crises. The benefits substantially exceed the potential output costs for a range of higher capital and liquidity requirements. In order to reduce pro-cyclicality of credit, Basel III introduces a counter-cyclical capital buffer (CCyB) and proposes a credit-to-GDP gap as a guide for setting it. The useful properties of this indicator are confirmed generally for a broad array of countries and a long time span, which includes the most recent crisis. Researchers usually apply AUC-ROC analysis.[1] Drehmann and Juselius (2014) find that the credit-to-GDP gap outperforms other measures at long time horizons in the set of developed countries. Deryugina and Ponomarenko (2019) confirm that the standard credit gap indicator performs satisfactorily for emerging markets. Notably, such an assessment is based on certain assumptions about the policymakers' relative aversion to making different types of errors (i.e. missing the crisis or issuing a false signal). These preferences in turn depend on the expected severity of the financial crisis. There are, however, reasons to expect that this characteristic may change once the static regulations recommended by Basel III are implemented.

A related strand of research examines Basel III's macroeconomic effect. Behn et al. (2016) measure the gains of capital regulation as the expected output increase associated with the reduction in the likelihood and severity of banking crises. Budnik et al. (2019), working from the estimation of the FAVAR model, show that an increase in capital ratios has a sharply different impact on credit and economic activity depending on the way the bank adjusts. Arregui et al. (2013) find that changes in the regulation affect the expected probability of a crisis. Popoyan et al. (2017) develop an agent-based model and find that Basel III's prudential regulation is the best policy mix to improve the stability of the banking sector and smooth output fluctuations. In the Basel Committee on Banking Supervision's review of the literature (BCBS 2019), the estimated marginal reduction in the annual probability of a crisis ranged across studies from as little as 0.03% points to as much as 1.7% points. Using quantile regressions applied to a panel dataset of advanced economies, Aikman et al. (2019) find that higher levels of banking system capital significantly improve GDP-at-risk in the medium term. Therefore, there are good reasons to believe that the Basel III regulation changes the financial cycle's characteristics and reduces the severity of banking crises. As a result, policymakers' preferences regarding the early warning systems' performance will change.

This paper develops the notion of credit gap performance as an early warning indicator (EWI) of a crisis under different policymakers' preferences and conjectures how its performance may change once the static part of the Basel III regulation is implemented.

---

[1]ROC (receive operating characteristics curve) is created by plotting the true positive rate against the false positive rate at various threshold settings; AUC is the area under the ROC curve.

2 Data

We use the cross-section of 21 countries (see Table 2 in Appendix 1). We use the Bank for International Settlements (BIS) database as the source for credit series (adjusted for breaks all sectors' credit to private non-financial sector). The availability of these data determines the composition of the dataset. We use the Organization for Economic Co-operation and Development (OECD) database for GDP and price series (GDP deflator, if available, and consumer prices, otherwise). All data are seasonally adjusted using an X-12 procedure. Crisis periods are borrowed from Laeven and Valencia (2018).

3 Evaluation Method

The predictive ability of EWIs is usually tested using ROC analysis, but this approach only presents an average measure of usefulness. A more comprehensive evaluation approach is the analysis based on the 'usefulness' measure, which is calculated conditionally on the policymaker's relative aversion to missed crises as opposed to false alarms. We believe that it is important to test the indicator's performance under different preferences. Notably, the introduction of Basel III macroprudential measures may change the macroeconomic performance, such as the probability and severity of banking crises, the credit gap and therefore possibly the policymaker's preferences. The 'usefulness' approach allows us to develop this idea. We apply the 'signals' approach first developed by Kaminsky et al. (1998). In order to examine the performance of the credit gap, it is useful to design the following matrix (Table 1).

In this matrix, A is the number of quarters in which the indicator issued a good signal; B is the number of quarters in which the indicator issued a bad signal; C is the number of quarters in which the indicator failed to issue a signal when the crisis occurred; and D is the number of quarters in which the indicator did not issue a signal when in fact there was no crisis. A warning signal is considered to be issued when the indicator exceed a threshold, which runs through the indicator's distribution percentiles.[2]

The loss function of the policymaker is defined as (see Alessi and Detken 2011):

$$L = \theta \frac{C}{A + C} + (1 - \theta) \frac{B}{B + D} \tag{1}$$

[2]Unlike Kaminsky et al. (1998), we follow Borio and Lowe (2002) and define the thresholds in terms of percentage point gaps. We examine 101 thresholds in these exercises in the range of [0; 1] in steps of 0.01.

Table 1 Signalling matrix

	Crisis (within 4–12 quarters)	No crisis (within 4–12 quarters)
Signal was issued	A	B
No signal was issued	C	D

θ is the parameter revealing the policymaker's relative risk aversion between Type I (missing crisis) and Type II (false alarm) errors; C/(A + C) is the share of Type I errors; and B/(B + D) is the share of Type II errors. Following the approach of Alessi and Detken (2011), we employ the "usefulness" indicator to assess the models:

$$U = \min(\theta, 1 - \theta) - L. \tag{2}$$

A central banker can always realize a loss of min[θ; 1 − θ] by disregarding the indicator (i.e. by issuing the signal either always or never). If θ is smaller than 0.5, the benchmark is obtained by ignoring the indicator, and never having any signals issued, so that A = B = 0. The resulting loss L is θ. If θ exceeds 0.5, the benchmark for the central bank is assuming that a signal is always issued C = D = 0. The resulting loss is 1 − θ. An indicator is then useful to the extent that it produces a loss lower than min[θ; 1 − θ] for a given θ – that is, relying on the indicator reduces the loss compared to a situation in which the indicator is ignored.

4 Results

Credit gap indicators are estimated by applying a one-sided Hodrick–Prescott filter (λ = 400,000) to the log of the credit-to-GDP ratio recursively over the expanding window (with the minimum size of 12 quarters). We expect the credit gap to start issuing the warning signal 12 quarters before the crisis (crisis periods are here as defined by Laeven and Valencia (2018)) and exclude from the analysis four observations before the crisis and all of them during it, because warning signals are not truly useful any longer.

We adopt the signal approach and find the optimal thresholds by minimizing the lost function for different values of preference parameter θ.[3] As shown by Alessi and Detken (2011), it appears more relevant to obtain the results when the optimal threshold is imposed to be the same for all countries, and not for each country individually. Thus, the calculations are conducted for the pool of 21 countries and 22 banking crisis episodes. In Fig. 1, we report the usefulness indicators θ for the optimal thresholds calibrated for various θ. The results show that the maximum value of the usefulness function is achieved when θ = 0.5. The preference parameter

[3]Calculations are provided for all θ in the range of [0.01, 0.99] in steps of 0.01 to construct the smoothest usefulness function.

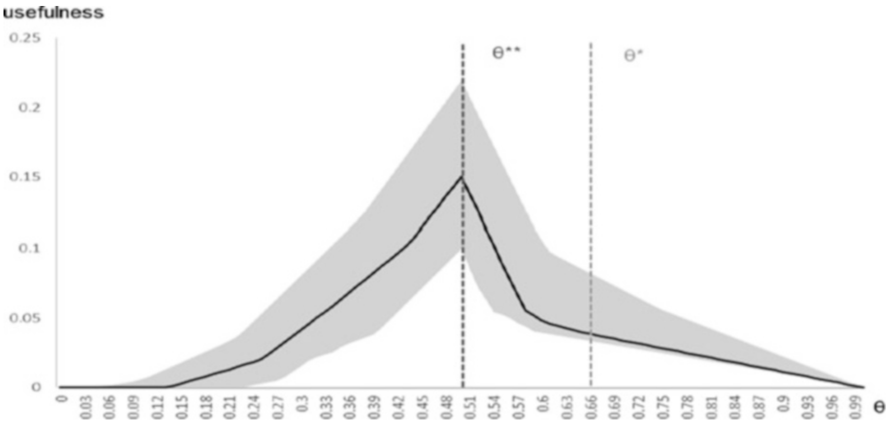


Fig. 1 Usefulness indicator for different preferences

of 0.5 represents a policymaker who is equally concerned about missing crises than issuing false alarms. The usefulness decreases with a shift in both directions from $\theta = 0.5$. This represents the increased difficulty of outperforming the static strategy in cases when the preferences are clear. In other words, the competitiveness of always (or never) issuing the signal strategy increases in case a policymaker clearly wants to avoid missing the crisis (or issuing a false signal).

Accordingly, the usefulness indicator is at its lowest for very low and very high θ. Its distribution is also asymmetric. A small increase in the preferences not to miss the crisis (θ rise from 0.5 to 0.58) sharply affects the usefulness, but a further increase in θ affects it only slightly. At the same time, as the policymaker's preferences to avoid false signals increases, the usefulness of the credit gap decreases more slowly, but from some point ($\theta < 0.13$), it makes no sense to use the signalling approach. We also calculate the confidence interval with bootstrap simulations. To do this, we construct 1000 samples, organized by randomly taking observations of the credit gap and the corresponding moment of the crisis (or its absence). For each obtained sample, we apply the signal approach, and, as shown earlier, we construct the usefulness function. The range of values is used to calculate the confidence band for Fig. 1. It shows significant uncertainty regarding the usefulness measure. After the severe financial crisis, bearing in mind the high costs of a financial crisis manifested in the form of large output losses, rising unemployment and huge public deficits, the literature conventionally assumes that decision-makers give the crisis detection preference a higher weight—that is, set $\theta > 0.5$ (see, for example, Detken and Smets 2004; Alessi and Detken 2018).

We may assume $\theta^* = 0.66$, meaning that the cost of missing a crisis is twice as high as the cost of issuing a false signal. At these preferences, the credit gap's usefulness is marginally positive. Let us now consider what will change after the introduction of the macroprudential regulation measures by Basel III and our assumptions about the credit gap performance in the new conditions. Presumably, the Basel III regulation (such as, for example, minimum static capital requirement

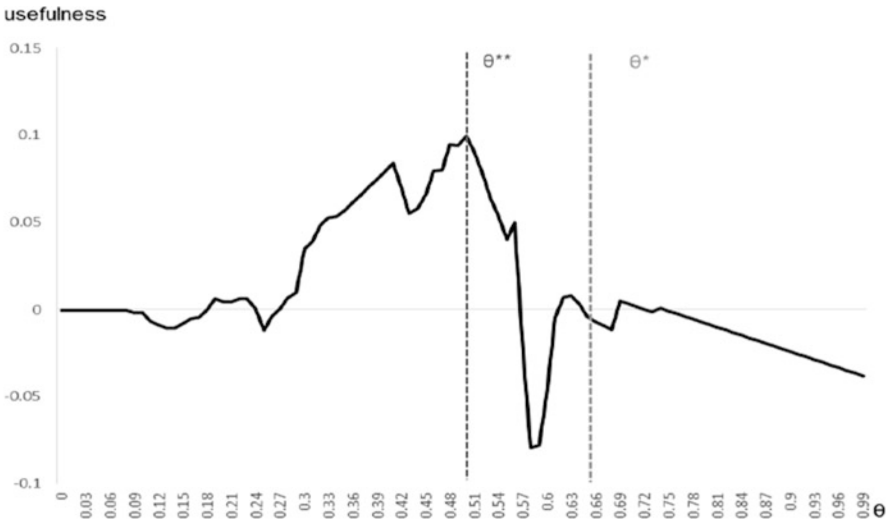


Fig. 2 Usefulness indicator for different preferences (out-of-sample)

CAR3) may reduce the severity of banking crises. For example, if we assume that the cost of a financial crisis will be halved after the implementation of Basel III's static capital requirements (see Appendix 2 for the justification of such assumption), we may assume $\theta^{**} = 0.5$. In these circumstances, the credit gap's usefulness increases significantly.

Next, we conduct a similar exercise to examine if the credit gap can be a good out-of-sample predictor of the crisis and apply the leave-one-out validation approach. Namely, we exclude data of one country (tested country) from the pool and estimate the optimal thresholds for each θ. These thresholds minimize losses for a truncated pool, and we expect that they are suitable for the tested country. Therefore, we evaluate A, B, C, and D (from Table 1) assuming that the signal is issued if the credit gap of the tested country exceeds the threshold estimated at the previous step and is not issued otherwise. This procedure is repeated 21 times, excluding all countries in turn. Finally, we estimate the loss function and usefulness function using the average values of the A, B, C, and D. The results are shown in Fig. 2. They show that the out-of-sample signals based on the credit gap indicators are useful only for θ from 0.28 to 0.56. According to these results, the gain in usefulness of the EWIs after the implementation of Basel III may be even more pronounced.

5 Conclusions

We analyse the ability of credit gap measures to predict banking crisis by estimating the usefulness measure conditionally on policymakers' preferences. The results show that the signals based on the credit gap indicators are most useful when the policymaker's preferences regarding Type I and Type II errors are approximately equal. However, according to the current consensus, the preferences to avoid missing a crisis are higher than issuing a false signal. The static Basel III measures may potentially lead to a decrease in the severity of the crises and, accordingly, reduce the cost of missing a crisis. Interestingly, this means that there is an interplay between static and dynamic Basel III regulation mechanisms and that the usefulness of the credit gap measure as an EWI and the efficiency of CCyB as a macroprudential measure in general are likely to increase once the static Basel III regulation measures are implemented.

Appendix 1: Dataset

Table 2 Cross-section of countries

Country	Time sample		Crises
	From	To	
Australia	Q1 1960	Q3 2016	–
Austria	Q1 1960	Q4 2016	2008–2012
Belgium	Q3 1970	Q3 2016	2008–2012
Canada	Q1 1960	Q3 2016	–
Denmark	Q1 1967	Q3 2016	2008–2009
Finland	Q3 1970	Q3 2016	1991–1995
France	Q3 1969	Q3 2016	2008–2009
Germany	Q1 1960	Q3 2016	2008–2009
Greece	Q1 1960	Q3 2016	2008–2012
Ireland	Q2 1976	Q3 2016	2008–2012
Italy	Q1 1960	Q3 2016	2008–2009
Japan	Q3 1964	Q3 2016	1997–2001
Korea	Q3 1962	Q3 2016	1997–1998
Netherlands	Q4 1960	Q3 2016	2008–2009
Norway	Q1 1960	Q3 2016	1991–1993
Portugal	Q1 1960	Q3 2016	2008–2012
Spain	Q4 1969	Q3 2016	1977–1981 2008–2012
Sweden	Q4 1960	Q3 2016	1991–1995 2008–2009
Switzerland	Q1 1960	Q3 2016	2008–2009
United Kingdom	Q4 1962	Q3 2016	2007–2011
United States	Q4 1951	Q3 2016	1988 2007–2011

Appendix 2: Modelling the Effect of Changes in Capital Requirements on Financial Crises' Severity

To assess the impact of the capital requirement introduction on the change in expected depth of recession or the severity of future crises, we use the model calibrated by Miles et al. (2013) to match historical experience dating back to almost 200 years.

The data are for the change in GDP per capita for a sample of 31 countries, and it starts, in some cases, in 1821 and lasts until 2008. The number of observations of annual GDP growth is almost 4500.

In line with Miles et al. (2013), we assume that the first difference of the log of per capita GDP (Y) follows a random walk with a drift and two random components. To capture capital requirement effect, we include an additional shock τ_t, which represents development banking insolvency as a response to the serious economic crisis. Like Miles et al. (2013), we assume that generalized falls in the value of bank assets are driven by changes in the level of incomes in the economy. Insolvency occurs when losses on bank assets exceed bank equity:

$$\log(A_t) = \log(A_{t-1}) + \gamma + u_t + v_t + \tau_t, \tag{3}$$

where A_t—income (or GDP), γ—average productivity growth. $u_t \sim N(0, \sigma^2)$ represents the standard shocks in normal times. v_t represents a financial shock. It equals zero in normal times, but make take a very large negative value -b with small probability p and symmetric shocks of lesser magnitude $\pm c$ with probability q:

$$v_t = \begin{cases} 0, \textit{with probability } (1-p-q); \\ -b, \textit{with probability } p; \\ +c, \textit{with probability } \dfrac{q}{2}; \\ -c, \textit{with probability } q/2. \end{cases} \tag{4}$$

The third shock τ_t represents the probability of an economic downturn becoming a full-scale systemic financial crisis. It links the value of capital adequacy ratio K and GDP losses. If banks have enough capital during a recession, the banking crisis does not occur ($\tau_t = 0$), but it will happen otherwise. We implement this assumption as follows:

$$\tau_t = \begin{cases} \delta * (\log(A_{t-1}) - \log(A_{t-2}) + K), \gamma + u_t + v_t + K < 0; \\ 0, \textit{oherwise}. \end{cases} \tag{5}$$

We set K = 3% for the benchmark specification. Other parameters are reported in Table 2.

Table 3 Model parameters

Description	Parameter	Value
Average productivity growth	γ	2.21×10^{-2}
Standard deviation of GDP growth	σ	3.5×10^{-2}
Annual probability of extreme financial shock	p	0.035×10^{-2}
Magnitude of extreme negative shock	b	-38×10^{-2}
Annual probability of standard financial shock	q	3.1×10^{-2}
Magnitude of standard financial shock	c	11×10^{-2}
Magnitude of financial crisis shock	δ	1.7×10^{-2}

Table 4 Statistics of artificial and empirical GDP growth rates

	Empirical	Artificial
Mean	1.81	1.85
Standard deviation	5.7	5.2
Skewness	−2.4	−2.6
Kurtosis	39	26

Under this parametrization, the model generates the distribution of GDP growth rates that is close to the empirical distribution reported by Miles et al. (2013). This comparison is reported in Tables 3 and 4.[4]

We proceed by conducting the following experiment. We change K from 3% to 10%, representing the increase in capital requirements in line with the Basel III recommendations. The new set of artificial GDP growth rates is computed, and several indicators of the severity of recessions in the alternative artificial datasets are compared.

The first indicator we calculate is the unconditional probability of observing a decline in GDP larger than a threshold P (we test P = 5% and P = 10%). The second indicator is the conditional probability of observing the decline larger than a threshold given that a recession takes place. The results are reported in Table 5. The estimates indicate that for P = 5%, the recession severity indicators are approximately halved when K is increased from 3% to 10%. The drop is even more significant if P = 10%. Arguably, these results may be regarded as a proxy for changes in the costs of a financial crisis under higher capital requirements. Accordingly, for the purpose of an early warning system's usefulness evaluation exercise, we assume that the losses associated with the Type I error (i.e. missing a crisis) may be twice as low under Basel III's capital requirements.

[4]The results presented in Tables 3 and 4 are based on 100,000 artificial observations.

Table 5 Severity of recessions under different capital requirements

	Unconditional probability of recession		Conditional probability of recession	
Threshold P	>10%	>5%	>10%	>5%
Empirical	0.025	0.07	0.092	0.258
Artificial (K = 3%)	0.026	0.057	0.096	0.212
Artificial (K = 10%)	0.006	0.033	0.023	0.121

References

Aikman, D., Bridges, J., Hoke, S. H., O'Neill, C., & Raja, A. (2019). *How do financial vulnerabilities and bank resilience affect medium-term macroeconomic tail risk? Bank of England Working Papers 824*. London: Bank of England.

Alessi, L., & Detken, C. (2011). Quasi real time early warning indicators for costly asset price boom/bust cycles: A role for global liquidity. *European Journal of Political Economy, 27*, 520–533.

Alessi, L., & Detken, C. (2018). Identifying excessive credit growth and leverage. *Journal of Financial Stability, 35*(C), 215–225.

Arregui, N., Benes, J., Krznar, I., Mitra, S., & Santos, A. (2013). *Evaluating the net benefits of macroprudential policy: A cookbook. International Monetary Fund IMF Working Paper No. 13/167*. Washington, DC: International Monetary Fund.

Basel Committee on Banking Supervision. (2010). *An assessment of the long-term economic impact of stronger capital and liquidity requirements*. Basel: BIS.

Basel Committee on Banking Supervision. (2019). *The costs and benefits of bank capital – a review of the literature. BIS Working Paper No 37*. Basel: BIS.

Behn, M., Gross, M., & Peltonen, T. (2016). *Assessing the costs and benefits of capital-based macroprudential policy. ESRB Working Paper Series No 17*. New York: Entertainment Software Rating Board.

Borio, C., & Lowe, P. (2002). *Asset prices, financial and monetary stability: exploring the nexus. BIS Working Papers No 114*. Basel: BIS.

Budnik, K., Affinito, M., Barbic, G., Ben Hadj, S., Chretien, E., Dewachter, H., Gonzalez, C. I., et al. (2019). *The benefits and costs of adjusting bank capitalisation: evidence from euro area countries. Working Paper Series 2261*. Frankfurt, Germany: European Central Bank.

Deryugina, E., & Ponomarenko, A. (2019). Determination of the current phase of the credit cycle in emerging markets. *Russian Journal of Money and Finance, 78*(2), 28–42.

Detken, C., & Smets, F. (2004). *Asset price booms and monetary policy. Working paper series 364*. Frankfurt, Germany: European Central Bank.

Drehmann, M., & Juselius, M. (2014). Evaluating early warning indicators of banking crises: satisfying policy requirements. *International Journal of Forecasting, 30*, 759–780.

Kaminsky, G., Lizondo, S., & Reinhart, C. (1998). *Leading indicators of currency crisis. Staff Paper 1*. Washington, DC: International Monetary Fund.

Laeven, L., & Valencia, F. (2018). *Systemic banking crises revisited. IMF Working Paper No WP/18/206*. Washington, DC: International Monetary Fund.

Miles, D., Yang, J., & Marcheggiano, G. (2013). Optimal Bank Capital. *The Economic Journal, 123*(567), 1–37.

Popoyan, L., Napoletano, M., & Roventini, A. (2017). Taming macroeconomic instability: Monetary and macro-prudential policy interactions in an agent-based model. *Journal of Economic Behavior & Organization, 134*(C), 117–140.

Does Only Volume Matter? A Stress Test for the Adequacy of International Currency Reserves for Russia

Renat Akhmetov, Vera Pankova, Oleg Solntsev, and Elizaveta Orlova

Abstract This paper seeks to determine the optimal volume of international currency reserves of the Bank of Russia to prevent adverse fluctuations of the Russian ruble exchange rate, causing a threat to financial stability. We create a system of models, taking into account the linkages between the dynamics of exchange rate and the behavior of economic agents—households, non-financial industries, and banks. Our empirical exercise allows to conclude that, with the occurrence of the most severe stress and the immediate provisions of currency liquidity by the Bank of Russia, the current volume of international reserves will be sufficient to eliminate its consequences. However, in case of retarded provisions of currency liquidity, the volume of highly liquid reserves will not be sufficient, forcing the Bank of Russia to sell a significant volume of foreign government securities. In this light, the Bank of Russia should change the structure of the international reserves in favor of highly liquid assets by reducing the share of securities and increasing the share of short-term deposits in foreign banks with high credit ratings. As for the volume of international reserves for Russia, including less liquid components, it is sufficient

All the authors are grateful to Dmitry Galimov, a leading expert of CMASF, for the preparation of methodology and implementation of calculations by industries. This study was prepared as part of the research project "Long-term targets for the development of the financial sector" made on request of the Bank of Russia.

R. Akhmetov · V. Pankova · O. Solntsev (✉)
The Center for Macroeconomic Analysis and Short-term Forecasting (CMASF), Moscow, Russia

The Institute of Economic Forecasting of the Russian Academy of Sciences (IEF RAS), Moscow, Russia

The Center for Fundamental Studies of National Research University Higher School of Economics, Moscow, Russia
e-mail: osolntsev@forecast.ru

E. Orlova
The Institute of Economic Forecasting of the Russian Academy of Sciences (IEF RAS), Moscow, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_13

to surpass the maximum possible stress in the foreign exchange market and to subsequently keep the Russian economy sustainable.

Keywords Exchange rate volatility · Stress testing · International currency reserves

JEL Classification E5 · G01 · G28

1 Introduction

After adopting the floating exchange rate regime at the end of 2014, the Bank of Russia's monetary policy envisages the application of operations with international reserves to ensure financial stability rather than to regulate the exchange rate. Its goal is to prevent an excessive (panic) demand in the foreign exchange market (unrelated to fundamental factors) and a subsequent surge in the ruble exchange rate volatility as well as shortages of foreign exchange liquidity. The shortage of foreign exchange liquidity has an irreversible and adverse impact on the financial sector, resulting in the loss of financial stability of a significant number of banks, the increase in the frequency of defaults on bank loans, and the rise in deposit dollarization.

Thus, since 2014, the accumulation of reserves by the Bank of Russia is driven by the so-called precautionary motive rather than mercantilist motive. This shift makes constructing models for the Russian economy, which are relevant to the current situation and based on cross-country panel data, much more challenging. This is due to the fact that most of the episodes of currency crises over the past 20 years have taken place in the countries that adjusted exchange rates and thereby followed the "mercantilist motive" of hoarding reserves (in particular, Asian countries).

The Bank of Russia does not directly intervene in the foreign exchange market, implementing only the tools of temporary provision or absorption of foreign currency liquidity (for example, foreign exchange REPO operations). It implies that in our study we focus on the role of the Bank of Russia as lender of last resort in terms of providing currency liquidity (e.g. Gopinath and Stein 2018). This paper seeks to determine the adequate volume of international currency reserves of the Bank of Russia to prevent detrimental fluctuations of the ruble exchange rate, which can undermine financial stability and cause a shortage of foreign exchange liquidity.

The recent currency crises of 2008–2009 and 2014–2015 in Russia show that under stress conditions panic demand for foreign currency is exhibited not only by banks, but also by non-financial companies and households. It implies that our models of demand for foreign currency should consider the behavior of several sectors of the economy in accordance with the approach of Čeh and Krznar (2009) as well as Gopinath and Stein (2018).

In addition, the experience of the recent crises emphasizes a highly significant role of large companies and banks both in aggravating the crisis and, conversely, in stabilizing the situation. The "switching" of large exporters and corporate debtors' operations in the foreign exchange market makes its dynamics nonlinear. Therefore, to ensure the relevance of our estimates, we build on the elements of agent-based modeling, following Aizenman and Lee (2007).

2 Literature Review

The studies on the adequacy of international reserves can be divided into two major strands, which differ with regard to the main motive for the accumulation of reserves.

The first line of research comes from the so-called mercantilist motive (Dooley et al. 2005, 2009; Aizenman and Lee 2008). This is about the accumulation of foreign exchange reserves to prevent the appreciation or depreciation of exchange rate in order to conduct industrial policy, e.g. to support export-oriented industries. This policy is typical for East Asian countries (Japan, South Korea, China, Hong Kong, etc.), which receive competitive advantages in foreign markets due to the undervaluation of their currencies. In line with such motive, the volume of international reserves is a consequence of exchange rate regulation, while the optimal amount of reserves per se is an insignificant issue.

In the second strand of research, the authors examine the so-called precautionary motive (Jeanne and Rancière 2011), i.e. the desire to accumulate foreign reserves as a safety net against the adverse impact of crises such as sudden stops of capital inflows, capital flights, and the increasing volatility of financial markets (Aizenman and Lee 2007). Researchers working in this field suggest that reserves serve as a cushion for the economy, which allows to reduce possible losses in the level of GDP, investment and welfare in case of currency volatility (Čeh and Krznar 2009). A number of studies find that the volume of international reserves is negatively correlated with the probability of crisis, i.e. the fact of accumulation of reserves reduces the likelihood of crisis events (Garcia and Soto 2006; Čeh and Krznar 2009; Gourinchas and Obstfeld 2012; Bussière et al. 2015).

The feature of this approach is that researchers are concerned about the optimal level of reserves, since an excessive amount of hoarding is associated with costs, and it can lead to the inefficient use of national savings.

The search for the optimal level of reserves can be carried out by comparing the costs and benefits of the country from owning them (Ben-Bassat and Gottlieb 1992; Garcia and Soto 2006; Čeh and Krznar 2009; Jeanne and Rancière 2011; Dabla-Norris et al. 2011; Calvo et al. 2012; Jeann and Sandri 2016). This approach involves minimizing the loss function of the following form:

$$\text{Loss} = \pi(R) \times f(R) + (1 - \pi(R)) \times \delta R, \tag{1}$$

where δR—opportunity costs from the allocation of reserves in reliable but low-income assets; $\pi(R)$—probability of a currency crisis or a crisis of external debt that is associated with the size of the accumulated reserves; $f(R)$—amount of economic losses caused by crisis (GDP, consumption, investment, etc.).

Caballero and Panageas (2008) proposed a different approach to finding the optimal level of reserves. The authors of this study develop a quantitative model of global equilibrium investigating the impact of the accumulation of reserves on the elimination of the consequences of a sudden stop of capital inflows. The researchers conclude that by using hedge financial instruments, countries are able to accumulate reserves more efficiently, i.e. avoiding excessive hoardings of reserves. In addition, the origin of a sudden stop and the correlation of its occurrence with other global events should be taken into account to determine the optimal level of reserves.

A separate research question in the "precautionary paradigm" refers to international reserves as a tool to maintain the liquidity of the banking system (i.e., the focus of this research is on the role of the Central Bank as lender of last resort). In this vein, Gopinath and Stein (2018) document that the size of foreign exchange reserves should be determined by the Central Bank, based on the level of dollarization of the banking system and the volatility of the exchange rate. The authors of this study consider a model which is populated with three types of economic agents—households, banks, and the Central Bank. It involves an assumption that in an economy with a larger share of imports denominated in foreign currency, the population's propensity to hoard foreign currency is higher, which in turn leads to a higher level of dollarization of the banking system. If the domestic currency depreciates, the probability of a banking crisis arises due to bank insolvency (for example, due to a mismatch between foreign assets and liabilities). Consequently, the central bank as lender of last resort should be able to provide the banking system with the necessary volume of foreign currency liquidity, which helps banks repay their liabilities. The higher the volatility of the exchange rate and the more adverse the repercussions of the crisis for households, the more the central bank should rely on the accumulation of international reserves.

Aizenman and Lee (2007) examine the sufficiency of international currency reserves, using an enhanced version of the model of bank runs (Diamond and Dybvig 1983). They consider an open developing economy, which is integrated into the international financial system and exposed to external liquidity shocks due to capital outflows. The authors show that under conditions in which banks finance long-term investment projects by attracting short-term deposits and when only the central bank acts as lender of last resort, the accumulation of international reserves is legitimized. This is due to the fact that the scale of potential costs caused by a liquidity shock and the subsequent outflow of deposits exceeds the opportunity costs for the central bank to store these reserves.

In case of Croatia, Čeh and Krznar (2009) find that central bank's strategy of reserve accumulation should depend on whether the "parent banks" of Croatian credit institutions with foreign capital will take on the role of a lender of last resort in the situation of a simultaneous sudden stop and a banking crisis.

In addition to the studies based only on one of the motives for international reserve accumulation, some researchers attempt to explain the long-term level of reserves by a wide range of economic, financial, and institutional factors using cross-country data (Cheung and Ito 2009; Obstfeld et al. 2010; Dominguez 2009; Dominguez et al. 2012). Herd behavior in the accumulation of reserves for certain groups of countries is also considered (Cheung and Qian 2009). Herd behavior means a more intensive accumulation of international reserves by the countries whose geographical neighbors are also hoarding reserves. Cheung and Qian (2009) uncover this effect for ten Asian economies. It is attributed to the aspiration of these countries to prevent a speculative attack on their national currencies in case of a crisis in neighboring countries.

3 A General Set-up of the Sufficiency Model of International Reserves

Since we focus on the cases of panic demand for foreign currency, we simulate only the demand, which significantly exceeds trend values, without considering the aggregate demand for foreign currency in the Russian foreign exchange market. This panic demand is understood as the effect of "fire sales" (Shleifer and Vishny 2011; Cont and Schaanning 2017) of assets denominated in rubles by economic agents (households and industries) to mitigate their losses.

The currency crisis in this research is modeled as an iterative process, so it allows us to evaluate the results of the Bank of Russia's currency liquidity injection. The analysis is based on stress-testing methodology, which assumes assessing the consequences of the occurrence of rare, but probable adverse shocks.

Our models study the consequences of both exogenous and endogenous stresses in the foreign exchange market, which occur due to a decline in oil prices. It is possible to distinguish between the fundamental consequences of an oil price shrinkage and currency depreciation expectations in the short-term period.

Two channels of panic demand for foreign currency liquidity are modeled. The first one is the conversion of previously accumulated ruble liquidity into foreign currencies by bank customers. The second is a surge of activity in foreign currency customer accounts (due to increased mistrust), which means the transfer of foreign currency from bank to bank (including foreign banks) and withdrawals. It is assumed that the adverse impact on financial stability through both of these channels can be offset by means of the short-term provision of foreign currency liquidity by the Bank of Russia. A plausible tool in this case is foreign exchange REPO transactions.

Thus, our approach involves a quantitative assessment of the volume of the Bank of Russia's liquidity provision, which is necessary to prevent the destructive effects of an explosive exchange rate depreciation. The short-term implications for the resilience of the Russian banking sector that arise from the crisis are also estimated. The most significant indicators of banking sector instability are the number of banks which become fragile (including the systemically important ones), the total deficit of equity in the banking system, and the total shortage of the liquidity.

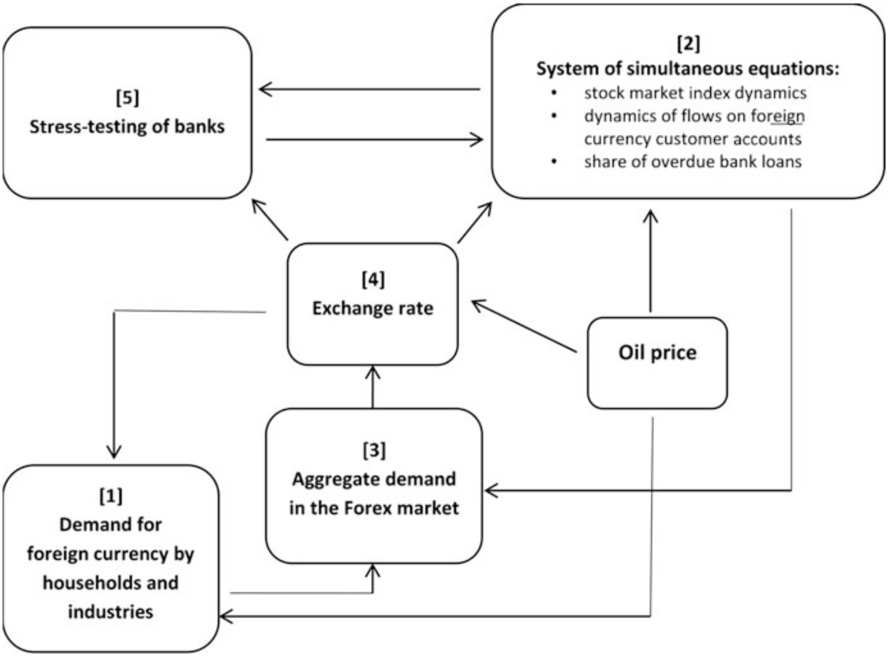


Fig. 1 Impact of stress in the foreign exchange market and an oil price shock on the demand for foreign currency by Russian industries, households, and the banking system

Figure 1 shows our approach to the simulation of stress in the Russian foreign exchange market and the short-term effects of this stress.

Block 1 refers to households' behavior and that of industries under our stress scenario. In this block, we construct the model, which captures the first channel of the adverse impact on the foreign currency liquidity of banks (the conversion of previously accumulated ruble liquidity into foreign currencies by bank customers). The model includes such individual characteristics of industries as profit, debt denominated in rubles and debt denominated in foreign currency, the volume of exports and imports, which determine the demand for foreign currency liquidity for each of the industries. If a sharp exchange rate depreciation occurs, then all industries exhibit panic demand for foreign currency, which leads to a further amplification of the exchange rate depreciation (the "fire sales" mechanism). Block 2 presents a system of simultaneous equations, which dissects the effect of exchange rate and oil price stresses on stock market dynamics, the share of overdue bank loans, the dynamics of foreign currency customer accounts in the banking system. The exchange rate and oil price shocks are exogenous parameters in our system of simultaneous equations. Block 3 in Fig. 1 assesses aggregate panic demand in the foreign exchange market, which is obtained as a result of the initial shock propagation through the two channels mentioned above. Block 4 is responsible for the exchange rate model, which accounts for: (1) aggregate panic demand for foreign currency liquidity, (2) reduction of oil prices, (3) monetary policy rate. Block 5 aims to stress test the banking system in terms of simultaneous ruble depreciation and falling oil prices. The stress test involves the "fire sales" mechanism for securities. If banks face a foreign currency liquidity shortage due to a mismatch between foreign assets and liabilities, they will try selling their securities (on the asset side) to buy foreign currency.

4 Modeling of the Stress Situation in the Foreign Exchange Market and its Short-Term Effects

4.1 Panic Demand for Foreign Currency by Households and Industries

Corporate demand for foreign currency is estimated, based on 15 key types of economic activity (industries). To perform econometric analysis for the period 2013–2017, we calculate the set of the following indicators, characterizing the differentiation of industries in terms of their elasticity to changes in the exchange rate:

- the share of exports in output (according to Russian Federal State Statistics Service data, Fig. 2);

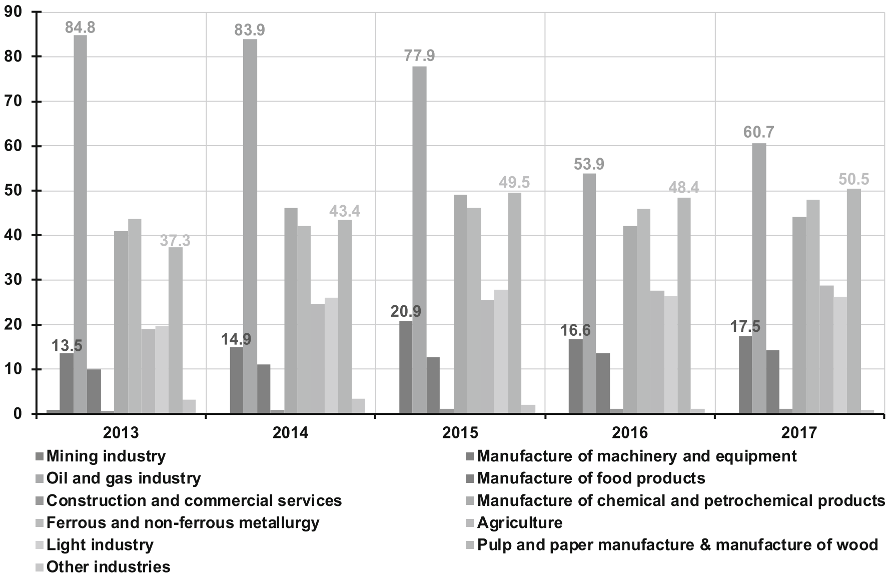


Fig. 2 Export share in output by industries, %

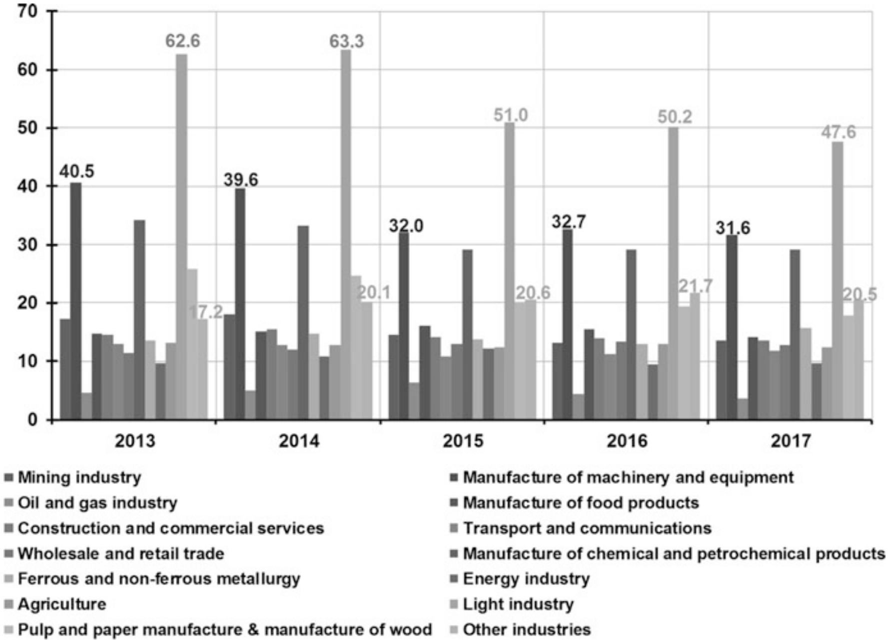


Fig. 3 Import share in tangible costs by industries, %

- the import share in tangible costs (according to Russian Federal State Statistics Service data, Fig. 3);
- the share of imports in fixed investment (according to Russian Federal State Statistics Service data, Fig. 4);
- corporate external debt—total debt on corporate eurobonds and international syndicated loans denominated in foreign currencies by industries (estimations based on the data from "Cbonds," Table 1).

The demand of companies and households for foreign currency is estimated by solving the optimization problem for each of these agents under the following assumptions:

1. The future annual expenses of the agents should not exceed the volume of future annual income. Expenditures include current expenses (tangible costs for companies and consumption for households), fixed investment (investments in fixed assets for enterprises and net investments in residential property for households), and upcoming repayments of debt.
2. Given a deterioration of the economic situation (ruble depreciation and fall in export prices) and a concomitant revaluation of the flow of forthcoming payments on foreign currency debt and changes in exports and imports, the expenses of companies may exceed their incomes. Consequently, companies and households should try to restore their balances, reducing their potential losses by purchasing foreign currency in advance. Enterprises purchase currency out of available funds

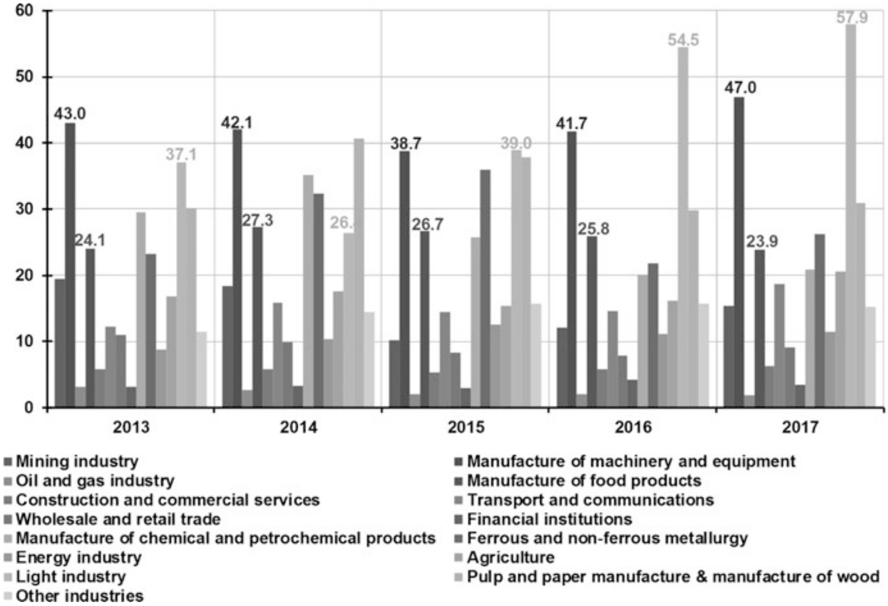


Fig. 4 Import share in fixed investment by industries, %

Table 1 Structure of corporate external debt[a] by industries, %

Industries	2013	2014	2015	2016	2017
Mining industry	4.6	5.4	6.2	4.2	6.1
Manufacture of machinery and equipment	0.5	0.5	0.6	0.5	0.6
Oil and gas industry	54.9	50.8	46.6	50	44.6
Manufacture of food products	0.1	0.1	0	0	0
Transport and communications	11.2	12.7	14	14.9	13.8
Construction and commercial services	0.9	1.1	1	1.1	0.9
Wholesale and retail trade	0.2	0.2	0.2	0.2	0.2
Financial institutions (except credit institutions)	2.6	2.1	1.5	1.3	1.8
Manufacture of chemical and petrochemical products	6.6	7.6	9.1	9.2	10.1
Ferrous and non-ferrous metallurgy	14.7	14.7	16.1	14.3	18.2
Energy industry	1.6	2.6	2.3	2.1	2.3
Agriculture	0.1	0	0	0	0
Light industry	0	0	0	0	0
Pulp and paper manufacture & manufacture of wood	0	0	0	0.2	0.5
Other industries	2.1	2.2	2.3	1.8	0.7

[a]External debt is debt to non-residents denominated in foreign currencies

generated by profits of the previous periods, and households—by converting short-term ruble deposits (up to 1 year) and a fraction of their ruble cash holdings, which are not used for current settlements.

Table 2 Estimation results for the model of the impact of current assets on the company's total revenue

Variables	Estimated coefficients
Current assets at the beginning of the year (coefficient *b*), mln. rub.	−0.150** (0.069)
Intercept	57.411*** (0.048)
Number of observations, thousands	13.9
R2 between	0.104

**Significant at 5% level; ***Idem, 1% level

3. Firms can use up to 100% of the profit of the previous period to purchase foreign currency without affecting their production process. A further increase in purchases of foreign currency leads to the reduction of current assets of companies:

$$\Delta\text{Current\_assets} = \left[\begin{array}{c} 0, \text{if purchase} \leq \text{profit}; \\ \text{purchase} - \text{profit, otherwise}; \end{array}\right. \tag{2}$$

where ΔCurrent _ assets—reduction in current assets; purchase—purchase of foreign currency by the company; profit—profit of the previous period.

Further, in the case of a reduction in current assets, enterprises face a loss. Moreover, incomes tend to decline even faster if the reduction further proceeds. We assume that the relationship between these two indicators at the firm level has a parabolic form specific to each industry.[1] Regression estimates on micro level with sectoral fixed effects are represented in Table 2. When constructing an aggregated loss function for each industry, we take into account the level of industry concentration:

$$\text{loss}_i = \Delta\text{Current}_{\text{assets}_i} * (\text{FE}_i + b * \text{Herfindahl}_i * \Delta\text{Current}_{\text{assets}_i}), \tag{3}$$

where loss_i—losses of industry i in case of current asset sales; FE_i—industry-specific fixed effect, i.e. reduction in income, which is proportional to the withdrawal of current assets; Herfindahl_i—Herfindahl index for industry i.

There are certain constraints on the purchase of foreign currency by households. Households cannot purchase foreign currency in excess of the volume of short-term ruble deposits (up to 1 year) and cash holdings (less current expenditures).

4. The panic demand of firms and households reinforces the initial ruble depreciation. It leads to a reassessment of the value of foreign trade transactions and payments on the foreign currency debt, and a possible further deterioration of expenses to income ratio. Then, the "fire sales" mechanism of ruble assets starts.

[1]The coefficients FE (include an intercept common to all industries and an individual effect of the industry) and *b* were estimated on the data for companies from 15 industries, with the following fixed effects model: revenue to current assets = FE – *b* * (current assets).

At each round of "fire sales," households and industries revise their demand for foreign currency to meet the following condition:

$$\begin{aligned} &\mathrm{Debt\_ser_{for}} * \mathrm{cur}_{\$} + \mathrm{Debt\_ser_{dom}} + \mathrm{import} * \mathrm{cur}_{\$} + \mathrm{invest} \\ &\quad \leq \mathrm{export} * \mathrm{prices_{export}} * \mathrm{cur}_{\$} + \mathrm{other} - \mathrm{loss} + \mathrm{purchase_{cur}} * \mathrm{cur}_{\$}; \end{aligned} \tag{4}$$

where $\mathrm{Debt\_ser_{for}}$—repayment of debt denominated in foreign currency; $\mathrm{Debt\_ser_{dom}}$—repayment of debt denominated in rubles; $\mathrm{cur}_{\$}$—USD/RUB exchange rate, growth rate; import—import volume; export—export volume; invest—volume of fixed investment, excluding imports of equipment; $\mathrm{prices_{export}}$—growth rate of export prices for raw materials; other—other income; loss—income loss in case of exaction of current assets; $\mathrm{purchase_{cur}}$—purchase of foreign currency.

4.2 *Modeling Secondary Shocks in the Financial Markets (the System of Simultaneous Equations)*

The system of simultaneous equations (SSE) allows to estimate the interactions of the following indicators: the dynamics of the Moscow Stock Exchange Index (IMOEX), the share of overdue bank loans, and the dynamics of foreign currency customer accounts in the banking system.[2]

The volume of "fire sales" of securities estimated in block 5, which is described above, is used to forecast the dynamics of the Moscow Stock Exchange Index (IMOEX). It helps to account for secondary effects of national currency depreciation. To this end, the indicator of "fire sales" of stocks on the Moscow stock exchange is included in the equation of the dynamics of IMOEX. This indicator is designed in such a way that when there are 1) an excess of trade volume on the Moscow stock exchange over its average level and 2) a drop in IMOEX by more than 5%, the indicator is equal to the value of this excess, and in all other cases it is equal to zero. The shocks in oil prices and the exchange rate are exogenous parameters in the SSE. These shocks lead to a fall of the stock index, an increase in the level of overdue bank loans, and a surge of activity in foreign currency customer accounts.

The SSE is estimated for the period November 2014–June 2018. The rationale behind the choice of this time span is that we aim to assess the effects solely for the floating exchange rate regime. The Central Bank of Russia adopted the floating exchange rate regime in November 2014. Accordingly, the Bank of Russia does not intervene to influence the ruble exchange rate under normal conditions and there is

[2]This indicator is calculated as growth rate of the volume of foreign currency customers' accounts in Russian banking system minus 3 standard deviations of this growth rate for a sample of banks at a given period. This indicator can be considered as a measure of the risk of outflow of foreign currency deposits for a bank.

Table 3 Estimation results for the system of simultaneous equations

Variables	Coefficients
Equation 1: Moscow stock exchange index (IMOEX), growth rate, %	
Oil price (Brent), growth rate, %	0.045*** (0.015)
Indicator of dynamics of foreign currency customer accounts in banking system ("–": Increase of activity, "+": Decrease of activity), %	0.077** (0.033)
Indicator of "fire sales" of stocks on the Moscow stock exchange, bln. rub.	–0.011*** (0.002)
MIACR rate[a] (lag = 1 month), %	0.148*** (0.013)
Equation 2: Share of overdue bank loans for banking system, %	
Moscow stock exchange index (IMOEX), growth rate, %	–0.141*** (0.021)
MIACR rate (lag = 1 month), %	0.835*** (0.006)
USD/RUB exchange rate, growth rate, %	0.093*** (0.018)
Equation 3: Indicator of dynamics of foreign currency customer accounts in banking system ("–": Increase of activity, "+": Decrease of activity), %	
Share of overdue bank loans for banking system, %	–0.438*** (0.011)
Moscow stock exchange index (IMOEX), growth rate, %	0.175*** (0.016)
Exchange rate USD/RUB, growth rate, %	–0.078*** (0.007)

The system was estimated with GMM (HAC)
**Significant at 5% level; ***Idem, 1% level
[a]Weighted Average Actual Rates on Moscow banks' loans

no need to keep additional international reserves for interventions. The results of the SSE estimation are represented below (Table 3).

4.3 Modeling the Volume of the Aggregate Panic Demand for Foreign Currency

We estimate the econometric model for the period November 2014–October 2018, but only crisis episodes[3] are considered (Table 4):

[3]We determine a currency micro-crisis as follows: more than 2.5% excess of the exchange rate of its trend level followed by 5% excess of currency demand of its trend level is considered as the beginning of the crisis (only if the crisis period has an economic interpretation). The end of the crisis is considered as the day of the last crisis event, after which there are no significant excesses of the exchange rate and currency turnover on the Moscow Exchange within 15 working days (2.5% and 5%, respectively). A month is considered a crisis month, if there is at least 1 day of crisis.

Table 4 Estimation results for the model of the aggregate panic demand for foreign currency

Variables	Estimated coefficients
Total demand of industries and households for foreign currency	0.262** (0.923)
Indicator of dynamics of foreign currency customer accounts in banking system, %	−5.804*** (1.542)
Number of observations	9
R2-adjusted	0.859

**Significant at 5% level; ***Idem, 1% level

Table 5 Estimation results for the model of the influence of panic demand for foreign currency on the exchange rate during periods of currency crises

Variables	Estimated coefficients
Panic demand, mln. dollars, ln	0.002*** (0.001)
Oil price (Brent), growth rate	−0.110 (0.079)
Weighted average actual rates on Moscow banks' ruble loans for one-day (lag = 1 day), %;	−0.002** (0.001)
Number of observations	97
R2-adjusted	0.12

**Significant at 5% level; ***Idem, 1% level

$$\text{demand}_t = 0.262\text{TD}_t - 5.804\text{volat}_t + \varepsilon_t, \tag{5}$$

where demand_t—aggregate panic currency demand; TD_t—total demand of industries and households for foreign currency; volat_t—dynamics of foreign currency customer accounts in banking system; ε_t—residual.

4.4 Modeling the Effect of Panic Currency Demand on the Ruble Exchange Rate

We estimate the regression of the following type:

$$\text{ex\_rate}_t = 0.002\text{demand}_t - 0.110\text{brent}_t - 0.002\text{MIACR}_{t-1} + \varepsilon_t, \tag{6}$$

where ex_rate_t—average daily exchange rate USD/RUB in the Moscow Exchange, growth rate; demand_t—panic currency demand (in logs); brent_t—oil price (Brent), growth rate; MIACR_{t-1}—weighted average actual rates on Moscow banks' ruble loans for one-day (MIACR); ε_t—residual.

The table below represents the estimation results (Table 5).

4.5 Stress Testing of the Banking System

The forecast of additional capitalization of the banking sector in case of a currency crisis is based on the modeling of bank balance sheets. The sample of banks includes more than 500 Russian credit institutions, representing about 95% of total assets of the banking sector.

Each bank's need for additional capital is estimated by sequentially calculating the following indicators:

1. Volume of securities portfolio—volume of government (GovSec) and corporate securities (shares and bonds, CorpSec), which are held by each bank for possible sale in order to offset the shortage of liquid funds. Besides, there are two types of securities—for trading and for investment.

A shrinkage in a corresponding market index (stock or bond) causes the revaluation of only trading securities and there are no changes for investment securities:

$$\begin{aligned} \mathrm{GovSec}_{i,t} &= \mathrm{GovSec}_{i,t_0} * (1 - \mathrm{GSIndexgr}_t); \\ \mathrm{CorpSec}_{i,t} &= \mathrm{CorpSec}_{i,t_0} * (1 - \mathrm{CSIndexgr}_t). \end{aligned} \tag{7}$$

2. Liquidity shortfall (LiqSh), which is defined as the volume of liquid assets—LA (absolutely liquid assets of a bank—AbsLA, and most liquid foreign assets—FA), less funds necessary for the bank to meet the minimum liquidity requirements (LiqRatio), as well as funds to cover a possible outflow of currency as a result of depreciation (FCDout).

The latter are calculated, taking into account the dynamics of foreign currency customer accounts in the banking system, obtained from the system of simultaneous equations (SSE).

3. Volume of "fire sales" of government and corporate securities (FireSales) should cover the deficit of liquid funds of a bank arising from the occurrence of currency depreciation and capital shock:

$$\mathrm{FireSales}_{i,t} = \left[\begin{array}{r} -\mathrm{LiqDef}_{i,t}, \text{if } \mathrm{GovSec}_{i,t} + \mathrm{CorpSec}_{i,t} \geq \left|\mathrm{LiqDef}_{i,t}\right|; \\ 0, \text{if else}. \end{array} \right. \tag{8}$$

4. Capital adequacy ratio, which is determined by the predictive values of equity (Capital) and risk-weighted assets (RWA). In this case, capital is adjusted for changes in the value of securities as a result of "fire sales."

Table 6 Parameters of the initial shock in stress scenarios

Initial shock/scenario	Significant shock	Maximum shock
Exchange rate growth, %	4.3	23.0
Oil price growth (Brent), %	−6.2	−16.4

In turn, risk-weighted assets are adjusted for the increase in the value of the bank's assets denominated in foreign currency (FCA) due to the national currency depreciation:

$$\mathrm{Capital}_{i,t} = \mathrm{Capital}_{i,t_o} + \Delta\mathrm{GovSec}_{i,t} + \Delta\mathrm{CorpSec}_{i,t};$$
$$\mathrm{RWA}_{i,t} = \mathrm{RWA}_{i,t_0} + \Delta\mathrm{FCA}_{i,t};$$
$$H1_{i,t} = {}^{\mathrm{Capital}_{i,t}}\!/_{\mathrm{RWA}_{i,t}} \tag{9}$$

5. Additional capital requirement—the forecast of this requirement for bank i in year t is calculated according to the formula below. In the meantime, it is expected that the bank's capital adequacy ratio should be at least 10%[4] in order to maintain an additional buffer for damping possible shocks:

$$\Delta\mathrm{Cap\bar{i}tal}_{i,t} = \mathrm{Capital}_{i,t} * \left({}^{10\%-H1_{i,t}}\!/_{H1_{i,t}}\right). \tag{10}$$

5 Results and Policy Implications

5.1 Description of Stress Scenarios

We consider two scenarios to assess the effects of a shock in the foreign exchange market coupled with an oil price shock. The first scenario involves a sizeable shock, while the second one assumes the strongest impact (Table 6). These types of shocks differ with respect to the scale of the initial shock and subsequent reaction of economic agents to it. These scenarios are based on the distribution of the daily exchange rate and oil price growth rates during the period of the floating exchange rate regime in Russia (since November 2014).[5]

[4] Capital adequacy ratio is equal to 8% under standard (non-crisis) conditions.

[5] The scenario with a sizeable currency stress assumes the depreciation of the ruble by 4.3%, which corresponds to the 99 percentile of the corresponding distribution, and the decline in oil prices by 6.2%, which corresponds to the first percentile of the appropriate distribution. In the scenario with the maximum currency stress, the dynamics of the corresponding indicators implies that their initial change is two times higher than the maximum/minimum growth rate.

After the initial shock, the oil price stabilizes at a new level and the exchange rate continues to change, in line with the models of economic agents' behavior described above, and the assumptions about the magnitude of the Bank of Russia liquidity provisions.

The iterative process of calculations involves the following steps: (1) assessing the initial stress in the foreign exchange and global oil market, (2) determination of the volume of the panic demand for foreign currency, (3) determination of the volume of the foreign currency liquidity provision of the Bank of Russia (through currency REPO operations and lombard lending in foreign currency), (4) determination of uncovered volume of panic demand for foreign currency, (5) change of the exchange rate. If the volume of foreign currency supply by the Bank of Russia fully covers the volume of panic demand, the ruble exchange rate returns to the level reached immediately after the initial shock.

5.2 Possible Duration of the Currency Shock Period

To assess how durable the initial stress in the foreign exchange market is, we look into historical episodes of similar currency shocks in emerging economies during 2008–2018. Currency micro-crises mean simultaneous depreciation of the exchange rate and excess of its trend level[6] by a certain country's specific value, taking into account its exchange rate dynamics in the historical perspective. This is a criterion for determination of the starting point of the crisis.

The end of the crisis is the last day of the crisis, after which, for 15 trading days, there are no significant deviations of the exchange rate from the fundamental values. Based on this criterion, five currency micro-crises are identified (Table 7). The average duration of the crisis is no more than a week (5.6 days).

5.3 Estimation of Highly Liquid International Reserves

The stress testing approach to estimate a sufficient amount of international reserves implies the most severe conditions for the currency crisis occurrence.

We assume that liquid international reserves consist of foreign currency cash excluding currency deposits of the Russian Government in the Bank of Russia, accounts and deposits in foreign banks with high credit ratings and securities of foreign governments. It is crucial to take into account the limited ability of the Bank of Russia to use the currency deposits of the Russian Government for the exchange rate stabilization.[7] In accordance with the international experience, currency crises

[6]Trend values of the exchange rates are calculated using the Hodrick-Prescott filter.

[7]Currency deposits of the Russian Government are additional oil and gas revenues and funds of the Russian National Wealth Fund, which are stored on special accounts in the Bank of Russia.

Table 7 Currency micro-crises around the world during 2008–2018

Country and crisis dates	Duration of crisis (trading day)	Criterion of national currency depreciation (%)	Criterion of exceeding the exchange rate its trend level (%)	Total depreciation of national currency (%)
Iceland, 1–10 October 2008	6	4	4	25
Argentina, 18 December 2015–11 January 2016	11	7	7	41.5
Mexico, 9–14 November 2016	4	4	10	11.5
Turkey, 13–14 August 2018	2	5	5	27.4
Argentina, 30 August–5 September 2018	5	7	7	23

characterized by fast propagation (6 days on average). Thus, the Bank of Russia may not have enough time to negotiate the use of this component of international reserves. According to our estimates, the volume of liquid foreign exchange reserves was about $220 billion for 1st January, 2019. Moreover, we provide the most conservative estimate of the international reserves (the volume of highly liquid reserves). This measure includes only the volume of foreign currency cash excluding currency deposits of the Russian Government in the Bank of Russia and accounts and deposits in foreign banks with high credit ratings. Securities of foreign governments are excluded, because they may become less liquid in case of their massive sales. The volume of highly liquid reserves was about $50 billion 1st January, 2019.

5.4 *Evaluation of the International Reserves Adequacy*

According to the results of stress testing (Table 8), in the situation of significant stress during 2018, and with late injections of foreign currency liquidity by the Bank of Russia, $92 billion would be required to curb the stress in the foreign exchange market. The volume of highly liquid foreign currency reserves (at the end of 2018 was about $50 billion) would be insufficient. However, a deficit of $10–15 billion in highly liquid reserves is not critical. For example, this deficit could be offset by selling securities of foreign governments without any risk of their depreciation. As for the maximum stress, in case of late provisions of foreign currency liquidity by the Bank of Russia, the volume of required international reserves ($140 billion) will be

Table 8 Stress test results for the 2018-year conditions

Indicators	In the absence of stress/non-crisis (01.07.2018)	The significant stress with immediate provisions by the Bank of Russia	The significant stress with late provisions by the Bank of Russia	The maximum stress with immediate provisions by the Bank of Russia	The maximum stress with late provisions by the Bank of Russia
The amount of highly liquid foreign currency reserves needed to overcome shocks, $ bln.	0	32.1	92.3	48.4	140.1
The growth rate of the USD/RUB, %	0.9	4.3	42.1	23.0	82.7
The volume of "fire sales" of securities carried out by banks to cover the shortage of liquid funds, bln. rubles	773	600	1049	785	1558
Number of banks that become unstable due to the disability to cover the liquidity and capital shortages	66	68	79	71	91
Uncovered shortage of capital of the banking system, bln. rubles	1727	1767	2335	1984	7912
Uncovered shortage of liquidity of the banking system, bln. rubles	184	172	212	184	243

two times higher than the current level of foreign currency reserves. This gap could not be closed by the sale of foreign government securities without the risk of their depreciation.

However, once the actions by the Bank of Russia are prompt, the need to provide the market with currency liquidity significantly decreases. The existing highly liquid foreign currency reserves of the Bank of Russia appear sufficient.

Under conditions of the maximum stress and immediate provisions of foreign currency liquidity by the Bank of Russia after the initial shock (the first iteration of the crisis process), only $48 billion of international reserves for elimination of the

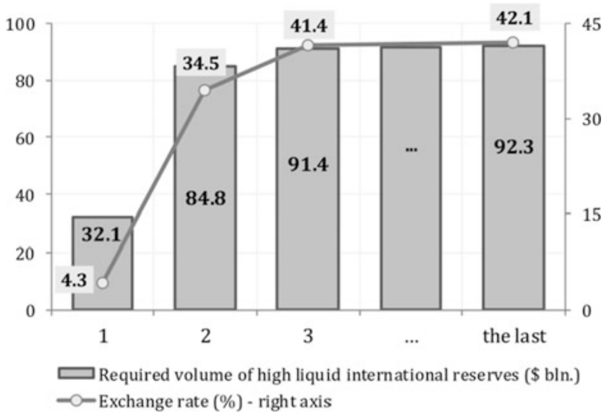


Fig. 5 Change in the USD/RUB exchange rate and the volume of highly liquid foreign currency reserves required to eliminate the shock effects, conditional on the iteration of crisis amplification (the sizeable shock scenario for conditions of 2018)

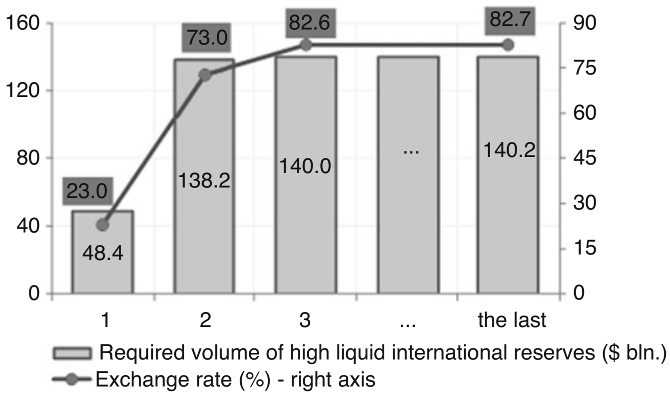


Fig. 6 Change in the USD/RUB exchange rate and the volume of highly liquid foreign currency reserves required to eliminate the shock effects, conditional on the iteration of crisis amplification (the maximum shock scenario for conditions of 2018)

consequences is to be provided. This is almost three times less than the volume of provisions in the last iteration (Table 8). If the regulator forbears the initial shock and the crisis process passes to the second iteration, the consequences are sharply amplified (Figs. 5 and 6). Thus, under conditions of the maximum stress, the costs of even an insignificant delay in the reaction of the Bank of Russia are very high.

The shift of international reserves towards more liquid assets (for example, short-term deposits in highly reliable foreign banks) is appropriate in case of the maximum stress and late reaction of the Bank of Russia. Based on the most conservative approach to the required volume of international reserves and after using a part of them to halt the crisis, the remaining part of international reserves should meet the sufficiency requirement on the basis of accepted international criteria. It will help to avoid the risk of a sovereign credit ratings decrease, as well as the occurrence of a secondary wave of the currency crisis, caused by a speculative attack on the ruble, due to expectations about the external insolvency of Russia. Based on the strictest benchmark—the Reddy criterion—the remaining part of the international reserves should be equal to $221 billion. Thus, the total volume of reserves (including less liquid components) which is sufficient to overcome both the maximum stress in the foreign exchange market and to meet the post-crisis requirements for their adequacy in accordance with international criteria, is estimated at $361 billion. The current volume of international reserves of Russia ($468 billion) is significantly higher than our estimates.

5.5 Estimation of the Stress Effects for the Banking System

Under the conditions of the sizeable stress and late provisions of liquidity by the Bank of Russia, 79 banks will face a lack of capital equal to 2.3 trillion rubles (see Table 8). This is by 0.6 trillion rubles more than a shortage of capital in the absence of stress. It means that an additional deficit will be equal to 6.4% of the total volume of capital of the Russian banking system. In the meantime, the number of affected banks will increase by 13, compared to a non-crisis situation. Four systemically important banks will need additional capitalization (by three banks more than in a non-crisis scenario). Uncovered shortage of capital for systemically important banks will be equal to 0.75 trillion rubles.

In case of the sizeable stress and immediate provisions of liquidity by the Bank of Russia, the magnitude of the necessary capitalization of the banking system will be comparable to the non-crisis scenario (total uncovered capital shortage will increase by only 40 billion rubles, with 75% of the shortage associated with two systemically important banks). Under the conditions of the maximum stress and late provisions of liquidity by the Bank of Russia, the total deficit of capital of the banking system will amount to 7.9 trillion rubles. This is 6.2 trillion rubles more than in the non-crisis scenario. The additional shortage of capital in this case will be equal to 65% of the total volume of capital of the Russian banking system, and 91 banks will need additional capitalization (including five systemically important credit institutions, with a total uncovered deficit of capital of 5.3 trillion rubles).

If the Bank of Russia provides foreign currency liquidity immediately in case of the occurrence of the maximum stress, the uncovered shortage of capital of the banking system will be equal to 2 trillion rubles. The need for additional funds for restoring financial stability will be experienced by 71 banks (including two systemically important ones).

6 Conclusion

The aim of this study is to determine the sufficient volume of international currency reserves of the Bank of Russia, which allows to prevent the adverse volatility of the ruble exchange rate, causing a threat to financial stability. We elaborate a system of models for the analysis of the relationship between the ruble exchange rate dynamics and economic agents' (households, non-financial companies and banks) behavior in the Russian foreign exchange market. We not only estimate the initial effects of stress in the foreign exchange and the global oil markets, but also examine the impact of the diffusion of these stress by quantifying changes in the demand for foreign currency of economic agents.

In this research, the development of iterative stress testing of banks plays a major role. This approach allows to capture the effect of "fire sales" of securities by banks with the shortage of liquidity or violation of capital adequacy of banks due to a mismatch between assets and liabilities denominated in foreign currency.

The stress testing reveals that the current volume of highly liquid international reserves is sufficient to avoid adverse consequences of the ruble depreciation, even in case of maximum stress. However, the current volume of highly liquid international reserves is inadequate, if the reaction of the Bank of Russia to the stress is retarded. In this case, the Bank of Russia will have to sell a significant amount of government securities of foreign issuers. It can lead to a price drop for these securities and additional losses in the reserves of the Bank of Russia.

In this light, the Bank of Russia should change the structure of its international currency reserves towards highly liquid assets, by reducing the share of securities and increasing the share of short-term deposits in foreign banks with high credit ratings. As for the volume of international currency reserves for Russia (including less liquid components), it is sufficient to overcome the maximum possible stress in the foreign exchange market and to subsequently maintain the solvency of the Russian economy.

References

Aizenman, J., & Lee, J. (2007). International reserves: Precautionary versus mercantilist views, theory and evidence. *Open Economies Review, 18*(2), 191–214.

Aizenman, J., & Lee, J. (2008). Financial versus monetary mercantilism: Long-run view of large international reserves hoarding. *The World Economy, 31*(5), 593–611.

Ben-Bassat, A., & Gottlieb, D. (1992). On the effect of opportunity cost on international reserve holdings. *The Review of Economics and Statistics, 74*(2), 329–332.

Bussière, M., Cheng, G., Chinn, M. D., & Lisack, N. (2015). For a few dollars more: Reserves and growth in times of crises. *Journal of International Money and Finance, 52*, 127–145.

Caballero, R. J., & Panageas, S. (2008). *A global equilibrium model of sudden stops and external liquidity management*. MIT Department of Economics Working Paper No. 08-05.

Calvo, G. A., Izquierdo, A., & Loo-Kung, R. (2012). *Optimal holdings of international reserves: Self-insurance against sudden stop*. NBER Working Paper No. 18219.

Čeh, A. M., & Krznar, I. (2009). *Optimal international reserves of the CNB with endogenous probability of crisis*. Croatian National Bank Working Paper No. W-21.
Cheung, Y.-W., & Ito, H. (2009). A cross-country empirical analysis of international reserves. *International Economic Journal, 23*, 447–481.
Cheung, Y.-W., & Qian, X. (2009). Hoarding of international reserves: Mrs Machlup's wardrobe and the joneses. *Review of International Economics, 17*(4), 824–843.
Cont, R., & Schaanning, E. (2017). *Fire sales, indirect contagion and systemic stress testing*. Norges Bank Working Paper 02/2017.
Dabla-Norris, E., Kim, J. I., & Shirono, K. (2011). *Optimal precautionary reserves for low income countries: A cost-benefit analysis*. IMF Working Paper No. 11/249.
Diamond, D. W., & Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. *Journal of Political Economy, 91*(3), 401–419.
Dominguez, K. (2009). International reserves and underdeveloped capital markets. *NBER International Seminar on Macroeconomics, 6*(1), 193–221.
Dominguez, K., Hashimoto, Y., & Ito, T. (2012). International reserves and the global financial crisis. *Journal of International Economics, 88*(2), 388–406.
Dooley, M., Folkerts-Landau, D., & Garber, P. (2005). *International financial stability: Asia, interest rates, and the dollar*. Deutsche Bank Global Research.
Dooley, M., Folkerts-Landau, D., & Garber, P. (2009). BrettonWoods II still defines the international monetary system. *Pacific Economic Review, 14*, 297–311.
Garcia, P., & Soto, C. (2006). Large hoardings of international reserves: Are they worth it? *External Vulnerability and Preventive Policies, 10*(1), 171–206.
Gopinath, G., & Stein, J. C. (2018). Trade invoicing, Bank funding, and central Bank reserve holdings. *AEA Papers and Proceedings, 108*, 542–546.
Gourinchas, P.-O., & Obstfeld, M. (2012). Stories of the twentieth century for the twenty-first. *American Economic Journal: Macroeconomics, 4*(1), 226–265.
Jeann, O., & Sandri, D. (2016). *Optimal reserves in financially closed economies*. NBER Working Paper No. 22139.
Jeanne, O., & Rancière, R. (2011). The optimal level of international reserves for emerging market countries: A new formula and some applications. *The Economic Journal, 121*(555), 905–930.
Obstfeld, M., Shambaugh, J. C., & Taylor, A. M. (2010). Financial stability, the trilemma, and international reserves. *American Economic Journal: Macroeconomics, 2*(2), 57–94.
Shleifer, A., & Vishny, R. (2011). Fire sales in finance and macroeconomics. *Journal of Economic Perspectives, 25*(1), 29–48.

Regulatory Measures Against Systemic Risk in Banking Sector: The Evidence for the Republic of Belarus

Svetlana Malykhina

Abstract This paper discusses the framework of systemic risk assessment and monitoring in the Belarusian banking sector. It involves comparisons with similar approaches in Russia and Kazakhstan, showing that these countries are generally keen to adopt the tools proposed by the Basel Committee on Banking Supervision. As for the Republic of Belarus, standard risk management instruments so far have sufficed to prevent risk propagation, while the need for a proper legal definition and the demarcation of systemic risk is emphasized.

Keywords Risk management · Systemic risk · Risk assessment · Capital buffer · Corporate governance · Supervisory board · Remuneration committee · Regulatory measures

JEL G01 · G32 · G34 · G38

1 Introduction

One of the lessons of the economic and financial crisis of the late 2000s was the increased attention of regulators to systemically important financial institutions that have had a profound impact on the banking sector of all countries and their economies as a whole. Such attention arises from the inability of banks to take into account the effect of their actions on other players in the banking system, which adversely affects the ability to manage systemic risk. Systemic risk was previously understood as the probability of contagion, causing a cascade of defaults. The crisis revealed that, apart from contagion, systemic risk is due to a common shock, leading to simultaneous defaults by several financial institutions, as well as a behavioral

S. Malykhina (✉)
Belarusian State Economic University, Minsk, Belarus

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_14

aspect, i.e. the dissemination of unfavorable information about one bank, causing an increase in the cost of refinancing for other banks.

In order to promote greater sustainability of the banking sector, the Basel Committee on Banking Supervision (BCBS) developed a comprehensive set of reforms, which attach particular importance to the measures aimed at limiting systemic risk. For example, a methodology was proposed to assess the degree of systemic significance for global and domestic systemically important banks (G-SIBs and D-SIBs), which introduced the requirement for such banks to maintain additional capital buffers within the Basel III standards. Thus, individual elements of the risk management system were introduced, namely: the assignment of banks to one or another group of systemic importance by the regulator, based on established criteria, systemic risk measurement and its mitigation, e.g. by creating an additional buffer for systemically important banks and forcing all banks to have a countercyclical capital buffer to prevent a systemic risk build-up during the periods of fast credit supply growth. The international standards regulating the activities of participants in financial markets have been introduced, taking into account the principle of proportionality, i.e. depending on the significance of a financial institution in the international or domestic market, the scale and nature of its activities, the risks taken and their materiality, as well as the state of the market. The introduction of new standards in relation to systemic risk in countries with developing economies, including the Republic of Belarus, revealed a number of difficulties in adopting systemic risk management in commercial banks. The conventional risk management involves a legal definition of risks. A specific feature of systemic risk is the difficulty of identifying it at the level of an individual participant in the financial market; therefore, the term systemic risk is interpreted widely from the standpoint of the sources and factors of its occurrence as well as market participants exposed to this risk. Below are the definitions of international regulatory organizations, for which the interpretation of this risk as a disruption of the financial system is common.

The International Monetary Fund (IMF), Financial Stability Board (FSB), and Bank for International Settlements (BIS) define systemic risk as the risk of disruption to financial services that is caused by an impairment of all or parts of the financial system and has the potential to have significant negative consequences for the real economy (IMF 2009). The European Systemic Risk Board (ESRB) defines systemic risk as the risk of disruption in the financial system with the potential to have serious negative consequences for the internal market and the real economy. All types of financial intermediaries, markets, and infrastructure may be potentially systemically important to some degree. Nonetheless, national regulators set requirements for risk management not for the financial market as a whole, but for its specific participants (primarily, banks), and determine them from the standpoint of costs (losses) for these participants from such disruptions. For example, the Bank of Russia established a requirement to reduce systemic risk of clearing institutions and legally defined it as the risk of costs (losses) of a clearing organization, for which the inability to meet its obligations by one or more financial market participants causes its inability to meet its obligations appropriately before other financial market participants (Bank of Russia 2015a).

2 Systemic Risk Assessment in the Financial Sector: Peculiarities and Challenges

In the Republic of Belarus, the definition of systemic risk in banking has not been established legally. In order to integrate management of this risk into the overall risk management system of a bank, it is proposed to define systemic risk as the risk of a bank having losses, due to the disruption of the sustainable functioning of the banking system caused by the poor financial conditions of its participants or the termination of their obligations. Identifying this risk involves identifying its main sources. Given the nature of systemic risk, i.e. impact on the banking sector and on an individual bank, it is feasible to apply two complementary approaches—macro and micro. The macro-approach focuses on the dynamics of macro-indicators to identify possible bubbles in the economy, while the micro-approach focuses on the state of individual banks. It should be noted that an individual bank can be both subject to systemic risk and serve as a source. Thus, based on the current practice, the main source of systemic risk in the banking sector of the Republic of Belarus can be attributed to:

- structural imbalance in the organization of the banking sector due to the existence of systemically important banks (SIBs);
- size (scale of activity) of individual SIBs;
- general exposure of individual banks to risks, arising from the common behavior of banks in the financial market;
- excessive risk taking by banks (credit, FX, liquidity) during the growth phase of the economic cycle;
- presence of direct or indirect ties that create the possibility for the transfer of risks from one bank to another (contagion effect).

The presence of the two mentioned approaches (macro- and micro-) in the identification of systemic risk implies different approaches to its assessment and monitoring. To assess the systemic risk exposure of the banking sector as a whole (macro-approach) and its monitoring, various sets of indicators and tools are used, which are now commonly referred to as macro-prudential. For example, in an IMF study of international experience in ensuring financial stability, approaches to monitoring systemic risk were studied in 63 countries (IMF 2011). The following most frequently used indicators were identified:

- share of overdue loans of the banking sector (NPL ratio);
- ratio of liquid banking assets and short-term liabilities;
- indicators of FX risk and risk related to capital flows with respect to emerging markets;
- financial leverage with respect to developed countries.

88% of countries use quantitative models and tools to identify, assess and analyze systemic risk, among which the following are commonly used:

- early warning systems for financial crises;
- asset pricing and real estate pricing models;
- models of “contagion effect” in the interbank lending market;
- models of macro-financial links;
- risk models of an individual institution (which are the most common, used in 55% of countries);
- stress testing.

Are these indicators used only for systemic risk assessment? Is there a reference set of systemic risk indicators suitable for all countries? Obviously not. Each country, when developing its own system of indicators, takes into account (or at least should take into account) the peculiarities inherent in its markets and banking sector. Thus, in the Republic of Belarus, the National Bank uses an aggregate systemic risk index (ISR-index) to obtain a quantitative assessment of banking sector vulnerability to the temporal dimension of systemic risk associated with the accumulation of discrepancies in the economy and the monetary sphere (National Bank of the Republic of Belarus 2017a). These discrepancies, in particular, include foreign trade imbalances, disruptions in the domestic foreign exchange market and excessive lending to the economy, exceeding the capacity of the real sector to pay back borrowed resources.

Therefore, the ISR-index incorporates such variables as credit gap (deviation of the current level of loans issued to the economy from the long-term equilibrium trend), the level of systemic liquidity (the ratio of interbank loans to deposits), financial leverage and capital flows ratio (ratio of banks’ liabilities from non-residents to existing claims to non-residents). The greater the positive value of the ISR-index, the more serious the imbalances accumulated in the economy and the higher the level of systemic risk of the banking sector. The negative value of the index indicates the absence of systemic banking risk, a zero value indicates that the systemic risk factors are on their equilibrium paths, and the situation can be regarded as a stable. The dynamics of the ISR-index are volatile, which is consistent with the financial cycle patterns in the Republic of Belarus. As at the end of 2017, the ISR-index was negative for the first time in 5 years (National Bank of the Republic of Belarus 2017b). This was due to the improved financial reputation of the Republic of Belarus, as evidenced by the revision in the Standard & Poor’s and Fitch Ratings agencies of the country’s credit rating to “B”, and Moody’s Investors Service to “B3”, “stable” forecast.

As for the systemic risk assessment and monitoring at the level of an individual bank (micro-approach), the Basel Committee recommended assessing the systemic importance of banks and establishing additional capital buffers for them to curb systemic risk. The Basel Committee founded to identify G-SIBs and D-SIBs laid the grounds for its assessment in different countries (Table 1).

Thus, in the Republic of Belarus, in view of the absence of G-SIBs, the regulator was guided by four categories of indicators used for D-SIBs (size, interconnectedness, substitutability, complexity). The four categories in the Belarusian methodology are made up of nine indicators that have different weights in the aggregate score.

Table 1 Assessment of the systemic importance of banks: a comparison of different banking indicators

Basel Committee on Banking Supervision		National Bank of the Republic of Belarus		National Bank of Kazakhstan (NBK)		Central Bank of the Russian Federation (Bank of Russia)	
Category Weight	Indicator Weight	Category Weight	Indicator Weight	Category Weight	Indicator Weight	Category Weight	Indicator Weight
Size 20%	Total exposures 20%	Bank scale 25%	Share of claims recorded on balance sheet accounts and liabilities recorded on off-balance sheet accounts of the bank (contingent liabilities and obligations under transactions) exposed to risks (assets and off-balance liabilities at risk) in total assets and off-balance liabilities at risk of banks 15%	Bank size 40%	Share of bank assets in the total volume of banks' assets 20%	Size of credit organization 50%	Percentage (share) of assets of a credit institution/ bank in the total assets of banking system 50%
			Share of regulatory capital of the bank in the total volume of regulatory capital of banks 10%		Share of bank liabilities in the total volume of banks' liabilities 20%		

(continued)

Table 1 (continued)

Basel Committee on Banking Supervision		National Bank of the Republic of Belarus		National Bank of Kazakhstan (NBK)		Central Bank of the Russian Federation (Bank of Russia)	
Interconnectedness 20%	Intra-financial system assets 6.67%	Interconnectedness with resident banks 16%	Share of funds placed by the bank in other banks in the total volume of funds placed by banks in other banks 8%	Interconnectedness of bank with participants in the financial market 20%	Share of interbank assets (including contingent assets) and bank investments in subsidiaries in the total volume of interbank assets and banks' investments in subsidiaries 5%	Interconnectedness with credit and other financial organizations funds placed 12.5%	Percentage (share) of a credit institution's claims on credit and other financial organizations in the total amount of funds placed in banks and other financial institutions 12.5%
	Intra-financial system liabilities 6.67%		Share of funds attracted by the bank from other banks in the total volume of funds attracted by banks 8%		Share of interbank liabilities (including contingent liabilities of the bank) to banks and pension assets of Unified Accumulation Pension Fund invested in bank deposits and securities	Interconnectedness with credit and other financial organizations funds raised 12.5%	Percentage (share) of a credit institution's liabilities to banks and other financial organizations in the aggregate amount of funds raised from banks and other

					issued by the bank in the total volume of banks' interbank liabilities and pension assets of that Fund invested in deposits in banks and in securities issued by banks 5%		financial organizations 12.5%
	Securities outstanding 6.67%				Share of deposits of households placed in the bank subject to guaranteeing by Kazakhstan Deposit Insurance Fund in the total volume of deposits of individuals placed in banks to be guaranteed by the Fund 10%	The volume of deposits of individuals 25%	Percentage (share) of deposits placed by individuals in a bank on the basis of a bank deposit agreement (including saving certificates) or a bank account agreement in the total volume of deposits of individuals placed in banks 25%

(continued)

Table 1 (continued)

Basel Committee on Banking Supervision		National Bank of the Republic of Belarus		National Bank of Kazakhstan (NBK)		Central Bank of the Russian Federation (Bank of Russia)
Substitutability/ financial institution infrastructure 20%	Payment activity 6.67%	Significance of bank for the economy 39%	Share of funds attracted by the bank from individuals in the total volume of funds attracted by banks from individuals 15%	Substitutability of bank 20%	The share of total bank payments made through the interbank money transfer system. the clearing system of Kazakhstan Interbank Settlement Center of the NBK. payments in the electronic banking services market (in the bank's network). payments and transfers made through correspondent accounts opened between the bank and its counterparties. through the	

					international money transfer systems (hereinafter referred to as non-cash payments). in the total volume of non-cash payments of banks 10%	
	Assets under custody 6.67%		Share of funds attracted by the bank from business entities in the total amount of funds raised by banks from business entities 9%		Share of bank loan portfolio in the total loan portfolio of banks 7%	
	Underwritten transactions in debt and equity markets 6.67%		Share of the bank's claims on business entities and individuals (requirements for the economy) in the total amount of banks'		Share of assets accepted by the bank for custodial services in the total volume of assets accepted by banks for custodial services 3%	

(continued)

Table 1 (continued)

Basel Committee on Banking Supervision		National Bank of the Republic of Belarus		National Bank of Kazakhstan (NBK)		Central Bank of the Russian Federation (Bank of Russia)
			claims on the economy 15%			
Complexity 20%	Notional amount of OTC derivatives 6.67%			Complexity of bank operations 20%	Share of contingent claims of a bank on derivative financial instruments and foreign currency in the total volume of conditional claims of banks on derivative financial instruments and foreign currency 5%	
	Trading and AFS securities 6.67%				Share of the bank's contingent liabilities on derivative financial instruments and foreign currency in the total volume of banks' conditional	

					liabilities on derivative financial instruments and foreign currency 5%	
	Level 3 Assets 6.67%				Share of the total amount of securities recorded by the bank at fair value through profit or loss and securities held by the bank available for sale in the total volume of securities recorded by banks at fair value through profit or loss and securities held by banks available for sale 10%	
Cross-jurisdictional activity 20%	Cross-juris-dictional claims 10%	Interconnectedness with non-resident banks 20%	Share of bank claims on non-residents in the total			

(continued)

Table 1 (continued)

Basel Committee on Banking Supervision		National Bank of the Republic of Belarus		National Bank of Kazakhstan (NBK)	Central Bank of the Russian Federation (Bank of Russia)
			volume of banks claims on non-residents 10%		
	Cross-jurisdictional liabilities 10%		Share of funds attracted by the bank from non-residents in the total volume of funds attracted by banks from non-residents 10%		

Sources: based on Basel Committee on Banking Supervision (2013), Bank of Russia (2015b), National Bank of the Republic of Belarus (2017c), National Bank of Kazakhstan (2014)

The category with the greatest weight is the importance of a bank for the country's economy (39%), which includes claims and liabilities of legal entities and individuals, which emphasizes the regulator's special attention to the influence of a bank on the credit risk and liquidity of the sector. The category with the lowest weight is interconnectedness with resident banks (16%), which characterizes the impact of a bank on domestic interbank lending operations. The size of the bank's activities (25%) includes the all balance sheet assets and off-balance sheet liabilities at risk, as well as regulatory capital. The interrelation with non-resident banks (20%) captures the involvement in international financial relations, characterized mainly by the attraction of foreign investment and the use of standard settlement instruments. Due to the relatively small volume of transactions in over-the-counter markets, transactions with complex derivatives, the provision of custodial and other special financial services, special indicators were not set for them.

If the aggregate score of systemic significance is more than 5%, the bank belongs to group I, if from 1 to 5%—to group II. These banks must maintain a buffer of systemic significance in addition to core capital adequacy ratio. The size of the buffer is set in line with the proportionality principle, i.e. 1.5 and 1% for these groups, respectively. The list of such banks is annually updated by the National Bank and posted on its website. In 2018, more than half of the 24 existing Belarusian banks were categorized as systemically important, including seven in group I, and six in group II. Since not only banks, but also other financial institutions can generate systemic risk, it is advisable for the regulator to extend to them the requirements to create a systemic importance buffer. In the Republic of Belarus, the National Bank has the right to establish a buffer of systemic importance for such organizations, based on reasoned supervisory judgment.

The approach to assessing the systemic importance of financial institutions used in Kazakhstan has a more detailed structure and a different calibration of weights in the aggregate score. Four categories consist of eleven indicators. Bank size has the largest weight (40%), which covers the amount of all assets and liabilities, while the other three categories have the same weight (20%). A specific indicator, such as the amount of deposits of individuals guaranteed by the Kazakhstan Deposit Insurance Fund (10%), is included in the category of interrelatedness of a bank with financial market participants, while interbank liabilities cover not only banks, but also the country's Unified Accumulation Pension Fund.

The category of the interchangeability of a bank includes three indicators—bank non-cash payments through all settlement (transfer) systems operating in the country, and the size of loan portfolio and assets accepted for custodial services. The three indicators included in the category of complexity of the operations carried out by the bank mean the active conduct of transactions with derivative financial instruments and securities by the country's banks. A bank is considered systemically important if the aggregate score is 10% or more. The list of such banks is released annually by the regulator, and they must maintain a systemic importance buffer in the amount of 1% of the assets and contingent risk-weighted liabilities, in addition to capital adequacy ratios.

In Russia, where financial markets are developed better, a less complex structure for assessing systemic significance is used. Its four categories are not split into separate indicators. The size of a credit organization has the largest weight (50%), followed by the volume of household deposits (25%). The categories of interconnectedness with financial organizations by funds placed and raised have the same weight—12.5% (see Table 1).

The Bank of Russia includes organizations whose aggregate score exceeds 1% of the total aggregate score of all banks and financial institutions into the list of systemically important entities. For banks included in the list, the minimum allowable buffer for systemic importance is 1% of risk-weighted assets, which should be provided from the basic capital to comply with the requirements for bank capital adequacy ratios.

Thus, by following the Basel Committee non-binding standards, such countries, as Belarus, set additional requirements (systemic importance buffer as well as countercyclical buffer) for their banks, thereby increasing the capital burden. Is this measure really effective against systemic risk? Will it be sufficient to support the sector if this risk is realized? Thanks to the policies of the National Bank of the Republic of Belarus, no systemic crisis has occurred in the banking system, though isolated cases of bank failures took place for individual reasons, mainly due to improper management. Thus, there has been no practical test for the effectiveness of the systemic importance buffer in Belarus.

3 Quantitative and Qualitative Tools to Limit Systemic Risk at the Bank Level

It is likely that in developing countries with a relatively small banking sector and a powerful regulator, other prudential tools can serve as effective precautionary measures, limiting conventional bank risks and preventing them from developing into the systemic one.

Thus, the National Bank of Belarus applies instruments to limit the indebtedness of individuals, which include two specific indicators:

Debt service ratio, representing the percentage ratio of monthly loan payments to the volume of the borrower's average monthly income. It is calculated prior to granting a consumption loan and should not exceed 40%. In case of excess, the debt on such loans should not exceed 10% of the total amount of debt to the bank on consumer loans;

Loan-to-value ratio (LTV ratio), representing the percentage of the loan relative to the value of the property accepted as collateral and/or the amount of other collateral in accordance with the contract. It is calculated before granting a loan to finance real estate and should not exceed 90%. In case of excess (up to 100%), the debt on such loans should not exceed 10% of the total amount of debt to the bank on loans to finance real estate (National Bank of the Republic of Belarus 2019).

In 2019, the National Bank of Belarus introduced measures to limit the systemic risk generated by business models of banks with increased risk appetite. For banks that implement such business models, increased regulatory requirements are applied with respect to capital adequacy as well as special and mandatory provisioning.

As an indicator of the increased business risk implemented by banks, the excess of interest rates set by banks on new deposits, loans and issued bonds over the corresponding estimated values of standard risk (EVSR) is used (National Bank of the Republic of Belarus 2019).

The EVSR is calculated monthly by the National Bank based on the average interest rates on six financial instruments offered by systemically important banks assigned to significance group I in domestic credit and deposit markets. These instruments including demand and term deposits of households (broken down into three groups by their placement periods, ranging from 1 month to over 1 year), new loans provided to households and legal entities. The calculated values of the EVSR are published on the National Bank website.

More subtle tools are used in countries with more developed financial markets. For example, in order to prevent excessive growth in household indebtedness and increase the resilience of banks to potential systemic risks, the Bank of Russia applies increased add-ons to risk weights for mortgage loans with low LTV values and for unsecured consumer loans. Similar add-ons apply to the risk weights for loans denominated in foreign currency (Bank of Russia 2019).

A set of measures to limit any type of risk involves not only the use of quantitative, but also qualitative tools. In terms of systemic risk, such an instrument was proposed by the Basel Committee, which stresses the need of establishing remuneration committees under the supervisory boards in systemically important banks. These committees should support the board in overseeing the remuneration system design and operation to ensure that remuneration is appropriate and consistent with the bank culture, long-term business and risk appetite, performance and control environment as well as with any legal or regulatory requirements (BCBS 2015).

It is emphasized that the remuneration committee should be constituted in a way that enables it to exercise competent and independent judgment on remuneration policies and the incentives they create. The remuneration committee works closely with the bank risk committee in evaluating the incentives created by the remuneration system. It also draws attention to the need for close interaction between this committee and the risk committee under the supervisory board. Without prejudice to the tasks of the remuneration committee, the risk committee should examine whether incentives provided by the remuneration system take into consideration risk, capital, liquidity and the likelihood and timing of earnings.

The recommendations of the Basel Committee are incorporated into the requirements of the National Bank of the Republic of Belarus to corporate governance procedures in financial institutions. Since 2017, the supervisory board is to create a remuneration committee in systemically important banks, which is headed by an independent director (see Fig. 1). The functions of the committee include monitoring the decisions made regarding the remuneration and compensation system (RCS), as well as evaluating the compliance of this system with established requirements.

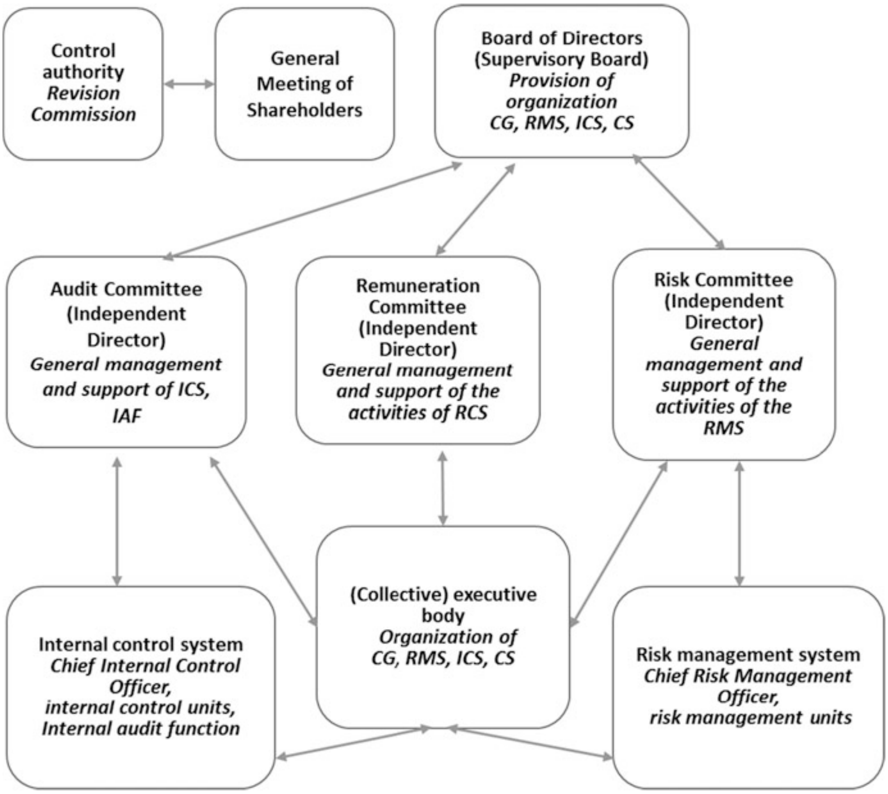


Fig. 1 Corporate governance organization in Belarusian banks. Source: National Bank of the Republic of Belarus (2012)

In addition, the remuneration committee defines a list of risk takers. The list includes bank employees who make decisions about the conduct of transactions, whose outcomes can significantly affect the level of risks taken by the bank and/or to lead to a situation threatening the sound functioning of the bank and the interests of its depositors and other creditors.

Therefore, the list must include the chair of the board and all members of the collective executive body of the bank, the chief accountant, the chief risk officer, the chief internal controller, and the chief auditor. The terms of remuneration to those included in this list are determined by the supervisory board and/or by the general meeting of the bank shareholders.

Thus, the regulator orders the management of systemically important banks to directly link the remuneration of those responsible for risk taking with their risk appetite. Such personalized measure applied to the employees of systemically important banks served as an effective addition to quantitative measures applied to curb the systemic risks of banks.

When organizing systemic risk management at the bank level, it becomes necessary to integrate it with the risk management of settlement (payment) systems in which the bank is participating. Unlike the systemic risk of a bank, the risks of payment systems are well studied. The main risks include operational, settlement, credit, liquidity, strategic, reputational, and systemic risk, into which the risks inherent in the payment system can grow. The bank should manage such risks at its own level. For these purposes, international regulators developed standards to manage financial market infrastructure, based on which national requirements are elaborated (BCBS 2006, IOSCO 2012a, b).

In the Republic of Belarus, the risk management process in the countrywide payment system is organized by the National Bank. It defined the systemic risk in the payment system as the risk of an event, when the inability of one participant of the payment system to fulfill its obligations causes the inability of other participants to do so in a timely manner (i.e. the domino effect). In such a situation, the inability of the payment system to settle transactions can lead to significant adverse consequences for the economy as a whole, and the loss of trust by legal entities and individuals (National Bank of the Republic of Belarus 2017a).

Since the most efficient approach to the risk management of payment systems is to integrate it into the bank overall risk management system, the National Bank has set requirements for payment system operators regarding integrated risk management practices.

Regulatory bodies of most countries supervise the financial market, including payment systems. With the development of IT, it has become possible to accumulate big data over a long period and it is expected to increase the effectiveness of current off-site supervision and forecasting adverse trends.

The Basel Committee on Banking Supervision in the Basel II Capital Agreement proposed a single supervisory mechanism that is characterized by cyclicality and the use of formalized assessments of the state of banks, which should be regularly revised in the course of supervision. The mechanism was further developed in the European Banking Authority (EBA) document on common procedures and methodologies for supervisory review and evaluation process (SREP) (EBA 2014). The SREP methodology includes the following pillars of assessment:

- categorization of a financial institution and its periodic review, i.e. financial institutions are divided into four categories, depending on their size, structure, internal organization, nature and complexity of activities, while also taking into account the level of systemic risk. The frequency, intensity, and details of the SREP evaluation should depend on the category of the institution;
- regular monitoring of key financial and non-financial indicators, which allows the supervisor to monitor changes in the financial state and risk profile of the institution, as well as facilitate updating the assessment of individual SREP elements when new information is received beyond the planned supervisory actions. These indicators include all prudential standards, risk indicators,

market-related indicators, as well as indicators identified by banks in recovery plans;

- business model analysis, focused on assessing the viability of the current business model of the institution and coherence of its strategic plans;
- assessment of internal management and control in the institution as a whole. It is necessary to make sure that internal management, including internal audit and control is consistent with the risk profile, business model, size and complexity of the institution, and also to assess the degree to which it complies with the standards of proper internal management and risk control;
- risk assessment with regard to capital, liquidity, and sources of funding. It should be focused on assessing the significant risks of the institution to which it is exposed or may be subject. It assesses both the quantitative aspect of the risk exposure and the quality of management and control applied to mitigate the impact of these risks. The supervisor must determine the scale of the potential impact of such risks on the institution;
- assessment of a financial institution's capital adequacy. Since the institution may face risks that are not covered or not fully covered by mandatory capital buffers, the supervisor must assess the size and composition of the additional capital required to cover such risks, as well as its ability to comply with capital requirements during the business cycle. In addition, the supervisor should assess the risk of vulnerability of the institution related to its size and capital structure;
- assessment of the liquidity resources sufficiency of a financial institution. It is necessary to ensure that the liquidity of the institution provides sufficient coverage of liquidity and funding risks, and to determine whether it is necessary to establish special liquidity requirements to cover the risks to which the institution is exposed or can be exposed. The supervisor should assess the risk of vulnerability of an institution related to its liquidity and funding profile;
- overall SREP rating. The supervisor needs to shape a comprehensive, holistic view of the risk profile and viability of a financial institution and formalize it through an overall rating. The overall SREP rating includes an assessment of each element within the framework of the holistic approach, taking the form of a numerical indicator, followed by certain rationale. The overall SREP rating should lay the ground for subsequent supervisory action;
- supervisory measures. Based on the overall rating, banks are divided into groups characterizing their state, followed by the inclusion into the Supervisory Inspection Program with the establishment of a certain periodicity.

The SREP methodology was developed, taking into account the proportionality principle. However, it is directly implemented in respect to SIBs only in the first pillar of the assessment, i.e. financial institution categorization, which affects the frequency, intensity, and detail of the SREP.

4 Conclusion

Thus, it appears crucial to work out general principles of systemic risk regulation at the bank level that would be integrated into the overall risk management system of a bank. The developers of such principles could be international financial institutions that specialize in setting standards for the activities of financial market participants based on the study of international best practices in this area. This would help improve the efficiency of risk management in the financial sector and, consequently, strengthen the sustainability of banking and payment systems.

References

Bank of Russia. (2015a). *On requirements to reduce the risks of clearing activities and requirements to the document (s) determining measures to reduce credit, operational and other risks, including risks associated with combining clearing activities with other activities*. Bank of Russia Regulation № 463-P, dated March 12, 2015. The Central bank of the Russian Federation [Electronic resource]. – Mode of access: Retrieved October 18, 2018, from https://www.cbr.ru/queries/xsltblock/file/48353?fileid=-1&scope=1638

Bank of Russia. (2015b). *On the methodology for determining systemically important credit institutions*. Note of the Central Bank of the Russian Federation № 3737-U dated 22 July 2015– Mode of access: Retrieved October 30, 2018, from https://www.cbr.ru/queries/xsltblock/file/48353?fileid=-1&scope=1667

BCBS. (2006). *The Joint Forum. High-level principles for business continuity, August 2006*. Bank for International Settlements. [Electronic resource] – Mode of access: Retrieved October 18, 2018, from https://www.bis.org/bcbs/publ/d328.pdf

BCBS. (2013). *Global systemically important banks: updated assessment methodology and the higher loss absorbency requirement, July 2013*. Bank for International Settlements. [Electronic resource] – Mode of access: Retrieved October 18, 2018, from https://www.bis.org/publ/bcbs255.pdf

BCBS. (2015). *Guidelines. Corporate governance principles for banks, July 2015*. Bank for International Settlements. [Electronic resource] – Mode of access: Retrieved October 18, 2018, from https://www.bis.org/bcbs/publ/d328.pdf

EBA. (2014). *Guidelines on common procedures and methodologies for the supervisory review and evaluation process (SREP)*. European Banking Authority [Electronic resource]. Mode of access: Retrieved October 15, 2018, from http://www.eba.europa.eu/documents/10180/935249/EBA-GL-2014-13.pdf

IMF. (2009). *Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations – Report to the G-20 Finance Ministers and Central Bank Governors*. International Monetary Fund [Electronic resource]. Mode of access: Retrieved October 18, 2018, from https://www.imf.org/external/np/g20/pdf/100109.pdf

IMF. (2011). *Macroprudential Policy: An Organizing Framework*. Background Paper, 2011. International Monetary Fund [Electronic resource]. Mode of access: Retrieved October 18, 2018, from https://www.imf.org/external/np/pp/eng/2011/031411a.pdf

IOSCO. (2012a). *Committee on Payment and Settlement Systems. Technical Committee of the International Organization of Securities Commissions. Principles for financial market infrastructures, April 2012*. Bank for International Settlements. [Electronic resource] – Mode of access: Retrieved October 21, 2018, from https://www.bis.org/cpmi/publ/d101a.pdf

IOSCO. (2012b). *Committee on Payment and Settlement Systems. Board of the International Organization of Securities Commissions*. Principles for financial market infrastructures: Disclosure framework and Assessment methodology, December 2012. Bank for International Settlements. [Electronic resource] – Mode of access: Retrieved October 21, 2018, from https://www.bis.org/cpmi/publ/d106.pdf

National Bank of Kazakhstan. (2014). *On approval of the Rules for classifying financial organizations as systemically important*. Resolution of the Board of the National Bank of Kazakhstan № 257 of 24 December 2014. – National Bank of Kazakhstan [Electronic resource]. Mode of access: https://tengrinews.kz/zakon/natsionalnyiy_bank_respubliki_kazahstan/finansyi/id-V14Y0010210/

National Bank of the Republic of Belarus. (2012). *On approval of the Instruction on organization of corporate governance of a bank, Development Bank of the Republic of Belarus JSC, non-bank financial institution*. Resolution of the Board of the National Bank of the Republic of Belarus № 557 dated October 30, 2012. – National Bank of the Republic of Belarus [Electronic resource]. – Mode of access: Retrieved October 18, 2018, from http://www.nbrb.by/engl/legislation/documents/PP_557_544.pdf

National Bank of the Republic of Belarus. (2017a). *Analytical Survey "Financial Stability in the Republic of Belarus", 2017*. [Electronic resource]. – Mode of access: Retrieved October 19, 2018, from http://www.nbrb.by/engl/Publications/FinStabRep/FinStab2017e.pdf

National Bank of the Republic of Belarus. (2017b). *On approval of the Instruction on the procedure for determining systemically important banks, non-bank credit and financial organizations*. Resolution of the Board of the National Bank of the Republic of Belarus № 180 dated May 18, 2017. – National Bank of the Republic of Belarus [Electronic resource]. Mode of access: Retrieved October 10, 2018, from http://www.nbrb.by/legislation/documents/PP_180.pdf

National Bank of the Republic of Belarus. (2017c). *On approval of the Strategy of risk management in the payment system of the Republic of Belarus*. Resolution of the Board of the National Bank of the Republic of Belarus № 155 dated April 29, 2017. – National Bank of the Republic of Belarus [Electronic resource]. – Mode of access: Retrieved October 10, 2018, from https://www.nbrb.by/payment/SuperVision/Documents/PP_155.pdf

National Bank of the Republic of Belarus. (2019). *Macroprudential regulation*. [Electronic resource]. Mode of access: Retrieved May 30, 2019, fromhttp://www.nbrb.by/finsector/FinancialStability/MacroprudentialRegulation/

Press releases. The Central bank of the Russian Federation [Electronic resource]. – Mode of access: https://www.cbr.ru/press/PR/?file= 07032019_095500NANT2019-03-07T09_47_49.htm

Regulation (EU) № 1092/2010 of the European Parliament and of the Council of 24 November 2010 on European Union macro-prudential oversight of the financial system and establishing a European Systemic Risk Board. Official Journal of the European Union [Electronic resource]. Mode of access: Retrieved October 19, 2018, from https://www.esrb.europa.eu/shared/pdf/ESRB-en.pdf

Real Effects of Financial Shocks in Russia

Vasilisa Baranova

Abstract Over the last 10 years, Russia has faced many external and internal challenges. Using the Financial Stress Index for Russia (the ACRA FSI), which indicates the proximity of the Russian financial system to crisis and its reaction to different events, I show that shocks in the Russian financial system have adverse effects on real economic activity. The VAR model and Toda–Yamamoto augmented Granger causality tests are my research tools. I also estimate a threshold structural VAR model, revealing that the impact of a financial shock is bigger and longer lasting for distressed periods compared to normal periods in the Russian financial system. All my findings are in line with other research studies for both emerging and advanced economies.

Keywords Financial crisis · Financial stress · Emerging markets · Threshold VAR

JEL G01 · G17 · G32

1 Introduction

The global financial crisis has become a trigger for the emergence of a large number of indicators that assess the state of a country's financial system or can even predict it. Examples of such indicators are financial stress and financial conditions indices (FSIs and FCIs). They measure the state of financial stability in a particular country or region. In practice, some central banks have adopted them (European Central Bank 2011; Hakkio and Keeton 2009; Kliesen and Smith 2010) to monitor financial stability and conduct monetary policy. Investors also rely on these indices when assessing the overall risk of investing in financial instruments of a country or region.

V. Baranova (✉)
Gazprombank, Moscow, Russia
e-mail: vasilisa.baranova@gazprombank.ru

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_15

Many international organizations and financial institutions use these indices (Bloomberg, Goldman Sachs, Citi, Bank of America, OECD, IMF, etc.).

Researchers also build FSIs and FCIs, using them for various purposes. For example, they apply these indices to analyze the interconnectedness of financial systems of different countries or to study whether financial stress transmits to the real economy. Indeed, recent financial crises showed that an increase in financial stress can dampen economic activity. Apparently, during the periods of financial instability, firms may decide to postpone their investments until better times. Financial stress is also deemed to increase the cost of borrowing, which leads to lower investment and economic growth. Some researchers suggest that the degree of financial stress transmission on economic activity differs between normal and stressful periods.

As to my knowledge, few of existing studies aim to discover the susceptibility of the Russian economy to financial stress and none of them uses ACRA financial stress index (ACRA FSI) as a proxy of financial stress. Against this backdrop, I decided to carry out such analysis for Russia. Thus, the primary goal of this paper is to analyze whether and to what extent shocks in financial system transmit to real economy in Russia. Meanwhile, I examine whether the effect of financial stress on economic activity differs for normal and distressed periods in the Russian financial system.

In the line with previous researches in this field, I use vector autoregression (VAR) and threshold structural vector autoregression (TSVAR) models and conduct Granger causality tests (Toda and Yamamoto 1995) to take into account non-stationarity of selected time series. I also perform robustness check to determine the credibility of obtained results. The proxy of financial stability in Russia is ACRA FSI, which was proposed several years ago (Kulikov and Baranova 2016, 2017). The rest of the paper is organized as follows. Firstly, I provide a literature review of existing ways to study the relationship between financial stress and economic activity, paying special attention to emerging markets. The next section introduces my research hypothesis and sets up methodological tools used for the analysis. Then, I do robustness check and discuss obtained results.

2 Literature Review

FSIs and FCIs are real-time indicators of financial stability. They can be applied for a variety of academic purposes. Many academic studies use them to examine financial stress transmission to real economy. They usually apply different econometric techniques, for instance, modifications of VAR model, Granger causality tests, and impulse response functions. Some of them take into account potential nonlinearity of shock transmission. In my analysis, I rely on two strands of literature. The first one examines financial stress transmission for advanced economies, the second one focuses on emerging markets. Within the latter strand, I pay special attention on the existing research on Russian economy.

Financial stress can cause recessions (Bloom 2009; Cardarelli et al. 2011). Indeed, when there is uncertainty in financial markets, economic agents may decide to postpone their investments and wait until better times. This can cause decreases in economic activity (Davig and Hakkio 2010). Another view on stress transmission to real economy is "financial accelerator" framework (Bernanke et al. 1999), when an increase in the financial stress (worsening financial conditions of firms) rises the cost of borrowing, lowers investments, and also leads to a decline in economic growth.

Many papers use FSIs and/or FCIs to examine the relationship between financial stress and economic activity for advanced economies. For example, the authors who proposed one of the most popular financial stress index for USA—Kansas City Fed FSI (KCFSI)—(Hakkio and Keeton 2009) examined whether a rise of financial stress entails any change of the net percent of US banks, which tightened their standards over the previous 3 months, and Chicago FED national activity index (CFNAI). They performed prediction tests by running regressions on the lagged values of dependent and independent variables. Their results indicate that financial stress can predict economic slowdown and changes in credit standards. Davig and Hakkio (2010) used KCFSI, CFNAI, and a regime-switching VAR model to illustrate that during the periods of increased financial stress, its effect on real economy is significantly higher than during normal times. Ubilava (2014) used the same variables as a proxy of financial stress (Kansas Fed FSI) and economic performance for USA (CFNAI), but applied a different version of nonlinear VAR, a vector smooth transition autoregressive (VSTAR) model, to account for a potentially greater degree of susceptibility of financial and economic activity during stressful periods. The author's findings are in line with Hakkio and Keeton (2009). The author of other financial stress index for USA (Monin 2017) also found out that financial stress (OFR FSI) can predict CFNAI. He used a modification of Granger causality test, proposed by (Toda and Yamamoto 1995). Illing and Liu (2003) used TVAR for Canadian FSI and obtained results consistent with Davig and Hakkio (2010). Roye (2011) used a Bayesian VAR model and impulse response functions to analyze the effects of financial stress on real economic activity for Eurozone and Germany, encompassing real GDP growth rates, inflation and short-term interest rate. He found adverse effects of financial stress on the real economy. The authors of financial stress index for the UK (Chatterjee et al. 2017) use a similar procedure to determine whether there is a relationship between financial stress and economic activity, building on the TVAR and generalized impulse response functions. According to their findings, the transmission of shocks in normal and stressful periods is different in the UK. Finally, Aboura and Roye (2017) used a Markov-switching model to show that stressful periods in France generate pronounced economic reactions, which are negligible otherwise.

In comparison with advanced economies, a smaller number of FSIs were constructed for emerging markets. However, they generally use similar econometric models to test how financial instability affects economic activity. For example, Aklan et al. (2016) computed a financial stress index for Turkey and found a significant adverse impact of financial instability on real economic activity by using VAR and Granger causality tests. Polat and Ozkan (2019) also constructed a financial stress index for Turkey and obtained quite similar results, using a structural VAR model. Tng and Kwek (2015) applied the same model to examine the interaction of financial stress and economic activity for the ASEAN-5 countries. Their results are consistent with other studies. Cevik et al. (2016) constructed the financial stress index for four Southeast Asian economies (South Korea, Malaysia, the Philippines, and Thailand) and exploited impulse response functions. According to their results, financial stress causes economic slowdowns. Stona et al. (2018) introduced a FSI for Brazil and used a Markov-switching VAR model to examine its nonlinear relationships with real activity, inflation, and monetary policy.

There are some papers that examine financial stress and economic activity interaction for Russia. For example, Stolbov and Shchepeleva (2016) used Granger causality tests based on the Toda and Yamamoto approach and found the effects of financial stress on industrial production in 9 out of 14 emerging markets, including Russia. Using bivariate VAR models and impulse response functions, Çevik et al. (2013) documented linkages between the fluctuations of economic activity and financial stress for Bulgaria, Czech Republic, Hungary, Poland, and Russia. In particular, they found a significant correlation between composite leading indicators (CLI) of economic activity calculated by OECD and the financial stress index for Russia. The bivariate VAR models showed a strong negative response of industrial production, investment, and foreign trade growth rates to the rise in financial stress for this country.

All the aforementioned studies found strong statistical evidence of the relationship between financial and real sector. However, relatively little attention is paid to the Russian economy. I contribute to the existing literature by addressing the financial stress interactions with economic activity in Russia, taking into account potential nonlinearity of shock transmission. Thus, I depart from Stolbov and Shchepeleva (2016) and Çevik et al. (2013) who also analyzed the impact of financial stress on the economic activity in Russia by using a different financial stress index (ACRA FSI) and applying a TSVAR model. As I use a nonlinear VAR, my analysis is close to Chatterjee et al. (2017), Davig and Hakkio (2010), Ubilava (2014), who adopted such methodology for different countries.

3 Hypothesis Development and Data

This paper has two goals. First, it aims to test for causality between financial stress and economic activity in Russia, Second, it examines a changing degree of response of economic activity to financial stress.

The scatter plot lends preliminary support to hypothesis about the existence of two regimes in the relationship between production index and ACRA FSI (Fig. 1). Indeed, economic activity measured by the production index and financial stress tends to move in opposite directions during distressed periods. There is also a negative relationship between these variables in the normal regime but to less extent. Two black diamonds illustrate the average values for ACRA FSI and production index in normal and stressed periods. They are (0.713; 1.03) and (2.381; −1.001), respectively.

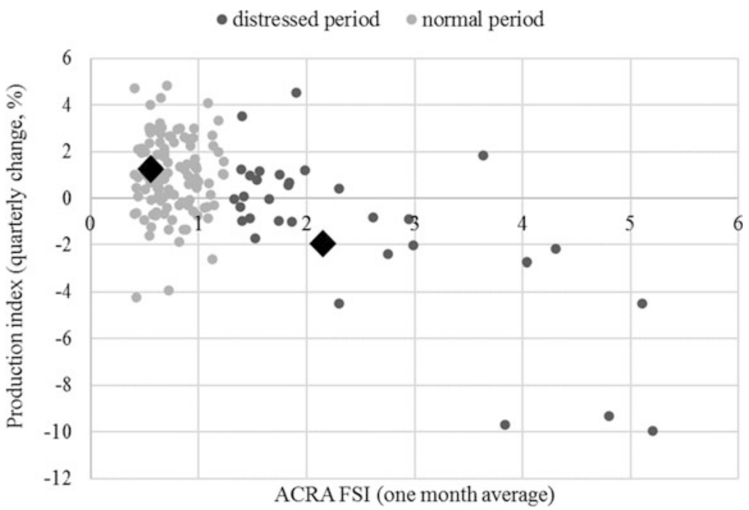


Fig. 1 Relationship between ACRA FSI and production index. Source: ACRA, author's calculations

Further, I discuss data used to test my hypotheses. As a proxy for financial instability in Russia, I take the ACRA FSI.[1] This index captures the financial system's proximity to a financial crisis. Twelve factors are used for the ACRA FSI calculation (Kulikov and Baranova 2016):

- Spread between money market interest rates and zero-coupon short-term OFZs[2] (3 months);
- Interest rate spread between large issues of liquid corporate bonds and zero-coupon OFZ rate (5 years);
- Stock market volatility;
- Financial sector stock price index;
- Divergence of financial institutions' stock returns;
- Spread between the interbank loan interest rate and 1-day liquidity interest rate offered by the Bank of Russia;
- Differential between crude oil spot and forward prices (1 year);
- Crude oil price volatility;
- Currency exchange rate volatility;
- Inflation;
- Velocity of the simultaneous stock prices drops of financial institutions and sovereign debt (flight to liquidity);

[1]Values of ACRA FSI are published on a daily basis on the official ACRA website: https://www.acra-ratings.com/research/index.

[2]OFZs are bonds issued by the Federal Russian government.

- Velocity of divergence between stock prices of financial institutions and quality lender-issued bonds (flight to quality).

Prior to weighting and summing up, the original factors are transformed in the way that makes their value increase along with financial stress. The transformed factors are normalized to ensure that each of their historical dynamics has a zero sample mean and a single-unit standard deviation within a fixed timeframe. Weights of normalized and transformed factors are calculated as the coordinates of the first principal component. Financial stress is an unobservable phenomenon. Thus, in order to make sure that it behaves correctly in case of financial shocks, one can observe index dynamics after these events. It strengthens the credibility of the indicator (Fig. 2).

I use monthly production index as a proxy of economic activity as a weighted average of six core industries (agriculture, industrial production, construction, retail trade, wholesale trade, and transport). I seasonally adjust it using the Census X-12 method. As the production index is available only at monthly frequency, I transform the ACRA FSI by taking average monthly values. The sample covers the period from January 2006 to June 2019.

4 Methodology and Estimation Results

In order to test the first hypothesis, I apply a VAR model and Granger causality tests (Granger 1969). First, I perform stationarity check for the time series. I use Augmented Dickey–Fuller (ADF) test and Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test (Kwiatkowski et al. 1991). The null hypothesis for ADF test is non-stationarity, while for the KPSS test it is stationarity. I use both tests, as null hypotheses for them are different and to ADF test is sensitive to structural breaks in time series. The results of the tests are presented in Table 1. The ADF test rejects the null hypothesis of non-stationarity at 5%, but does not reject it at 1%. The production index is stationary at any reasonable significance level as well as its first difference. According to the KPSS tests, both series are stationary even in levels. Thus, the results are inconclusive at the 5% significance level for the ACRA FSI.

Hence, I conclude that unit root tests results suggest the maximum level of integration is one. That is, one lag of the ACRA FSI and production index should be included as exogenous variables into a VAR model to perform the Toda–Yamamoto Granger causality test. For this test specification, there is no need for all variables to be stationary or cointegrated. Next, I choose an optimal lag length for VAR model. The results are in Table 2. According to Schwarz Criterion (SC) and Hannan–Quinn information criterion (HQ), three lags should be selected. Thus, in order to perform the Toda–Yamamoto Granger causality test, I should use the VAR model with four lags. The results of the causality tests are in Table 3. Autocorrelation LM test indicates the absence of serial correlation in the residuals of model. Small p-values reject the null hypothesis of no causality when production index is the

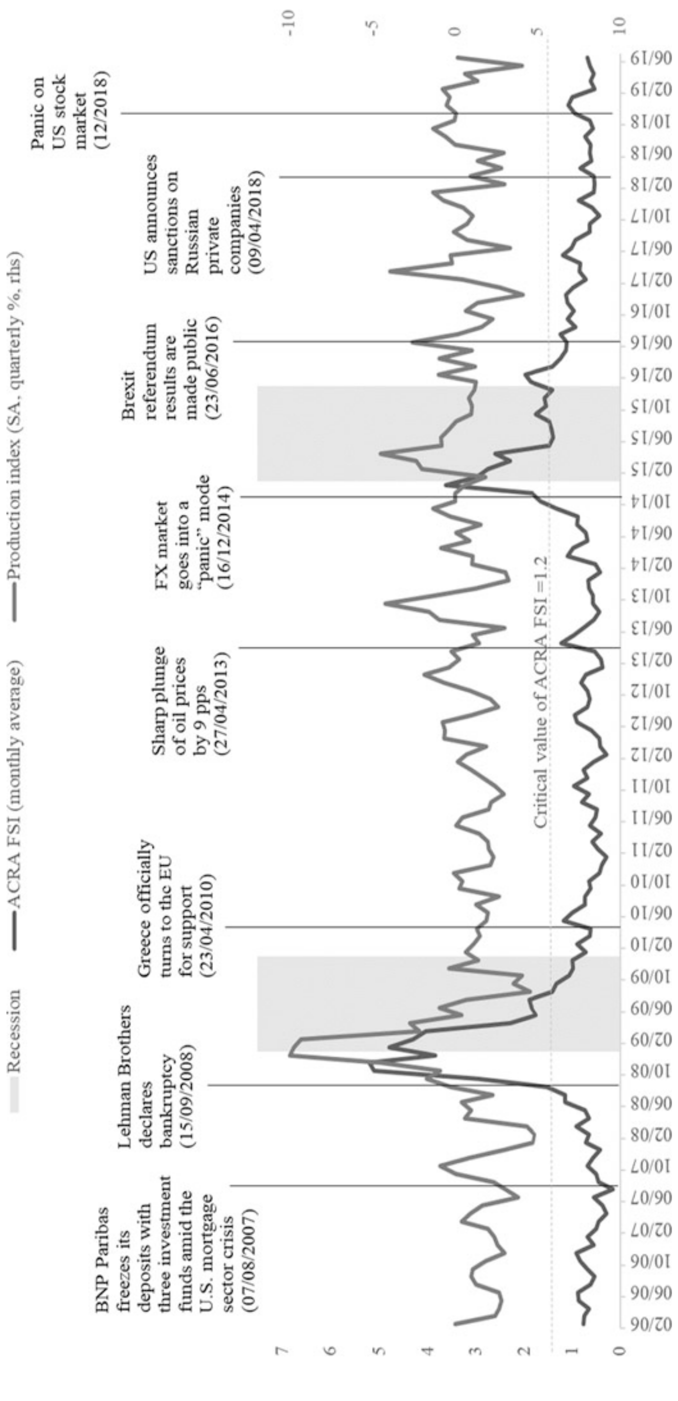


Fig. 2 Dynamics of the ACRA FSI and production index after selected events. Source: ACRA, author's calculations

Table 1 Results of stationarity tests

	ACRA FSI	Production index
p-values for the ADF test (null hypothesis: non-stationarity)		
Level	4.87%	0.00%
First difference	0.00%	0.00%
KPSS tests (null hypothesis: stationarity)		
Level	Not reject H_o	Not reject H_o
First difference	Not reject H_o	Not reject H_o

Source: ACRA, author's calculation

Table 2 Optimal lag length selection

Lags	AIC	SC	HQ
0	6.674	6.713	6.690
1	4.796	4.913	4.844
2	4.809	5.004	4.888
3	4.582	4.855*	4.693*
4	4.567	4.919	4.710
5	4.557*	4.987	4.731

Source: ACRA, author's calculation

Table 3 Toda–Yamamoto Granger causality test

VAR model for Production index and ACRA FSI (lags = 3)	
Does ACRA FSI help to predict production index?	Yes
p-value for Granger causality test (dependent variable production index)	0.00
Does production index help to predict ACRA FSI?	No
p-value for Granger causality test (dependent variable ACRA FSI)	0.14

Source: ACRA, author's calculation

dependent variable. The generalized impulse response function (GIRF) with 95% confidence band of production index to one standard deviation shock of ACRA FSI is represented in Fig. 3. It captures a negative response of economic activity to financial stress. This result is consistent with Stolbov and Shchepeleva (2016) and Çevik et al. (2013). I perform robustness check by narrowing the sample from January 2012 to June 2019. The results of the robustness check are in line with the full-sample analysis.

Finally, I estimate a TSVAR model. This type of VAR model captures non-linearities and structural breaks in time series. There are two states of stability of the financial system: normal and distressed regimes. The threshold value of ACRA FSI is 1.2 (Fig. 3). GIRFs for TSVAR model are illustrated in Fig. 4. They suggest that for both types of periods financial stress and real economy are negatively correlated. However, for distressed periods the impact of financial shock is much more pronounced and longer lasting. Indeed, GIRFs show that for these periods one standard deviation increase in financial stress leads to a significant decline of economic activity over several months. This result is consistent with the existing literature.

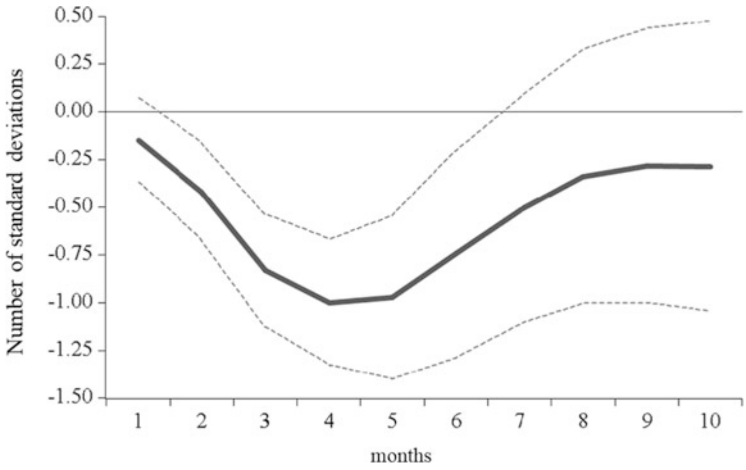


Fig. 3 Generalized impulse response function of the production index to one standard deviation change in ACRA FSI. Source: ACRA, author's calculation

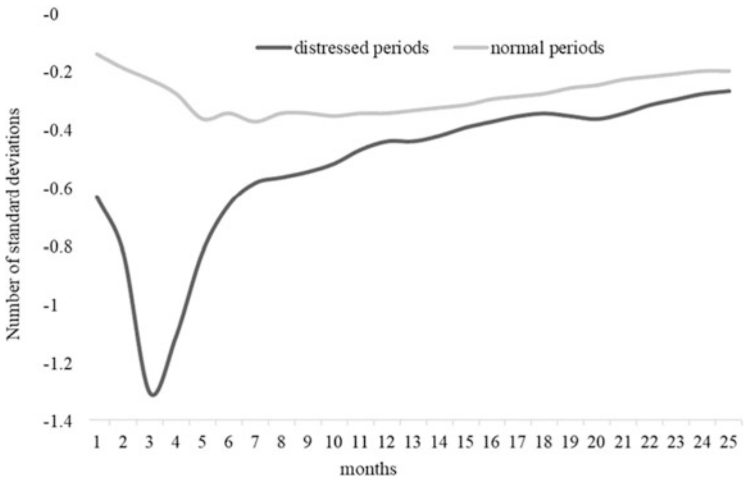


Fig. 4 GIRFs for TSVAR model (response of Production index to one standard deviation shock of ACRA FSI). Source: ACRA, author's calculation

5 Conclusions

The results of the performed tests indicate that there is a strong statistical evidence that the ACRA FSI helps predict changes in the production index. It implies that the inclusion of financial stress index can significantly improve forecasting accuracy of production and real economic growth. However, one needs to be very cautious when interpreting the results of Granger causality tests, as it does not necessarily indicate causality. That is, the definition of causality is related to the idea of the cause-and-effect relationship, while "Granger causality" is rather a statistical concept, which does not necessarily imply it. Overall, the findings of this research are consistent with those of the aforementioned studies. Thus, the ACRA FSI can contribute to the improvement of forecasting models for economic growth in Russia, as it is a real-time indicator.

The paper also finds that there are different responses of economic activity during normal and stressful regimes. These results are in line with the results for financial stress indices for advanced economies (USA, UK, Canada) and emerging markets (Brazil).

References

Aboura, S., & van Roye, B. (2017). Financial stress and economic dynamics: The case of France. *International Economics, 149*(C), 57–73.

Aklan, N., Çinar, M., & Akay, H. (2016). Financial stress and economic activity relationship in Turkey: Post-2002 period (Türkiye'de Finansal Stres ve Ekonomik Aktivite İlişkisi: 2002 Sonrasi Dönem). *Yönetim ve Ekonomi: Celal Bayar Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 22*(2), 567–580.

Bernanke, B., Gertler, M., & Gilchrist, S. (1999). The financial accelerator in a quantitative business cycle framework. In *Handbook of Macroeconomics* (pp. 1341–1393). Cambridge, MA: National Bureau of Economic Research.

Bloom, N. (2009). The impact of uncertainty shocks. *Econometrica, 77*(3), 623–685.

Cardarelli, R., Elekdag, S., & Lall, S. (2011). Financial stress and economic contractions. *Journal of Financial Stability, 7*(2), 78–97.

Çevik, E., Dibooglu, S., & Kutan, A. (2013). Measuring financial stress in transition economies. *Journal of Financial Stability, 9*(4), 597–611.

Cevik, E. I., Dibooglu, S., & Kenc, T. (2016). Financial stress and economic activity in some emerging Asian economies. *Research in International Business and Finance, 36*(C), 127–139.

Chatterjee, S., Chiu, J., Hacioglu-Hoke, S., & Duprey, T. (2017). A financial stress index for the United Kingdom. *SSRN Electronic Journal, 9*, 697.

Davig, T. A., & Hakkio, C. S. (2010). What is the effect of financial stress on economic activity. *Economic Review, 2*, 35–62.

European Central Bank. (2011). Financial Stability Review June 2011.

Granger, C. W. J. (1969). Investigating causal relations by econometric models and cross-spectral methods. *Econometrica, 37*(3), 424–438.

Hakkio, C. S., & Keeton, W. R. (2009). Financial stress: What is it, how can it be measured, and why does it matter? *Economic Review, 2*, 5–50.

Illing, M., & Liu, Y. (2003). *An index of financial stress for Canada* (pp. 3–14). Ottawa: Bank of Canada.

Kliesen, K., & Smith, D. (2010). Measuring financial market stress. *Economic Synopses, 2010*, 2.

Kulikov, D., & Baranova, V. (2016). Principles of calculating the financial stress index for the Russian Federation. Retrieved from https://www.acra-ratings.com/criteria/129

Kulikov, D., & Baranova, V. (2017). Financial stress index for Russian financial system. *Money and Credit, 6*, 39–48.

Kwiatkowski, D., Phillips, P. C. B., & Schmidt, P. (1991). *Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?* (p. 979). London: Elsevier.

Monin, P. (2017). The OFR financial stress index. *Risks, 7*(1), 25.

Polat, O., & Ozkan, I. (2019). Transmission mechanisms of financial stress into economic activity in Turkey. *Journal of Policy Modeling, 41*(2), 395–415.

Stolbov, M., & Shchepeleva, M. (2016). Financial stress in emerging markets: Patterns, real effects, and cross-country spillovers. *Review of Development Finance, 6*(1), 71–81.

Stona, F., Morais, I. A. C., & Triches, D. (2018). Economic dynamics during periods of financial stress: Evidences from Brazil. *International Review of Economics & Finance, 55*(C), 130–144.

Tng, B. H., & Kwek, K. T. (2015). Financial stress, economic activity and monetary policy in the ASEAN-5 economies. *Applied Economics, 47*(48), 5169–5185.

Toda, H. Y., & Yamamoto, T. (1995). Statistical inference in vector autoregressions with possibly integrated processes. *Journal of Econometrics, 66*(1–2), 225–250.

Ubilava, D. (2014). On the relationship between financial instability and economic performance: Stressing the business of nonlinear modelling. *Macroeconomic Dynamics, 23*, 80–100.

van Roye, B. (2011). *Financial stress and economic activity in Germany and the Euro area* (p. 1743). Kiel: Institute for the World Economy.

Part V
Estimating and Managing Financial Risks: Topical Trends in Emerging Capital Markets

Innovation in Developing Countries' Risk Estimation and Management

Andrey Egorov and Dmitry Pomazkin

Abstract Today, an increasingly important role in the economy is being acquired by informatization processes. Digital technologies simplify information transfer and accelerate these processes. The effective use of various complex banking technologies, as well as the use of information and communication technologies in banking operations, can improve the organization of financial products and various tools that are key ways to stimulate the needs and preferences of customers. Various financial innovations, including Internet banking, ATMs, and mobile banks, are increasingly becoming a vital force for diversification, revenue generation, and cost reduction for both banks and customers. The article is devoted to the problem of development of innovations in the field of financial technologies in developing countries. It is shown that, despite a significant increase in innovation activity in the field of the financial sector of the economy, a large number of developments in this area, some gaps remain associated with the practical application, implementation of financial innovations, the use of innovative tools in the field of financial technologies in developing economies, the definition of the role and places of financial innovation in the overall structure of the financial sector. This article aims to fill these gaps. There is a connection between financial innovation and the efficiency of the banking industry for both developed and developing countries around the world. The banking sector in a developing economy is growing thanks to financial innovations in various payment systems, including the use of ATMs, mobile banking, and electronic banking. The aim of the work is to analyze the innovation of risk assessment and risk management in developing countries, for which the following tasks were set: consider the features of information technology, financial instruments, and services; explore diffusion models of banking innovations; identify features of digital solutions in developing countries.

Keywords Innovation · Financial technologies · Banking · Digital technologies

A. Egorov · D. Pomazkin (✉)
National Research University Higher School of Economics, Faculty of Economic Sciences, Moscow, Russia
e-mail: ayegorov@hse.ru

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_16

JEL G00 · G17 · G32

1 Information Technologies and Financial Instruments and Services

Due to the economic growth in the middle of the twentieth century, many countries were faced with the problem of increasing growth potential. But during around the 1970s and 1980s, communication technology and computer technology converged. The former were associated with the transfer of information, and the latter with the processing of information, which made it possible to create a single system for processing and exchanging information. The integration has affected a wide range of areas and activities, including management information systems, professional databases, transaction clearing systems, online requests, email, etc. The changes for the financial industry were significant. This has been accompanied by the liberalization of financial markets and capital flows in many parts of the world. (Nakaso 2016).

With introduction of new technologies, new approaches in financial research have appeared. In the 1990s, financiers began to apply new concepts such as chaos theory, fuzzy sets, and neural networks. New methods for analyzing time series of high-frequency data have been developed to analyze huge data sets. All this has found its application in the research departments of banks and other financial institutions. However, this development has made it more difficult to manage the portfolio and risks in these institutions.

Today informatization processes play an increasingly important role in the economy. Digital technologies simplify information transfer and accelerate these processes. Therefore, the financial services industry can be considered as an «information industry». Financial services, such as payments and settlements, investment decisions, and risk management, are based on large-scale information processing. Today, there are many different financial technologies (FinTech). But there are three categories of technologies that are most in demand.

The first category includes «blockchain» and distributed ledger technology (DLT), which appeared in 2008 with the concept of «bitcoin». These technologies minimize the risk of loss or forgery of information. Blockchain and DLT can significantly affect the «money» and «ledgers», which are the basic infrastructure for financial activities. This situation raises many questions from the point of view of economic theory. However, today the majority of efforts to introduce these technologies into practice are still at the experimental stage. Blockchain technology will also allow to use «smart contracts» for various purposes. One example is the continuous adjustment of car insurance fees in accordance with the behavior of each policyholder driving. Smart contracts can have the potential to overcome «moral hazard» through the use of new information technologies.

The second category includes artificial intelligence and big data analytics which are developing due to a sharp increase in computing power. This category is of great importance in risk management both in banks and other financial institutions. Real-time, high-precision big data analytics minimize risk.

The third category includes other technological innovations, such as mobile phones and smartphones. Today they have become a new means of access to financial services. Many companies are now competing with each other for the provision of financial services through applications on smartphones. Mobile phones and smartphones are now spreading rapidly, not only in advanced economies, but also in developing countries, where financial services have not yet become widespread. FinTech has opened up the possibility of provision basic financial services with these new tools. Mobile phones and smartphones have the characteristics of «personalized» tools which allow you to analyze individual customers. As a result, FinTech makes it easier for the industry to provide more specialized services.

Although FinTech has many benefits, it brings new challenges. First of all, FinTech is changing the structure of settlements and other financial services. For example, in the case of non-bank P2P credit companies, it is difficult to obtain sufficient information on financial intermediation from their balance sheets. In addition, imposing restrictions on these balances may not be very effective in influencing their P2P lending activities. Financial authorities should consider how they can obtain the necessary information to maintain financial stability. Also, innovations in information technology simultaneously have led to the emergence of various new tactics for cyber threats.

Innovations in information technology and FinTech increase financing efficiency and contribute to economic development. FinTech-driven financial inclusion clearly illustrates the positive feedback between finance and economics. People in developing countries gain access to financial services through FinTech and expanding e-commerce. FinTech contributes to the economic development of the country. However, in developed countries, where basic financial services are already widespread, it is rather difficult to quantify the impact of FinTech on the economy.

We can highlight some of the opportunities that digital technologies provide to economic entities. The development of the digital economy is a factor:

– increasing the competitiveness of financial institutions in the global market;
– cheapening and simplifying the solution of standard tasks implemented by conducting large volumes of operations;
– ensuring interaction between economic entities in order to provide financial services directly, without the participation of intermediaries;
– creating new jobs and increasing labor productivity;
– the emergence of new goods and services.

The effective use of various complex banking technologies, as well as the use of information and communication technologies (ICT) in banking operations, can improve the organization of banking products and various tools, which are key ways to stimulate the needs and preferences of customers. Various financial innovations, including Internet banking, ATMs, and mobile banks, are increasingly becoming a vital force to diversify banks, generate revenue, and reduce costs for both banks and customers (Abubakar and Tasmin 2012).

With the advent of new technologies, banks began to improve payment processing in the back office, correspondence mechanisms, and risk management procedures. Also, financial institutions around the world began to think about strategies for jointly limiting the risks of the payment system. At present, banking sector is increasingly guided by its main strategic role—to balance its indicators through various profitability ratios, such as return on capital, return on assets, profit before tax, etc.

Now the market is looking for new approaches and business models. Many managers see FinTech companies as a risk to their business, as they may lose most of their profits. Therefore, financial institutions are beginning to restructure their behavioral patterns and open up to cooperate with the FinTech industry. Often, large companies have difficulties changing their strategy or implementing innovative solutions. The main reason for this is the bureaucratic component within large organizations. To solve this problem, they establish partnerships with FinTech companies and startups. Possible options for such processes are the introduction of artificial intelligence, blockchain, reorientation of the company's development direction with the establishment of partnerships with FinTech companies. It is worth noting the regulatory component of FinTech. For example, regulators in the USA, the Netherlands, Singapore, and Hong Kong are putting forward initiatives such as "sandboxes." They aim to create a unified environment in which financial institutions, payment service providers, and other financial market participants can implement and test innovative models. The regulator also monitors processes and works with participants on regulatory aspects. This model was the origin of the FinTech regulatory process itself—RegTech. Companies are being created to help FinTech projects meet regulatory requirements. The leaders in this area are the United Kingdom and the USA.

2 Diffusion Models of Banking Innovations

The current stage of development of the global banking system goes through a crisis and increased competition in the financial markets. One of the main factors in the development of banks is the desire to constantly introduce innovations. Currently, innovation is a key factor in the stability, competitiveness, and sustainable growth of banks and other financial institutions. The globalization of financial markets is driving the transition to a more homogeneous banking market. This leads to the development and implementation of innovative technologies to obtain a competitive advantage. Due to globalization, banking content has changed. Every day it became more and more complex and diversified. Banks also faced new risks and attracted new groups of customers. In particular, the Russian banking sector is a typical catch-up economy. Innovative development is mainly developed through the

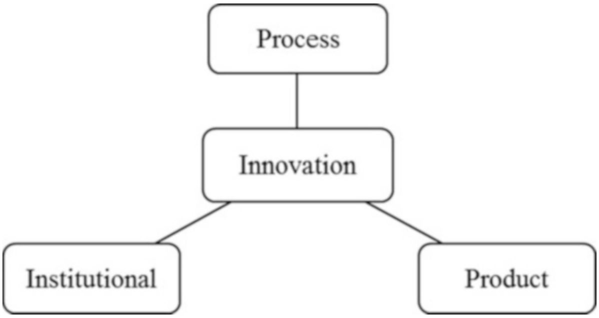


Fig. 1 Types of innovation

implementation of existing international experience, which is transferred between banking systems in the process of diffusion (Jdanova and Karminsky 2013).

Technological advances create opportunities for new profits due to increased investment by financial institutions in new innovative products. Effective banking sector activity based on this theory is oriented primarily to the client. Because of this, banks seek to provide high customer value, achieve lower transaction costs, and increase their market share and financial performance (Barney 1991). Innovative banking products also improve administrative efficiency and help lower transaction costs for customers.

Innovation is the realization of a new market idea. Financial innovation is the introduction of new financial instruments in the financial markets with the help of new technologies. This category includes process, products, and institutional innovation (Fig. 1).

The process is new ways of doing business and implementing information technology (Abor 2005). For example, mobile banking, online banking, and much more. Product innovation is new financial products such as securitized assets, derivatives, foreign currency mortgages, hedge funds, stock funds, private equity and structured retail products, and more. Institutional innovation is the process of introducing new types of financial firms, such as discount brokerage firms, Internet banking, specialized credit card companies, etc.

All these types of innovations improve the financial systems used in borrowing and lending, which ultimately shortens interactions with customers. Also included are innovations in technology, capital generation, and risk transfer. This increases affordable credit for borrowers and provides financial institutions with a new and inexpensive way to raise capital. (Tahir et al. 2018).

Disseminating and embracing innovation is as important as creating innovation. Without them, innovation becomes unclaimed.

Spread of innovation is the speed of spread of an innovative product. When innovations enter the market, they become visible and accessible. At the same time, the simpler the introduction of innovations and the faster the distribution on the market, the stronger the economic growth in the corresponding segment. The high rate of diffusion of innovation implies a great socio-economic benefit from the initial investment. The appearance of a new product or technology always leads to a

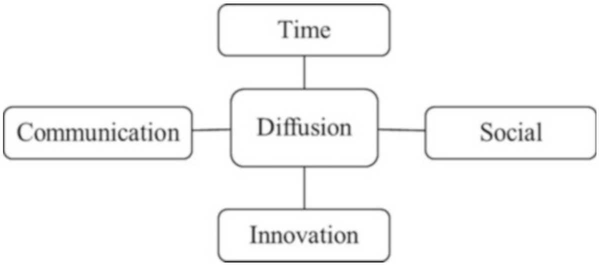


Fig. 2 Elements in the spread of the innovation process

disruption of the existing balance in the market. This allows you to get an «innovative rent». It represents an additional income. It can only be obtained by banks that are the first to innovate. Other banks become «follower banks». Followers should pursue a policy of reactive and evolving innovations to increase their productivity and competitiveness. This allows you to increase market share and gain temporary monopoly power. But the faster the spread of innovation, the easier it is for banks to lose their competitive advantage and monopoly influence. The speed of distribution depends on the current stage of the innovation life cycle, industry, and a number of other factors. The leading catalyst for many innovation processes is the globalization of financial markets. It is expressed in a directed change in the markets for banking products and services, as well as in changes in consumer behavior.

Distribution and introduction of synonyms and often use them interchangeably. Distribution is a macroprocess associated with the distribution of a new product from its source to the consumer. Implementation is a microprocess. It focuses on the steps that an individual consumer goes through when deciding whether to accept or reject a new product.

Research on the spread of innovation in the field of implementation is more focused on the attributes of the end user. They take into account such decision-making factors as perceived utility; perceived ease of use; relative advantages; social norms of personal innovation; estimated risks and costs.

Technological innovation process consists of three separate stages: the process of invention; the innovation process, when ideas are transformed into market products; distribution process (Suriñach et al. 2009).

Diffusion theory explores the nature of the distribution of innovation at all stages of its life cycle. It represents one of the most fundamental formations of a system-institutional approach to the description of the economy of innovation. The goal of any diffusion model is to explain the temporal picture of the diffusion process of new technology on the market. The first model for diffusion of innovation was proposed by Rogers. She suggested that diffusion is the process by which innovation is transmitted through specific channels between members of the social system (Rogers 2003). Several basic concepts of the theory of innovation can be distinguished: innovation, communication channels, time, and the social system (Fig. 2).

Innovation is ideas, practices, or objects that are perceived by a person as new. Communication channels are the means by which messages are transmitted from one person to another. Time as an element of the model includes the process of adoption of innovations; the relative time for which innovation is taken by society; and speed of innovation. The social system is a set of interconnected units participating in a joint decision-making process to achieve a common goal. The speed of diffusion of innovation is the relative speed with which innovation is adopted by members of this social system. In this process, each person goes through five stages: knowledge, belief, decision, implementation, and confirmation. The probability of introducing innovations and the speed of this process is influenced by such factors as relative advantages compared with the previous analogy; compatibility level; complexity of use; the possibility of pilot testing a new technology; observability, transparency, and accessibility for other users. During the spread of any innovation, there is a «critical point». This is the moment when innovation reaches a critical mass and, after its further spread, demonstrates independent behavior.

In the model of Frank Bass, a temporary picture of sales of a new innovative product in the early stages was considered (Bass 1969). Sales reach a peak and then stabilize at a level slightly below the peak. This is due to the relative increase in replacement sales and lower initial sales. There are two categories of people: innovators or decision makers; and followers, the number of which depends on the number of innovators who have adopted innovation in previous times. The former make decisions on introducing innovations to optimize their business practices, increase wealth, and maintain a competitive advantage. Innovators make decisions solely in accordance with a changing environment, legislation, or any new outside knowledge, including information in the media. Followers, in turn, make a decision based on the experience of innovators who are in the same social system.

The acceleration of the diffusion process depends on various macroeconomic conditions and demographic changes. The active participation of banks in innovation policy increases their competitive advantage and strengthens their financial position (Roberts and Amit 2003). The decision of banks to innovate is primarily influenced by their previous experience with other innovations and the degree to which they are associated with technology companies in other industries (Pennings and Harianto 1992). The introduction of innovations in business activities is often caused by awareness about a new technology; the possibility of its use and adaptation; the profitability of introducing a new technology. Also, to ensure their possible implementation, an important point is the behavior of suppliers of new innovative technologies, with their improvement, and with a decrease in their value over time (Suriñach et al. 2009).

A country's technological potential determines the extent to which these technologies are incorporated into everyday economic life (Burns 2009). Also, spatial effects (distance from an innovative country), problems of regional integration, and globalization contribute to the spread between countries. Various studies also emphasize the importance of economic openness as a determinant of the spread of innovation in developing countries (Ang and Kumar 2014). This is because there is a strong positive correlation between openness and technology adoption (Almeida and Fernandes 2008). The dissemination and implementation of technologies rely on significant and targeted technological efforts, as well as on the country's human capital and financial potential. Dissemination and implementation of technologies require appropriate institutions and policies to stimulate and facilitate the process (Fu et al. 2011). At the same time, financial innovations are spreading rapidly to increase shareholder return on investment.

3 Peculiarities of Digital Decisions in Developing Countries

In developing countries, banks strive to improve their financial performance while being able to improve and maintain their efficiency and market activity (Kamau and Oluoch 2016). The Bank's efficiency and activity are measured by various financial ratios that allow banks to access financial indicators from their resources. The most commonly used ratios are return on assets, return on equity, and others. In this era of globalization and technological progress, the number of TRANS national banking institutions has increased. This has led to an increase in the level of complexity in the form of financial products used by banks to serve their clients (Victor et al. 2015).

There is a link between financial innovation and the performance of the banking industry for both developed and developing countries around the world. The banking sector in the emerging economy is being strengthened by financial innovations in various payment systems, including the use of ATMs, mobile banking, and e-banking. This progress has increased competition in the banking sector in many developing countries. This has a positive effect on Bank performance and customer satisfaction (Nkem and Akujinma 2017).

Financial innovation is considered one of the significant forces of banks activities. They have an impact on consumers and can improve the efficiency and profitability of the banking industry. These innovations are a product that banks use to reduce costs and improve the industry as a whole. Financial innovation is a vital force and has critical potential for improving banking performance (Kane 1981; Silber 1983). The effectiveness of banks can be measured by the capacity and ability of banks to generate the optimal level of revenue from their resources. Therefore, the role of information and communication technologies (ICT) cannot be ignored when considering financial innovation products in the banking sector (Kamau and Oluoch 2016).

Technological progress is considered one of the driving forces for creating new opportunities for the development of the banking sector in developing countries. Technological innovations are important for gaining a competitive advantage, and in the modern world this has changed the perspectives and approaches of the banking sector compared to traditional banking services (Shabbir et al. 2016).

Many studies on the introduction of financial innovation products in developing countries are based on the Fred Davis technology adoption model (TAM). This model was based on the theory of intelligent action (TRA) and allows us to explain the determinants of consciously intended behavior (Chuttur 2009). Acceptance and rejection of a technology can be predicted by the perceived ease of use and perceived utility of the technology. The model of technology adoption is consistent with

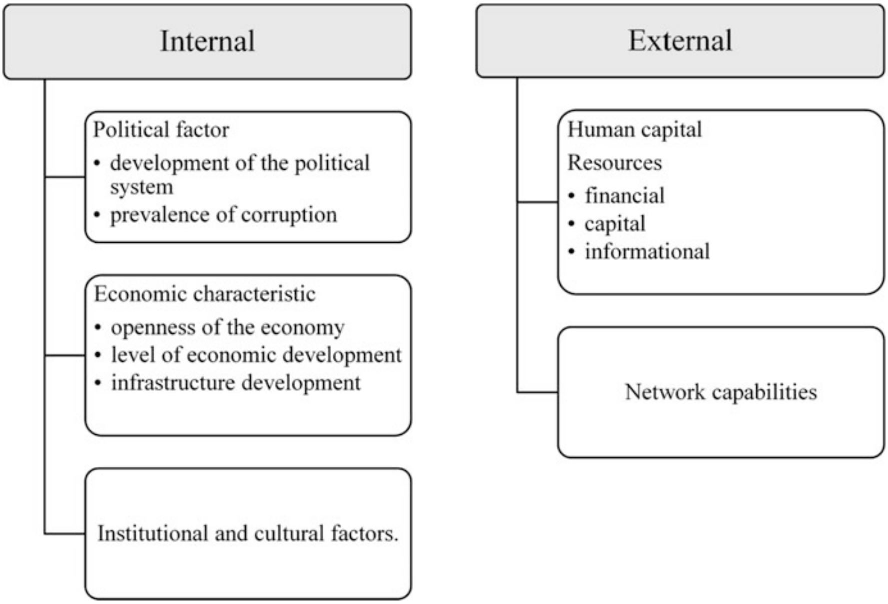


Fig. 3 Internal and external factors/barriers to the spread of innovation

Rogers'theory of the spread of innovation. Technology adoption is a function of many factors, including relative advantage and ease of use (Kabanda 2014).

For most developing countries, internal and external barriers and factors influence the creation and dissemination of innovation: (1) political factors, which include the development of the political system and the prevalence of corruption; (2) economic characteristics, including the openness of the economy, the level of economic development, and inadequate infrastructure; (3) institutional and cultural factors. The main internal factors are: (1) lack of human capital, (2) resources, mainly financial, capital, and information, and (3) network capabilities. In addition, regulation also affects the spread of innovation. Historically, Bank regulators have supported the slower spread of financial innovation and regulations that may hinder innovation (Forrer and Forrer 2014).

The level of technology in countries reflects the pace of technology diffusion within countries. Therefore, the transfer, adoption, and adaptation of knowledge to low-income countries are an important challenge for economic growth and global development. Developing countries face a number of «external» factors that act as barriers or amplifiers to the spread of innovation (Fig. 3).

Because innovation is expensive, risky, and dependent on many factors, many disruptive innovations are concentrated in rich countries and among a small number of firms. The ability of a developing country to absorb and apply foreign technologies depends on the extent to which it is exposed to foreign technologies, i.e. there are cross-country effects (World Bank 2008). The spread of knowledge in developing countries is determined by the degree of openness of the economy and the characteristics of the host country. Most of the technological progress in developing countries has been achieved through the absorption and adaptation of existing and new technologies on the market. The level of technology in developing countries reflects the pace of technology diffusion within countries. Technology diffusion transfers include trade, foreign direct investment, and social networks, while a country's absorption capacity depends on the public and business climate, technological literacy, and financing of innovative firms.

4 Conclusion

The current stage of development of the global banking system goes through a crisis and increased competition in the financial markets. One of the main factors in the development of banks is the desire to constantly introduce innovations. Currently, innovation is a key factor in the stability, competitiveness, and sustainable growth of banks and other financial institutions. Innovations in information technology and FinTech increase financing efficiency. Ultimately, this contributes to economic development. FinTech-driven financial inclusion clearly illustrates the positive link between finance and the economy. People in developing countries access financial services through FinTech. This leads to the expansion of e-commerce. FinTech contributes to the economic development of the country. However, in developed countries, where basic financial services are already widespread, it is rather difficult to quantify the impact of FinTech on the economy.

In general, innovations in developing countries contribute to increasing the competitiveness of financial institutions in the world market, reducing the cost and simplification of financial transactions, creating new jobs, and increasing labor productivity, as well as the emergence of new goods and services.

The globalization of financial markets is driving the transition to a more homogeneous banking market. This leads to the development and implementation of innovative technologies to obtain a competitive advantage. Due to globalization, banking content is changing and becoming more complex and diversified.

Technological progress is considered one of the driving forces for creating new opportunities for the development of the banking sector in developing countries. Technological innovations are essential for gaining a competitive advantage, and in the modern world this has changed the perspectives and approaches of the banking sector compared to traditional banking services.

References

Abor, J. (2005). Technological innovations and banking in Ghana: An evaluation of customers' perceptions. *IFE Psychologia: An International Journal, 13*(1), 170–187.

Abubakar, A. A., & Tasmin, R. B. H. (2012). The impact of information and communication technology on banks' performance and customer service delivery in the banking industry. *International Journal of Latest Trends in Finance and Economic Sciences, 2*(1), 80–90.

Almeida, R., & Fernandes, A. M. (2008). Openness and technological innovations in developing countries: Evidence from firm-level surveys. *Journal of Development Studies, 44*(5), 701–727. https://doi.org/10.1080/00220380802009217.

Ang, J. B., & Kumar, S. (2014). Financial development and barriers to the cross-border diffusion of financial innovation. *Journal of Banking & Finance, 39*, 43–56. https://doi.org/10.1016/j.jbankfin.2013.10.011.

Barney, J. (1991). Firm resources and sustained competitive advantage. *Journal of Management, 17* (1), 99–120.

Bass, F. M. (1969). A new product growth for model consumer durables. *Management Science, 15* (5), 215–227.

Burns, A. (2009). Technology diffusion in the developing world. In *The World Bank and OECD, innovation and growth: Chasing a moving frontier*. Cambridge, MA: OECDiLibrary. https://doi.org/10.1787/9789264073975-10-en.

Chuttur, M. Y. (2009). Overview of the technology acceptance model: Origins, developments and future directions. *Working Papers on Information Systems, 9*(37), 9–37.

Forrer, A. S., & Forrer, D. A. (2014). Analysis of the relationship between economic cycle swings and adoption rate models of financial innovation diffusion. *Journal of Business & Economics Research, 13*(2), 103–114.

Fu, X., Pietrobelli, C., & Soete, L. (2011). The role of foreign technology and indigenous innovation in the emerging economies: Technological change and catching-up. *World Development, 39*(7), 1204–1212.

Jdanova, O. R., & Karminsky, A. (2013). The diffusion of banking innovations: Bank cards on Russian market. *Innovative Marketing, 9*(3). https://businessperspectives.org/journals/innovative-marketing/issue-137/the-diffusion-of-banking-innovations-bank-cards-on-russian-market

Kabanda, G. (2014). Technology affordances and diffusion for mobile connectivity and applications in Zimbabwe. *International Journal of Emerging Technology and Advanced Engineering, 4*(6), 13–23.

Kamau, D. M., & Oluoch, J. (2016). Relationship between financial innovation and commercial Bank performance in Kenya. *International Journal of Social Sciences and Information Technology, 2*(4), 34–47.

Kane, E. J. (1981). Accelerating inflation, technological innovation, and the decreasing effectiveness of banking regulation. *The Journal of Finance, 36*(2), 355–367.

Nakaso, H. (2016). FinTech – Its Impacts on Finance, Economies and Central Banking. *Bank of Japan*. https://www.boj.or.jp/en/announcements/press/koen_2016/data/ko161118a.pdf.

Nkem, I. S., & Akujinma, A. F. (2017). Financial innovation and efficiency on the banking sub-sector: The case of deposit money banks and selected instruments of electronic banking, Asian journal of economics. *Business and Accounting, 2*(1), 1–12.

Pennings, J. M., & Harianto, F. (1992). The diffusion of technological innovation in the commercial banking industry. *Strategic Management Journal, 13*(1), 29–46.

Roberts, P. W., & Amit, R. (2003). The dynamics of innovative activity and competitive advantage: The case of Australian retail banking, 1981 to 1995. *Organization Science, 14*(2), 107–122.

Rogers, E. M. (2003). Diffusion of innovations. In D. Stacks & M. Salwen (Eds.), *An integrated approach to communication theory and research* (pp. 1962–1995). Abingdon, UK: Routledge.

Shabbir, M. S., Rehma, A. K., & Shabbir, T. (2016). Combine effect of automated services and traditional services quality on customer satisfaction: Evidence from banking sector of Pakistan. *International Journal of Economics & Management Sciences, 5*(327), 1–8.

Silber, W. L. (1983). The process of financial innovation. *The American Economic Review, 73*(2), 89–95.

Suriñach, J., Autant-Bernard, C., Manca, F., Massard, N., & Moreno, R. (2009). *The diffusion/adoption of innovation in the internal market*. Brussels, Belgium: Directorate-General for

Economic and Financial Affairs Publications, European Commission. https://doi.org/10.2765/40088.

Tahir, S., Shah, S., Arif, F., Ahmad, G., Aziz, K., & Ulla, M. (2018). Are financial innovations improving performance? Analysis of technological innovations used in Pakistan. *Journal of Innovation Economics and Management, 27*(3), 195–214. https://doi.org/10.3917/jie.027.0195.

Victor, O. I., Obinozie, H. E., & Echekoba, F. N. (2015). The effect of information communication technology and financial innovation on performance on Nigerian commercial banks 2001–2013. *International Journal of Accounting Research, 42*(2437), 1–16.

World Bank. (2008). *Global economic prospects 2008: Technology diffusion in the developing world*. http://siteresources.worldbank.org/INTGEP2008/Resources/complete-report.pdf

Dynamic Fractal Asset Pricing Model for Financial Risk Evaluation

Bruno de Conti, Vladimir Gisin, and Irina Yarygina

Abstract This article is dedicated to the assessment of the dynamic fractional asset pricing model for financial risk evaluation and the use of the fractal markets theory to mathematically predict the price dynamics of assets as part of a financial risk management strategy. The article identifies recommendations for assessing financial risk based on mathematical methods for forecasting economic processes. Theoretical and empirical research methods were used. The article reveals the features of mathematical modeling of economic processes related to asset pricing in a volatile market. It is shown that financial mathematics in banking contributes to the stable development of the economy. The mathematical modeling of the price dynamics of financial assets is based on a substantive hypothesis and supported by fractal pair pricing models in order to reveal the specific market relations of business entities. According to the authors, the prospects of using forecast models to minimize the financial risks of derivative financial instruments are positive. The authors conclude that the considered methods contribute to managing financial risks and improving forecasts, including operations with derivatives.

Keywords Banking · Asset valuation · Economic and mathematical methods · Financial risk management · Hedging

JEL Classification G01 · G12 · G32 · G34 · G38

B. de Conti
Universidade Estadual de Campinas, Campinas, Brazil

V. Gisin
Financial University, Moscow, Russia

I. Yarygina (✉)
Financial University, Moscow, Russia

MGIMO University, Moscow, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_17

1 Introduction

The fractional market hypothesis (FMH) is important for assessing the dynamic fractional asset pricing model for financial risk evaluation. However, the mathematical apparatus of this theory (a model based on fractional Brownian motion) did not keep up with the substantial concept. The lack of an adequate mathematical fractional dynamics pricing model at the time of the fractional market hypothesis formation prevented the creation of a meaningful theory. Attempts to revise the classical theory were prompted by the peculiarities of market relations and by stylized facts. Stylized facts describe features of the statistical description of price dynamics that do not fit into the framework of the classical theory, namely, the excess volatility of asset returns and the heavy tails of price distributions. EMH underestimates the probability of extreme events and the asymmetry of the left and right tails of the return distributions (Sornette 2014). The autocorrelation of asset returns takes place and homogeneous assets can exhibit the absence of the dependence of profitability increments and the existence of a significant long-term memory. Some other aspects such as the clustering of volatility and the correlation of volume and volatility are worth mentioning in this respect. Trading volume and volatility show the same type of "long-term memory" behavior. The study of these phenomena began in the 1980s. However, the mathematical modeling of individual stylized facts was first carried out by researchers of the twentieth and twenty-first century (Cont 2001). It has been shown that market development features are directly related to risk assessment and the need to use predictive mathematical models for adequate asset management solutions to stabilize economic processes. A universal mathematical model of market asset price dynamics has not yet been developed. For example, there were studies aimed at finding a theoretical model that explains the market relations phenomena conducted within the European Central Bank (ECB) in 2014 and based on the data analysis of the developed economies of the EU countries (Hiebert et al. 2018). It is not possible to use this approach to predict the processes of emerging markets. The study of cryptocurrency price dynamics by the representatives of the European mathematical school in 2017 showed that the forecasting of price dynamics in cyberspace has some points in particular (Hiebert et al. 2018).

In this regard, the observation made in 2019 in stochastic financial mathematics is interesting (Restocchi et al. 2019). Analyzing the stylized facts of economic development on a large amount of statistical data, the authors concluded that emerging markets behave like markets where various political forecasts are implemented. This confirms the role of general and specialized information in banking. An attempt to connect the stylized facts of market phenomena and the behavior of economic agents involves multi-agent models, including those using artificial intelligence, where market participants implement a relatively rational asset management strategy to manage profit and risk (Pruna et al. 2016; Dhesi and Ausloos 2016). However, criticisms of multi-agent forecasting models, especially in emerging markets, which present particularities absent in developed markets, remain valid. In complex forecasting models of highly volatile non-traditional markets, the use of "non-standard" models is promising. The main asset-pricing theorem was proved for markets where mathematical modeling was not possible (Acciaio et al. 2016).

Dolinsky and Neufeld (2018) introduced the concept of fully incomplete markets and proposed a mathematical prediction of an asset hedging strategy. To calculate asset price dynamics and manage financial risks, it is necessary to use comprehensive information about real prices and virtual derivative financial instruments.

The variety of methods and models used in modern financial mathematics show that a unifying concept that generalizes the classical one and explains the stylized facts of market relations has not been found. The most systematic and consistent explanation of the stylized facts of economic development is obtained within the concept of a fractal market, involving the dependence of the predicted value of asset price dynamics on the history of market development. This article analyzes this approach.

The key assumption of FMH is the self-similarity of the price series of assets. As a rule, the price dynamics of financial market assets are modeled using self-similar processes. This is supported by statistical observations and economic arguments. Self-similarity is conditioned by a large number of market participants with different investment horizons and acting in the same environment. Market participants behave in the same way with respect to their investment horizon. This provides the invariance of market characteristics relative to the time scale. The Hurst index, H, is the statistical characteristic of scale invariance, with values ranging from 0 to 1. For Brownian motion, underlying the classical models of a volatile market, the Hurst index is 0.5. A value of H in the range of 0.5–1 indicates persistent (trend-stable) dynamics in the time series. A value in the range 0–0.5 indicates anti-persistent dynamics in the time series and demonstrates the tendency to return to the average value.

The mathematics describing self-similar random processes was developed by Kolmogorov (1940). Methods for accurate forecasts related to asset pricing have been developing for about half a century. However, no decisive results, such as the Black-Scholes-Merton model, have been found yet. The reason is that using fractional Brownian motion for asset price modeling in the stock market faces a difficult problem. Unlike classical models, the models based on fractional Brownian motion have arbitrage opportunities that cannot be described by the rational pricing theory.

For a long time, researchers believed that the existence of arbitrage opportunities was inextricably linked with autocorrelation and the memory of financial time series. A deeper penetration into the mathematics of the fractional market shows that arbitrage, autocorrelation, and self-similarity are due to various factors. Cheridito (2004) provides examples of Gaussian random processes which have the same long-term memory as the processes based on fractional Brownian motion with $H > 0.5$, and that lead to arbitrage-free market models. To build the price model, Rostek and Schobel (2013) applied the idea of a moving average. The methods proposed in Cheridito (2004) effectively connect the mathematical techniques with the market realities understood by financiers.

Thus the article is structured in four further sections. Section 2 discusses fractional Brownian motion and market models. In Sect. 3, the index of fractality is presented. Section 4 develops the analysis of volatility forecasting based on fractal characteristics. Section 5 presents some directions for further research in this field.

2 Fractional Brownian Motion and Market Models

Most researchers find it more promising to use fractional Brownian motion to build a market model. Replacing Ito integration with Wick integration can solve the problem of arbitrage opportunities. A significant disadvantage of Wick integration is the lack of a convincing economic interpretation. Therefore, mathematical models using Wick integration should be treated with caution. To solve the problem of minimizing financial risks using the mathematical modeling of price indicators of derivative financial instruments, a more complete account of trading financial instruments in a concrete financial market is required. One of the most promising areas is related to taking into account transaction costs. The fractional market with proportional transaction costs is arbitrage-free. Basically, the exact pricing of financial derivatives in such a market is fundamentally impossible. The non-arbitrage price of a financial instrument is determined not as a point, but as a value within a price range. It is possible to establish more or less accurate boundaries of this range. However, the fractional market hypothesis attracts participants by the opportunity to minimize the financial risks of asset management.

Classical predictive models suggest that a stochastic process with underlying Brownian motion describes risky asset price dynamics. Namely, let $S(t)$ be the price of the risky asset at time t. Then the return over a short interval of time $[t, t + \Delta t]$ can be decomposed as follows:

$$\frac{S(t+\Delta t)-S(t)}{S(t)}=\mu\Delta t+\sigma\Delta W(t), \tag{1}$$

where Δt is a time increment, $\mu+\frac{\sigma^2}{2}$ is the expected return, σ is the return volatility, $\Delta W(t) = W(t + \Delta t) - W(t)$, and $W(t)$ is a Wiener process (Brownian motion).

The decomposition of return presented in Eq. (1) is economically reasonable. The systematic part is presented by $\mu\Delta t$ and the random part is presented by $\sigma\Delta W(t)$. The increment $\Delta W(t)$ is considered to be normally distributed with an average of 0 and variance Δt. It is assumed that $W(t_2) - W(t_1)$ and $W(s_2) - W(s_1)$ are independent unless the time intervals $[t_1, t_2]$ and $[s_1, s_2]$ overlap.

Wiener processes belong to the class of self-similar stochastic processes. Generally speaking, a stochastic process is self-similar if a change in the time scale leads to a change in the spatial scale, keeping the probabilistic characteristics of the process unchanged. More precisely, a random process $X(t)$, $t \geq 0$, is called self-similar if, for any $a > 0$, we can find $b > 0$ such that stochastic processes $X(at)$ and $bX(t)$ have the same probabilistic characteristics. If there exists H such that $b = a^H$ for all $a > 0$, then H is the Hurst index and it is said that the process is self-similar with the Hurst index H. Given a Wiener process, its Hurst index is 0.5.

If changes in the return over non-overlapping time intervals are independent, it is reasonable to model return dynamics by Levy processes. Models based on Levy processes provide a good approximation of real price time series, sometimes much better than classical models (Shoutens 2003). They allow the consideration of such features as asymmetry and the heavy tails of probability distributions, and thereby more adequately assess risks (for example, ignoring heavy tails leads to underestimation of the risks associated with extreme events, which may be particularly important in emerging markets). To a large extent a better approximation is achieved due to a larger number of parameters. Typically, to describe a Levy process four parameters are used. Two of them are similar to the parameters of the Wiener process: the shift parameter (similar to the average value which may be determined in the Levy process) and the scale parameter (similar to the average deviation which cannot be determined for the Levy process). The other two parameters consider the features of time series not captured by Wiener processes.

Shoutens (2003) and Ferger et al. (2017) show that using Levy processes to describe the returns of world stock indices provides satisfactory results. Using Levy processes, it is possible to take into account the dynamic features of financial time series missing in classical models. Similar results are obtained regarding the Russian market in Gisin et al. (2012).

The predictive ability is an important property of the model. To be considered qualitative and predictively valuable, the model should be sufficiently stable with respect to small fluctuations in the initial data and relatively small shifts along the time axis. In this regard, increasing the number of parameters allows for a more accurate calibration on historical data, but the stability of the estimates is problematic. Data analysis shows that models with a normal distribution show good results for periods of 1–2 months. With a forecast period of more than 200 days, both classical models and those based on Levy processes are not entirely reliable. Finally, for periods of 100–150 days, models based on Levy processes provide the best results. Using non-classical models for the Russian market is more significant. For example, for the DJA, the distributions in the corresponding Levy processes are close to normal, and both are consistent with empirical data. It is no longer the case for the RTS index due to high transaction costs (we also include the costs due to insufficient liquidity). Fractional Brownian motion is a basic example of a self-similar random process with dependent increments. This dependence makes it possible to simulate processes with long-term memory using fractional Brownian motion. The phenomena related to trend formation are explained within such models.

Applying financial time series models based on self-similar processes can face fundamental difficulties, either with dependent or independent increments. In the classical Black-Scholes-Merton model, pricing is based on the fact that this model has an equivalent martingale probability measure. Substantially, this measure can be interpreted as some rational forecast, and the price of a derivative instrument is determined considering this forecast with respect to its future prices. In general, there is an infinite family of "rational forecasts" for self-similar processes with independent increments. Accordingly, there is an interval of prices interpreted as "fair." It is sometimes possible to estimate the boundaries of these intervals, which are often shallow. In models using fractional Brownian motion, with a Hurst index other than 0.5, there is no "rational forecast" (equivalent martingale measure), and there are arbitrage opportunities. Building pricing models within such models is only possible considering the features of real financial markets. Transaction costs are among these features. Boundaries of the fair price interval were investigated in Gerhold et al. (2014) and Guasoni and Weber (2017), see also (Guasoni et al. 2019). The authors connected trading volumes and the liquidity and dynamic parameters of price movement and got estimates allowing for optimal trading strategies. These papers make relevant the issue of the consistent use of the so-called market time in models. This concept has been used in many works. The results obtained in Gerhold et al. (2014) open up new possibilities for the Tobin tax. In our opinion, studies clearly indicate that in financial market models it is advisable to link time with financial events, and not just with the rotation of the Earth around the Sun. Using the fractional modeling method is promising for the management of financial risks in difficult market conditions when forecasting asset price dynamics.

Fractional Brownian motion with the Hurst index $0 < H < 1$ is a random process $\{B^H(t)\}$, where $B^H(0) = 0$, random variables $B^H(t)$ are normally distributed for all t, the mean value of $B^H(t)$ is 0 for any t, and the covariance of $B^H(t)$ and $B^H(s)$ is as follows:

$$E\left[B^H(t)B^H(s)\right] = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right). \tag{2}$$

Equivalently, we can assume that the variance of $B^H(t)$ is proportional to t^{2H} (in the Wiener process, it is proportional to t).

The trajectory of fractional Brownian motion is a fractional object with a fractional dimension. Using fractional Brownian motion, it is possible to build market models with many important properties, whose manifestation is demonstrated by real markets. We call such models "fractional markets." One of the most important and well-studied is a model similar to Eq. (1), where the risky asset price dynamics is described as follows:

$$\begin{gathered} D = 2 - H; \\ \frac{S(t+\Delta t) - S(t)}{S(t)} = \mu\Delta t + \sigma\Delta B^H(t). \end{gathered} \tag{3}$$

The behavior of the autocovariance function with the lag τ (assuming it is sufficiently large) is similar to the behavior of the function $2H(2H - 1)\tau^{2H-2}$. For all values of the Hurst index, autocorrelation tends to 0 with an increase in the time lag. At $H > 0.5$, autocorrelation is positive and decreases more slowly, the higher the value of H. For example, at $H = 0.8$, autocorrelation remains quite noticeable (approximately 0.2) even at $\tau = 10$. This case corresponds to persistence. At $H < 0.5$, autocorrelation becomes negative at $\tau < 1$, reaches its minimum value, and then tends to zero with increasing lag. This case corresponds to anti-persistency.

These properties of the Hurst index are associated with crisis phenomena. Empirical observations allow us to conclude that a decrease in the fractal dimension of the price trajectory precedes large changes in the markets. The fractional characteristics of markets in the period up to 2014 were analyzed in Guasoni and Weber (2017). With this in mind, studying the dynamics of the Hurst index becomes relevant. Navascués et al. (2016) and Dubovikov et al. (2004) studied this problem, and the concept of the index of fractality μ associated with the Hurst index by the relation:

$$H \approx 1 - \mu. \tag{4}$$

The dynamics of μ allows a statistically reliable description and, due to this, can be used for forecasting. In Putko et al. (2014) promising econometric approaches were proposed to describe the dynamics of the Hurst index. Similar method of Index determination was proposed in 2019 (Savitskii 2019; Song et al. 2019), but it did not include a transparent economic interpretation.

3 Index of Fractality

Dubovikov et al. (2004) introduced the variation index μ. Since the trajectories of fractional Brownian motion have a topological dimension of one, the index of variation coincides with the index of fractality (the difference between fractal dimension and topological dimension). In this section, we provide a definition of the index of fractality and describe how it can be calculated. In what follows, the index of fractality is used to model volatility.

Consider a time interval δ. Denote by $h(\delta)$ and $l(\delta)$, respectively, the maximum and the minimum price in this interval. Let $A(\delta) = h(\delta) - l(\delta)$. We use the amplitude $A(\delta)$ as a measure of volatility over an interval δ.

Now let δ_0 and δ_c be the time intervals such that $\delta_c = 2^n\delta_0$ for some $n > 0$. Let $\delta = 2^k\delta_0$, where $0 \leq k \leq n$. Consider a time interval $[t - \delta_c, t]$. It can be divided into 2^{n-k} intervals of length δ. The total of the amplitudes at these intervals is denoted by $V(\delta)$. Consider the regression:

$$\log V(\delta) = \alpha - \mu \log(\delta). \tag{5}$$

Dubovikov et al. (2004) show that regression (4) has a very high coefficient of determination. It almost coincides with 1 in a wide range (the authors considered δ varying from $8\delta_0$ through $1{,}024\delta_0$ with $\delta_c = 2^n\delta_0$). Thus, the estimate of μ is practically independent of the choice of divisors of δ_c, and we can consider the dynamic characteristics $\mu(t, \delta_0, \delta_c)$ and $\alpha(t, \delta_0, \delta_c)$. As a rule, $\delta_0 = 1$, and these characteristics are denoted by $\mu_{\delta_c}(t)$ and $\alpha_{\delta_c}(t)$.

Note that μ (unlike α) does not depend on the base of the logarithm in Eq. (5) and is an intrinsic characteristic of the fractal structure of the financial time series. Following Dubovikov et al. (2004), we say that μ is the index of fractality. When δ_c is small, the index of fractality is close to $D - 1$, where D is the fractal dimension of the stochastic price process. Since the convergence to $D - 1$ is very fast, we can estimate the fractal dimension using a small number of observations.

As a consequence of the very high coefficient of the determination of regression (4), we can use simplified estimates μ_s, α_s of μ, α. Assuming $\delta_0 = 1$ we have:

$$\alpha_s = \log_{\delta_c} V(\delta_0); \mu_s = \log_{\delta_c} V(\delta_0) - \log_{\delta_c} V(\delta_c). \tag{6}$$

Equations (6) provide the following decomposition of volatility with respect to δ_c:

$$\log_{\delta_c} V(\delta_c) = \alpha_s - \mu_s \approx \alpha - \mu. \tag{7}$$

4 Volatility Forecasting Based on Fractal Characteristics

Following Dubovikov et al. (2004), regression models for μ and α can be considered. By forecasting μ and α, we can forecast the volatility. This forecast has a distinctive feature. In most models, the future value itself is predicted, but usually for a short interval. The fractional model allows us to predict only the direction of growth of the values α and μ, but for a sufficiently long interval (from 1 to 8 months). The dependence of μ on t has a well-defined quasi-cyclic structure. This is the basis for building an econometric model. The quasi-cyclicality of fractal characteristics (in particular, the Hurst series dynamic) has been pointed out and discussed at the qualitative level before. Thus, it is logical to use periodic functions to model the index of fractality. So we present μ as follows:

$$\widehat{\mu}(t) = \sum_{i=1}^{k} [a_i \sin(\omega_i t) + b_i \cos(\omega_i t)]. \tag{8}$$

The econometric model corresponding to Eq. (8) is constructed as follows. First, we fix a time interval $[T_0, T_1]$. Let $\Delta = T_1 - T_0$ be the window width. Then the following equation is considered:

$$\mu(t) = x + b_1 \sin(\omega t) + b_2 \cos(\omega t) + \varepsilon(t). \tag{9}$$

We set the frequency ω to run values $0.0001k$, $k = 0, \ldots, 10000$. For each value of ω, the coefficient of determination $R^2(\omega)$ is determined. The maximum values of

$R^2(\omega)$ are well defined. The lowest maximum that has the highest value of $R^2(\omega)$ gives the main trend frequency. In addition, there are three or four maximums. They present the frequencies of the quasi-cycles.

At large intervals of T_0, Δ, the situation does not change qualitatively and is subject to only small quantitative changes. This confirms the quasi-cyclicality of the structure. Some values of T_0 cause phase transitions. The main trend frequency in Eq. (9) is bifurcated with the subsequent "overflow," the damping of the original "hump" and the increase of the new one.

These ideas were used in Putko et al. (2014) to predict trends in the ruble exchange rate. The regressions had a sufficiently high coefficient of determination R^2~0, 7 ÷ 0, 75. Backtests of the model showed that the direction of the exchange rate trend is predicted correctly in 60–70% of cases. The situation with the 2008 crisis turned out to be very well coordinated with the model.

Bertrand et al. (2018) and Ikeda (2017a, b) provide a large amount of data on the study of the values of the Hurst index on the stock market, which generally confirm this pattern. In this regard, the increase in the Hurst index in the Russian oil sector observed in 2019 is alarming. The Hurst index values close to 0.6 are typical for the Russian stock market (Aeroflot 0.58–0.63; Gazprom 0.53–060; Sberbank 0.57–0.64; Rosneft 0.53–0.57) in 2014–2018. These values were replaced in the first half of 2019 by higher ones (Tatneft 0.70; Surgutneftegaz 0.77; Rosneft 0.72).

In this regard, we refer to a study by the Utrecht University Faculty of Science, which provides estimates of the "normal" values of the Hurst index for various sectors: information technology 0.50–0.67; Finance 0.38–0.62; raw materials sector 0.38–0.63 (Guennoun et al. 2018). In the fractional market there is no martingale measure and, accordingly, there are arbitrage opportunities. The latter is related to the properties of the Ito integral. Mathematically, the situation may be corrected using Wick integration. However, this method of integration has not received adequate economic interpretation. This approach is easy to explain using a discrete approximation of the fractional Brownian motion, which serves as the main tool for calculations. We give a brief description of the discrete approximation.

Let the time interval $[0; T]$ be divided into n equal intervals. Let ξ_i, $i = 1, \ldots, n$, be random variables such that $\xi_i \in \{-1; 1\}$ and $P(\xi_i = 1) = P(\xi_i = -1) = \frac{1}{2}$. For each n, we can calculate the coefficients $k_{l,i}^{(n)}$, $l = 1, \ldots, n$, $i = 1, \ldots, l$, so that the sum:

$$\sum_{i=1}^{l} k_{l,i}^{(n)}\ \xi_i \tag{10}$$

approximate $B^H(t)$ for $t = l \cdot \frac{T}{n}$. Then we have:

$$\Delta B^H(t) = k_{l+1,l+1}^{(n)}\ \xi_{l+1} + \sum_{i=1}^{l} \left(k_{l+1,i}^{(n)} - k_{l,i}^{(n)}\right)\xi_i. \tag{11}$$

Equation (11) allows an approximation of the risky asset price in a fractional market at sufficiently large n. Let $\Delta t = \frac{T}{n}$, $S_0 = S(0)$, and μ be expected return in time interval Δt. Then.

$$S(\Delta t) = S_0\left(1 + \mu\Delta t + k_{1,1}^{(n)}\,\xi_1\right);$$
$$S(2\Delta t) = S(\Delta t)\left(1 + \mu\Delta t + k_{2,2}^{(n)}\,\xi_2 + \left(k_{2,1}^{(n)} - k_{1,1}^{(n)}\right)\xi_1\right), \tag{12}$$

and so on.

The Ito integration corresponds to the usual multiplication of terms. The Wick integration corresponds to the multiplication where the terms containing ξ_i^2 are discarded. So by the Ito integration we get:

$$S_I(2\Delta t) = S_0(1 + \mu\Delta t)^2 + k_{1,1}^{(n)}\left(k_{2,1}^{(n)} - k_{1,1}^{(n)}\right) + \left(k_{1,1}^{(n)}\,\xi_1 + k_{2,2}^{(n)}\,\xi_2\right)$$
$$\times (1 + \mu\Delta t) + k_{1,1}^{(n)} k_{2,2}^{(n)} \xi_1\xi_2, \tag{13}$$

and by the Wick integration:

$$S_W(2\Delta t) = S_0(1 + \mu\Delta t)^2 + \left(k_{1,1}^{(n)}\,\xi_1 + k_{2,2}^{(n)}\,\xi_2\right)(1 + \mu\Delta t) + k_{1,1}^{(n)} k_{2,2}^{(n)}{}_1\xi_2, \tag{14}$$

Therefore the Ito approximation and the Wick approximation differ in the trend component by:

$$k_{1,1}^{(n)}\left(k_{2,1}^{(n)} - k_{1,1}^{(n)}\right). \tag{15}$$

There is still no economically reasonable explanation concerning this difference. So we should be cautious about the results obtained by the Wick integration.

5 Directions for Further Research

Let us focus on the results related to pricing in markets with transaction costs. Kabanov and Safarian (2010) and Karp and Van Vuuren (2019) found an approach to describe the optimal strategies in markets with transaction costs. Under general assumptions, the ratio of capital invested in the risk component should be within the boundaries:

$$\pi_{-} = \frac{\rho - \lambda}{\gamma\sigma^2} \text{ and } \pi_{+} = \frac{\rho + \lambda}{\gamma\sigma^2} \tag{16}$$

where ρ is the excess return; γ is the relative risk aversion; ε is the spread between bid and ask prices, and λ is defined as follows:

$$\lambda = \gamma\sigma^2\left(\frac{3}{4\gamma}{\pi_*}^2(1-\pi_*)^2\right)^{1/3}\varepsilon^{1/3} + O(\varepsilon) \tag{17}$$

with $\pi_* = \frac{\rho}{\gamma\sigma^2}$.

For example, calculations using Eqs. (16) and (17) for Sberbank's ordinary shares in early 2014 gave:

$$\pi_{-} = 45.6\%, \pi_{+} = 48.2\%. \tag{18}$$

The liquidity premium calculated using the method from Kabanov and Safarian (2010) was equal to 0.04%. For assets with lower liquidity, boundaries were significantly lower, and the liquidity premium increased sharply. For example, for Bank "Primorye," the liquidity premium was 0.15%.

Recently, a significant number of studies have been devoted to modeling volatility using fractional Brownian motion. Within the framework of the constructed models, it is possible to explain the effects of short-term and long-term memory, the volatility smile, and some other features (Nika and Rasonyi 2018). The concept of rough fractional stochastic volatility (RFSV) has become widespread (Guennoun et al. 2018; Bayer et al. 2016). RFSV generalizes models with stochastic volatility that have been used for more than 20 years (see Gatheral et al. 2018). In the standard model of stochastic volatility described by the equations:

$$\frac{dS(t)}{dt} = \mu(t, S(t))dt + \sigma(t)dW^{(1)}(t); \tag{19}$$

$$d(\ln\sigma(t)) = k(\theta - \ln\sigma(t))dt + \gamma dW^{(2)}(t), \tag{20}$$

where $W^{(i)}(t)$ is a Wiener process, $i = 1, 2$ and it is proposed to use a fractional Brownian motion instead of $W^{(2)}(t)$. Research in this direction was stimulated by the fact that a stable pattern was empirically revealed using high-frequency data: the volatility dynamics are fractional, the Hurst index of the process $W^{(2)}(t)$ is 0.1 (Bayer et al. 2016). This value of the Hurst index corresponds to a very high volatility with a tendency to return to the mean. This observation makes it possible to significantly improve volatility forecasts, and, most importantly, to describe the possible risks and implied volatility of asset price dynamics much more accurately than using other models. The proposed approach is also promising for forecast models of the price dynamics of derivative financial instruments (Guennoun et al. 2018). In addition, fractal volatility parameters demonstrate predictive power relative to extreme events in the financial sector. An example is the collapse of Lehman Brothers and other US investment banks in 2008, which caused the global financial and economic crisis (Guennoun et al. 2018).

The justification of the feasibility of fractional models of asset price dynamics and their practical application in the financial sphere can help to minimize risks and strengthen the stable development of market relations.

References

Acciaio, B., Beiglbock, M., Penkner, F., & Schachermayer, W. (2016). A model-free version of the fundamental theorem of asset pricing and the super-replication theorem. *Mathematical Finance, 26*(2), 233–251.

Bayer, C., Friz, P., & Gatheral, J. (2016). Pricing under rough volatility. *Quantitative Finance, 16* (6), 887–904.

Bertrand, P. R., Combes, J. L., Dury, M. E., & Hadouni, D. (2018). Overfitting of Hurst estimators for multifractional Brownian motion: A fitting test advocating simple models. *Risk and Decision Analysis, 7*(1–2), 31–49.

Cheridito, P. (2004). Gaussian moving averages, semimartingales and option pricing. *Stochastic Processes and their Applications, 109*(1), 47–68.

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. *Quantitative Finance, 1*, 223–236.

Dhesi, G., & Ausloos, M. (2016). Modelling and measuring the irrational behavior of agents in financial markets: Discovering the psychological soliton. *Chaos, Solitons & Fractals, 88*, 119–125.

Dolinsky, Y., & Neufeld, A. (2018). Super-replication in fully incomplete markets. *Mathematical Finance, 28*(2), 483–515.

Dubovikov, M. M., Starchenko, N. V., & Dubovikov, M. S. (2004). Dimension of minimal cover and fractal analysis of time series. *Physica A: Statistical Mechanics and its Applications, 339* (3–4), 591–608.

Ferger D., Manteiga W. G., Schmidt T., & Wang J. L. (2017). *From statistics to mathematical finance* (440 p.). Springer.

Gatheral, J., Jaisson, T., & Rosenbaum, M. (2018). Volatility is rough. *Quantitative Finance, 18*(6), 933–949.

Gerhold, S., Guasoni, P., Muhle-Karbe, J., & Schachermayer, W. (2014). Transaction costs, trading volume, and the liquidity premium. *Finance and Stochastics, 18*(1), 1–37.

Gisin, V. B., Konnov, V. V., & Sharov, V. F. (2012). *Probabilistic models in analysis of asset pricing* (152 p.). Moscow: Financial University (In Russ.).

Guasoni, P., Nika, Z., & Rasonyi, M. (2019). Trading fractional Brownian motion. *SIAM Journal on Financial Mathematics, 10*(3), 769–789.

Guasoni, P., & Weber, M. (2017). Dynamic trading volume. *Mathematical Finance, 27*(2), 313–349.

Guennoun, H., Jacquier, A., Roome, P., & Shi, F. (2018). Asymptotic behavior of the fractional Heston model. *SIAM Journal on Financial Mathematics, 9*(3), 1017–1045.

Hiebert, P., Jaccard, I., & Schüler, Y. (2018). Contrasting financial and business cycles: Stylized facts and candidate explanations. *Journal of Financial Stability, 38*, 72–80.

Ikeda, T. (2017a). Fractal analysis revisited: The case of the US industrial sector stocks. *Economics Bulletin, 37*(2), 666–674.

Ikeda, T. (2017b). A fractal analysis of world stock markets. *Economics Bulletin, 37*(3), 1514–1532.

Kabanov, Y., & Safarian M. (2010). *Markets with transaction costs*. Springer.
Karp, A., & Van Vuuren, G. (2019). Investment implications of the fractal market hypothesis. *Annals of Financial Economics, 14*(1), 1–27.
Kolmogorov, A. N. (1940). Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *Reports of the Academy of Sciences of the USSR, 26*, 115–118.
Navascués, M. A., Sebastián, M. V., & Latorre, M. (2016). Stock indices in emerging and consolidated economies from a fractal perspective. In *Time series analysis and forecasting* (pp. 113–122). Cham: Springer.
Nika, Z., & Rasonyi, M. (2018). Log-optimal portfolios with memory effect. *Applied Mathematical Finance, 25*(5–6), 557–585. https://doi.org/10.1080/1350486X.2018.1542323.
Pruna, R. T., Polukarov, M., & Jennings, N. R. (2016). An asset pricing model with loss aversion and its stylized facts. In *IEEE symposium series on computational intelligence (SSCI)* (pp. 1–8). IEEE.
Putko, B. A., Didenko, A. S., & Dubovikov, M. M. (2014). Exchange rate volatility model (RUR/USD) using fractal characteristics of the financial series. *Applied Econometrics, 36*, 79–87 (In Russ.).
Restocchi, V., McGroarty, F., & Gerding, E. (2019). The stylized facts of prediction markets: Analysis of price changes. *Physica A: Statistical Mechanics and its Applications, 515*, 159–170.
Rostek, S., & Schobel, R. (2013). A note on the use of fractional Brownian motion for financial modeling. *Economic Modelling, 30*, 30–35.
Savitskii, A. V. (2019). Method for estimating the Hurst exponent of fractional Brownian motion. *Doklady Mathematics, 100*(3), 564–567.
Shoutens, W. (2003). *Levy processes in finance. Pricing financial derivatives* (170 p.). Wiley.
Song, W., Li, M., Li, Y., Cattani, C., & Chi, C. H. (2019). Fractional Brownian motion: Difference iterative forecasting models. *Chaos, Solitons & Fractals, 123*, 347–355.
Sornette, D. (2014). Physics and financial economics (1776–2014): Puzzles, Ising and agent-based models. *Reports on Progress in Physics, 77*(6), 062001.

Network Effects in Retail Payments Market: Evidence from Individuals

Egor Krivosheya

Abstract This paper evaluates empirically the effect of network externalities on individual behavior in the Russian retail payments market. Specifically, the effects of direct and indirect network externalities for cardholding and usage probabilities are examined. Using a representative sample of 1500 individuals across Russian regions, this paper finds significant robust evidence of a positive association between the degree of both types of network externalities and individuals' activity in the Russian retail payments market. Results are economically significant: a standard deviation increase in network effects leads to a 2.5–4 percentage points increase in the probability of cardholding and usage. The findings suggest a need to account for network effects that play an important role in the payment behavior before implementing any payment stimulating programs in Russia aimed at cardholders or users.

Keywords Retail payments · Payment cards · Network effects · Cardholders' behavior · Financial services

JEL G21 · D53 · E42 · L14

1 Introduction

There is an obvious trend towards a cashless economy in the modern world. On the one hand, financial regulators favor it and tend to stimulate this phenomenon both at the level of individual users and at the national level. In addition, there are other market participants, apart from the government, who contribute actively to the proliferation of non-cash payments. The retail payments market is a two-sided market. To have a payment settled with a payment card, two groups of end-users

E. Krivosheya (✉)
Moscow School of Management SKOLKOVO, Moscow, Russia

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_18

have to be involved: buyers (cardholders) and sellers accepting cards (merchants). Despite the substantial growth of the retail payments market in the high growth emerging markets regions, a significant share of users still prefer cash for transactions (Semerikova 2019; Plaksenkov et al. 2015). Whilst some measures stimulating the cashless economy have been efficient, the effect of government policy on retail payments market development is still limited (Krivosheya 2020; Krivosheya et al. 2015; Krivosheya et al. 2017; Krivosheya 2018). It may be attributed, in part, to the presence of network externalities which cannot be explicitly impacted by the stimulating proposals and initiatives.

A network effect is an effect that occurs when the utility that a user extracts from the consumption of a product or service increases with the number of other agents consuming the same product or service. There are two types of network externalities for cardholders in the retail payments market due to the two-sided nature of the retail payments market. Direct network externalities show how the probability of holding and using a card by an individual depends on the decisions of other cardholders. The indirect network effect, similarly, shows how the intention to hold and pay by card depends on the level of card acceptance by merchants. Therefore, the main purpose of this paper is to analyze empirically the effect of both types of network externalities for cardholding and card usage demands in Russia. This research aims to contribute to the literature on the determinants of cashless payments instrument holding and usage (Arango-Arango et al. 2018; Bagnall et al. 2014; Bounie and Francois 2006; Bounie et al. 2016; Carbó-Valverde and Liñares-Zegarra 2011; Gresvik and Haare 2008). The thoroughly investigated factors include transaction characteristics (e.g., cost of the purchased goods/services, type of goods, day of the week), merchants (store type, size, etc.), and socio-demographic characteristics (income, education, age, sex, employment status, etc.). However, few studies have evaluated empirically the presence of network externalities for customers, and those that have, do not distinguish between direct and indirect network externalities, especially for the card usage probability. These two types of network externalities affect the behavior of the individuals via different mechanisms and, hence, need to be separated in the empirical research.

The results of the research are important from the practical point of view as they help to understand the degree of potential influence different stimulating measures might have on the behavior of the individuals in the retail payments market, in particular, cardholding and usage. The effect of network externalities cannot be explicitly changed by any incentive programs or with other government or private sector interventions. There is therefore some probability that cannot be affected by any financial market policies. It would be valuable for the practitioners involved in the development of the financial services market such as the Central Bank of Russia, commercial banks, and payment systems to understand the degree of influence they could have on the individuals in the retail payments market. Besides, understanding the degree of network effects contributes to the understanding of the organic market growth resulting from the multiplicative effect of increased payment activity across two market sides.

Following this introduction, there are five sections in this Paper. In the subsequent section, the theoretical mechanisms of the effect of direct and indirect network externalities on cardholders' holding and usage probability will be explained. The subsequent sections explain the empirical set-up which consists of data, the empirical model description, and the estimation method. Section 7 explains the main results from a statistical and economic point of view. Section 8 identifies limitations, outlines directions for further research, and draws conclusions.

2 Theoretical Foundations of the Network Externalities in the Retail Payments Market and Hypotheses

The aim of this paper is to analyze the effect of network externalities (effects) on the probability of card holding and usage in the retail payments market. In general, network effects occur when the utility that a user extracts from consumption of the product or service increases with the number of other agents consuming this product or service. In the context of the retail payments market, this effect can be separated into direct and indirect effects (Katz and Shapiro 1985).

2.1 Individuals Benefits

A decision by an individual to hold and use a payment card is based on the relative size of the benefits and costs associated with holding and using cashless payments (Baxter 1983; Bedre-Defolie and Calvano 2013; Bolt and Chakravorti 2008; Krivosheya and Korolev 2016; Krivosheya 2020; Rochet and Tirole 2002, 2003, 2006). In any model of the retail payments market equilibrium, an individual chooses to engage in the market if the size of the net benefits (benefits associated with cashless payments compared to cash payments less any costs attributed to the cashless payment methods compared to the cash-based ones) exceeds zero (Baxter 1983; Bedre-Defolie and Calvano 2013; Guthrie and Wright 2007; Krivosheya and Korolev 2016; Rochet and Tirole 2002; Wright 2004). Direct and indirect network effects can change the size of the benefits and fees (Bedre-Defolie and Calvano 2013; Bolt and Chakravorti 2008; Krivosheya and Korolev 2016). To begin with, it is important to define both concepts in the context of the work.

Individuals make two decisions in the retail payments market: first, they choose whether to hold a card and, then, they choose whether to use a card in payment for goods and services (e.g., Baxter 1983; Bolt and Chakravorti 2008; Krivosheya and Korolev 2016). As a result, benefits are usually separated into fixed and variable (Bedre-Defolie and Calvano 2013; Krivosheya and Korolev 2016). Variable benefits represent the benefits arising from each particular transaction. Such benefits may be manifested, for instance, in the form of increased speed of transactions, satisfaction from paying with a card compared to cash, ability to defer payments, lower risk of fraud, or easier personal financial management (Baxter 1983; Bedre-Defolie and Calvano 2013; Grauwe et al. 2002; Guthrie and Wright 2007; Krivosheya and Korolev 2016). Fixed benefits represent the benefits from holding a card instead of holding alternative methods of payment (e.g., cash or cheques). They, therefore, do not depend on the number of transactions. Examples of fixed benefits include the improved security and protection against robberies and the ability to consume more due to easier usage (e.g., no withdrawal costs, no need to calculate the necessary amount of cash holdings before transactions) (Bedre-Defolie and Calvano 2013; Grauwe et al. 2002; Hunt 2003; Krivosheya and Korolev 2016).

In this context, a person will have a card issued to him or her if his/her fixed benefits are greater than the costs of being issued with a card. A person will use the card for payments for goods and services if their variable benefits are greater than the variable costs of using the card, which are usually zero for Russian market. Network externalities may affect the value of all these four parameters, thereby altering the demand for cardholding and card usage.

2.2 *Direct Network Effects*

Direct network effects, in the context of this study, result from the increased activity (cardholding and card usage) of the cardholders. Direct network effects are associated with the increase in the demand for the issuing bank services, which may increase the interest on the remaining account balances and other bonuses (e.g., business passes to the airport lounges, concierge services, etc.) for holding money on the card account when the number of cardholders rises (Borzekowski et al. 2008; Ching and Hayashi 2010; Hayashi 2009; Humphrey 2010). In cases where the number of cardholders is lower, issuing banks can easily segment the potential cardholders and find its own niche among the individuals without payment cards (Hasan et al. 2012; Meadows and Dibb 1998; Todd and Lawson 2003). Segmentation of the potential customers allows issuers to charge higher fees than if they have to compete for existing cardholders with other issuing banks (Hasan et al. 2012; Todd and Lawson 2003). The quality of services and the level of fees are among the key factors for cardholders for choosing a bank (Arango-Arango et al. 2018; Bagnall et al. 2014; Borzekowski et al. 2008; Bounie and Francois 2006; Bounie et al. 2016). Taking this fact into the account, issuing banks are likely to change the quality of services without increasing the fees levied on the individuals or decrease the fees without decreasing the quality of services (Baxter 1983; Bedre-Defolie and Calvano 2013; Hasan et al. 2012; Rochet and Tirole 2002), as has been shown by Russian banks recently (Chernikova et al. 2015; Chizhikova 2013; Krivosheya and Korolev 2016).

Another important factor for the size of fixed benefits for individuals is the perception of holding a payment card (Baxter 1983; Bedre-Defolie and Calvano 2013; Krivosheya and Korolev 2016). A payment card, especially of premium type, may be considered as a signal of status (Arango-Arango et al. 2018; Roberts and Jones 2001; Souvignet et al. 2014). The larger the share of cardholders, the more likely other cardholders are to recognize the difference between payment instrument types. In addition, cardholders are subject to herd behavior: once individuals see others with payment cards, they start associating it with lower risks, higher benefits, and an overall more positive experience (Bagnall et al. 2014; Darban and Amirkhiz 2015; Shy 2011).

Finally, an increase in the share of cardholders leads to higher payment systems spending on anti-fraud systems and other aspects of security owing to the economies of scale in the industry (Kadhiwal and Zulfiquar 2007; Kim et al. 2010). Security of cashless payments has been an issue of particular focus for the payment systems during past few decades due to the increase in cyber risks and data breaches (Kim et al. 2010). As part of the response, payment systems have started more heavily to invest in the anti-fraud systems, especially in regions of higher cashless usage and holding. This has also led to the standardization of the fraud management systems across banks (Kadhiwal and Zulfiquar 2007; Kim et al. 2010). Overall, all of the mechanisms outlined above suggest that the direct network effects should be associated positively with the cardholding demand. The first hypothesis is, therefore:

H1: The probability of cardholding increases with a larger share of cardholders and users of cashless payments.

In order to investigate the effect of direct network externalities on card usage demand by cardholders we need to analyze how a bigger number of cardholders and card users affects net variable benefits. Similarly, an increase in the number of card users and cardholders is equivalent to the increase in the demand for the issuing banks' services, which may result in better loyalty (e.g., cashback and bonus) reward programs or other incentives activated per each transaction (Bedre-Defolie and Calvano 2013; Carbó-Valverde and Liñares-Zegarra 2011; Hasan et al. 2012; Krivosheya and Korolev 2016; Rochet and Tirole 2002). Once the number of cardholders rises, issuing banks start to compete for the existing card users with other issuers, thereby improving the quality of services for the same or lower usage fees (providing better stimulating programs and cashbacks) (Bedre-Defolie and Calvano 2013; Hasan et al. 2012; Rochet and Tirole 2002).

The perception of card usage by cardholder may also be altered as a result of the increased number of card users. Once a cardholder sees that more people are paying by card for the transactions, he or she will start to think that it may be safer to use payment cards (Arango-Arango et al. 2018; Darban and Amirkhiz 2015; Gresvik and Haare 2008; Humphrey et al. 1996). This is similar to the herd behavior outlined above.

Payment systems invest more funds in the processing systems to increase the transaction speed with a larger number of cardholders (Asokan et al. 2000; Massoth and Bingel 2009; Teo et al. 2015). Payment systems respond to the fact that the network becomes busier (more users—longer processing) by improving constantly the processing infrastructure in more active regions (Asokan et al. 2000; Massoth and Bingel 2009). In fact, they do not allow the processing speed to drop below the initial level as a result of the platform competition in order not to decrease the quality of services (Asokan et al. 2000; Teo et al. 2015). More active card usage in some regions fosters payments innovations (Ali et al. 2014; Milne 2006; Rysman and Schuh 2017). Payments may become more convenient as a result of these innovations (e.g., Apple Pay/Samsung Pay/Android Pay, other wallets and contactless payments, etc.) (Au and Kauffman 2008; de Kerviler et al. 2016; Mas and Radcliffe 2010; Slade et al. 2013; Souvignet et al. 2014; Wang 2008). Providers of such services (e.g., issuing banks, startups, technological firms) find it profitable to enter a particular region if the number of potential users allows them to break even (Hasan et al. 2012; Milne 2006; Rysman and Schuh 2017).

Finally, the higher share of the individuals engaged with the payments market may foster the creation of cardholders' associations aimed at protecting and improving the cardholders' welfare (Chernikova et al. 2015; Krivosheya and Korolev 2016; Rochet and Tirole 2002). Their bargaining power is usually higher than that of each particular individual, making them more effective in protecting cardholders interests (e.g., by putting pressure on tariffs or voting against the interchange fee cuts, etc.) (Carbo-Valverde and Liñares-Zegarra 2012; Malaguti and Guerrieri 2014; McGinnis 2012; Weiner and Wright 2005). The more cardholders there are in the issuing banks' portfolios, the larger the bargaining power of such associations and the more favorable the conditions in the retail payments market for the cardholders. Overall, all of above appear to suggest that the direct network effects should be associated positively with the card usage demand. The second hypothesis is, therefore:

H2: The probability of card usage increases with the larger number of cardholders and users of cashless payments.

2.3 Indirect Network Effects

Indirect network effects in the context of this paper are associated with the higher acceptance rate at the merchants' side of the market. First of all, payments product diversity increases as a result of higher merchants' acceptance rates. The stores can offer co-branded cards (Arango and Taylor 2008a; Manchanda and Saqib 2008; Worthington 1999). This type of cards usually takes the form of a merchant's bonus or loyalty card with a payment function provided by some bank. The probability that a particular individual finds a suitable payment product from the merchant increases when the number of shops that accept cashless transactions increases (Arango-Arango et al. 2018; Bounie and Francois 2006; Gresvik and Haare 2008). Besides, the co-branded card products and co-branded loyalty programs tend to be associated with a better quality of loyalty programs (Manchanda and Saqib 2008), which may translate into higher fixed and variable benefits for cardholders (Krivosheya and Korolev 2016). In addition, the overall development of the payment network resulting from higher acceptance rates leads to the emergence of more sophisticated products (payment innovations, etc.) offered by banks (Ali et al. 2014; Hasan et al. 2012; Milne 2006; Rysman and Schuh 2017). As a result, potential cardholders can find a product that is more suitable for their needs and preferences. Some banks are also likely to be both acquirers and issuers (Bolt and Chakravorti 2008; Chizhikova 2013; Krivosheya and Korolev 2016; Rochet and Tirole 2002), hence, as a result of higher acceptance rates they may redistribute funds within the departments of the bank and promote cardholding more actively (Krivosheya 2018). Such active promotion may, again, lead to a better quality of service for the same or smaller fees charged by the banks. All in all, the higher share of accepting merchants is likely to translate into higher net fixed benefits levels. In other words, indirect network externalities are likely to be positively associated with the probability of cardholding. The third hypothesis is, hence:

H3: The probability of cardholding increases with the bigger number of accepting merchants.

Finally, the card usage demand might also be impacted by higher merchant acceptance. Importantly, cardholders have a better chance of using cashless payments when more merchants accept cards. As a result, the option value to pay with a card increases for each particular individual, thus increasing his/her benefits value (Bedre-Defolie and Calvano 2013). The cashiers become more skilled and better trained when acceptance rates are higher (knowing how to operate a POS terminal becomes a job requirement for the cashiers) (Arango and Taylor 2008b; Humphrey et al. 2003; Jonker 2011). Besides, equipment gets more innovative when more merchants accept cards, increasing further the benefits associated with paying by card (Ali et al. 2014; Rysman and Schuh 2017). Some loyalty programs are conditional on the type of merchants and particular merchant brands. For instance, some banks provide higher cashback for some merchant categories or assign more bonuses for a transaction at the partner merchants' locations (Bolton et al. 2000; Carbó-Valverde and Liñares-Zegarra 2011; Ching and Hayashi 2010). The probability that a particular store where an individual uses his or her payment card is a participant in some kind of banking loyalty program is higher when more stores accept payment cards. To summarize, higher acceptance by merchants is likely to increase the probability of card usage because of the increased net variable benefits.

H4: The probability of card usage increases with the bigger number of accepting merchants.

3 Data

The principal data is collected from the proprietary sources provided by the Centre for Research in Financial Technologies and Digital Economy SKOLKOVO-NES (formerly Finance, Payments, and e-Commerce chair) of Moscow School of Management SKOLKOVO. The Centre conducted the national survey of Russian cardholders in 2013–2014. The survey is representative of the Russian economy as a whole as well as Russian regions, and includes quotas for age, gender, and regions to ensure that the valid proportion of different groups of individuals (in terms of income, age, gender, and geographical area) is sampled. The survey was organized as face-to-face interviews and included individuals who are at least 18 years old and reside in cities with a population of at least 500,000. Three stage probability sampling was performed in order to guarantee sample representativeness. The questionnaire includes sections on the individual's payment behavior and socio-demographic profile (age, education, gender, income, location, and work).

The survey also includes a separate data sample of 800 traditional (offline) merchants focused on their profile and behavior in the retail payments market. The latter is used for the calculation of indirect network effects. The final sample for the analysis includes 1500 individuals. In line with the official Russian statistics for 2013–2014, 44.4% of all respondents are female, and 55.6% are male. 26.7% of the respondents come from Moscow, 11.3% from Saint-Petersburg, and the remaining 62% are from other Russian regions. 73.5% of all the respondents hold at least one payment card, whereas 26.5% do not have any cashless payment instruments at all. 75% of all the cardholders use cards to pay for their transactions and the remaining 25% always pay by cash. In order to mitigate the selection bias problem, we include both individuals who hold and do not hold a card. The representativeness for the Russian retail payments market (major characteristics of the sample outlined above coincide with the official Russian statistics for the population) ensures that the selection bias is minimized. The sample is further reduced based on the availability of control variables.

4 Model

In order to test the hypotheses developed in the previous section we construct the following models for cardholding and card usage probabilities:

$$\begin{cases} \text{Holding}_i = \alpha + \beta * \text{DNE}_i + \gamma * \text{INE}_i + +\theta * \text{Age}_i + \tau * \text{ED}_i + \zeta * \text{SD}_i + \\ \qquad \eta * \text{Income}_i + \phi * \text{Travel}_i + \varepsilon_i; \\ \text{Usage}_i = \widetilde{\alpha} + \widetilde{\beta} * \text{DNE}_i + \widetilde{\gamma} * \text{INE}_i + +\widetilde{\theta} * \text{Age}_i + \widetilde{\tau} * \text{ED}_i + \widetilde{\zeta} * \text{SD}_i \\ \qquad + \widetilde{\eta} * \text{Income}_i + \widetilde{\phi} * \text{Payment characteristics}_i, \end{cases} \tag{1}$$

where i refers to each individual. Holding_i is a binary variable, which takes the value 1 if an individual has at least one card, and 0 otherwise. Usage_i denotes a dummy variable, which attains the value of 1 if an individual who has a card uses it to pay for goods and services, and 0 otherwise. Data on dependent variables is available from the surveys. DNE_i represents the vector of direct network externalities while INE_i represents the vector of indirect network effects. ED_i stands for the vector of the education-related characteristics. SD_i is a vector of social and demographic characteristics of the individual. Travel_i denotes the vector of characteristics related to the travel frequency. Payment characteristics_i is a vector of variables reflecting payment behavior and contract details. Finally, α, β, γ, θ are the vectors of coefficients, and ε_i refers to the error term. The first step is independent from the second one. In order to mitigate potential selection bias arising from the fact that individuals can only pay with a card when they are the cardholders (i.e., Holding variable is 1 for all the potential users of the card in the sample), the second model is not independent from the first one, being step two of the estimations. Some unreported robustness checks are performed with the assumption that the models are independent.

5 Independent Variables

5.1 Explanatory Variables

There are two key categories of the explanatory variables in the models: direct and indirect network effects. In order to measure the network effects we adapt the measures developed in Bounie et al. (2017) who used the survey data on French (2014) and European (2017) cardholders and merchants. Their measures of network externalities included the average value of purchases in a particular merchant industry and the estimates of the probability that the purchase will be paid for by card given a particular merchant type and transaction value. These measures, however, do not separate the direct and indirect network effects. Besides, they depend on a number of assumptions and calculations performed by authors on the proprietary central bank data (Bounie et al. 2017). The separation of effects was not possible because their surveys of merchants and cardholders were conducted in different years. The sample used in this paper allows the potential problems of not separating network effects and possibly unrealistic assumptions necessary for the calculations to be mitigated. As the individuals and merchants surveys were conducted within the same timeframe and geographic regions, the adapted measures of the previous studies are applied to the actual average individuals and merchants payment activity in the region. Geographic regions include eight federal districts and thirty three regions. Direct network effect is measured in four possible variations. It could be either the regional or federal district average holding of cards. Both are calculated as the average share of cardholders (number of cardholders relative to the number of individuals in the region) in the survey sample in a particular region. The latter measure is preferable because the sample was constructed in such a way as to represent federal districts. Data on regions may, sometimes, be over—or underestimated due to the absence of quotas at regional level. However, regional variables are used for robustness checks. At the same time, direct network externalities can be measured as either the regional average usage of cards or federal district average usage of cards. Unlike holding of cards, the usage of cards is observed by other cardholders, which may better reflect some of the theoretical mechanisms outlined in the previous sections (e.g., regarding the psychological factors).

Although, in theory, the impact of the direct network effects may be subject to reverse causality issue because it is calculated as an average occurrence of dependent variable in the sample, this is not the case in the data. Individual decisions to hold or use a card are unlikely to affect aggregate outcomes because of the size of the industry. In each of the 8 federal districts there are at least 70 individuals with most of the districts containing more than 100 individuals (except eastern and southern federal districts). The central federal district contains more than 400 individuals. Therefore, individuals cannot affect aggregate outcomes. There are at least 30 people sampled in each of the regions, with some regions having more than 100 individuals. Similarly, the aggregate outcomes are unlikely to be affected by the individual decision at either regional or federal district levels.

Indirect network effect measures are based on the sample of 800 traditional (offline) merchants surveyed in the same time period. The nation-wide survey included quotas for merchant types and federal districts to ensure sample representativeness for the Russian merchants' market. There are two possible ways of measuring these network externalities. The first one is regional average card acceptance rate by merchants. The second one is federal district average card acceptance rate by merchants. Again, the latter is preferred as the data was sampled to be representative at federal district level, while the former is used for the robustness checks.

5.2 Control Variables

In order to isolate the effect of network effects from the potential effects of other variables that have been found to influence the payment behavior of the individuals we have introduced a number of control variables. The key control variables identified in the previous studies include socio-demographic characteristics of an individual, education, and income levels, travel frequency, and the details of a contract with an issuer (Arango-Arango et al. 2018; Bagnall et al. 2014; Bounie and Francois 2006; Bounie et al. 2016; Gresvik and Haare 2008; Humphrey et al. 1996). The set of controls chosen for the models follows Krivosheya and Korolev (2016) who used the same data sample in order to estimate the effect of an individual's benefits level on his/her payment frequency. In a number of unreported robustness checks we have also added the regional level characteristics. Although the main outcome for the effect of the network externalities does not change, these are not included in the main analysis because of the significant sample reductions due to limited availability of regional level data on such relevant characteristics as the share of shadow economy and the intensity of tax evasion practices. The first set of control variables indicates the respondent's age. This data is available directly from the survey. We follow Krivosheya and Korolev (2016) who used age group dummies instead of a direct age variable. Previous studies have found that people of older age are less active in the retail payments market, however, the relationship is non-linear because young people are often less well paid and, hence, do not always have enough funds to maintain card balances by themselves (Arango-Arango et al. 2018; Gresvik and Haare 2008). Age is measured as dummy variables: 18–24 years old, 25–34 years old, 35–44 years old, 55–64 years old, 65+ years old. 45–54-year-olds are chosen as a reference category. Another factor affecting the probability of holding and using payment cards is education. Education might reflect the level of financial literacy of a respondent, which links to the level of an individual's awareness of retail payments (Bagnall et al. 2014; Bounie and Francois 2006; Bounie et al. 2016; Gresvik and Haare 2008). Education is evaluated by basic professional, middle professional, and higher professional dummies. School is set as a reference category. Social and demographic measures include married status dummy, children dummy, advanced PC user dummy. Marital status and the number of children can affect the probability of holding and using payment cards because of improved family financial management provided by the basic banking services associated with the payment card account (e.g., SMS notifying balances) (Bagnall et al. 2014; Bounie et al. 2016; Humphrey et al. 1996). Besides, partners and children may have several payment instruments linked to one account balance, improving the transfer and uses of income across family members (Bagnall et al. 2014; Bounie et al. 2016; Krivosheya and Korolev 2016). The level of technology proficiency relates to the person's ability to conduct cashless payments using some basic software and hardware (e.g., digital payments, POS terminals) (Bounie and Francois 2006; Krivosheya and Korolev 2016). Technology adoption is proxied by the self-assessment of the level of computer proficiency provided by the individual during the survey. The level of income reflects the ability of an individual to cover fees and expenses associated with payment card issuance and usage (Bagnall et al. 2014; Bounie and Francois 2006; Krivosheya and Korolev 2016). The data on income level of an individual was collected during a survey using the standard sociological FOM (public opinion fund) guidelines regarding the income-related questions. Income level is determined by low income and high income dummies with middle income as a reference category. The cost of cash increases outside of the domestic region because of the foreign exchange risks and additional search and transaction costs related to the currency exchange (Arango-Arango et al. 2018; Bagnall et al. 2014; Gresvik and Haare 2008). This argument is especially pertinent for foreign travel. Besides, individuals tend to get cashless payment instruments more often for uses outside their home region because of the access to larger sums of money than they brought with them in cash (Bounie and Francois 2006; Gresvik and Haare 2008; Krivosheya and Korolev 2016; Wang 2008). In fact, the use in travel is one of the top reasons for issuing a card in the sample. Travel frequency is controlled in the holding model and is excluded from the usage model in order to allow for the differences necessary for the model estimation. Some robustness checks are performed including travel frequency and excluding other control variables groups. The results stay the same. Travel frequency is evaluated using three distinct dummies: frequent travel within Russia dummy, frequent travel within the neighboring foreign countries dummy, and frequent travel around the world dummy. The reference category is no traveling. Finally, the characteristics relating to the payment behavior and the contract with an issuer are controlled for in the usage (second stage) model. This vector of controls consists of three dummies: participation in the loyalty program, credit card, and the absence of fees for a payment card dummy. Loyalty programs provide additional motivation for using payment cards in order to be reimbursed in bonuses or cashback (Carbó-Valverde and Liñares-Zegarra 2011; Ching and Hayashi 2010; Krivosheya and Korolev 2016). Withdrawals on credit cards are charged additional fees, making it more expensive for a credit card holder to use cash (Krivosheya and Korolev 2016; Rochet and Wright 2010; Wang 2008).

6 Statistic and Econometric Methods

The first (holding) model can be used independently of the usage model. Such dependence of the usage model is estimated using the two-stage Heckman selection model. Following Schuh and Stavins (2010) and Krivosheya and Korolev (2016), we use the probit model to estimate cardholding probability and then the two-stage Heckman selection model to estimate the card usage probability. The results of probit model estimation for cardholding probability are also used as a selection equation in card usage probability modeling.

Krivosheya and Korolev (2016), who use the same dataset as this research, suggest that the Heckman two-stage model outperforms the alternatives when used to estimate a Russian individual's payment behavior. There are also some drawbacks that need to be accounted for during the second stage of the modeling. These can include the potential multicollinearity of the explanatory variable in the second stage leading to inconsistent estimates. To solve this problem, we need to add at least one extra predictor in the first step. In the usage model, we exclude the travel related control variables and add the payment characteristics vector instead. We use robust standard errors in all of the models to account for potential heteroscedasticity as well as other error related issues. Marginal effects allow the economic significance of the effects to be examined. This study also uses the two-step Heckman probit specification in the number of unreported robustness checks. The results are in line with the main analysis. Table 1 provides a descriptive statistic of the variables used in the main analysis. Cross-correlations are available at request. Most of the correlation coefficients point to the absence of multicollinearity as the correlations are less than 50%, except for the relationship between federal and regional variables. These variables are not used in most of the regression specifications simultaneously. Some specifications towards the end of the paper include these variables simultaneously using the aggregated factors obtained from the results of the principal component analysis (PCA) to mitigate the multicollinearity problem. These factors are provided in Table 1 as well.

Table 1 Descriptive statistics

		Mean	S.D.	Min	Max
(1)	18–24 y.o.	0.16	0.37	0.00	1.00
(2)	25–34 y.o.	0.21	0.41	0.00	1.00
(3)	35–44 y.o.	0.17	0.37	0.00	1.00
(4)	55–64 y.o.	0.16	0.37	0.00	1.00
(5)	65+ y.o.	0.13	0.34	0.00	1.00
(6)	Basic professional	0.09	0.29	0.00	1.00
(7)	Middle professional	0.32	0.46	0.00	1.00
(8)	Higher professional	0.34	0.48	0.00	1.00
(9)	Married	0.60	0.49	0.00	1.00
(10)	Have children	0.38	0.49	0.00	1.00
(11)	Advanced PC user	0.71	0.45	0.00	1.00
(12)	Low income	0.11	0.31	0.00	1.00
(13)	High income	0.31	0.46	0.00	1.00
(14)	Frequent travel within Russia	0.05	0.22	0.00	1.00
(15)	Frequent travel within neighboring foreign countries	0.02	0.13	0.00	1.00
(16)	Frequent travel around the world	0.03	0.16	0.00	1.00
(17)	Participates in a loyalty program	0.16	0.36	0.00	1.00
(18)	Credit card	0.04	0.20	0.00	1.00
(19)	No fees for card	0.30	0.46	0.00	1.00
(20)	Regional Component: Usage	0.00	1.20	−3.09	3.29
(21)	Federal District Component: Usage	0.00	1.26	−3.44	3.29
(22)	Federal District Average Holding of Cards	0.73	0.06	0.49	0.79
(23)	Federal District Average Usage of Cards	0.55	0.06	0.46	0.72
(24)	Regional Average Holding of Cards	0.73	0.11	0.42	1.00
(25)	Regional Average Usage of Cards	0.55	0.13	0.05	0.88
(26)	Federal District Average Acceptance Rate	0.50	0.06	0.39	0.63
(27)	Regional Average Acceptance Rate	0.52	0.09	0.20	0.70

7 Empirical Testing of the Hypotheses About Network Externalities

7.1 Cardholding Probability

We begin by analyzing the determinants of the cardholding probability using the probit estimation method for the first model developed in the previous section. Table 2 presents the results. These results address hypotheses H1 and H3 regarding the effect of network externalities on cardholding probability.

To begin with, the initial specification (1) is the baseline model with the factors outlined in previous studies (e.g., Krivosheya and Korolev 2018). According to Table 2 significant variables and their signs are the same as expected and correspond to the previous studies (Krivosheya and Korolev 2016). Predictive power of the

Table 2 Correlations of network effects and cardholding probability

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	Baseline Model	Direct NE: Regional Holding	Direct NE: Federal District Holding	Direct NE: Regional Usage	Direct NE: Federal District Usage	Indirect NE: Regional Acceptance	Direct NE: Federal District Acceptance
Network effects							
Regional Average Holding of Cards		3.227 *** (0.359)					
Federal District Average Holding of Cards			2.687 *** (0.612)				
Regional Average Usage of Cards				1.777 *** (0.311)			
Federal District Average Usage of Cards					1.555** (0.692)		
Regional Average Acceptance Rate						1.103** (0.524)	
Federal District Average Acceptance Rate							2.000** (0.782)
Age							
18–24 y.o.	−0.0813 (0.142)	−0.0998 (0.147)	−0.0706 (0.143)	−0.0816 (0.146)	−0.0767 (0.143)	−0.214 (0.173)	−0.218 (0.173)
25–34 y.o.	0.137 (0.135)	0.111 (0.140)	0.151 (0.135)	0.125 (0.137)	0.140 (0.135)	0.155 (0.161)	0.153 (0.161)
35–44 y.o.	0.110 (0.139)	0.0810 (0.143)	0.105 (0.140)	0.0924 (0.141)	0.120 (0.140)	0.121 (0.164)	0.131 (0.165)
55–64 y.o.	0.0429 (0.130)	−0.0369 (0.132)	0.0232 (0.130)	0.0264 (0.131)	0.0384 (0.130)	0.149 (0.160)	0.143 (0.161)
65+ y.o.	−0.646 *** (0.135)	−0.655 *** (0.138)	−0.653 *** (0.136)	−0.634 *** (0.137)	−0.632 *** (0.135)	−0.501 *** (0.172)	−0.490 *** (0.172)

Education							
Basic professional	0.293** (0.135)	0.273** (0.138)	0.289** (0.135)	0.306** (0.136)	0.308** (0.135)	0.376** (0.163)	0.350** (0.162)
Middle professional	0.452 *** (0.0994)	0.391 *** (0.103)	0.436 *** (0.100)	0.435 *** (0.101)	0.442 *** (0.0999)	0.485*** (0.126)	0.472 *** (0.127)
Higher professional	0.465 *** (0.102)	0.481 *** (0.106)	0.472 *** (0.103)	0.494 *** (0.104)	0.473 *** (0.102)	0.503*** (0.126)	0.503 *** (0.126)
Social & demographic characteristics							
Married	0.0706 (0.0819)	0.0805 (0.0840)	0.0892 (0.0827)	0.0655 (0.0834)	0.0676 (0.0825)	−0.0715 (0.101)	−0.0755 (0.101)
Have children	0.0712 (0.0885)	0.0894 (0.0910)	0.0489 (0.0896)	0.112 (0.0898)	0.0594 (0.0889)	−0.0353 (0.103)	−0.0574 (0.104)
Advanced PC user	0.589 *** (0.101)	0.547 *** (0.103)	0.560 *** (0.102)	0.561 *** (0.103)	0.580 *** (0.101)	0.741*** (0.128)	0.741 *** (0.128)
Income level							
Low income	−0.191 (0.118)	−0.204* (0.119)	−0.217* (0.119)	−0.218* (0.119)	−0.220* (0.118)	−0.295* (0.157)	−0.291* (0.156)
High income	0.0589 (0.0890)	0.0904 (0.0920)	0.0789 (0.0893)	0.0807 (0.0899)	0.0611 (0.0889)	0.0863 (0.104)	0.101 (0.105)
Travel frequency							
Frequent travel within Russia	0.222 (0.208)	0.253 (0.216)	0.214 (0.209)	0.240 (0.208)	0.234 (0.206)	0.142 (0.220)	0.160 (0.220)
Frequent travel within neighboring foreign countries	−0.159 (0.341)	−0.119 (0.337)	−0.142 (0.339)	−0.115 (0.338)	−0.158 (0.341)	0.243 (0.404)	0.218 (0.403)
Frequent travel around the world	−0.350 (0.271)	−0.408 (0.263)	−0.359 (0.267)	−0.336 (0.266)	−0.333 (0.268)	−0.661** (0.304)	−0.617** (0.304)

(continued)

Table 2 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Constant	−0.0670 (0.131)	−2.345 *** (0.290)	−2.014 *** (0.460)	−1.025 *** (0.219)	−0.910 ** (0.398)	−0.736** (0.318)	−1.143 *** (0.416)
Observations	1500	1500	1500	1500	1500	1019	1019
Pseudo R2	0.145	0.191	0.157	0.165	0.149	0.149	0.150

Robust standard errors in parentheses. *** denotes significance at 1% significance level. ** -- at 5%. * -- at 10% significance level models is also similar to the previous studies in this area (e.g., Arango-Arango et al. 2018; Krivosheya and Korolev 2018).

Models (2)–(5) add different measures of the direct network effect to the baseline model. In model (2) the direct network effect is measured as the regional average holding of cards. The positive effect is significant at 1% significance level. In model (3) we change the direct network effects measure for the federal district average holding of card. The result stays similar to model (2). Model (4) introduces regional average usage of card, which is observable to the cardholders in the region and, hence, may introduce distinct mechanisms outlined in the theoretical framework section. The result is, again, significant at the 1% significance level and the effect is positive.

Finally, model (5) uses federal district average usage of card. As in all of these cases the direct network externality effect is positive and significant. Other controls also exhibit the same significance and direction of the effects as in the baseline model. These results support hypothesis H1 stated in the theoretical framework meaning that the positive influences for the cardholding probability are indeed present in Russian retail payments market. According to the marginal returns, a standard deviation increase in the federal district usage increases cardholding probability by 2.9 percentage points.

Indirect network effects are analyzed in models (6) & (7). Again, the indirect network effect is measured as either the regional average acceptance rate or the federal district average acceptance rate by merchants. Similarly to the direct effect, the indirect network effect is always positively significant and increases the demand for cardholding. Model (6) shows that at 5% significance level the regional average acceptance rate increases the probability of cardholding. Similarly, Model (7) introduces the main measure of indirect network effects at the federal district level and concludes the same: at the 5% significance level there is a positive association between cardholding probability and indirect network effects. So, the hypothesis H3 that probability of cardholding increases with the bigger share of accepting merchants is also not rejected.

From the economics point of view, one standard deviation increase in the federal district average acceptance rate increases the probability of cardholding by 3.79 percentage points, holding all other parameters fixed. Similarly, having basic professional education increases the probability of cardholding by 8.33 percentage points by comparison with the "school" level of education. The result is economically significant.

A number of further robustness checks add network effects simultaneously. In this case indirect network externality becomes insignificant. Potential explanation may be linked to the multicollinearity problem between direct and indirect network effects at the same level of aggregation (correlation coefficients between acceptance and holding (usage) are 0.69 (0.82) at the federal district level). In order to mitigate the multicollinearity problem and get valid results we use principal component analysis (PCA) based on (federal) regional usage and (federal) regional acceptance levels to construct an aggregate factor. Both federal and regional components are significant at 5% significance level.

Simultaneously network effects account for a smaller share of probability than the simple sum of two separate contributions. This happens because some of the underlying mechanisms coincide for both externality types. One standard deviation increase in an aggregate factor at the federal district level results in a 3.13 percentage point increase in the cardholding probability. The result is significant economically as well as statistically.

7.2 Card Usage Probability

To test the remaining two hypotheses regarding the card usage probability we present the results of the analysis using the two-step Heckman model. Results are outlined in Table 3. Selection equations presented in models (2) and (7) are equivalent to the results of the baseline model estimation in the previous subsection and represent the first step of the Heckman two-step procedure. Mills ratio is presented on the line lambda.

Model (1) provides the results of the baseline model estimation without network effects. Most of the controls remain as in probit models but we also include payment behavior details instead of travel frequency. The significance and signs of the controls are the same as in previous studies (Arango-Arango et al. 2018; Krivosheya and Korolev 2018).

As before, we begin by adding direct network effects into the baseline model. Direct network effects are evaluated by the same average holding and usage levels as before at both regional and federal district levels. Models (3) and (4) suggest that the average holding levels are not significant for the card usage probability. As outlined in the theoretical framework, some of the mechanisms behind the influence at the average cardholding levels are not strong enough for the variable net benefits as the cardholding decisions are rarely evident to the individuals and more often affect only the behavior of issuing banks.

In model (7) and for further robustness checks we add the average regional and federal district usage levels instead of holding levels. The impact of direct network effects becomes positive and significant at any reasonable significance level. From the economic point of view, a standard deviation increase in the average federal district usage of cards results in a 3.34 percentage point increase in the card usage probability by each particular merchant. By comparison, being a high income instead of a middle income individual increases card usage probability by 2.9 percentage points.

Therefore, hypothesis H2 is not rejected and the direct network effects increase the probability of the card usage even when controlled for other individual characteristics and potential selection bias.

In order to test hypothesis H4 we add the indirect network effects to the models. Some reduction in the number of observations happens due to the availability of data on merchants' acceptance. In contrast to the direct network externalities results, the indirect network effects are always positive and significant for the card usage

Table 3 Correlations of network effects and card usage probability

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Variables	Baseline Model	Baseline Model	Direct NE: Regional Holding	Direct NE: Regional Holding	Direct NE: Federal District Holding	Direct NE: Federal District Holding	Direct NE: Regional Usage
Network effects							
Regional Average Holding of Cards			−0.103 (0.0969)				
Federal District Average Holding of Cards					−0.176 (0.206)		
Regional Average Usage of Cards							0.522*** (0.0838)
Age							
18–24 y.o.	0.0779* (0.0423)	−0.0813 (0.144)	0.0774* (0.0418)	−0.0813 (0.144)	0.0768* (0.0420)	−0.0813 (0.144)	0.0785* (0.0412)
25–34 y.o.	0.0723* (0.0392)	0.137 (0.134)	0.0754* (0.0390)	0.137 (0.134)	0.0731* (0.0390)	0.137 (0.134)	0.0686* (0.0383)
35–44 y.o.	0.0199 (0.0397)	0.110 (0.136)	0.0215 (0.0393)	0.110 (0.136)	0.0211 (0.0394)	0.110 (0.136)	0.0186 (0.0387)
55–64 y.o.	−0.0724* (0.0393)	0.0429 (0.129)	−0.0683* (0.0391)	0.0429 (0.129)	−0.0706* (0.0391)	0.0429 (0.129)	−0.0804 ** (0.0384)
65+ y.o.	−0.154 (0.124)	−0.646 *** (0.138)	−0.164 (0.124)	−0.646 *** (0.138)	−0.162 (0.124)	−0.646 *** (0.138)	−0.161 (0.122)
Education							
Basic professional	−0.0873 (0.0643)	0.293** (0.138)	−0.0850 (0.0639)	0.293** (0.138)	−0.0840 (0.0641)	0.293** (0.138)	−0.0682 (0.0630)
Middle professional	−0.0188 (0.0717)	0.452 *** (0.0989)	−0.0123 (0.0717)	0.452 *** (0.0989)	−0.0149 (0.0717)	0.452 *** (0.0989)	−0.0124 (0.0703)

(continued)

Table 3 (continued)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Higher professional	0.0126 (0.0721)	0.465 *** (0.102)	0.0164 (0.0719)	0.465 *** (0.102)	0.0159 (0.0720)	0.465 *** (0.102)	0.0315 (0.0707)
Social & demographic characteristics							
Married	0.000124 (0.0262)	0.0706 (0.0828)	0.000491 (0.0259)	0.0706 (0.0828)	0.000278 (0.0260)	0.0706 (0.0828)	−0.00534 (0.0256)
Have children	0.00970 (0.0260)	0.0712 (0.0857)	0.00951 (0.0257)	0.0712 (0.0857)	0.0104 (0.0258)	0.0712 (0.0857)	0.0170 (0.0254)
Advanced PC user	0.0700 (0.0937)	0.589 *** (0.102)	0.0760 (0.0935)	0.589 *** (0.102)	0.0752 (0.0937)	0.589 *** (0.102)	0.0798 (0.0919)
Income level							
Low income	−0.00695 (0.0514)	−0.191 (0.119)	−0.00926 (0.0511)	−0.191 (0.119)	−0.00808 (0.0512)	−0.191 (0.119)	−0.0153 (0.0503)
High income	0.0647** (0.0252)	0.0589 (0.0893)	0.0655 *** (0.0250)	0.0589 (0.0893)	0.0657 *** (0.0251)	0.0589 (0.0893)	0.0654 *** (0.0246)
Payment behavior details							
Participates in the loyalty program	0.192*** (0.0267)		0.190*** (0.0267)		0.189*** (0.0269)		0.179*** (0.0263)
Credit card	0.147*** (0.0461)		0.147*** (0.0460)		0.145*** (0.0461)		0.138*** (0.0453)
No fees for card	0.352*** (0.0215)		0.355*** (0.0216)		0.353*** (0.0215)		0.325*** (0.0215)
Travel frequency							
Frequent travel within Russia		0.222 (0.205)		0.222 (0.205)		0.222 (0.205)	
Frequent travel within neighboring foreign countries		−0.159 (0.355)		−0.159 (0.355)		−0.159 (0.355)	

Frequent travel around the world		−0.350 (0.268)		−0.350 (0.268)		−0.350 (0.268)	
Constant	0.560** (0.243)	−0.0670 (0.134)	0.618** (0.248)	−0.0670 (0.134)	0.676** (0.278)	−0.0670 (0.134)	0.257 (0.243)
Lambda		−0.224 (0.301)				−0.209 (0.301)	
Observations	1500	1500	1500	1500	1500	1500	1500
P-value of comparison test	0	0	0	0	0	0	0

Robust standard errors in parentheses. *** denotes significance at 1% significance level. ** -- at 5%. * -- at 10% significance level probability. Hypothesis H4 that the probability of card usage indeed increases with the higher share of accepting merchants is also not rejected. This result persists when we use regional average card acceptance levels instead of the federal district level.

Finally, we repeat the final step of the probit analysis and add both direct and indirect network effects into the baseline model. As in probit, only PCA analysis provides us with two valid specifications showing that the combined network effects are positively associated with the card usage probability and significant at 1% significance level. Once the network externalities are included separately, the impact of indirect network effects disappears. This is, again, explained by the high correlation between the explanatory variables and, therefore, supports the robustness of the presented results regarding hypotheses H2 and H4.

From the economics point of view, a standard deviation in the federal district component increases cashless payment usage probability by 3.96 percentage points.

8 Conclusion

This paper empirically evaluates the effect of direct and indirect network externalities for cardholding and card usage probabilities in Russia. A representative sample of 1500 individuals from across Russian regions was used. This paper finds significant and robust evidence in favor of positive association between the degree of both types of network externalities and the individuals' activity in the Russian retail payment market. Besides, the results are significant from the economics point of view. This paper aims to contribute to the body of research on the determinants of cashless payments instruments' holding and usage (Arango-Arango et al. 2018; Bagnall et al. 2014; Bounie and Francois 2006; Bounie et al. 2016; Carbó-Valverde and Liñares-Zegarra 2011; Gresvik and Haare 2008).

Few of the studies analyze the presence of network externalities for the customers empirically, and those that do fail to distinguish between direct and indirect network externalities. Besides, none of the papers outlines the network externalities on the high growth retail payments market. Also, none of the studies describes an empirical investigation of the effects of network externalities on the cardholding probability and looks only at the effect on usage. This paper fills these gaps by analyzing empirically the effect of network externalities in the Russian retail payments market in the context of cardholding and card usage probabilities of an individual. The results of the paper are important both from theoretical and practical points of view. Financial entities implement different incentives aimed at stimulating cardholding and usage behavior. However, the degree of the potential impact depends on the magnitude of the network effects which cannot be explicitly changed by public or private sector intervention. Accounting for this, the real degree of influence could be measured and forecasted by Central Bank of Russia, commercial banks, and payment systems. Although the extensive analysis is focused on the Russian retail payments market, the results can be extrapolated to other high growth emerging retail payments markets, such as Turkey, China, India, Latin America, and so on.

The Russian retail payments market is slightly different from other markets in the region because most of the services as well as financial innovations are supplied by the traditional financial services (Jdanova and Karminsky 2013; Semerikova 2019; Krivosheya et al. 2017; Krivosheya 2020). However, this should not affect the importance of the network effects. There are some limitations in this research that may suggest directions for further research. First of all, whilst we analyzed network effects in Russia only, it could vary from country to country. In developing countries there could be no network effects due to the early stage of their market development. Other countries could be analyzed both separately and together to investigate the effect of cross-border payments and the presence of network externalities among groups with smaller degree of communication. Secondly, the data was collected from the cities with a population of at least 500,000 but there are also smaller cities, where the degree of network externalities may be smaller. Although this restriction does not undermine the representativeness of the data, it is worth considering them either separately or as a part of a similar national study. Thirdly, the latest available data is for 2013–2014. Although the direction and presence of network externalities should not differ much, the association between network effects and demand for card holding and acceptance may intensify with the evolution of payment technologies and innovation. Future studies could test this hypothesis empirically.

References

Ali, R., Barrdear, J., Clews, R., & Southgate, J. (2014). *Innovations in payment technologies and the emergence of digital currencies*. Rochester, NY: Social Science Research Network.

Arango, C., & Taylor, V. (2008a). *Merchant acceptance, costs, and perceptions of retail payments: A Canadian survey* (Discussion Paper No. 08–12). Bank of Canada.

Arango, C., & Taylor, V. (2008b). Merchants' costs of accepting means of payment: Is cash the least costly? *Bank of Canada Review, 2008–2009*(Winter), 17–25.

Arango-Arango, C. A., Bouhdaoui, Y., Bounie, D., Eschelbach, M., & Hernandez, L. (2018). Cash remains top-of-wallet! International evidence from payment diaries. *Economic Modelling, 69*, 38–48.

Asokan, N., Janson, P., Steiner, M., & Waidner, M. (2000). State of the art in electronic payment systems. In B. M. V. Zelkowits (Ed.), *Advances in computers* (Vol. 53, pp. 425–449). Elsevier.

Au, Y. A., & Kauffman, R. J. (2008). The economics of mobile payments: Understanding stakeholder issues for an emerging financial technology application. *Electronic Commerce Research and Applications, 7*(2), 141–164.

Bagnall, J., Bounie, D., Huynh, K. P., Kosse, A., Schmidt, T., Schuh, S. D., & Stix, H. (2014). *Consumer cash usage: A cross-country comparison with payment diary survey data*. Rochester, NY: Social Science Research Network.

Baxter, W. F. (1983). Bank interchange of transactional paper: Legal and economic perspectives. *The Journal of Law & Economics, 26*(3), 541–588.

Bedre-Defolie, O., & Calvano, E. (2013). Pricing payment cards. *American Economic Journal: Microeconomics, 5*(3), 206–231.

Bolt, W., & Chakravorti, S. (2008). *Economics of payment cards: A status report* (pp. 15–27). Q IV: Economic Perspectives.

Bolton, R. N., Kannan, P. K., & Bramlett, M. D. (2000). Implications of loyalty program membership and service experiences for customer retention and value. *Journal of the Academy of Marketing Science, 28*(1), 95–108.

Borzekowski, R., Elizabeth, K. K., & Shaista, A. (2008). Consumers' use of debit cards: Patterns, preferences, and price response. *Journal of Money, Credit and Banking, 40*(1), 149–172.

Bounie, D., & Francois, A. (2006). *Cash, check or bank card? The effects of transaction characteristics on the use of payment instruments*. Rochester, NY: Social Science Research Network.

Bounie, D., François, A., & Hove, L. V. (2016). Consumer payment preferences, network externalities, and merchant card acceptance: An empirical investigation. *Review of Industrial Organization*, 1–34.

Bounie, D., François, A., & Van Hove, L. (2017). Consumer payment preferences, network externalities, and merchant card acceptance: An empirical investigation. *Review of Industrial Organization, 51*(3), 257–290.

Carbó-Valverde, S., & Liñares-Zegarra, J. M. (2011). How effective are rewards programs in promoting payment card usage? Empirical evidence. *Journal of Banking & Finance, 35*(12), 3275–3291.

Carbo-Valverde, S., & Liñares-Zegarra, J. M. (2012). *Payment card interchange fees: Assessing the effectiveness of antitrust investigations and regulation in Europe*. Rochester, NY: Social Science Research Network.

Chernikova, L. I., Faizova, G. R., Egorova, E. N., & Kozhevnikova, N. V. (2015). Functioning and development of retail banking in Russia. *Mediterranean Journal of Social Sciences, 6*(6 S4), 274.

Ching, A. T., & Hayashi, F. (2010). Payment card rewards programs and consumer payment choice. *Journal of Banking & Finance, 34*(8), 1773–1787.

Chizhikova, E. S. (2013). The current payment system of the Russian Federation. *Middle-East Journal of Scientific Research, 14*(2), 244–247.

Darban, M., & Amirkhiz, H. (2015). *Herd Behavior in Technology Adoption: The Role of Adopter and Adopted Characteristics*. 2015 48th Hawaii International Conference on System Sciences (cc. 3591–3600).

de Kerviler, G., Demoulin, N. T. M., & Zidda, P. (2016). Adoption of in-store mobile payment: Are perceived risk and convenience the only drivers? *Journal of Retailing and Consumer Services, 31*, 334–344.

Grauwe, D., Paul, & Rinaldi, L. (2002). *A model of the card payment system and the interchange fee*. Rochester, NY: Social Science Research Network.

Gresvik, O., & Haare, H. (2008). *Payment habits at point of sale. Different methods of calculating use of cards and cash in Norway*. Norges Bank Working Paper.

Guthrie, G., & Wright, J. (2007). Competing payment schemes. *The Journal of Industrial Economics, 55*(1), 37–67.

Hasan, I., Schmiedel, H., & Song, L. (2012). Returns to retail banking and payments. *Journal of Financial Services Research, 41*(3), 163–195.

Hayashi, F. (2008). Do U.S. Consumers Really Benefit from Payment Card Rewards?

Hayashi, F. (2009). Do US consumers really benefit from payment card rewards. *Economic Review, 94*(1), 37–63.

Humphrey, D., Willesson, M., Lindblom, T., & Bergendahl, G. (2003). What does it cost to make a payment? *Review of Network Economics, 2*(2).

Humphrey, D. B. (2010). Retail payments: New contributions, empirical results, and unanswered questions. *Journal of Banking & Finance, 34*(8), 1729–1737.

Humphrey, D. B., Pulley, L. B., & Vesala, J. M. (1996). Cash, paper, and electronic payments: A cross-country analysis. *Journal of Money, Credit and Banking, 28*(4), 914.

Hunt, R. M. (2003). An introduction to the economics of payment card networks. *Review of Network Economics, 2*(2).

Jdanova, O. R., & Karminsky, A. (2013). The diffusion of banking innovations: Bank cards on Russian market. *Innovative Marketing, 9*(3).

Jonker, N. (2011). *Card acceptance and surcharging: The role of costs and competition*. De Nederlandsche Bank Working Paper No. 300.

Kadhiwal, S., & Zulfiquar, A. U. S. (2007). Analysis of mobile payment security measures and different standards. *Computer Fraud & Security, 2007*(6), 12–16.

Katz, M. L., & Shapiro, C. (1985). Network externalities, competition, and compatibility. *The American Economic Review, 75*(3), 424–440.

Kim, C., Tao, W., Shin, N., & Kim, K.-S. (2010). An empirical study of customers' perceptions of security and trust in e-payment systems. *Electronic Commerce Research and Applications, 9*(1), 84–95.

Krivosheya, E. (2018). Evaluating Efficient Multilateral Interchange Fees: Evidence from End-User Benefits. Higher School of Economics Research Paper No. WP BRP, 66.

Krivosheya, E. (2020). Determinants of benefits in the retail payments market: Evidence from Russian consumers. *Journal of Promotion Management*, 1–20.

Krivosheya, E., & Korolev, A. (2016). Benefits of the retail payments card market: Russian cardholders' evidence. *Journal of Business Research, 69*(11), 5034–5039.

Krivosheya, E., & Korolev, A. (2018). Benefits of the retail payments card market: Evidence from Russian merchants. *Journal of Business Research, 88*, 466–473.

Krivosheya, E., Korolev, A., & Plaksenkov, E. (2015). *Measures for stimulating a cashless economy in Russia*. Moscow School of Management SKOLKOVO working paper.

Krivosheya, E., Semerikova, E., Korolev, A. & Tarusova, E. (2017). *Cashless economy in Russia 2030: Scenarios for market and industry*. Moscow School of Management SKOLKOVO Working Paper.

Malaguti, M. C., & Guerrieri, A. (2014). Multilateral Interchange Fees: Competition and regulation in light of recent legislative developments. ECRI Research Report No. 14, 31 January 2014.

Manchanda, R. V., & Saqib, N. (2008). Consumers' evaluations of co-branded products: The licensing effect. *Journal of Product & Brand Management, 17*(2), 73–81.

Mas, I., & Radcliffe, D. (2010). *Mobile payments go viral: M-PESA in Kenya*. Worldbank, CGAP.

Massoth, M., & Bingel, T. (2009). *Performance of Different Mobile Payment Service Concepts Compared with a NFC-Based Solution*. In 2009 Fourth International Conference on Internet and Web Applications and Services (cc. 205–210).

McGinnis, P. C. (2012). Misguided regulation of interchange fees: The consumer impact of the Durbin amendment. *Loyola Consumer Law Review, 25*, 285.

Meadows, M., & Dibb, S. (1998). Assessing the implementation of market segmentation in retail financial services. *International Journal of Service Industry Management, 9*(3), 266–285.

Milne, A. (2006). What is in it for us? Network effects and bank payment innovation. *Journal of Banking & Finance, 30*(6), 1613–1630.

Plaksenkov, E., Korovkin, V., & Krivosheya, E. (2015). *Cashless economy in Russia: Tendencies, perspectives, opportunities*. Moscow School of Management SKOLKOVO Working Paper.

Roberts, J. A., & Jones, E. (2001). Money attitudes, credit card use, and compulsive buying among American college students. *Journal of Consumer Affairs, 35*(2), 213–240.

Rochet, J.-C., & Tirole, J. (2002). Cooperation among competitors: Some economics of payment card associations. *The Rand Journal of Economics, 33*(4), 549–570.

Rochet, J.-C., & Tirole, J. (2003). Platform competition in two-sided markets. *Journal of the European Economic Association, 1*(4), 990–1029.

Rochet, J.-C., & Tirole, J. (2006). Two-sided markets: A progress report. *The Rand Journal of Economics, 37*(3), 645–667.

Rochet, J.-C., & Wright, J. (2010). Credit card interchange fees. *Journal of Banking & Finance, 34* (8), 1788–1797.

Rysman, M., & Schuh, S. (2017). New innovations in payments. *Innovation Policy and the Economy, 17*, 27–48.

Schuh, S., & Stavins, J. (2010). Why are (some) consumers (finally) writing fewer checks? The role of payment characteristics. *Journal of Banking & Finance, 34*(8), 1745–1758.

Semerikova, E. (2019). Payment instruments choice of Russian consumers: Reasons and pain points. *Journal of Enterprising Communities: People and Places in the Global Economy, 14*(1),22–41.

Shy, O. (2011). A short survey of network economics. *Review of Industrial Organization, 38*(2), 119–149.

Slade, E. L., Williams, M. D., & Dwivedi, Y. K. (2013). Mobile payment adoption: Classification and review of the extant literature. *The Marketing Review, 13*(2), 167–190.

Souvignet, T., Hatin, J., Maqua, F., Tesniere, D., Léger, P., & Hormière, R. (2014). Payment card forensic analysis: From concepts to desktop and mobile analysis tools. *Digital Investigation, 11* (3), 143–153.

Teo, A.-C., Ooi, K.-B., Yew, K.-T., Tan, G. W.-H., & Hew, T.-S. (2015). The effects of convenience and speed in m-payment. *Industrial Management & Data Systems, 115*(2), 311–331.

Todd, S., & Lawson, R. (2003). Consumer preferences for payment methods: A segmentation analysis. *International Journal of Bank Marketing, 21*(2), 72–79.

Wang, Y. (2008). Determinants affecting consumer adoption of contactless credit card: An empirical study. *Cyberpsychology & Behavior, 11*(6), 687–689.

Weiner, S., & Wright, J. (2005). Interchange fees in various countries: Developments and determinants. *Review of Network Economics, 4*(4), 1–34.

Worthington, S. (1999). Branding and relationships issues in affinity and co-branded credit cards. *Journal of Brand Management, 6*(3), 188–197.

Wright, J. (2004). The determinants of optimal interchange fees in payment systems. *The Journal of Industrial Economics, 52*(1), 1–26.

Conclusion: Instruments of Financial Sustainability in Emerging Markets

Alexander Karminsky, Paulo Emilio Mistrulli, Mikhail Stolbov, and Yong Shi

Abstract The conclusion summarizes the findings obtained by the authors of the monograph with respect to different dimensions of risk management in emerging markets.

Keywords Banking · Basel Accords · Emerging markets · Financial regulation · Ratings · Risk management

Risk management has stormed into the financial world. Today it is difficult to work out more advanced approaches for business, which would set the problems as well as offer timely and adequate solutions to them. This volume of the series contributes to theoretical and empirical literature on risk management in emerging markets, developing practical tools of risk management in commercial banks. The team of authors, combining academic analysis and practical experience of risk modeling and assessment, presents to the reader their new elaborations and the results of their implementation.

The book covers five broad research programs. Each of them is dedicated to an important aspect of theory and/or practice of risk management and reflects solely the authors' professional view on this or that problem.

First, special features of banking system development in emerging countries and their regulation are comprehensively reviewed. A comparative analysis is conducted to uncover the trends of banking sector development in various countries as well as in the banking regulation. Possible effects from the implementation of the latest-

A. Karminsky (✉) · M. Stolbov
Moscow, Russia
e-mail: akarminsky@hse.ru

P. E. Mistrulli
Rome, Italy

Y. Shi
Omaha, NE, USA

A. M. Karminsky et al. (eds.), *Risk Assessment and Financial Regulation in Emerging Markets' Banking*, Advanced Studies in Emerging Markets Finance,
https://doi.org/10.1007/978-3-030-69748-8_19

generation Basel Accords are evaluated. The initiatives of G-20 and the Financial Stability Board as regards the improvement of financial regulation are critically examined. A controversial character of these initiatives and difficulties of their implementation at the national and global levels are revealed.

Second, the monograph features the experience of using ratings to measure bank risks. These results constitute an important line of research conducted by NRU Higher School of Economics. In particular, it is shown that in order to improve the quality of ratings evaluation, apart from making rating scales for various economic entities, it is reasonable to use clusterization based on the pattern technology and to take into consideration such macrofinancial factors as credit cycles. Besides, we conduct the comparative analysis of approaches to making rating assessment in Russia and other countries with the methodology used by the big three rating agencies. At the same time, it is shown that internal ratings are of special importance as they lay grounds for the development of risk management in commercial banks in line with the Basel Accords.

Third, we present elaborations as regards the adaptation of advanced instrumental methods and models of risk management to the realities of emerging markets, including Russia. In particular, it is shown which adjustments are necessary for an appropriate use of default probability models, recovery rate and loss given default assessments in terms of developing countries. Besides, we determine the advantages of using searches in Google for market risk prediction. Apart from that, we offer new approaches to stress-testing of the liquidity risk in the Russian banking sector.

Fourth, the monograph considers a range of macrofinancial stability issues. For example, the experts from the Bank of Russia offer the methodology, which evaluates the accuracy of banking crisis early-warning systems from the perspective of the Basel III regulatory requirements. The structural model of macrofinancial relationships in Russia calibrated with empirical data is offered to stress-test the sufficiency of international reserves of the Bank of Russia. Besides, on the basis of advanced vector autoregression models, it is shown that financial stress in the Russian economy has long run real effects, suppressing the dynamics of the production index.

Fifth, the monograph places into the spotlight the role of leading-edge financial technologies based on the digitalization in banks and the potential of mathematical models for the innovative development of risk management.

We believe it is important to proceed with the studies falling under the mentioned research programs, both from the theoretical and practical points of view. Thus, this monograph will be interesting not just only for academic specialists, university professors, students, and post-graduate students, but also for practitioners employed by banks and financial institutions of developing countries. The book will also be of interest for specialists from related sectors of economics and finance, as it vividly demonstrates the achievements and prospects of contemporary financial risk management.