National Bank of Serbia

WORKING PAPERS BULLETIN

September

2021

01

National Bank of Serbia

WORKING PAPERS BULLETIN

September
2021

01

NATIONAL BANK OF SERBIA
Belgrade, Kralja Petra 12
Tel: 011/3027-100
Tel: 011/333-8000
www.nbs.rs

Number of copies: 60
ISSN 2787-3315

Issued by:

National Bank of Serbia

Printed by:

Institute for Manufacturing Banknotes and Coins – Topčider

Governor's Foreword

As the central monetary authority in the country, the National Bank of Serbia is committed to numerous activities and with its work it influences many spheres of the life of our citizens and businesses. Some of those activities are widely recognised and frequently communicated to the public, but there are many more that have so far not been visible enough or presented to the broader public adequately. One of them is the research work done by different organisational units of the National Bank of Serbia, which helps us achieve our main objectives and functions.

Seeking to make this part of our efforts more visible, but also to make a contribution to discussion within the academic and wider community, we are launching a new publication – the NBS Working Papers Bulletin. The Bulletin will feature working papers – predominantly by our staff, analysing the causal links in the economy, key factors affecting economic indicators, and numerous other topics relevant for the functioning of a central bank and the financial system at large. One of the main ideas is to present the activities and functions of the National Bank of Serbia that have not been visible enough thus far – from the theoretical and practical aspects of monetary policy implementation, FX operations and FX reserves management, regulating and supervising financial institutions, foreign exchange matters, strengthening the stability of the financial system, domestic and cross-border payment transactions, financial services consumer protection, to all other areas within the remit of the National Bank of Serbia.

The Bulletin will also include analyses of current trends in different areas of central banking. We believe it is important to understand how economic interconnections and laws reflect on our everyday lives – with this knowledge we can recognise potential risks and challenges on time and prepare an appropriate response.

The Bulletin will also look into contemporary trends in international research. We will keep an eye on all dilemmas and issues faced by the current economic policy makers, and analyse the latest developments in practice and compare them with theory.

The results achieved in the previous years impose an obligation on us to continue in the same way in the future, but also to acquaint the professional and broader public more closely with the issues faced in our day-to-day work, as well as with the key challenges of modern central banking. I find it rather important to understand how economic developments and causal links and economic laws reflect on the everyday lives of citizens and the performance of the corporate sector so that we could prepare better for the challenges ahead. Almost two years of fighting against the pandemic best shows the ever evolving nature of the challenges faced by the entire world. It is our task to counter those challenges as readily as possible and to respond with the right measures and activities. What helps us in making this a reality is by all means practice, experience, but also scientific awareness that trends change.

Looking ahead, our measures and decisions will continue to be well-timed, appropriate, carefully weighed and mindful of local and global trends, while the working papers published in the Bulletin will show to the broader public a part of the analyses and activities on which the National Bank of Serbia works on a daily basis in an effort to build a better future for our citizens.

Dr Jorgovanka Tabaković, Governor

WORKING PAPERS:

**INTEREST RATE PASS-THROUGH IN SERBIA:
EVIDENCE FROM INDIVIDUAL BANK DATA** 5

**A DSGE MODEL WITH FINANCIAL DOLLARIZATION
– THE CASE OF SERBIA** 33

**ASSESSMENT OF THE REPUBLIC OF SERBIA'S SYSTEMIC RISK
AND THE LIKELIHOOD OF A SYSTEMIC CRISIS** 73

GLOBAL DEVELOPMENT TRENDS IN PAYMENT CARD INDUSTRY 119

Working Papers describe research in progress by the author(s) and are published to encourage discussion and suggestions for future work.

National Bank of Serbia

INTEREST RATE PASS-THROUGH IN SERBIA: EVIDENCE FROM INDIVIDUAL BANK DATA

Mirjana Miletić, Aleksandar Tomin, Anđelka Đorđević

The views expressed in the papers constituting this series are those of the author, and do not necessarily represent the official view of the National Bank of Serbia.

Economic Research and Statistics Department

THE NATIONAL BANK OF SERBIA

Belgrade, 12 Kralja Petra Street

Telephone: (+381 11) 3027 100

Belgrade, 17 Nemanjina Street,

Telephone: (+381 11) 333 8000

www.nbs.rs

Interest rate pass-through in Serbia: evidence from individual bank data

Mirjana Miletić, Aleksandar Tomin, Andjelka Djordjević

Abstract: The paper considers interest rate pass-through in Serbia, based on evidence from individual bank data. Analysis was conducted for the period from September 2010 to May 2021 using panel cointegration tests and estimates obtained by the fully modified ordinary least square method (FMOLS), dynamic ordinary least square method (DOLS), pooled mean group method (PMG) and mean group method (MG).

Estimation results suggest that there is a significant long-run relationship between bank lending rates in national currency and rates in the domestic money market. Interest rate pass-through from money market to rates on dinar loans is complete for both corporate and household loans, whereas the reaction is stronger and faster in case of corporates, as they have more alternative sources of finance than households. Estimates obtained by the FMOLS, DOLS, PMG and MG methods are quite similar, indicating the robustness of the results. The pass-through estimation is also performed for the shorter period – September 2010 to end-2014, with results suggesting the interest rate channel gained more strength over time, thanks to the increasing interbank competition, higher economic growth, and more favorable macroeconomic prospects of the economy. Statistically significant impact of the risk premium measured by EMBI on dinar corporate loans is also confirmed. Given the fact that around two thirds of loans are FX-indexed, we have estimated the influence of 3M and 6M EURIBOR to rates on euro-indexed corporate and household loans. Long-run relationship and statistically significant impact of country risk premium on euro-indexed interest rates is also confirmed, along with the high pass-through of EURIBOR.

In addition, we tested whether interest rate pass-through is affected by some individual bank characteristics such as size, strength of deposit base, liquidity, quality of credit portfolio, capital position and the share of dinar loans in total loans.

Key words: interest rate pass-through, panel, monetary transmission mechanism.
JEL Code: C32, C33, E43.

[Shorter version of the paper was presented at the conference XVIth ESCB Emerging Markets Workshop in Rome, November 2018]

Non-Technical Summary

In developed market economies the interest rate channel is the most important transmission mechanism channel. Therefore, it is important for every central bank to assess its efficiency – the speed at which bank lending and deposit rates adjust to changes of the monetary policy rate. Generally, monetary transmission process encompasses two phases. In the first phase, change in the reference rate is transmitted to the money market rates and is largely dependent on yield curve stability. In the second phase, money market rates carry forward to lending and deposit interest rates as they represent opportunity cost or the cost-of-funds. In theory, in the long run the change in the policy rate should completely transmit to a change in lending interest rates. However, different factors such as asymmetric information, imperfect substitution, level of competition among banks, macroeconomic conditions etc. can hamper its full transmission.

Many empirical studies on interest rate pass-through are available and they differ in scope, geographical dimension, estimation method, time dimension, selection of exogenous variable, etc. However, all of them tackle the same two questions – the degree and the speed of the pass-through – and the results vary across countries and bank products. This heterogeneity could be explained by different factors – degree of competition among banks, banking system ownership, monetary policy regime, development of money market and financial system, openness of the economy, legal and cultural differences etc.

Comparable and consistent time series of bank lending rates on new business loans for Serbia start from September 2010. That was the main reason why we opted for panel method in examining the pass-through of money market interest rates to various bank interest rates. Though interest rate pass-through in Serbia had been already touched, this is the first time it has been examined from individual bank level. This was an additional stimulus for us as this kind of analysis is quite scarce in the available empirical literature. So far it was conducted for some individual euro area and non-euro area countries (Germany, Italy, Belgium, Poland, and Turkey). Generally, they found almost complete long-run pass-through, but incomplete and heterogeneous size and speed of adjustment in the short run. Our findings are similar, suggesting that there is a significant long-run relationship between bank lending rates in national currency and money market rates.

We also investigated the transmission process of this segment of credit market by examining the relationship between EURIBOR and interest rates on new euro-indexed business loans. The results are similar to those for dinar loans. High long-term coefficients and almost complete pass-through are found for both corporate and household loans, with significant impact of the risk premium, measured by EMBI. Looking by the type of loans, the pass-through is stronger for corporate loans as current assets and investment loans have higher coefficients than housing loans.

The size and speed of the pass-through are also determined by individual bank characteristics (size, strength of deposit base, quality of credit portfolio, capital position, liquidity and share of dinar loans in total loans), which we examined in the second phase of our analysis, by grouping banks into two clusters. The results suggest that the long-term adjustment is complete for corporate and household lending rates, with mixed results between clusters in terms of the speed of adjustment. In the case of household sector, adjustment is faster for banks with smaller balance sheet assets, while well-capitalized banks, banks with higher non-performing loans and a higher share of deposits in total liabilities tend to adjust more slowly. As for corporates, the adjustment is faster for less capitalized, more liquid banks, banks with a higher deposit base and lower non-performing loan ratios.

Contents

1 Introduction **10**
2 Theoretical background of interest rate pass-through **11**
3 An overview of the empirical literature **13**
4 Econometric methodology **14**
5 Data description and definition of variables **17**
6 Empirical results **19**
6.1 Testing the relationship for dinar interest rates 19
6.2 Testing the relationship for euro-indexed interest rates 22
6.3 Testing impact of different individual bank characteristics on interest rate pass-through 23
7 Concluding remarks **24**
Appendix **26**
References **30**

1 Introduction

Inflation targeting is the main monetary policy strategy in large number of countries. As central banks of those countries tend to achieve the inflation target by changing the interest rate applied in their main monetary policy operations, the analysis of the monetary transmission mechanism through different channels becomes crucial for monetary authorities, while estimating the interest rate pass-through effect on the real economy has an important role.

Interest rate pass-through could be defined as the degree to which changes of central bank key policy or market rates transmit onto retail bank rates (lending and deposit bank rates). A higher interest rate pass-through indicates a more effective interest rate channel, where complete pass-through means that changes in policy rates are fully transferred to retail bankrates. Besides, the size of the interest rate pass-through is an indicator of the degree of competition among commercial banks in the credit market. Thus, the interest rate pass-through is important not only for monetary policy, but also for financial stability.

Many empirical studies based their analysis on the assumption that a long-run equilibrium among market and monetary policy rates exists. The underlying assumption of this relation is that banks set their interest rates in relation to the marginal cost of funding, which is approximated by the money market rate. In the next step, the relationship between lending interest rates and money market interest rates is been estimated. However, a lot of empirical studies conducted on different countries' data have found incomplete interest rate pass-through with the imperfection of markets, lower degree of competition among banks, presence of asymmetric information, etc. as a possible explanation of this phenomenon.

Our paper contributes to the literature of the bank lending channel and interest rate pass-through by being the first to analyse the interest rate pass-through in Serbia based on a micro dataset.

The basis of this analysis is testing the long-run relationship between monetary policy rate and different bank lending interest rates in Serbia and estimating how much of the changes in bank lending rates can be attributed to changes in the key monetary policy rate and money market interest rates. Having in mind a substantial share of euro-indexed loans in total loans, we have also examined the pass-through effect of EURIBOR to interest rates on euro-indexed loans. The dataset consists of per annum average interest rates on new business corporate and household national currency and euro-indexed loans for a sample of 19 banks operating in the Serbian market on monthly basis, covering the period from September 2010 to May 2021. Empirical testing is done using panel estimation techniques.

In order to empirically investigate whether interest rate pass-through is affected by individual banks characteristics, two bank clusters were set up. Based on the distribution of each indicator such as size, liquidity, strength of deposit base, quality of credit portfolio and capital position, for both clusters long-run and short-run pass-through rates are estimated and compared.

The rest of the paper is structured as follows. In Section 2 we discuss theoretical background and summarize different explanations for the possible sluggish and incomplete interest rate pass-through. In Section 3 we provide an overview of the literature and empirical findings relating to interest rate pass-through in Central and Eastern European countries (CEECs), as well as industrial countries. Section 4 provides a description of econometric methodology that was used in empirical analysis, followed by a description of the data sample in Section 5. Empirical findings of the pass-through effect for the whole sample of banks, as well as for the effect of bank characteristics on interest rate pricing are presented in Section 6, while Section 7 summarizes the main conclusions.

2 Theoretical background of interest rate pass-through

Crucial for testing the effectiveness of monetary policy is monitoring how changes in the key monetary policy rate are transmitted onto money market rates at the longer maturity, in the first stage, and how much bank deposit and lending rates are affected by the changes in money market rates, in the second.

The first stage of transmission from the monetary policy rate to money market rates depends on the yield curve stability. The connection between short-term and long-term (market) nominal interest rates is provided by the term structure of interest rates. The slope and dynamics of the yield curve is usually determined by three main theories: expectations (long-term interest rates are obtained as the average of current and expected short-term interest rates), liquidity preference (investors require liquidity premium for holding less liquid assets) and segmentation (interest rates for different segments can be determined individually, according to specific demand and supply factors). If the term structure remains unchanged over time, changes in the policy rate will result in the proportionate shift in the yield curve. If, for some reason, the shape of the yield curve changes, the size of the pass-through can also change.

The second stage concerns how the change in money market rates is reflected in bank lending and deposit rate changes. In line with the *cost-of-funds approach* [see de Bondt (2002, 2005)] banks set their retail rates according to their marginal costs, which are approximated by money market rates. The corresponding market rate is assumed to represent opportunity cost or the cost-of-funds against which the bank sets its retail rate with an addition of a mark-up aimed to compensate bank for the interest rate risk and credit risk. Additionally, the selection of market rates of similar maturity also reflects the degree of competition between traditional banking products (loans and deposits) and non-bank products (capital market-based products). The abovementioned links between these rates can be described through following interconnections. In funding their short-term loans, banks often rely on money market instruments, which makes bank loan rates linked to money market rates. On the other hand, as investment in bonds stands as an alternative to their lending activity, the yields on long-term government securities can be viewed as opportunity costs for banks. Similar link exists between market and deposit rates since households and non-financial sector can hold their assets not only in deposits but also in government securities. Additionally, deposits can be viewed as a source of banks' funding alternative to money market instruments.

A similar approach, known as *monetary policy approach,* comes from Sander and Kleimeier (2004a) who say that if the assumption of a stable yield curve is fulfilled (monetary policy rate affects simultaneously short- and long-term rates), the relationship between monetary policy rate and retail (lending and deposit) rates can be observed directly.

Coming back to the cost-of-funds approach, the relationship between market rates and bank lending and deposit rates can be illustrated using a marginal cost pricing model, where the interest rate set by the bank $i^{l/d}$ equals the marginal cost of funding approximated by a market interest rate i^{m} and a constant mark-up α [see de Bondt (2002)]:

$$i^{l/d} = \alpha + \beta i^{m}$$

The degree of pass-through in the previous equation is represented by the coefficient β, where elasticity lower than one results in an incomplete pass-through from money market rates on lending/deposit rates ($\beta < 1$).

Many factors may influence the strength and speed of the interest rate pass-through [see Egert and MacDonald (2006), Horváth, Krekó and Naszódi (2004)]. Complete interest rate pass-through may not prevail in the presence of **asymmetric information** (adverse selection and moral hazard). Stiglitz and Weiss (1981) explain how the existence of asymmetric information between lenders and borrowers may cause an upward stickiness of lending rates. Any increase in lending rates may result in adverse selection or moral hazard, or both. In adverse selection more risky projects are favored to safer, which in such a set of circumstances are regarded as not profitable. Moral hazard arises when borrowers choose to invest in riskier projects due to high rates of return, or when even safer projects fail to pay credit back, knowing that potential costs will be borne, partially or completely, by others (most often by the government budget). In order to avoid these situations, banks may opt to adjust lending rates disproportionally to the rise in market rate, setting them at lower levels, below the equilibrium rate. However, asymmetric information can also result in amplified pass-through ($\beta > 1$) in the case when banks charge disproportionally higher interest rates in an attempt to compensate for higher risks resulting from adverse selection and moral hazard [see de Bondt (2005)].

The **structure of financial system** and the availability of non-bank financing options may also affect the pass-through. In developed capital and money markets, companies have more options for alternative non-bank sources of finance, which makes the loan demand more sensitive to changes in interest rates. Thus, the **imperfect substitution** between bank deposits and other money market and capital market instruments may cause incomplete interest rate pass-through.

Level of banks competition may also influence interest rate pass-through. Usually, a higher degree of banks competition results in a higher interest rate pass-through [see Kot, (2004)]. This effect might differ depending on the direction of change in the key policy rate. For example, Mojon (2000) concludes that sharper competition among banks contributes to a faster and more symmetric adjustment of bank rates, while Weth (2002) showed that if competition is weak, hikes of the key policy rate result in quicker changes of lending than deposit rates, while opposite holds true in a situation of reducing the key policy rate.

Capital and **liquidity position** of the bank could influence interest rate elasticity too. Less liquid and less capitalized banks will adjust their rates faster and to a larger extent than well-capitalized and banks with a better liquidity position, since they have less ability to offset the effects of changes in market rate.

Macroeconomic conditions also have impact on the interest rate pass-through [see Egert et al. (2007) and Egert and MacDonald (2009)]. For example, interest rate pass-through is usually more rapid in the period of higher economic growth, due to the fact that more favorable economic conditions for enterprises and households enable banks to pass more easily the changes in the interest rate onto their lending and deposit rates. Higher market rate volatility is usually connected with increased uncertainty that may lessen the size and the speed of the pass-through. Rotenberg and Saloner (1987) explain price rigidity by formulating the *menu costs theory*, according to which banks will change their lending rates only when the benefits from doing so are greater than the costs of changing the rates. Hence, if the monetary policy rate change is perceived as small and temporary, and the costs associated with changing retail rates are higher than the benefits, banks may opt to delay the retail rate changes.

Quality of credit portfolio can also influence the interest rate pass-through. Banks with a higher share of NPLs would benefit from an expansive monetary policy to strengthen their liquidity and their financial health rather than increase their credit portfolio by cutting their interest rates [Saborowski and Weber, 2013]. Therefore, higher NPL ratios are expected to reduce the pass-through.

3 An overview of the empirical literature

In the past two decades numerous studies examining the interest rate pass-through have been conducted. Despite the diversity of approaches, the majority of the studies conclude that the degree and speed of pass-through differ considerably across countries, as well as across banking products, especially in the short run. The evidence of whether there is full pass-through in the long run is more scattered and so far, no clear consensus has emerged [Sorensen and Werner (2006)].

Some of the first research papers on interest rate pass-through in advanced economies [Cottarelli and Kourelis (1994), Borio and Fritz (1995)] found mostly complete pass-through in the long run and incomplete adjustment in the short run. Research, based on aggregate level data for EU countries such as in Mojon (2000), Donnay and Degruse (2001), Toolsema et al. (2001), Sander and Kleimeier (2004a), de Bondt (2005), etc. in general find incomplete and sluggish pass-through, with differences in coefficients among countries. The main findings of these analyses are: the rates on loans tend to react faster with more complete pass-through than rates on deposits and that the reaction is more complete and faster for short term loans then for those of longer maturities. The heterogeneity in pass-through is explained by a different degree of competition among banks, banking system ownership, monetary policy regime, development of money market and financial system, openness of the economy, etc.

A number of interest rate pass-through analyses have been performed for CEE countries [Horvath et al. (2004), Wrobel and Pawlowska (2002), Crespo-Cuaresma et al. (2004, 2007), Sander and Kleimeier (2004b), Tieman (2004), Petrevski and Bogoev (2012), Saborowski and Weber (2013), etc.]. In most of them, results indicate that the pass-through in CEE economies is faster than in the euro area and higher for loans than for deposits. Among loans, usually lending rates for households react less with slower adjustment than those for corporate loans. Heterogeneity among CEE countries is present and can be explained by different macroeconomic factors and financial structure. Crespo-Cuaresma et al. (2007) found evidence on declining pass-through in five CEE countries (Czech Republic, Hungary, Poland, Slovakia and Slovenia) in the period from mid 1990s up to mid-2000s, which can be attributed to inflation slowdown, declining competition among banks and greater reliance on foreign funding. Petrevski and Bogoev (2012) analysed the pass-through from money market rates to lending rates in three SEE economies with rigid exchange rate regimes – Macedonia, Bulgaria and Croatia. They found evidence of complete pass-through in the long run only for Macedonia, and incomplete for Bulgaria and Croatia, while in the short run the adjustment is incomplete and sluggish. Saborowski and Weber (2013) named the high liquidity, NPL ratios and loan dollarization as factors that weaken the pass-through in the group of Eastern European economies.

The panel method is less represented in the research on interest rate pass-through so far. In their panel, based on aggregate data, Sorensen and Werner (2006) reported a high degree of heterogeneity of the pass-through of market interest rates to bank interest rates in the euro area. Different long-run multipliers and speed of adjustment coefficients among countries can be primarily attributed to the different degree of competition in the banking sector across countries. Panel analyses on bank level data, as used in this paper, are quite scarce. Studies concerning some individual eurozone countries comprise those by Weth (2002), Gambacorta (2008) and De Graeve et al. (2004, 2007), finding almost complete pass-through in the long run for Germany and Italy, and heterogeneous results for Belgium. All of them found incomplete and heterogeneous size and speed of adjustment in lending rates in the short run. As for the non-euro area countries, research based on bank level data has been done by Chmielewski (2003), Horvath et al. (2004), Aydin (2007), Stanislawska (2014). In Turkey [see Aydin (2007)], the pass-through is higher for household loans then for corporate loans. Among household loans, the rates on cash and automobile loans move proportionally with the policy rate, while housing loans in the period of rapid credit growth display excessive sensitivity to the policy rate changes. For Poland [see Stanislawska (2014)] the results suggest complete pass-through effect for corporate deposits and some categories of household deposits. Also, completeness is found for consumer credit rates. As far as the influence of individual bank characteristics on the pass-through effect is concerned, results point out that they affect long-run multipliers only to a limited degree.

Several studies also examine the issue of an interest rate pass-through process depending on the direction of key policy rate change. The response of bank rates to changes in the key policy rate or money market rates seems to depend in some cases on whether market interest rates are rising or falling [Aydin (2007), Yildirim (2013), Mojon (2000)] or whether bank interest rates are below or above equilibrium levels [Hofmann (2000), Kleimeier and Sander (2000)].

4 Econometric methodology

Prior to testing long-run relationship between monetary policy rate and different kinds of lending rates, different panel unit root tests had been applied [(Levin, Lin and Chu (2002), Im, Peseran and Shin (2003), Maddala and Wu (1999), Choi (2001)].

After panel unit root test to determine whether time series of interest rates are nonstationary, we checked long-run relationship using Peter Pedroni (1997) panel cointegration tests. Although we presented all seven Pedroni tests, the decision relating to cointegration was made based on group *ADF*, panel *ADF*, and panel ρ statistics, that is, if at least one of those statistics confirms it. Specifically, we had in mind the results of Pedroni (2004) that showed that for values of T larger than 100, all seven statistics that were proposed do fairly well and are quite stable, while for smaller samples (T is lower than 20), group *ADF* statistics is the most powerful, followed by panel *ADF* and panel ρ statistics. We chose to use the non-weighted instead of the weighted panel Pedroni statistics due to their better performance in small samples. We also implemented Westerlund (2006) panel cointegration tests, which take into account cross-sectional dependence. Two of those tests are designed to test the alternative hypothesis that the panel is cointegrated as a whole (panel tests), while the other two (group mean tests) are designed to test the alternative hypothesis that at least one unit is cointegrated. Those test statistics are all normally distributed.

The long-run relationship between relevant macroeconomic variables was estimated by fully modified OLS (FMOLS), dynamic OLS (DOLS), pooled mean group (PMG), and mean group (MG) estimator techniques.

FMOLS estimation allows for serial correlation in the residuals and for endogeneity of regressors in the cointegrating regression, and results in an asymptotically efficient estimation of the cointegrating vector. The pooled FMOLS coefficients can be computed in two different ways: within a dimension and between dimensions. Here, we will present only the between-dimension group FMOLS estimator of the mean panel cointegration parameter from the equation:

$$y_{it} = \alpha_{1i} + \beta x_{it} + \mu_{it}$$

$$x_{it} = \alpha_{2i} + x_{it-1} + \epsilon_{it}$$

where is $\xi_{it} = (\mu_{it}, \epsilon_{it,})$ vector error process with the asymptotic covariance matrix Ω_i, where is Ω_{11} scalar long run variance of the residual μ_{it}, Ω_{22} is $m \times m$ long run covariance among the ϵ_{it}, and Ω_{21} is $m \times 1$ vector of long run covariance between μ_{it} and each of ϵ_{it}.

Cointegration parameter is given as:

$$\beta^*_{GFM} = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)^2\right)^{-1}\left(\sum_{t=1}^{T}\left(x_{it} - \bar{x}_i\right)y^*_{it} - T\hat{\gamma}_i\right),$$

$$y^*_{it} = y_{it} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}}\Delta x_{it},$$

$$\hat{\gamma}_i = \hat{\Gamma}_{21i} + \hat{\Omega}^0{}_{21i} - \frac{\hat{L}_{21i}}{\hat{L}_{22i}}\left(\hat{\Gamma}_{22i} + \hat{\Omega}^0{}_{21i}\right)$$

where x_{it} is the m-dimensional vector of explanation variables, and $\hat{L}_i$ is the lower triangular decomposition of a consistent estimator of the asymptotic covariance matrix $\Omega_i = \Omega_i^0 + \Gamma_i - \Gamma_i'$, with $\hat{L}_i$ normalized such that $\hat{L}_{22i} = \hat{\Omega}_{22i}^{-1/2}$, Γ_i is the weighted sum of autocovariances and the serial correlation adjustment parameter $\hat{\gamma}_i$. The FMOLS estimator is distributed normally [see Pedroni 1997, p. 103].

The expression following the summation over *i* is identical to the conventional time series FMOLS estimator, and the between-dimension estimator can be constructed simply as the average FMOLS estimator for each panel member. Likewise, the associated t statistics for the between-dimension estimator can be constructed as:

$$\bar{t}_{\beta^* GFM} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \hat{L}_{11i}^{-2} \left(\sum_{t=1}^{T} \left(x_{it} - \bar{x}_i \right)^2 \right)^{-1/2} \left(\sum_{t=1}^{T} \left(x_{it} - \bar{x}_i \right) y_{it}^* - T \hat{\gamma}_i \right) \rightarrow N(0,1)$$

In order to obtain an unbiased estimator of the long-run parameters, DOLS estimator uses parametric adjustment to the errors by including the past and the future values of the differenced regressors. The DOLS estimator is obtained from the following equation:

$$y_{it} = \alpha_i + x_{it}' \beta_i + \sum_{j=-q}^{q} c_{ij} \Delta x_{it+j} + \upsilon^*_{it}$$

where c_{ij} is the coefficient of a lead or lag of the first differenced explanatory variables. The estimated coefficient of DOLS is given by:

$$\hat{\beta}_{GDOLS} = \frac{1}{n} \sum_{i=1}^{n} \left(\sum_{t=1}^{T} \left(z_{it} z_{it}' \right) \right)^{-1} \left(\sum_{t=1}^{T} z_{it} \hat{y}_{it}^* \right),$$

where $z_{it} = \left[x_{it} - \bar{x}_i, \Delta x_{i,t-q}, \ldots, \Delta x_{i,t+q} \right]$ is $2(q+1) \times 1$ vector of regressors and $\hat{y}_{it}^* = y_{it} - \bar{y}_i$.

The Pooled Mean Group (PMG) estimator, introduced by Pesaran, Shin and Smith (1997), involves pooling and averaging and allows the intercepts, short-run coefficients and error variances to differ freely across groups, but the long-run coefficients are constructed to be the same. They propose estimating the following autoregressive distributed lag (ARDL) model of order p and q:

$$\Delta y_{it} = \mu_i + \varphi_i (y_{i,t-1} - \Theta x_{it}) + \sum_{j=1}^{p-1} \lambda_{ij} \Delta y_{i,t-j} + \sum_{j=0}^{q-1} \delta_{ij} \Delta x_{i,t-j} + \epsilon_{ij}$$

The dependent variable in the first differences is regressed on the lagged values of the dependent and independent variables in the levels and first differences. The long-run coefficients, Θ, are defined to be the same across groups. Testing statistical significance of the error correction term from the pooled mean group estimator can be used as a test of cointegration. A negative and statistically significant error correction term, φ, confirms presence of long-run relationship between y_{it} and x_{it}. The equation is estimated using the maximum likelihood procedure.

In this setup, Θ shows the degree of pass-through in the long run, while δ_{i0} stands for short-term pass-through (effect on bank lending interest rates within a month). Error correction term displays how fast banks respond to monetary policy decisions on policy rate, where average adjustment period in terms of months is calculated as $(1 - \delta_{i0})/\varphi$.

The kind of estimation where panel coefficients were obtained by averaging individual cross-sectional coefficients is called Mean Group (MG) estimation. The MG allows that all of the parameters can differ across units. Pesaran, Shin and Smith (1997) suggest using Hausman test (1978) to test long-run homogeneity. Rejection of the test would suggest that panel is too heterogeneous for imposing long-run homogeneity, in which case PMG method is inadequate.

5 Data description and definition of variables

For dinar loans, the dataset consists of per annum average interest rates on new business corporate (*C_NB*) and household (*H_NB*) loans on a monthly basis covering the period from September 2010 to May 2021. As a proxy of monetary policy rate we have used the central bank key policy rate (IR), as well as interbank money market rates with one-week maturity (BELIBOR1W) and with three-month maturity (BELIBOR3M). The main monetary policy instrument is the key policy rate applied in the National Bank of Serbia (NBS) main open market operations – notably the 1-week reverse repo transactions[1] – to temporarily change the liquidity conditions of the banking system. As of December 2012, main open market operations of withdrawing liquidity have been conducted at variable interest rate auctions, with the key policy rate indicating the maximum rate that could be accepted.

In the period observed on average approximately 37% of new business loans to the private sector were in dinars. Out of that, with 25% of new business loans in dinars on average, the dinarization is less pronounced in the corporate sector, while this share for households is much higher (66%). Bearing in mind the level of euroization of loans we have tested the effects of the three- and six-month EURIBOR to interest rates on euro-indexed loans, too. Since the reserve requirement is used as a supportive monetary policy instrument that influences the price of banks' sources of finance and has been amended on several occasions in the period under review[2], we have used EURIBOR series adjusted by effective foreign currency reserve requirement rate (RR), calculated as *EURIBOR/(1-RR)*.

[1] 2-week repo transactions until July 2012.

[2] By decreasing foreign currency reserve requirement ratio and by increasing its portion allocated in dinars. As of February 2016, ratios for foreign currency liabilities stood at 20% and 13%, for maturities up to and over two years, with 38% and 30% of foreign currency reserve requirement allocated in dinars. Dinar reserve requirement rates stood at 5% and 0%, depending on maturity.

The interest rate data came from statistics collected by the NBS. During the period observed the number of banks in the Serbian market declined from 33 to 24. For testing the pass-through effect on lending rates for households we have included 19 banks (accounting on average for 98.2% of the total banking sector assets) and for corporates we have included 17 banks (accounting on average for 92.2% of the total banking sector assets), as they have reported interest rate data for the whole period. The number of banks included in the analysis for testing the reaction of different loan categories of household and corporate lending rates is lower and it is dependent upon whether the banks granted a certain type of loan during the whole period observed.

Chart 1 shows the movement of the NBS repo rate and interbank money market rates, Chart 2 displays weighted average rates by type of dinar loans, and Chart 3 and Chart 4 the movement of money market and dinar lending rates for the corporate and household sector, respectively.

Chart 1. **Monetary policy and interbank market rates**
(in %)

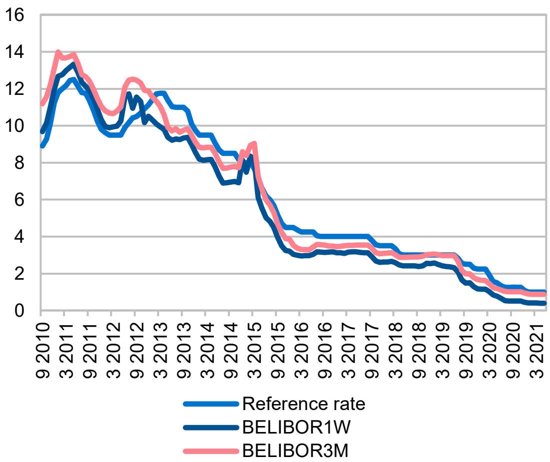


Chart 2. **Weighted average dinar lending rates to households and corporates by purpose**
(in %)

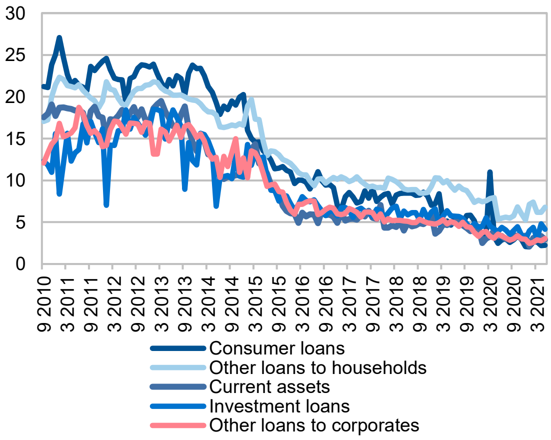


Chart 3. **Interest rates for corporate dinar loans and BELIBOR3M**
(in %)

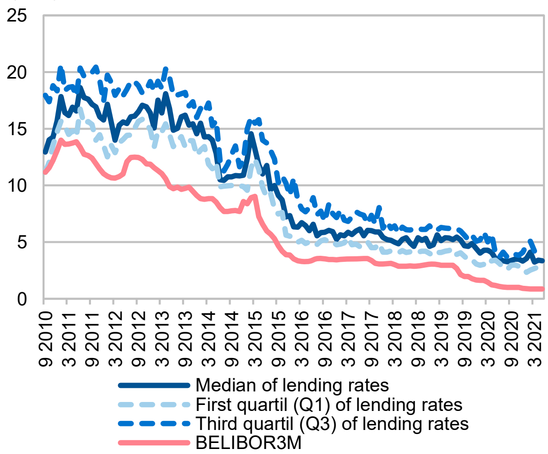
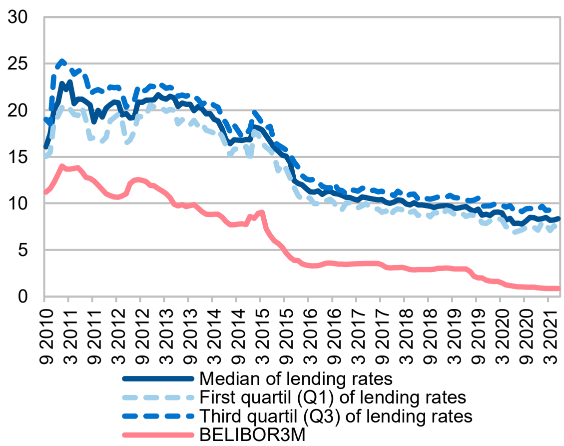


Chart 4. **Interest rates for household dinar loans and BELIBOR3M**
(in %)



Source: NBS.

Chart 1 confirms that short-term money market rates generally mirrored the key policy rate movements. The relationship between lending rates on household and corporate loans and interbank money market rates is evident (Chart 3 and Chart 4), too. However, for a more precise conclusion, we need to test and estimate the long-run relationship.

Although this period is characterized by phases of monetary policy tightening and relaxing, the period of relaxation is prevailing. The latest cycle of monetary policy relaxation started in May 2013 and since then a significant fall in corporate and household lending rates has been recorded, with the exception of early 2015, when temporary dinar liquidity shortage in the banking sector caused higher volatility of interbank money market rates, and consequently lending rates. However, the fall in interest rates has accelerated since March 2015 along with the speed-up in monetary policy relaxation.

In order to address the problem of heterogeneity of the pass-through effect across banks, we take into account several bank characteristics: **size, quality of credit portfolio, structure of financing, capital position and liquidity**. Each feature is considered separately. According to the median value of each indicator, banks are divided into two groups. We used total assets as a size indicator and the share of NPLs in total loans as a portfolio quality indicator. Structure of financing was measured by the share of non-financial sector deposits in total liabilities, capital position by capital adequacy ratio (CAR) and liquidity as the share of liquid[3] in total assets.

6 Empirical results

6.1 Testing the relationship for dinar interest rates

We started the empirical analysis by testing for the presence of a unit root in a series of interest rates on dinar loans. According to the results of panel unit root tests, non-stationarity could not be strongly rejected at 5% significance level in almost all cases, which led us to the conclusion that variables are non-stationary in levels (see Table 1 in Appendix).

The cointegration relationship between interest rates on new business corporate dinar loans and money market interest rates BELIBOR1W and BELIBOR3M is confirmed by all Pedroni tests for the model with individual intercepts. Long-run relationship is also confirmed for interest rates on new household dinar loans and money market interest rates by all Pedroni tests. To control cross-sectional dependence, we employ Westerlund panel cointegration test. Overall, the results strongly reject the null hypothesis of no cointegration (see Table 2 and Table 3 in Appendix).[4]

In order to estimate the long-run relationship between dinar lending and money market rates, we employed FMOLS and DOLS methods. The results obtained for the whole sample

[3] Liquid assets comprise claims on NBS, claims under repo transactions and investment in government securities.

[4] The unit root and panel cointegration tests were performed for EURIBOR and interest rates on euro-indexed loans too, but, for practical reasons, only the results of those carried out on dinar loans are presented.

suggest that there is a significant long-run relationship (Table 1) and that the pass-through effect is complete for both total corporate and household loans. In both cases, the pass-through effect is more pronounced for new business corporate than for household loans, which is in line with the findings of Horvath et al. (2004), Crespo-Cuaresma et al.(2007) and Sander and Kleimeier (2004b) and opposite to those of Aydin (2007).

Table 1 **Estimates of the long-run pass-through of market rates to bank lending rates**

	FMOLS		DOLS	
	Corporate loans	**Household loans**	**Corporate loans**	**Household loans**
BELIBOR1W	1.29***	1.19***	1.29***	1.19***
BELIBOR3M	1.24***	1.15***	1.24***	1.15***

Note: Grouped *FMOLS* and grouped *DOLS* with automatic lags selection based on *SIC*. * refers to statistical significance at 10%, ** refers to statistical significance at 5%, and *** refers to statistical significance at 1%.

In order to test the change in the interest rate channel strength, we have estimated coefficients in two periods, 2010–2014 and 2010–2021. The coefficients obtained proved the strengthening of the interest rate channel over time (Table 2), as for the whole period of analysis (2010–2021) the results suggest complete interest rate pass-through, thanks to the increasing interbank competition in the loan market, higher economic growth and reduced internal and external imbalances and therefore, significantly reduced macroeconomic uncertainty.

Table 2 **Estimates of the long run pass-through of market rates to bank lending rates - DOLS method**

	2010-2014		2010-2021	
	Corporate loans	**Household loans**	**Corporate loans**	**Household loans**
BELIBOR1W	0.91***	0.68***	1.29***	1.19***
Number of observations	867	965	2.166	2.421
BELIBOR3M	0.88***	0.68***	1.24***	1.15***
Number of observations	867	965	2.166	2.421

Note: Grouped *DOLS* with automatic lags selection based on *SIC*. * refers to statistical significance at 10%, ** refers to statistical significance at 5%, and *** refers to statistical significance at 1%.

Further, we tried to examine the impact of the country risk premium, measured by EMBI, on the interest rate pass-through. The results confirmed a relationship with statistically significant coefficients hovering around 0.5 (Table 3), indicating that lower risk premium in addition to monetary policy relaxation of the NBS also affect the fall of dinar lending interest rates for corporate sector. The inclusion of EMBI has brought the interest rate pass-through coefficient closer to 1. For the household sector, a statistically significant relationship between lending rates and EMBI was not found.

Table 3 **Estimates of the long run pass-through of market rates to bank lending rates**

	FMOLS	DOLS
	Corporate loans	**Corporate loans**
BELIBOR1W	1.12***	1.13***
EMBI	0.52***	0.48***
BELIBOR3M	1.07***	1.08***
EMBI	0.54***	0.50***

Note: Grouped *FMOLS* and grouped *DOLS* with automatic lags selection based on *SIC*. * refers to statistical significance at 10%, ** refers to statistical significance at 5%, and *** refers to statistical significance at 1%.

The long-run relationship is confirmed by PMG and MG methods (Table 4), too, also indicating complete pass-through effect for corporate and household loans. Hausman test indicates insignificant difference between the results obtained by MG and PMG methods for corporates, while in the case of households, using MG method is preferable. Overall, results obtained by PMG and MG methods show that the long-run interest rate pass-through effect for corporate loans ranges from 1.18 to 1.30, while the result obtained for household loans ranges between 1.19 and 1.23, depending on the money market rate used.

Table 4. **PMG and MG estimates of the pass-through of market rates to bank lending rates**

	PMG				**MG**			
	Corporate loans		**Household loans**		**Corporate loans**		**Household loans**	
	estimate	p value	estimate	p value	estimate	p value	estimate	p value
	BELIBOR1W, monthly average				**BELIBOR1W, monthly average**			
Q_i	1.235	0.000	1.344	0.000	1.296	0.000	1.225	0.000
f_i	-0.505	0.000	-0.210	0.000	-0.574	0.000	-0.254	0.000
d_{i0}	-0.377	0.019	-0.071	0.592	-0.504	0.000	-0.123	0.369
d_{i1}			0.234	0.030			0.185	0.040
λ_{i1}			0.019	0.636			0.037	0.351
m_i	1.430	0.000	1.370	0.000	1.463	0.000	1.829	0.000
Average adjustment period	2.7		5.1		2.6		4.4	
Hausman	1.210	0.271	3.460	0.060				
	BELIBOR3M, monthly average				**BELIBOR3M, monthly average**			
Q_i	1.184	0.000	1.302	0.000	1.247	0.000	1.192	0.000
f_i	-0.519	0.000	-0.221	0.000	-0.593	0.000	-0.270	0.000
d_{i0}	-0.424	0.000	0.053	0.701	-0.583	0.000	-0.011	0.942
d_{i1}			0.222	0.126			0.158	0.256
λ_{i1}			-0.002	0.954			0.017	0.636
m_i	1.220	0.000	1.337	0.000	1.200	0.000	1.791	0.000
Average adjustment period	2.7		4.3		2.7		3.7	
Hausman	1.390	0.238	3.200	0.070				
Number of observations	2,176		2,432		2,176		2,432	

Table 4 also presents estimates of the short-term adjustment of dinar lending rates. In line with expectations, the speed of adjustment term is negative and statistically significant in all cases. About 50–60% of the money market change after one month is transmitted to corporate lending rates and about 21–27% is transmitted to household lending rates. This result is not surprising as corporates have more finance alternatives relative to households. The average period of adjustment in the interest rate channel for household loans (4–5 months) is longer than in the case of corporate loans (2–3 months). This can be linked to the fact that companies have more alternative sources of finance than households. However, immediate reaction is negative, although statistically insignificant in the case of household loans.

Analysis was conducted for different types of dinar loans too – in the case of corporate sector analysis covers interest rate on current assets loans and for households it covers interest rates for cash loans (Table 5). Estimation results indicate that dinar corporate loans for current assets display excessive sensitivity to interbank money market rates, while cash loans are slightly less sensitive than other types of household loans in dinars. One of the possible reasons could be that cash loans are largely refinanced and that banks compete among themselves not only by lowering interest rates, but also by extending loan maturity.

Table 5. **DOLS estimates of the pass-through of market rates to different types of bank lending rates**

	Corporate current asset loans (15 banks)	**Household cash loans (19 banks)**
BELIBOR1W	1.31***	1.12***
BELIBOR3M	1.26***	1.09***

Note: Grouped *DOLS* with automatic lags selection based on *SIC*. * refers to statistical significance at 10%, ** refers to statistical significance at 5%, and *** refers to statistical significance at 1%.

6.2 Testing the relationship for euro-indexed interest rates

We have also estimated the long-term pass-through from EURIBOR to euro-indexed loans, both for total and for certain types of loans - current asset and investment loans for corporates, and housing loans for households (Tables 6 and 7). As in the analysis of dinar loans, we employed the FMOLS and DOLS methods. The obtained results suggest pass-through that is almost complete both in the case of corporates and households, while a somewhat stronger relationship was found with EURIBOR3M than with EURIBOR6M.

Table 6 **Estimates of the long run pass-through between euro-indexed bank lending rates, money market rates and risk premium**

	FMOLS		DOLS	
	Corporate loans	Household loans	Corporate loans	Household loans
EURIBOR3M	0.89***	0.99***	0.94***	1.02***
EMBI	0.94***	0.82***	0.91***	0.81***
EURIBOR6M	0.90***	0.88***	0.86***	0.91***
EMBI	0.99***	0.81***	0.88***	0.79***

Note: Grouped *FMOLS* and grouped *DOLS* with automatic lags selection based on *SIC*. * refers to statistical significance at 10%, ** refers to statistical significance at 5%, and *** refers to statistical significance at 1%.

The country risk premium, measured by EMBI, declined and has been hovering around its historic lows at the end of 2019, thanks to the improvement in macroeconomic fundamentals. The NBS supported this by delivering low and stable inflation, preserving financial stability, and contributing to the improved investment climate. The drop in risk premium allowed for the decline in interest rates, as we found a strong and significant influence of the country risk premium on the euro-indexed lending rates, on both corporate and household loans. This relationship proves to be stronger in the case of corporate loans, with statistically significant coefficients around 0.9, as corporates are more exposed to general macroeconomic conditions, while for household loans they move around 0.8.

Looking by the type of loans, we found stronger connection for current asset loans (with coefficients between 0.96 and 1.05) compared to investment loans (between 0.74 and 0.9). The impact of risk premium is also significant but more pronounced for current asset loans, as those loans have a more dispersed use and are granted to a broader range of different clients and therefore might be perceived as riskier by banks. The obtained lower coefficients for housing loans (0.6–0.7) can be partially explained by the presence of subsidized lending programs in the part of the period observed. Although significant, the impact of the risk premium on this type of loans is also lower (0.5), which is understandable as the vast majority of these loans, besides being adequately collateralized, is also insured with the National Mortgage Insurance Corporation.

Table 7 **Estimates of the long run pass-through between euro-indexed bank lending rates, money market rates and risk premium**

	FMOLS			DOLS		
	Corporate Current Assets (16 banks)	Corporate investment loans (13 banks)	Housing loans (15 banks)	Corporate Current Assets (16 banks)	Corporate investment loans (13 banks)	Housing loans (15 banks)
EURIBOR3M	1.05***	0.83***	0.61***	1.04***	0.90***	0.66***
EMBI	1.05***	0.84***	0.56***	1.04***	0.81***	0.54***
EURIBOR6M	0.96***	0.74***	0.54***	0.97***	0.80***	0.59***
EMBI	1.02***	0.83***	0.55***	1.00***	0.79***	0.53***

Note: Grouped FMOLS and grouped DOLS with automatic lags selection based on *SIC*. * refers to statistical significance at 10%, ** refers to statistical significance at 5%, and *** refers to statistical significance at 1%.

6.3 Testing impact of different individual bank characteristics on interest rate pass-through

To test whether interest rate pass-through is affected by individual bank characteristics, a separate equation was estimated for each bank characteristic (size, strength of deposit base, quality of credit portfolio, capital position, liquidity, dinarization criterion). Banks with the indicator value above the median (except for the share of NPLs, where it is opposite) are categorized in Cluster I, while banks with a lower-than-median indicator value belong to Cluster II (Table 8). The tests are performed only for dinar loans. The results are presented in Appendix, and they suggest that individual bank characteristics influence more long-term adjustment of household lending rates in national currency.

Table 8 **Average values of indicators by clusters during the observed period**

Bank characterstics	Median
Size criterion - total assets (bln RSD)	131.7
Portfolio quality - share of NPLs to total loans, gross principle*	1.7% - corporates 3.0% - households
Deposit base - share of deposits in total liabilities	48.8%
Capital position - Capital adequacy ratio	22.0%
Liquidity - share of liquid assets** to total assests	23.5%
Dinarization criterion - share of dinar in total loans	41.8% - households 17.4% - corporates

* Average values of indicator in the last 12 months, cluster I refers to values below the median

** Claims on NBS, claims under repo transactions and investment in government securities

As for the size criterion (Table 5 in Appendix), bigger banks adjust their interest rate on corporate loans somewhat faster than smaller banks, but the difference is not significant. Bigger banks have easier access to cheaper sources of funding (as this group mainly comprises subsidiaries of foreign banks) and they usually serve better quality corporate clients that have a broader range of possible sources of finance. The opposite holds true in the case of household loans – smaller banks adjust faster. A possible explanation could be that the smaller value of a single credit lot allows for stronger competition in the household segment.

As for the quality of the credit portfolio, the reaction of lending rates on corporate loans to money market rates BELIBOR1W is almost the same for both clusters (Table 6 in Appendix), while to BELIBOR3M it is stronger for banks with lower NPL ratios, which is in line with intuition. In the case of households, the reaction is stronger for the banks with lower NPL ratios too, while differences among the coefficients between clusters are higher than in the case of corporates.

Differences between the clusters are also found in the case of sources of bank financing. Banks with a higher share of deposits from non-financial sector recorded a slower adjustment to market interest rates in the case of household loans, which is in line with expectations. A statistically significant difference was also found for corporate loans, but contrary to expectations, the lower long-run coefficient was recorded for banks with a lower share of deposits (Table 6 in Appendix). One of the reasons could be that this cluster is predominantly made of foreign banks’ subsidiaries that have access to cheaper funds from abroad, either within the group or through participation in international credit lines.

The results indicate almost the same long-term reaction of interest rates on corporate loans for both clusters of banks based on capital adequacy ratios (Table 8 in Appendix), while in the case of household loans, banks with lower capital adequacy ratios tend to adjust more. Slower adjustment of big and well-capitalized banks is in line with the credit channel view, as those banks have more capacity to avoid transferring the higher cost of the sources of funding onto clients in the case of monetary tightening. However, this classification into clusters is tentative, as all banks in Serbia are well-capitalized, with individual ratios far above the minimum capital adequacy requirement.

As for the liquidity criterion, the theoretical prediction that less liquid banks adjust faster to interest rate changes is not confirmed, given that more liquid banks follow money market interest rates faster than less liquid ones (Table 8 in Appendix). That could partly be explained by the prevalence of monetary policy easing and the reduction in money market interest rates during the sample period, which more liquid banks used to increase their market position. Also, it should be borne in mind that this cluster classification, as in the case of capitalization criterion, is only tentative, since all banks in Serbia are liquid.

The results indicate stronger long-term reaction of banks with a higher share of dinar loans in total loans (Table 9 in Appendix) for both corporate and household loans, which is in line with expectations, while in the case of corporates, the difference between clusters is not significant.

7 Concluding remarks

The aim of this paper was to test the long-run relationship between monetary policy rate and money market rates and different bank lending dinar interest rates in Serbia and to test the strength of the interest rate channel. Analysis was conducted for the period from September 2010 to May 2021 by the panel data methods. The novelty of the paper lies in presenting evidence of interest rate pass-through for the Serbian economy, based on the average pass-through across individual banks.

The results of empirical analysis can be summarized as follows:

1. **The confirmed statistically significant long-run relationship between monetary policy rates/money market rates and dinar lending rates shows that the interest rate pass-through effect is complete**.

2. **The interest rate channel gained more strength over time**, as confirmed by the higher long-run relationship coefficient in the whole period compared to the period 2010–2014.

- Some of the possible explanations for the strengthening of the interest rate pass-through are a higher level of competition, decline in the risk premium and interest rate volatility in recent years, rising reliance on domestic sources of funding, speed-up in economic recovery and strengthening of macroeconomic fundamentals, etc.
- Interest rate pass-through for dinar loans appears to be complete for both corporate and household lending rates, with long-run coefficients exceeding 1.
- The long run pass-through estimates using different methods (FMOLS, DOLS, and MG) are very close to each other, indicating robustness of the results. Also, results of PMG and MG methods pointed to a homogenous reaction for the corporate sector.

- o Average period of adjustment in the interest rate channel for household loans is longer than in the case of corporate loans and could be linked to the fact that companies have more alternative sources of finance than households.

3. A statistically significant relationship between lending rates for corporates, money market rates and risk premium, measured by EMBI, indicates **that lower risk premium, in addition to monetary policy relaxation of the NBS, also affects dinar lending interest rates for the corporate sector**. The inclusion of EMBI has brought the interest rate pass-through coefficient closer to 1. An assessment done for different types of dinar loans indicates that **corporate loans for current assets display excessive sensitivity to interbank money market rates, while consumer loans are slightly less sensitive than other types of household loans**.

4. **Individual bank characteristics affect the interest rate pass-through.**

- o In the case of the household sector, banks with a higher capital adequacy ratio, higher NPL share, higher share of deposits in total liabilities and higher share of dinar loans in total loans tend to adjust more slowly to changes in the reference rate, though with a complete pass-through. Contrary to expectations, the pass-through is higher for banks with smaller than for banks with bigger balance sheet assets.
- o As for corporates, the adjustment is faster for banks with lower NPL ratios, which is in line with intuition. Contrary to expectations, adjustment is faster for banks with a higher share of liquid assets in total assets and a higher deposit base, but, as in the case of households, the pass-through is complete for both bank groups. As for other criteria, the differences across banks in respect of the adjustment of dinar corporate loans interest rates are minimal.

5. The assessed pass-through between EURIBOR and new business euro-indexed loans suggest complete pass-through in the case of household loans by both FMOLS and DOLS methods, while for corporate loans it is close to complete. The strong connection between lending rates and the risk premium is also found. Looking by the type of loans, the pass-through effect of EURIBOR is stronger for current asset loans than for investment loans. The pass-through effect is lower for housing loans compared to total euro-indexed household loans.

Appendix

Table 1 **Panel unit root test results**

	Corporate loans	Household loans	Current assets loans	Cash loans
Levin, Lin & Chu t*	-0.769 (0.211)	0.271 (0.607)	0.110 (0.544)	0.219 (0.587)
Im, Pesaran and Shin W	2.400 (0.991)	2.13(0.983)	2.418 (0.992)	2.232 (0.987)
ADF - Fisher X^2 - stat	11.815 (0.99)	23.931 (0.963)	10.215 (0.99)	17.351 (0.998)
ADF - Choi Z - stat	2.674 (0.996)	2.211 (0.986)	2.679 (0.996)	2.331 (0.990)

Note: P values are given in parentheses. Model with individual effects was used.
The number of lags included in the model is chosen according to the Schwarz information criterion.

Table 2 **Results of Pedroni panel cointegration test**

	[C_NB, BELIBOR1W]	[H_NB, BELIBOR1W]	[C_NB, BELIBOR3M]	[H_NB, BELIBOR3M]
Panel v	9.310***	12.391***	6.730***	7.731***
Panel rho	-56.400***	-25.948***	-56.470***	-25.831***
Panel PP	-24.121***	-14.538***	-24.262***	-14.777***
Panel ADF	-15.130***	-11.227***	-15.063***	-11.349***
Group rho	-50.491***	-18.202***	-50.470***	-17.704***
Group PP	-27.350***	-12.283***	-27.634***	-12.525***
Group ADF	-18.051***	-11.513***	-17.958***	-11.609***

Note: Model with individual intercepts. In all tables, * refers to statistical significance at 10%, ** refers to statistical significance at 5%, and *** refers to statistical significance at 1%. The number of lags included in the model is chosen according to the Schwarz information criterion.

Table 3 **Results of Westerlund panel cointegration tests**

	[C_NB, BELIBOR1W]	[H_NB, BELIBOR1W]	[C_NB, BELIBOR3M]	[H_NB, BELIBOR3M]
Gt	-5.444 (0.000)	-4.432 (0.000)	-5.486 (0.000)	-4.554 (0.000)
Ga	-51.119 (0.000)	-33.104 (0.000)	-52.643 (0.000)	-34.886 (0.000)
Pt	-21.556 (0.000)	-19.414 (0.000)	-21.606 (0.000)	-19.862 (0.000)
Pa	-46.026 (0.000)	-31.729 (0.000)	-47.202 (0.000)	-32.288 (0.000)

Note: Values in parentheses are bootstrapped p-values, robust in the presence of common factors in the time series.

Table 4 **MG estimates of the pass-through of market rates to bank lending rates according to size criterion**

	Corporate				Households			
	Cluster I		Cluster II		Cluster I		Cluster II	
	estimate	p value	estimate	p value	estimate	p value	estimate	p value
	BELIBOR1W, monthly average				BELIBOR1W, monthly average			
Q_i	1.301	0.000	1.288	0.000	1.199	0.000	1.254	0.000
f_i	-0.609	0.000	-0.535	0.000	-0.242	0.000	-0.269	0.000
d_{i0}	-0.573	0.002	-0.424	0.034	-0.098	0.639	-0.150	0.414
d_{i1}					0.279	0.064	0.081	0.553
λ_{i1}					0.076	0.050	-0.005	0.944
m_i	1.347	0.000	1.593	0.000	1.870	0.000	1.782	0.000
Average adjustment period	2.6		2.7		4.5		4.3	
Number of observations	1,152		1,024		1,270		1,143	
	BELIBOR3M, monthly average				BELIBOR3M, monthly average			
Q_i	1.252	0.000	1.242	0.000	1.170	0.000	1.217	0.000
f_i	-0.626	0.000	-0.555	0.000	-0.258	0.000	-0.281	0.000
d_{i0}	-0.533	0.005	-0.640	0.000	0.039	0.859	-0.066	0.737
d_{i1}					0.293	0.106	0.008	0.969
λ_{i1}					0.046	0.227	-0.014	0.836
m_i	1.067	0.000	1.351	0.000	1.856	0.000	1.718	0.000
Average adjustment period	2.4		3.0		3.7		3.8	
Number of observations	1,152		1,024		1,270		1,143	

Note: Cluster I refers to banks with bigger total assets, cluster II to banks with total assets less than 131.7 billion of dinars.

Table 5 **MG estimates of the pass-through of market rates to bank lending rates according to quality of portfolio criterion**

	Corporate				Households			
	Cluster I		Cluster II		Cluster I		Cluster II	
	estimate	p value	estimate	p value	estimate	p value	estimate	p value
	BELIBOR1W, monthly average				BELIBOR1W, monthly average			
Q_i	1.329	0.000	1.247	0.000	1.268	0.000	1.178	0.000
f_i	-0.565	0.000	-0.588	0.000	-0.259	0.000	-0.250	0.000
d_{i0}	-0.383	0.032	-0.678	0.001	-0.198	0.058	-0.039	0.886
d_{i1}					0.166	0.202	0.207	0.218
λ_{i1}					0.072	0.256	-0.001	0.976
m_i	1.514	0.000	1.390	0.000	1.758	0.000	1.906	0.000
Average adjustment period	2.4		2.9		4.6		4.2	
Number of observations	1,280		896		1,270		1,143	
	BELIBOR3M, monthly average				BELIBOR3M, monthly average			
Q_i	1.280	0.000	1.200	0.000	1.233	0.000	1.147	0.000
f_i	-0.577	0.000	-0.615	0.000	-0.274	0.000	-0.263	0.000
d_{i0}	-0.475	0.001	-0.737	0.000	-0.061	0.647	0.045	0.872
d_{i1}					0.091	0.641	0.233	0.263
λ_{i1}					0.058	0.335	-0.027	0.516
m_i	1.237	0.000	1.148	0.000	1.713	0.000	1.877	0.000
Average adjustment period	2.6		2.8		3.9		3.6	
Number of observations	1,280		896		1,270		1,143	

Note: Cluster I refers to banks with the share of NPLs for corporates less than 1.7% and for households less than 3%, cluster II to banks with the share of NPLs for corporates over 1.7% and for households over 3%.

Table 6 **MG estimates of the pass-through of market rates to bank lending rates according to deposit base criterion**

	Corporate				Households			
	Cluster I		Cluster II		Cluster I		Cluster II	
	estimate	p value	estimate	p value	estimate	p value	estimate	p value
	BELIBOR1W, monthly average				BELIBOR1W, monthly average			
Q_i	1.361	0.000	1.237	0.000	1.184	0.000	1.271	0.000
f_i	-0.550	0.000	-0.596	0.000	-0.225	0.000	-0.303	0.000
d_{i0}	-0.311	0.159	-0.676	0.000	0.157	0.363	-0.434	0.010
d_{i1}					0.067	0.448	0.317	0.093
λ_{i1}					0.040	0.438	0.035	0.598
m_i	1.249	0.000	1.653	0.000	1.764	0.000	1.901	0.000
Average adjustment period	2.4		2.8		3.7		4.7	
Number of observations	1,024		1,152		1,270		1,143	
	BELIBOR3M, monthly average				BELIBOR3M, monthly average			
Q_i	1.312	0.000	1.189	0.000	1.156	0.000	1.233	0.000
f_i	-0.568	0.000	-0.615	0.000	-0.225	0.000	-0.318	0.000
d_{i0}	-0.460	0.023	-0.693	0.000	0.288	0.097	-0.343	0.073
d_{i1}					0.031	0.768	0.300	0.269
λ_{i1}					0.020	0.697	0.016	0.792
m_i	0.978	0.003	1.340	0.000	1.748	0.000	1.838	0.000
Average adjustment period	2.6		2.8		3.2		4.2	
Number of observations	1,024		1,152		1,270		1,143	

Note: Cluster I refers to banks with the share of deposits in total liabilities over 48.8%, cluster II to banks with the share of deposits in total liabilities less than 48.8%.

Table 7 **MG estimates of the pass-through of market rates to bank lending rates according to capital position criterion**

	Corporate				Households			
	Cluster I		Cluster II		Cluster I		Cluster II	
	estimate	p value	estimate	p value	estimate	p value	estimate	p value
	BELIBOR1W, monthly average				BELIBOR1W, monthly average			
Q_i	1.293	0.000	1.296	0.000	1.147	0.000	1.311	0.000
f_i	-0.560	0.000	-0.581	0.000	-0.249	0.000	-0.261	0.000
d_{i0}	-0.617	0.005	-0.478	0.003	0.089	0.696	-0.358	0.001
d_{i1}					0.044	0.796	0.343	0.000
λ_{i1}					-0.023	0.709	0.145	0.016
m_i	1.827	0.000	1.253	0.000	2.003	0.000	1.634	0.000
Average adjustment period	2.9		2.5		3.7		5.2	
Number of observations	1,024		1,152		1,270		1,143	
	BELIBOR3M, monthly average				BELIBOR3M, monthly average			
Q_i	1.245	0.000	1.257	0.000	1.121	0.000	1.272	0.000
f_i	-0.567	0.000	-0.606	0.000	-0.264	0.000	-0.274	0.000
d_{i0}	-0.663	0.010	-0.553	0.000	0.185	0.542	-0.228	0.055
d_{i1}					-0.039	0.866	0.378	0.070
λ_{i1}					-0.036	0.531	0.077	0.001
m_i	1.570	0.000	0.984	0.000	1.983	0.000	1.577	0.000
Average adjustment period	2.9		2.6		3.1		4.5	
Number of observations	1,024		1,152		1,270		1,143	

Note: Cluster I refers to banks with capital adequacy ratio over 22%, cluster II to banks with CAR less than 22%.

Table 8 M**G estimates of the pass-through of market rates to bank lending rates according to liquidity criterion**

	Corporate				Households			
	Cluster I		Cluster II		Cluster I		Cluster II	
	estimate	p value	estimate	p value	estimate	p value	estimate	p value
	BELIBOR1W, monthly average				BELIBOR1W, monthly average			
Q_i	1.351	0.000	1.232	0.000	1.236	0.000	1.213	0.000
f_i	-0.579	0.000	-0.569	0.000	-0.234	0.000	-0.278	0.000
d_{i0}	-0.401	0.061	-0.620	0.000	-0.136	0.521	-0.108	0.549
d_{i1}					0.295	0.049	0.063	0.636
λ_{i1}					0.109	0.011	-0.042	0.505
m_i	1.267	0.000	1.683	0.000	1.780	0.000	1.882	0.000
Average adjustment period	2.4		2.8		4.9		4.0	
Number of observations	1,152		1,024		1,270		1,143	
	BELIBOR3M, monthly average				BELIBOR3M, monthly average			
Q_i	1.300	0.000	1.187	0.000	1.207	0.000	1.176	0.000
f_i	-0.581	0.000	-0.606	0.000	-0.249	0.000	-0.291	0.000
d_{i0}	-0.392	0.017	-0.798	0.000	-0.010	0.966	-0.012	0.952
d_{i1}					0.326	0.069	-0.028	0.892
λ_{i1}					0.076	0.085	-0.047	0.416
m_i	0.963	0.002	1.467	0.000	1.763	0.000	1.821	0.000
Average adjustment period	2.4		3.0		4.1		3.5	
Number of observations	1,152		1,024		1,270		1,143	

Note: Cluster I refers to banks with the share of liquid assets over 23.5%, Cluster II refers to banks with the share of liquid assets less than 23.5%.

Table 9 **MG estimates of the pass-through of market rates to bank lending rates according to dinarization criterion**

	Corporate				Households			
	Cluster I		Cluster II		Cluster I		Cluster II	
	estimate	p value	estimate	p value	estimate	p value	estimate	p value
	BELIBOR1W, monthly average				BELIBOR1W, monthly average			
Q_i	1.300	0.000	1.290	0.000	1.316	0.000	1.124	0.000
f_i	-0.557	0.000	-0.594	0.000	-0.264	0.000	-0.245	0.000
d_{i0}	-0.415	0.022	-0.605	0.003	-0.028	0.870	-0.228	0.302
d_{i1}					0.162	0.053	0.211	0.294
λ_{i1}					0.057	0.152	0.016	0.833
m_i	1.568	0.000	1.345	0.000	1.888	0.000	1.762	0.000
Average adjustment period	2.5		2.7		3.9		5.0	
Number of observations	1,152		1,024		1,270		1,143	
	BELIBOR3M, monthly average				BELIBOR3M, monthly average			
Q_i	1.252	0.000	1.242	0.000	1.278	0.000	1.097	0.000
f_i	-0.580	0.000	-0.607	0.000	-0.286	0.000	-0.250	0.000
d_{i0}	-0.455	0.009	-0.728	0.000	0.107	0.510	-0.142	0.574
d_{i1}					0.104	0.353	0.219	0.426
λ_{i1}					0.034	0.373	-0.002	0.998
m_i	1.325	0.000	1.060	0.001	1.888	0.000	1.682	0.000
Average adjustment period	2.5		2.8		3.1		4.6	
Number of observations	1,152		1,024		1,270		1,143	

Note: Cluster I refers to banks with the share of dinar loans to households over 41.8% and to banks with the share of dinar loans to corporate sector over 17.4%.

References

Aydin, H.I., 2007. “Interest Rate Pass-Through in Turkey”, The Central Bank of the Republic of Turkey, Working Paper no. 07/05.

Aziz, A. and Sh. Rummun, 2013, “Interest rate pass-through in the UK: Has the transmission mechanism changed during the financial crises”, Economic Issues, Vol. 18, Part I.

Bai, J. and Ng, S. 2004. “A PANIC Attack on Unit Roots and Cointegration”, Econometrica 72, 1127-1177.

Borio, C.E.V., Fritz, W., 1995. “The response of short-term bank lending rates to policy rates: a cross country perspective”. BIS Working Paper 27, Basel.

Chang, Y., 2003. “Nonlinear IV Unit Root Tests in Panels with Cross Sectional Dependency”, Journal of Econometrics, 110, 261–92.

Chmielewski, T., 2003. “Interest rate pass-through in the Polish banking sector and bank-specific financial disturbances”. In: Paper Presented at the ECB Workshop on Asset Prices and Monetary Policy, Frankfurt, 11–12 December.

Choi, I., 2001. “Unit Root Tests for Panel Data”, Journal of International Money and Finance 20 249-272.

Choi C. Y, L. Hu and M. Ogaki, 2004. “Robust Estimation of Structural Spurious Regressions and A Hausman-type Cointegration Test,” Working paper, Department of Economics, the Ohio University.

Coricelli, F.; Égert, B.; MacDonald, R. 2006. “Monetary transmission mechanism in Central and Eastern Europe: gliding on a wind of change”. Bank of Finland, BOFIT discussion papers, Vol. 8, p. 59.

Cottarelli, C., Kourelis, A., 1994. “Financial structure, bank lending rates, and the transmission mechanism of monetary policy”. IMF Staff Papers 41, 587–623.

Crespo-Cuaresma, J., Egert, B., Reininger, T., 2004. “The interest rate pass-through in new EU member states: the case of the Czech Republic, Hungary and Poland”, William Davidson institute, Working paper 671.

Crespo-Cuaresma, J., Egert, B., Reininger, T., 2007. “Interest rate pass-through in Central and Eastern Europe reborn from ashes merely to pass away?”, Journal of Policy Modeling 29, 209–225.

de Bondt, G. 2002. “Retail Bank Interest Rate Pass-Through: New Evidence at the Euro Area Level”. ECB Working Paper No. 136.

de Bondt, G., 2005. “Interest rate pass-through: empirical results for the Euro Area”. German Economic Review 6, 37–78.

De Graeve, F., De Jonghe, O. and Vander Vennet, R., 2004. “The Determinants of Pass-Through of Market Conditions to Bank Retail Interest Rates in Belgium”, Ghent University, Working Paper, 261.

De Graeve, F., De Jonghe, O., Vander Vennet, R., 2007. “Competition, transmission and bank pricing policies: evidence from Belgian loan and deposit markets”. Journal of Banking and Finance 31, 259–278.

Donnay, M., Degryse, H., 2001. “Bank lending rate pass-through and differences in the transmission of a single EMU monetary policy”, Center for Economic Studies Discussion Paper Series 01.17, Katholieke Universiteit Leuven.

Egert, B, Crespo-Cuaresma, J. and Reininger, T. 2007, Interest Rate Pass-Through in Central and Eastern Europe: Reborn from Ashes Merely to Pass-Away, Journal of Policy Modeling, Vol. 29, pp. 209-225.

Egert, B. and MacDonald, R. 2009, Monetary Transmission Mechanism in Central and Eastern Europe: Surveying the Surveyable, Journal of Economic Survey, Vol. 23, No. 2, pp. 277-327.

Enders, W. and Siklos, P.L. 2001. Cointegration and threshold adjustment. Journal of Business and Economic Statistics, 19, 166-176.

Gambacorta, L., 2008. “How do banks set interest rates?” European Economic Review 52, 792–819.

Horvath, C., Kreko, J., Naszodi, A., 2004. “Interest rate pass-through: the case of Hungary”. MNB Working Paper 2004/8, Budapest.

Im, K-S., Pesaran M. H and Shin Y., 1997. “Testing for Unit Roots in Heterogeneous Panels.” WP 9526, DAE, University of Cambridge. Revised June 1997.

Kot, A. 2004. “Does Banking Sector Competitiveness Affect Interest Rate Pass-Through?”. Paper presented at the 3rd Macroeconomic Policy Research Workshop, 29-30 October, National Bank of Hungary.

Lapinskas, V. 2011. “Interest rate pass-through in Lithuania”, EkonomikaVol. 90(2), 28-46

Levin, A, C-F. Lin, C-Sh. J. Chue, 1993. “Unit Root Tests in Panel Data: Asymptotic and Finite Sample Properties”, Journal of Econometrics 108, 1-24

Maddala, G.S and S. Wu, 1999. “A Comparative Study of Panel Data Unit Root Tests and a Simplified Test.” Oxford Bulletin of Economics and Statistics, Special Issue, 0305-9049.

Mojon, B. 2000. “Financial structure and the interest rate channel of ECB monetary policy”. ECB Working Paper, No. 40.

Moon, H. R. and B. Perron, 2004. “Testing for a Unit Root in Panels with Dynamic Factors”. Journal of Econometrics 122, 81–126.

Mueller-Spahn, S. 2005. The Pass-Through From Market Interest Rates to Retail Bank Rates in Germany, Centre for Finance and Credit Markets, Nottingham, Working Paper 08/05

Pedroni, Peter. 1997. “Fully modified OLS for heterogeneous cointegrated panels and the case of purchasing power parity.” http://web.williams.edu/Economics/wp/pedroniaie.pdf.

Pedroni, Peter. 2004. “Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis.” Econometric Theory, 20: 597-625.

Pesaran, Hashem, Yongcheol Shin, and Ronald Smith. 1998. “Pooled mean group estimation of dynamic heterogeneous panels.” ESE Discussion Paper 16.

Petrevski, G., Bogoev, J., 2012. “Interest rate pass-through in South East Europe: An empirical analysis,” Economic Systems 36.

Rotemberg, J. and Saloner, G. 1987, The Relative Rigidity of Monopoly Pricing, American Economic Review, Vol. 77, pp. 917-926.

Sander, H., Kleimeier, S., 2004a. “Convergence in euro-zone retail banking? What interest rate pass-through tells us about monetary policy transmission, competition and integration”. University of Maastricht METEOR Research Memoranda 045.

Sander, H., Kleimeier, S., 2004b. “Interest rate pass-through in an enlarged Europe: The role of banking market structure for monetary policy transmission in transition economies”. Journal of International Money and Finance 23, 461–492.

Sorensen, C.K. and T. Werner, 2006, “Bank Interest Rate Pass-Through in the Euro Area: A Cross Country Comparison “, ECB Working Paper, 580.

Stiglitz, J., Weiss, A., 1981. “Credit rationing in markets with imperfect information”. American Economic Review 71, 393–410.

Saborowski, C., Weber S., 2013. “Assessing the Determinants of Interest Rate Transmission Through Conditional Impulse Response Functions”, IMF Working Paper, WP/13/23.

Stanislawska , E., 2014. “Interest rate pass-through in Poland. Evidence from individual bank data”, NBP Working Paper No.179.

Toolsema, L., Sturm J-E., de Haan, J. 2001.Convergence of monetary transmission in EMU: new evidence. CESifo Working Paper 465, Munich

Weth, M.A. 2002. “The Pass-Through from Market Interest Rates to Bank Lending Rates in Germany”. Deutsche Bundesbank Discussion Paper No. 11.

Wrobel, E., Pawlowska, M., 2002. “Monetary transmission in Poland: some evidence on interest rate and credit channels”. National Bank of Poland, Materials and Studies Paper 24, Warsaw.

Yildirim, D. 2012. Interest rate pass-through to Turkish lending rates: A threshold cointegration analysis, ERC Working Papers in Economics 12/07.

A DSGE MODEL WITH FINANCIAL DOLLARIZATION – THE CASE OF SERBIA

Mirko Đukić, Tibor Hlédik, Jiří Polanský, Ljubica Trajčev, Jan Vlček

Available at www.nbs.rs

The views expressed in the papers constituting this series are those of the author, and do not necessarily represent the official view of the National Bank of Serbia.

Economic Research and Statistics Department

THE NATIONAL BANK OF SERBIA

Belgrade, 12 Kralja Petra Street,

Telephone: (+381 11) 3027 100

Belgrade, 17 Nemanjina Street,

Telephone: (+381 11) 333 8000

www.nbs.rs

A DSGE Model with Financial Dollarization – the Case of Serbia

Mirko Đukić, Tibor Hlédik, Jiří Polanský, Ljubica Trajčev, and Jan Vlček[5]

Abstract: We modify a DSGE model of a small open economy by adding financial euroization in order to capture the main channels of the monetary transmission mechanism in matching with the Serbian data. In contrast to the standard DSGE workhorse, the model encompasses commercial banks and foreign exchange-denominated deposits and loans. Given these features, the model is well-suited to evaluating effects of the nominal exchange rate on the financial wealth and consumption of households. The model structure, including optimization problems and first-order conditions, is provided in the paper. The model properties are tested to match the stylized facts of dollarized economies. Specifically, the model is calibrated to the Serbian data, and a model-consistent multivariate filter is used to identify unobserved trends and gaps.

Key words: DSGE model, financial dollarization.
JEL Code: E44, F41, F47.

[5] At the time of writing this paper, Tibor Hlédik, Jiří Polanský, and Jan Vlček were working at the Czech National Bank, and Mirko Đukić and Ljubica Trajčev at the National Bank of Serbia.

Non-Technical Summary

This paper presents the dynamic stochastic general equilibrium (DSGE) model for Serbia, developed as a result of a technical cooperation project between the Czech National Bank and the National Bank of Serbia (NBS). The model reflects the stylized facts and macroeconomic dynamics of a small open economy experiencing partial euroization.

As such, the paper responds to the growing interest in macro-financial linkages in emerging and developing market countries with financial euroization, such as Serbia. These countries modernize their policy frameworks by either moving towards inflation targeting (IT) or allowing for higher nominal exchange rate flexibility. However, due to financial euroization, the common DSGE model structure and transmission channels are not sufficient to capture the effects of exchange rate dynamics on the financial wealth of households, not least because these models omit the financial sector, banks in particular.

Fot that reason, we extend the existing DSGE workhorse model by including financial sector variables and introduce explicit banking at relatively low costs. Rather, it should be considered an attempt to build up a model which is rich enough to analyze the macroeconomic effects of euroization in Serbia. To introduce financial frictions, the model assumes two types of households – net borrowers and net lenders. Net borrowers have to finance a part of their expenditures using loans, borrowing against their wage income. Commercial banks play the role of financial intermediaries, providing loans to households on the asset side and collecting deposits on the liability side. Financial euroization is represented by foreign exchange (FX) denominated deposits and loans on commercial bank balance sheets. The share of FX loans and deposits is assumed to stay constant, given the insurance motivation of financial euroization in Serbia.

Banks are subject to regulatory requirements in the model. The requirements consist of capital requirements, approximated by the loan-to-deposit ratio. A penalty is applied to banks whenever they deviate from the loan-to-deposit ratio set by the central bank. Apart from the financial block, the structure of the model is consistent with the common DSGE workhorse for a small open economy. The paper provides a detailed description of the optimization problems of economic agents and the corresponding first-order conditions.

The model is calibrated to match the actual data of Serbia. Serbia is characterized by euroization at about 60 percent of total deposits and loans. Unlike banks, households and firms are not hedged against currency risks. As their revenues are mainly in dinars and most of their liabilities are in FX-indexed instruments, their financial position and wealth are affected by exchange rate dynamics.

The model properties are presented in the paper and discussed with regard to the stylized facts about Serbia. Specifically, an unexpected depreciation, an increase of the country risk premium, and a hike of the central bank rate are reported and described using the model transmission mechanism. The paper concludes with filtration of historical data for Serbia. As the core of the model is derived in stationary form, trends are exogenous to the model. This allows us to match the observed data without de-trending them. Note however that the model is not designed to be a forecasting model to replace the existing QPM at the National Bank of Serbia (NBS).

Contents

1 Introduction 38

1.1 General Overview of the Relevant Literature 39

2 Stylized Facts about the Serbian Economy – Support for the Implementation of the Model 41

3 Structure of the Model 45

3.1 Model Overview 45
3.2 Households 46
3.2.1 Net Savers 47
3.2.2 Net Borrowers 48
3.3 Production 50
3.3.1 Intermediate Producers 50
3.3.2 Final Goods Producers and Sticky Price Setting 51
3.4 Banking Sector 52
3.4.1 Wholesale banks 52
3.4.2 Retail Banks 55
3.5 Central Bank 56
3.6 Government 56
3.7 Rest of the Model 57
3.7.1 Exports 57
3.7.2 Net Foreign Asset Accumulation 58
3.7.3 Risk Premium 58

4 Model Properties 59

4.1 Calibration and Sensitivity Analysis 59
4.2 Impulse Responses 60
4.2.1 Unexpected Depreciation 60
4.2.2 Risk Premium Shock 62
4.2.3 Increase in Policy Rate 63
4.2.4 Inflation (Cost-push) Shock 64
4.3 Putting the Model to the Serbian Data 65
4.3.1 Filtering the Data 65
4.3.2 Historical Simulations 66

5 Conclusion 67

Appendix A: Calibrated Values of Parameters 69

References 71

1 Introduction

Although macroeconomic stabilisation has prompted a declining trend in financial euroisation in Serbia in recent years, it can still be characterized as relatively high. At end-2019 around two thirds of loans and deposits were FX-indexed. This has strong implications on transmission mechanism of the monetary policy, as well as on the impact of various shocks on the economy as a whole, which is very important for the monetary policy makers in the inflation targeting regime, pursued by the NBS since 2008. That was our main motivation to begin developing a model that would incorporate all these facts.

The main forecasting tool in the IT process is the QPM, a New Keynesian gap type of model, described in Đukić et al. (2011). Since the introduction of the IT regime, the model has been used as a key analytical tool for generating inflation forecasts and quantifying appropriate monetary policy responses. The QPM has proved to be a useful tool for producing and explaining inflation forecasts to the general public. Although the QPM has shown good forecasting performance over the past years, it lacks some important features of the Serbian economy, especially the ones related to monetary policy transmission in a euroized economy and the channels through which the domestic economy is affected by the ECB's monetary policy decisions.

The importance of dynamic stochastic general equilibrium (DSGE) models in central banks has been growing in recent years. In fact, most central banks in both developed and emerging market countries have introduced DSGE models as a core forecasting tool within their FPAS[6], while some of them are still in the implementation process. The use of such models for monetary policy (MP) purposes is mainly a response to the need for deeper and internally consistent macroeconomic analysis in line with recent advancements in the macroeconomic forecasting field (Christiano et al., 2005; Smets and Wouters, 2003, 2007). Although central bank staff can generate efficient forecasts with QPM-style macroeconomic models[7], a full-fledged DSGE framework offers mechanisms for answering more detailed policy questions, identifying initial conditions, and analyzing structural shocks due to its theoretical foundations and consistent structure (national accounts, relative prices, etc.)[8]. Following the experience of other central banks, the NBS initiated a project, realized with technical assistance from the Czech National Bank, to introduce a DSGE model for the Serbian economy.

DSGE models have strong theoretical micro foundations. This is generally viewed as their main advantage over reduced-form gap models. They are derived based on the assumption that representative households, firms, and banks optimize their behavior in order to maximize utility and profits.

In this paper we present the structure of the DSGE model for the Serbian economy. It is a medium-scale DSGE model based on the framework of Roger and Vlcek (2011). The model

[6] Forecasting and policy analysis system (Berg et al., 2006)

[7] Quarterly Projection Model (Berg et al., 2006)

[8] See, for example, the discussion in Bruha et al. (2013).

embodies the main principles under which the IT regime works, i.e., a monetary policy rule keeping inflation on target. Furthermore, it contains several standard features of New Keynesian DSGE models (e.g. rational expectations and price rigidity) and it also incorporates some ”non-standard” features important for emerging market economies, such as euroization. Euroization is accomplished through the introduction of an explicit banking sector into the model.

Some of the modeling choices (constant steady-state shares of financial euroization, exogenous exports, etc.) reflect the preference of the authors to keep the model tractable. Tractability was a high priority in order to be able to bring the model to the data and operationalize policy work at a later stage. The paper is therefore expected to inspire economists working at policy-making institutions in countries with a high level of euroization (dollarization) and searching for tractable policy analysis tools. Adding extra features to a DSGE model usually dramatically increases its size. Given that the paper tries to find a balance between including new features in the model and ensuring tractability, it focuses primarily on euroization and financial intermediation. As a result, we do not introduce capital and investment into our model and we capture fiscal policy in a very stylized form.

Regarding its role in policy making at the NBS, the DSGE model is to be used mainly as a policy analysis tool rather than for forecasting, which is still done by the QPM model due to its satisfactory performance. The main objective of developing the DSGE model is to contribute to the decision-making process in the NBS by deepening its analyses of the monetary policy transmission mechanism and the financial system. The new model enables us to analyze new features of the transmission mechanism not captured by the QPM model and provides a more sound framework from the theoretical perspective.

The model is calibrated taking into account recent data and specific features of the Serbian economy. This is a natural choice since the Serbian data are characterized by many underlying problems, such as a lack of long time series and frequent structural changes.

The paper is structured as follows. After presenting the literature survey, which covers DSGE models, we turn to presenting the stylized facts about the Serbian economy in Section 2. This is followed by in-depth analysis of the model structure in Section 3, which discusses agents’ optimization problems and the first-order conditions. Parameter identification and the model’s features as given by its impulse response functions are described in Section 4. Section 5 concludes.

1.1 General Overview of the Relevant Literature

From the 1970s onwards, many researchers strove to enhance the features of the neo-Keynesian models which were prevalent at that time. Research focused on the development of new macroeconomic models derived from microeconomic foundations and based on rational expectations to address the Lucas critique. As a result of those endeavors, DSGE models appeared to be able to successfully account for most of the research demands at the time.

Having in mind that Serbia is a highly euroized economy, we put a special emphasis on the literature covering this topic. Euroization/dollarization can be official (de jure), when foreign currency is legally used in parallel to, or instead of, the domestic currency, or unofficial (de facto), i.e. the partial use of foreign currency by domestic agents when it is not a legal tender. There are three main types of unofficial dollarization recognized in the literature: transaction dollarization (or currency substitution, when foreign currency is used as a medium of exchange), price dollarization (when prices are indexed to changes in the exchange rate), and financial dollarization (when foreign currency is used as a store of value).9

Here, we deal with financial dollarization, which means that there is a non-zero share of foreign currency assets and liabilities in the financial system. The term dollarized stems from the fact that most of the countries dealing with this issue come from Latin America, where the dollar is used as a means of payment and a store of value in parallel with the national currency. Today, this problem is also recognized in Central and Eastern European countries, where foreign-currency, mostly euro, denominated assets and liabilities make up a significant part of residents' financial portfolios. So, we can talk about euroization in the case of Serbia, and we use euroization and dollarization interchangeably in the paper.

Yeyati (2006) gives a good overview of the implications of financial dollarization for the transmission mechanism and monetary policy. He explains that models that try to capture financial dollarization usually do this through three different channels: 1) the portfolio view, which sees financial dollarization as the optimal choice for a given distribution of real returns in each currency in order to minimize the variance portfolio returns; 2) the market failure view, where financial dollarization is seen as a response to market imperfections, meaning that the default risk of a borrower indebted in the local currency at a high interest rate exceeds the risk of borrowing in foreign currency, as the local currency collapse risk is low; 3) the institutional view, which emphasizes how institutional failures can foster financial dollarization, i.e. if the government does not commit to low inflation, the expectations that the government will inflate away the debt burden lead to high domestic interest rates and inflation bias. Debt dollarization can therefore be seen as an option for averting inflation bias. Yeyati (2006) shows that financially dollarized economies tend to face a higher inflation rate, a greater propensity to suffer systemic banking crises, and slower and more volatile output growth.

The recent financial crises brought to the fore the importance of financial frictions and their impact on the real economy. Since then, numerous research papers have striven to incorporate them into macroeconomic models. One can generally distinguish three approaches: 1) the financial accelerator mechanism, which is based on information asymmetry between borrowers and lenders; 2) models with collateral constraints in which loans are secured by borrowers' disposable collateral; 3) the introduction of banks explicitly into the model, usually with an assumption that banks face operating costs or regulatory requirements.

Ize and Parrado (2006) investigate the causes of real and financial dollarization in the context of a small open economy model. They find that financial dollarization rises in the presence of volatility of domestic inflation and recedes during episodes of exchange rate volatility. Another example of this strand of the literature is Rennhack and Nozaki (2006), who

[9] Castillo et al. (2006)

find that financial dollarization is a rational response to increased macroeconomic uncertainty related to high inflation and exchange rate variability.

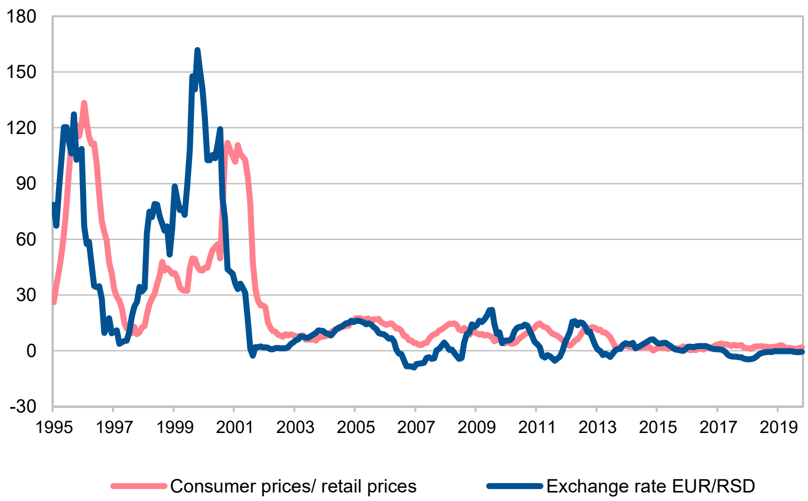
Extending models developed by Goodhart, Osorio and Tsomocos (2009) and Martinez and Tsomocos (2012), Urošević and Grga (2014) developed a DSGE model for Serbia, incorporating financial dollarization and foreign ownership of banks. In their paper they emphasize strong connection between currency risk and default risk. Depreciation of the local currency increases default risks, for which (in the model) economic agents pay penalty, which hurts economic activity. Unlike our model based on IT regime, Urošević and Grga modeled the central bank as a partial liquidity provider whose actions don't depend on the state of the system (there is no active monetary policy). The authors make a conclusion that, due to the large share of FX deposits and loans, interest rate channel in Serbia is weak, while the FX channel is strong.

The DSGE model of the National Bank of Romania, Copaciu et al. (2016), introduces euroization through two types of firms, depending on the currency at which they borrow. As a large share of firms is exposed to the FX risk, extending the model in this way introduces balace sheet effects, which results in contractionary effects of the depreciation of the local currency, through lower investments.

2 Stylized Facts about the Serbian Economy – Support for the Implementation of the Model

The high level of euroization in Serbia is a persistent problem whose sources date decades back. After recessionary and inflationary 1980s, the crises culminated at the beginning of the 1990s with war in the former Yugoslavia and economic sanctions against Serbia. Economic activity collapsed during this period, while fiscal dominance over monetary policy resulted in one of the highest rates of hyperinflation ever recorded, peaking in late 1993 and early 1994. Simultaneously, the banking system collapsed and, as a consequence, households' savings

Figure 1. **Inflation and Exchange Rate**
(y-o-y, in %)



Source: NBS.

disappeared. This led to a protracted crisis of confidence in the banking system, with savings being kept ”under the mattresses” rather than in banks. During this period, the Deutsche mark replaced the dinar almost completely, taking over the functions of money.

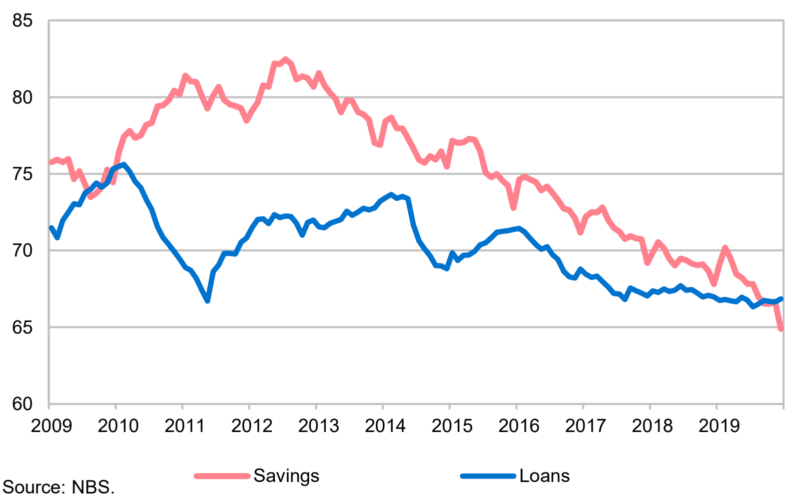
After a monetary reform in 1994, inflation was temporarily brought down into single digits, but soon thereafter the fiscal dominance again resulted in a surge in inflation, sometimes exceeding even 100%. Although confidence in the banking system was restored during the 2000s (household savings exceded EUR 11 billion in 2019, total for domestic and foreign currencies) and inflation was brought down to moderate levels (10% on average till 2013, and around 2% on average since 2014), euroization remained very high.

Generally, there are three types of euroization: transaction, price and financial euroization. The former two are basically no longer present in Serbia. Almost all transactions are made in local currency and, although the pass-through is still relatively high (0.15 in the short run), most prices are not directly FX-indexed (with few exceptions, such as prices of real estate, cars, and package holidays).

However, despite the decling trend, financial euroization is still very high. In 2019, 78% of household deposits and 65% of firms and households’ deposits in foreign currencies (mostly euro) despite a high spread between dinar and euro interest rates. An additional source of euroization is foreign ownership of commercial banks, which draw FX funds from abroad (typically by borrowing from their parent banks). Having liabilities in FX (FX savings and FX borrowing), banks extend most of their loans (67% to firms and households) in FX-indexed instruments. The euroization ratios (the share of FX savings and the share of FX loans) have been stable over the past decade (Figure 2).

Note the declining trend in euroization in recent years (Figure 2), which, not by chance, coincides with the reduction in the inflation rate to a low level, which confirms the thesis that the main cause of the euroization in Serbia is insurance against macroeconomic instability.

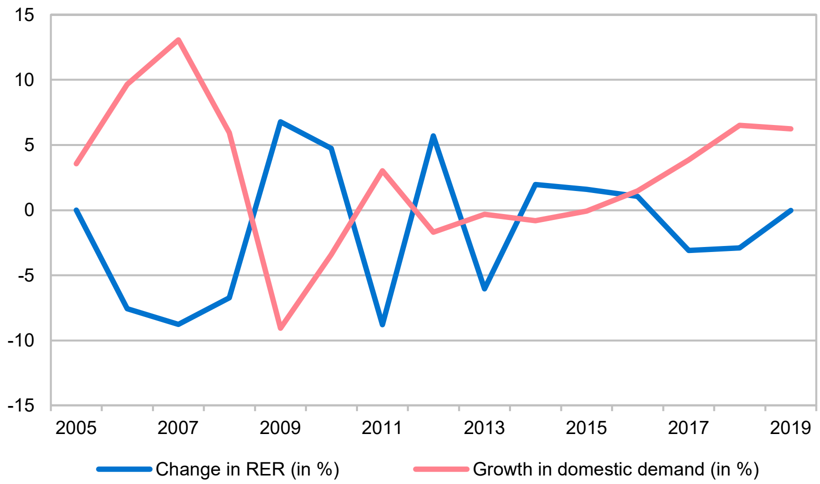
Figure 2. **Share of Indexed and FX Receivables and Payables in Total Bank Receivables and Payables from Households and Enterprises**
(in %)



Source: NBS.

While similar FX-shares on the asset and the liability side make banks hedged against currency risk, that is not the case with firms and households. As most of their revenues are in dinars and most of their liabilities to banks are in FX-indexed terms, their financial position is affected by exchange rate fluctuations. This means that depreciation of the dinar has a negative effect on domestic demand, as FX-indebted firms, households, as well as government, face higher installments (expressed in local currency), leaving them with less resources for consuming and investing. Figure 3 illustrates the negative correlation between real exchange rate depreciation and domestic absorption in Serbia. When the negative effect of a depreciation on domestic demand is stronger than its positive effect on net exports, economic activity will fall as a result. Such depreciation also raises the ratio of foreign (FX) debt to GDP (61.5% in 2019).

Figure 3. **Real Exchange Rate Depreciation and Domestic Absorption**
(annual change, in %)

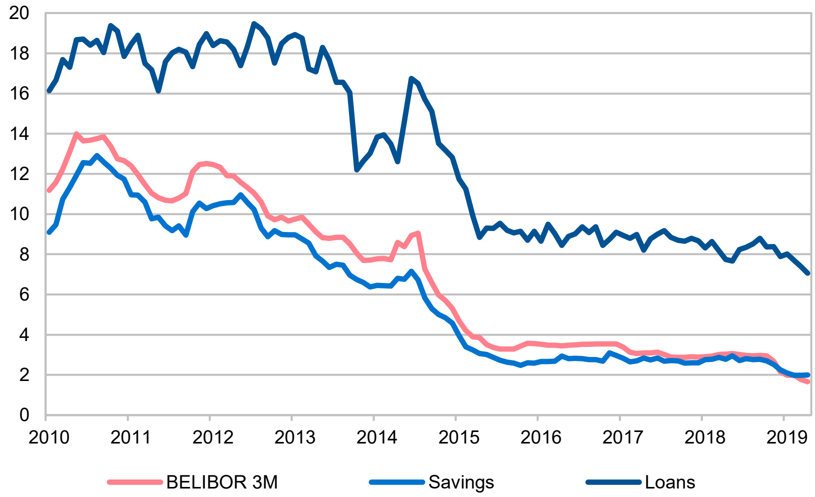


Source: NBS ans SORS.

When conducting monetary and exchange rate policy, the central bank therefore has to take into account its effects on financial stability along with the primary objective of price stability.

Another obvious challenge for monetary policy in a euroized economy is a weak interest rate channel. Although there is a strong pass-through from the NBS policy rate to RSD interest rates, as shown in Figure 4, interest rates on FX loans (70%) and deposits (73%) are not affected by the NBS's policy rate. These interest rates are driven by monetary conditions in the eurozone and the country risk premium of Serbia (which also depends on global risk aversion).

Figure 4. **Interest Rates on RSD Loans and Deposits**
(in %)



Source: NBS.

Reducing the level of euroization was one of the reasons why the NBS introduced inflation targeting (IT) at the end of 2008. The inflation target was initially set at 10±2% with a gradual linear decline to 4±1.5% till end-2012. In 2017 it was reduced to 3±1.5%, where it currently stands. The main instrument of monetary policy is *de facto* the one-week repo rate, which propagates through various transmission channels, the exchange rate channel being the strongest and fastest.

As a large part of the economy is euroized, the required reserves ratio (RRR) is also used as a monetary policy instrument, mainly in order to affect FX and FX-indexed loans. As part of the de-euroization strategy, FX liabilities are subject to a much higher RRR (13–20% depending on maturity) than dinar liabilities (0–5%). This instrument has also a financial stability purpose, as it makes savings safer, though at the cost of higher interest rates on loans. In addition, banks are required to hold a certain proportion of FX required reserves in dinars (currently 30–38% depending on the maturity of the funding sources). The main rationale for using this instrument is twofold. First, it affects exchange rate dynamics through its cash-flow effects. Second, it also influences interest rates on FX loans through different opportunity costs of holding reserves in dinars versus euros.

Finally, the NBS uses direct interventions in the foreign exchange market to smooth the volatility of the exchange rate. As such, the FX policy of the NBS is defined as a managed float regime. The interventions, however, are used only to reduce the short-term volatility of the exchange rate, not to influence the FX trend.

Following the introduction of the IT regime, the main challenges for monetary policy were high volatility of food prices and depreciation pressures arising mainly as a consequence of the global financial crisis. Because of food-price and risk-premium shocks, inflation in this period was very volatile and, on average, above the target. Only since late 2013 has the NBS managed to stabilize inflation at a relatively low level, around 2%.

Serbia is a small open economy exposed to external shocks. In the pre-crisis period, growth was driven mostly by rising domestic demand, leading to unsustainable external imbalances.

After the collapse of Lehman Brothers, slowdown in the domestic demand, depreciation of the dinar, as well as rising investments in the exporting sector, led to a significant reduction in external imbalances. Already in 2009 the current account deficit fell to less than a third of the 2008 level and has mostly moved within the 6-9% of GDP range ever since. In 2019, exports were 144% and imports 53% above the pre-crisis level. Despite that, the ratio of imports to GDP (61%) remained higher than that of exports (51%).

3 Structure of the Model

3.1 Model Overview

The model is based on the framework of Roger and Vlček (2011). It has been extended to incorporate several features which are not embedded in the original model. These mainly include euroization and a set of equations for model-consistent filtering of gaps. The original model is developed to match the stylized facts of emerging market economies. Besides several widely used features of DSGE models (e.g. real and nominal rigidities), it contains an explicit banking sector allowing the credit channel and the share of dollarization in the economy to be captured.

The model contains two types of households (savers and borrowers),[10] intermediate goods-producing firms, final goods-producing firms, a labor bundle, exporters, retail banks, wholesale banks, and monetary and fiscal authorities.[11] Households consume final consumption goods, save deposits at, or take loans from, commercial banks, and supply labor. Intermediate firms use labor and imports to produce intermediate goods. An assumption that firms finance a constant share of their production through commercial banks' loans is used to motivate demand for loans from firms. Monopolistically competitive retailers use intermediate goods to produce final goods, facing a Calvo signal to change their prices. Final goods are consumed by households and the government. Exporters are assumed to be independent of the domestic intermediate sector and face exogenous terms of trade.

Commercial banks collect deposits from households and borrow from abroad on the liability side while extending loans on the asset side. Deposits and foreign liabilities are assumed to be perfect substitutes. Banks also have to maintain a minimum loan-to-deposit ratio.[12] If they deviate below this ratio, they face penalty charges. The monetary authority targets year-on-year inflation four periods ahead via an interest rate rule. The government finances its spending by issuing government bonds and collecting lump-sum taxes. The ratio of nominal government spending to nominal private consumption is assumed to be constant in the long run.

---

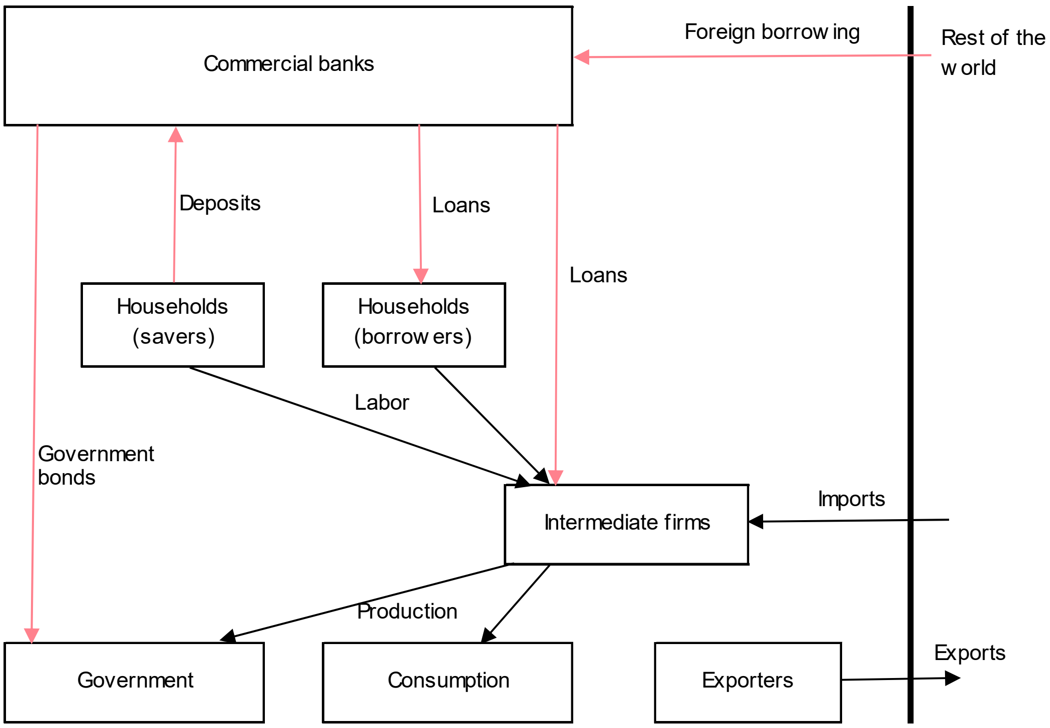
[10] Similar to the Gerali et al. (2010) framework with patient and impatient households.

[11] Final goods producers, wholesale banks, and the labor bundle are introduced for technical reasons to simplify the optimization problem.

[12] One can think of this ratio as a regulatory requirement set implicitly by the policy authority by imposing restrictions on the capital adequacy ratio.

The model captures the interaction between real and nominal financial variables in a consistent way. The structure of the model is presented in Figure 5.

Figure 5 **Structure of the Model**



Note: Black lines - real flows, red lines - financial flows.

3.2 Households

Households in the model are divided into (net) savers and (net) borrowers. This distinction is particularly important in dollarized economies, as the wealth effects of nominal depreciation of the local currency are positive for savers and negative for borrowers. Furthermore, the reaction of borrowers to a depreciation is typically more pronounced, because they immediately suffer from higher monthly interest payments, which leaves their disposable income lower. FX savers are expected to react more slowly to their higher income (expressed in domestic currency). We address this fact by calibrating different habit coefficients for the two groups (which will be described in Section 4.1).

Households do not optimize with respect to the level of euroization, as it represents insurance against macroeconomic volatility rather than search for yield.

There is a continuum of households indexed on a unit interval i. Households maximize their utility function separable into two arguments (consumption and leisure). The utility function contains external habit formation Abel (1990) and does not contain real money balances (due to full endogeneity of money). The maximization problem has the form:

$$\max_{C^h_{t(i)},N^h_{t(i)}} E_t \sum_{t=0}^{\infty} \beta^t [(1-\chi^h) \log (C^h_t(i) - \chi^h \bar{C}^h_{t-1}) - N^h_t(i)], \quad (1)$$

where $h = \{s, b\}$ in the superscript refers to net savers or net borrowers. C^h denotes (real) consumption, $\bar{C}^h_{t-1}$ is the past aggregate consumption of each group of households, N^h is the labor effort (hours worked), β is the discount factor, and χ^h is the habit persistence smoothing parameter. The share of net saver households is γ^s.

3.2.1 Net Savers

Net saver households consume final consumption goods, supply labor, and save deposits denominated in domestic and foreign currency at commercial banks. Their income comprises labor income, revenues from deposits, lump-sum net payments from the government, and dividends from ownership of firms and banks.

The previously mentioned maximization problem is set given the budget constraint:

$$P_t C^s_t(i) + D^d_t(i) + S_t D^f_t(i) \leq W^s_t N^s_t(i)$$
$$+(1+i^{dd}_{t-1}) D^d_{t-1}(i) + \left(1+i^{df}_{t-1}\right) S_t D^f_{t-1}(i) + \gamma^s \Gamma_t + \Pi_t, \quad (2)$$

where P_t is the price level,[13] D^d_t are deposits in the domestic currency, D^f_t are deposits in the foreign currency, S_t denotes the nominal exchange rate, W^s_t is the nominal wage, i^{dd}_t is the nominal interest rate on deposits in the domestic currency, i^{df}_t is the nominal interest rate on deposits in the foreign currency, $\gamma^s \Gamma_t$ denotes lump-sum net payments from the government that go to net savers, and Π_t are dividends (profits).

Financial dollarization is introduced through the assumption that households can save deposits denominated in both local and foreign currencies at commercial banks. The share of foreign currency deposits in total deposits is exogenous and equal to parameter $\lambda \in [0,1]$, where a higher parameter value implies a higher share of dollarization. Thus, the use of a constant nominal share of foreign and domestic currency deposits implies a slightly modified optimization problem of households

$$\max_{C^s_t(i),D_t(i),N_t(i)} E_t \sum_{t=0}^{\infty} \beta^t \{[(1-\chi^s) \log (C^s_t(i) - \chi^s \bar{C}^s_{t-1}) - N^s_t(i)]$$
$$-\Lambda^s_t [P_t C^s_t(i) + D_t(i) - W^s_t N^s_t(i) - (1+i^{dd}_{t-1})(1-\lambda) D_{t-1}(i)$$
$$-\left(1+i^{df}_{t-1}\right) \dot{S}_t \lambda D_{t-1}(i) - \gamma^s \Gamma_t - \Pi_t]\}, \quad (3)$$

[13] For simplicity, we do not distinguish between the price level and the consumption price deflator.

where D_{t-1} denotes total deposits and $\dot{S}_t = \frac{S_t}{S_{t-1}}$ is the gross rate of nominal exchange rate change, $\dot{S}_t > 1$ refers to depreciation and $\dot{S}_t < 1$ to appreciation. Λ_t^s denotes the shadow price of wealth.

Assuming flexible wages, the solution to the above-mentioned optimization problem results in the following equations:[14]

$$\frac{1-\chi^s}{C_t^s - \chi^s C_{t-1}^s} = \Lambda_t^s P_t, \tag{4}$$

$$\Lambda_t^s = \beta \Lambda_{t+1}^s \left[(1-\lambda)(1+i_t^{dd}) + \lambda(1+i_t^{df})\dot{S}_{t+1}\right], \tag{5}$$

$$\Lambda_t^s W_t^s = 1. \tag{6}$$

The first- order condition with respect to labor (6) can be substituted into the other two to eliminate the shadow price of wealth Λ_t^s. The optimality conditions have the form

$$\frac{C_t^s - \chi^s C_{t-1}^s}{1-\chi^s} = \frac{W_t^s}{P_t}, \tag{7}$$

$$\frac{C_{t+1}^s - \chi^s C_t^s}{C_t^s - \chi^s C_{t-1}^s} = \beta\left[(1-\lambda)(1+i_t^{dd}) + \lambda(1+i_t^{df})\dot{S}_{t+1}\right]\frac{1}{\dot{P}_{t+1}}, \tag{8}$$

where $\dot{P}_{t+1} = \frac{P_{t+1}}{P_t}$ is the expected gross inflation rate. The first equation equalizes the real wage and the marginal rate of substitution between consumption and leisure. The second equation is a modified version of the Euler equation balancing current and future discounted marginal utilities from consumption. Within the model, the two interest rate yields, i.e. $(1 + i^{dd})$ and $(1 + i^{df})\dot{S}$, are equal in the steady state, but can deviate from each other along business cycles.

3.2.2 Net Borrowers

On the other side there is a fraction of households which are net borrowers. Their objective is similar to that of savers, but with the difference that they can boost their consumption by borrowing from commercial banks. Thus, their budget constraint is:

$$\begin{gathered} P_t C_t^b(i) + (1+i_{t-1}^{ld})L_{t-1}^{hd}(i) + \left(1+i_{t-1}^{lf}\right)S_t L_{t-1}^{hf}(i) \le W_t^b N_t^b(i) + L_t^{hd}(i) \\ +S_t L_t^{hf}(i) + (1-\gamma^s)\Gamma_t, \end{gathered} \tag{9}$$

where L_t^{hd} and L_t^{hf} refer to loans to households granted by commercial banks in domestic and foreign currency, respectively. The share of foreign currency loans in total loans (L_t^h) is exogenous and equal to parameter λ, the same as for deposits held by savers. Every period, borrowers repay all their loans at the previously negotiated nominal rate for domestic (i_{t-1}^{ld})

[14] Symetric equilibrium is assumed after deriving the first-order conditions, e.g. $\bar{C}_t^s = C_t^s(i) = C_t^s$. Thus, none of the optimality equations in the paper contains indices for the continuum of households, firms, or banks, or "bars" for aggregate variables.

and foreign (i_{t-1}^{lf}) currency loans and take out new ones. Thus, their income comprises labor income, new loans, and lump-sum transfers.

We introduced a borrowing constraint similar to the debt-to-income ratio which is approximated by

$$L_t^h \leq mW_t^b N_t^b. \tag{10}$$

This means that the loan amount that households borrow is proportional to their wage.[15] Substituting (10) in the budget constraint eliminates loans from the maximization problem, which now has the following form

$$\begin{aligned}\max_{C_t^b(i),N_t^b(i)} E_t \sum_{t=0}^{\infty} \beta^t \{ & [(1-\chi^b)\log(C_t^b(i)-\chi^b\,\bar{C}_{t-1}^b)-N_t^b(i)] \\ & -\Lambda_t^b[P_t C_t^b(i)+(1+i_{t-1}^{ld})(1-\lambda)mW_{t-1}^b(i)N_{t-1}^b(i)+\left(1+i_{t-1}^{lf}\right)\dot{S}_t\lambda mW_{t-1}^b(i)N_{t-1}^b(i) \\ & -W_t^b N_t^b(i)(1+m)-(1-\gamma^s)\Gamma_t]\},\end{aligned} \tag{11}$$

where Λ_t^b stands for the borrowers' shadow price of wealth, and m for the fraction of the wage up to which loans can be extended.

Solving the maximization problem results in the following equations:

$$\frac{1-\chi^b}{C_t^b-\chi^b C_{t-1}^b}=\Lambda_t^b P_t, \tag{12}$$

$$\frac{\Lambda_t^b}{\Lambda_{t+1}^b}=\frac{1}{\Lambda_{t+1}^b W_t^b(1+m)}+\beta W_t^b m\left[(1-\lambda)(1+i_t^{ld})+\lambda\left(1+i_t^{lf}\right)\dot{S}_{t+1}\right]. \tag{13}$$

Equalizing the marginal rate of consumption and labor, we get a relationship between consumption, the real interest rate, and the real wage.

$$\begin{aligned}\frac{C_{t+1}^b-\chi^b C_t^b}{C_t^b-\chi^b C_{t-1}^b}= & \ \beta W_t^b m\left[(1-\lambda)(1+i_t^{ld})+\lambda\left(1+i_t^{lf}\right)\dot{S}_{t+1}\right]\frac{1}{\dot{P}_{t+1}} \\ & +\frac{1}{1+m}\frac{P_t}{W_t^b}\frac{C_{t+1}^b-\chi^b C_t^b}{1-\chi^b}\end{aligned} \tag{14}$$

From (14) we can see a positive relationship between future consumption and the loan-to-income ratio (m).

[15] In Iacoviello (2005) collateral constraints are tied to housing value, but Serbian banks have a regulatory rule that the monthly installment cannot exceed a certain fraction of the client's wage.

3.3 Production

3.3.1 Intermediate Producers

The competitive domestic intermediate goods producers combine households' labor and imported goods through Cobb-Douglas constant-returns-to-scale technology. They produce the intermediate goods Y_t according to the following production function:

$$Y_t = A_t M_t^{\alpha} N_t^{1-\alpha}, \tag{15}$$

where A_t is the technology shock, M_t denotes imported goods, $N_t$refers to labor input, and α is the share of imported goods in production. For simplicity, there is no capital in the model.[16] The total factor productivity shock is assumed to be persistent and defined via the exogenous process

$$A_t = \rho_\alpha A_{t-1} + \varepsilon_t^a. \tag{16}$$

Firms maximize profits given the prices of inputs, i.e. the nominal wage W_t and the foreign price P_t^* adjusted by the nominal exchange rate S_t. Also, they have to finance a constant fraction of their production via bank loans in either domestic or foreign currency. For simplicity, the share of dollarization is constant and equal to the share in deposits and loans in the households' problem (λ). The profit maximization has the form

$$\max_{N_t, M_t, L_t^{fd}, L_t^{ff}} \mathcal{L}_t = E_0 \sum_{t=0}^{\infty} \Xi_{0,t}[\, P_t^Y Y_t - W_t N_t - P_t^* S_t M_t + L_t^{fd} + L_t^{ff} S_t$$

$$-(1 + i_{t-1}^{ld}) L_{t-1}^{fd} - (1 + i_{t-1}^{lf}) L_{t-1}^{ff} S_t], \tag{17}$$

where P_t^Y is the price of intermediate goods, $W_t N_t$ and $P_t^* S_t M_t$ denote labor and imported goods costs,[17] L_t^{fd} are loans in domestic currency, L_t^{ff} are loans in foreign currency, i_t^{ld} and i_t^{lf} are the corresponding lending rates, and $\Xi_{t,s} = \beta^{s-t} \frac{\Lambda_s}{\Lambda_t}$ is the nominal pricing kernel (Anderle et al., 2009), as firms are owned by savers.

Assuming a loans-in-advance constraint

$$L_t^{fd} + L_t^{ff} S_t = L_t^f \ll \kappa P_t^Y Y_t, \tag{18}$$

where κ is the fraction of nominal production required to be financed through bank loans and L_t^f denotes the total volume of loans to firms, the optimization problem becomes

[16] For empirical work we replace consumption with the sum of consumption and investement in order to avoid excluding investment from our analysis.

[17] Intermediate firms use aggregate labor, $N_t = \gamma^s N_t^s + (1 - \gamma^s) N_t^b$, and $W_t$is the weighted average of these two ($W_t = \gamma^s W_t^s + (1 - \gamma^s) W_t^b$).

$$\max_{N_t,M_t} \mathcal{L}_t = E_0 \sum_{t=0}^{\infty} \Xi_{0,t} [P_t^Y A_t M_t^{\alpha} N_t^{1-\alpha} - W_t N_t - P_t^* S_t M_t + \kappa P_t^Y A_t M_t^{\alpha} N_t^{1-\alpha}$$

$$-(1+i_{t-1}^{ld})(1-\lambda)\kappa P_{t-1}^Y A_{t-1} M_{t-1}^{\alpha} N_{t-1}^{1-\alpha} - (1+i_{t-1}^{lf})\lambda \kappa P_{t-1}^Y A_{t-1} M_{t-1}^{\alpha} N_{t-1}^{1-\alpha} \dot{S}_t]. \tag{19}$$

The first-order conditions for optimal input demands are

$$W_t N_t = (1-\alpha) P_t^Y Y_t \left[1 + \kappa - \kappa \Xi_{t,t+1} \left[(1+i_t^{ld})(1-\lambda) + \left(1+i_t^{lf}\right)\lambda \dot{S}_{t+1}\right]\right], \tag{20}$$

and

$$P_t^* S_t M_t = \alpha P_t^Y Y_t \left[1 + \kappa - \kappa \Xi_{t,t+1} \left[(1+i_t^{ld})(1-\lambda) + \left(1+i_t^{lf}\right)\lambda \dot{S}_{t+1}\right]\right]. \tag{21}$$

3.3.2 Final Goods Producers and Sticky Price Setting

Final goods producers, or retailers, set their prices according to Calvo contracts, so their prices are sticky.[18] A continuum j of monopolistically competitive final goods-producing firms uses domestic intermediate goods to produce final goods, which are consecutively consumed by households and the government. The production function has a simple form

$$\mathscr{Y}_t(j) = Y_t(j), \tag{22}$$

where $\mathscr{Y}_t(j)$ is the final output of the j-th producer. The profit maximization has the form

$$\max_{P_t(j),Y_t(j)} \sum_{t=0}^{\infty} \xi_p^t [P_t(j)\mathscr{Y}_t(j) - P_t^Y Y_t(j)] \tag{23}$$

where P_t is the price of final goods and $(1-\xi_p)$ is the constant probability of receiving a signal to re-optimize prices. Final goods producers face a downward-sloping demand curve for their production[19]

$$Y_t(j) = \left(\frac{P_t(j)}{P_t}\right)^{-\vartheta_p} \bar{Y}_t, \tag{24}$$

where $\bar{Y}_t$ is aggregate demand for intermediate goods as defined in equation (27) below.

Assuming full backward indexation for producers who do not receive a signal to re-optimize their prices, the first-order condition implies a hybrid version of the Phillips curve

---

[18] The Calvo-Yun setup is used, introducing staggered prices into the model. Calvo (1983) develops a model in continuous time in which each firm is allowed to change its price only when it receives a random signal. Firms follow a perfect-foresight equation for the price-setting. In the Yun (1996) extension, firms maximize the present value of real profits when they set prices.

[19] The demand curve is a result of the profit maximization of a perfectly competitive (zero-profit) aggregator who puts together differentiated goods using the Dixit and Stiglitz (1977) constant-returns-to-scale production function. The assumptions of the optimization problem are described in Erceg et al. (2000). Similar assumptions are used for other monopolistically competitive sectors in the model.

$$\log \dot{P}_t = \frac{1}{1+\beta}\log \dot{P}_{t-1} + \frac{\beta}{1+\beta}\log \dot{P}_{t+1} + \frac{(1-\xi_p)(1-\xi_p\beta)}{\xi_p(1+\beta)}\log\left(\frac{P_t^Y * \text{markup}}{P_t}\right) + \varepsilon_t^p, \quad (25)$$

where the price of final goods is (in the long run) set as a constant markup over nominal marginal costs (P_t^Y). Thus, the expression $\log\left(\frac{P_t^Y * \text{markup}}{P_t}\right)$ can be interpreted as the real marginal cost, which can be linearized as follows:

$$\begin{aligned} rmc_t &= (1-\alpha)(w_t + \Lambda_t) + \frac{1-\alpha}{1-\chi}(c_t - \chi c_{t-1}) + \alpha z_t \\ &+ \Psi\left((1-\lambda)(r_t^{ld} - r_t^{dd}) + \lambda\left(r_t^{lf} - r_t^{df}\right)\right) - a_t. \end{aligned} \quad (26)$$

Final goods are used for domestic absorption only, i.e. they are not exported.[20]

$$\bar{Y}_t = C_t + G_t. \quad (27)$$

3.4 Banking Sector

The deposit and credit flows in the model are carried out by the banking sector, i.e. firms and households cannot borrow directly from international markets. Similarly, net saver households are only allowed to hold deposits in domestic banks.

The model includes two layers of banks: wholesale and retail ones.[21] First, a representative wholesale bank collects deposits from households, borrows from abroad, extends loans to retail banks, and purchases government bonds. Households' deposits and foreign borrowing are substitutes on the wholesale banks' liability side. Similarly, households' loans and firms' loans are perfect substitutes on the asset side.

Second, retail banks purchase loans from wholesale banks, differentiate them, and extend them to firms and households. Retail banks face a rare opportunity to change the price – the interest rate – of loans and deposits. Retail interest rates are staggered.

3.4.1 Wholesale banks

To illustrate wholesale banks' role in the model, we show the decomposition of banks' balance sheet and banks' asset and liability flows in the following Tables 1 and 2. All variables specific to the wholesale banking sector are denoted by tilde.

[20] Note also the difference between domestic absorption $\bar{Y}_t$ and Y_t, which represents overall GDP.

[21] This assumption is more of a technical one. It just simplifies the derivation of the model and does not represent the segmentation of banks in the Serbian financial market.

Table 1 **Banks' Balance Sheet**

Assets			Liabilities
Loans to households in local currency	$\tilde{L}_t^{hd}$	$\tilde{D}_t^{d}$	Deposits from households in local currency
FX loans to households	$\tilde{L}_t^{hf}S_t$	$\tilde{D}_t^{f}S_t$	FX deposits from households
Loans to firms in local currency	$\tilde{L}_t^{fd}$	$\tilde{F}_t^{B}S_t$	Borrowing from abroad
FX loans to firms	$\tilde{L}_t^{ff}S_t$		
Government bonds	$\tilde{B}_t$		
Required reserves	$rrS_t(\tilde{D}_t^{f}+\tilde{F}_t^{B})$		

Table 2 **Banks' Asset and Liability Flows**

Revenues	Expenses
$(1+i_{t-1}^{bd})\tilde{L}_{t-1}^{hd}-\tilde{L}_t^{hd}$	$(1+i_{t-1}^{dd})\tilde{D}_{t-1}^{d}-\tilde{D}_t^{d}$
$(1+i_{t-1}^{bf})\tilde{L}_{t-1}^{hf}S_t-\tilde{L}_t^{hf}S_t$	$(1+i_{t-1}^{df})\tilde{D}_{t-1}^{f}S_t-\tilde{D}_t^{f}S_t$
$(1+i_{t-1}^{bd})\tilde{L}_{t-1}^{fd}-\tilde{L}_t^{fd}$	$(1+i_{t-1}^{*})(1+Prem_{t-1})\tilde{F}_{t-1}^{B}S_t-\tilde{F}_t^{B}S_t$
$(1+i_{t-1}^{bf})\tilde{L}_{t-1}^{ff}S_t-\tilde{L}_t^{ff}S_t$	$fc\left(\frac{\tilde{L}_t+\tilde{B}_t}{\tilde{D}_t}\right)$
$(1+i_{t-1}^{bd})\tilde{B}_{t-1}-\tilde{B}_t$	
$rr(\tilde{D}_{t-1}^{f}S_t+\tilde{F}_{t-1}^{B}S_t)-rr(\tilde{D}_t^{f}S_t+\tilde{F}_t^{B}S_t)$	

The wholesale banks maximize their profits

$$\begin{aligned}\max_{\tilde{D}_t^d,\tilde{D}_t^f,\tilde{F}_t^B,\tilde{L}_t^{hd},\tilde{L}_t^{hf},\tilde{L}_t^{fd},\tilde{L}_t^{ff},\tilde{B}_t}\sum_{t=0}^{\infty}\Xi_{0,t}\{&\tilde{D}_t^{d}+\tilde{D}_t^{f}S_t+\tilde{F}_t^{B}S_t-\tilde{L}_t^{hd}-\tilde{L}_t^{fd}-\left(\tilde{L}_t^{hf}+\tilde{L}_t^{ff}\right)S_t-\tilde{B}_t\\
&-rr\left(\tilde{D}_t^{f}S_t+\tilde{F}_t^{B}S_t\right)+rr\left(\tilde{D}_{t-1}^{f}S_t+\tilde{F}_{t-1}^{B}S_t\right)\\
&+(1+i_{t-1}^{bd})(\tilde{L}_{t-1}^{hd}+\tilde{L}_{t-1}^{fd})+(1+i_{t-1}^{bf})(\tilde{L}_{t-1}^{hf}+\tilde{L}_t^{ff})S_t+(1+i_{t-1}^{bd})\tilde{B}_{t-1}\\
&-(1+i_{t-1}^{dd})\tilde{D}_{t-1}^{d}-\left(1+i_{t-1}^{df}\right)\tilde{D}_{t-1}^{f}S_t-(1+i_{t-1}^{*})(1+Prem_{t-1})S_t\tilde{F}_{t-1}^{B}\\
&-fc\left(\frac{\tilde{L}_t+\tilde{B}_t}{\tilde{D}_t}\right)\},\end{aligned} \tag{28}$$

subject to the balance sheet constraint

$$\tilde{L}_t^{hd}+\tilde{L}_t^{hf}S_t+\tilde{L}_t^{fd}+\tilde{L}_t^{ff}S_t+\tilde{B}_t+rr\left(\tilde{D}_t^{f}S_t+\tilde{F}_t^{B}S_t\right)=\tilde{D}_t^{d}+\tilde{D}_t^{f}S_t+\tilde{F}_t^{B}S_t \tag{29}$$

On the liability side of the balance sheet, banks borrow from abroad ($\tilde{F}_t^B$) and collect deposits from households in both foreign currency ($\tilde{D}_t^f S_t$) and domestic currency ($\tilde{D}_t^d$). On the asset side, banks extend loans to firms denominated in domestic ($\tilde{L}_t^{fd}$) and foreign currency ($\tilde{L}_t^{ff}S_t$), as well as loans to households denominated in domestic ($\tilde{L}_t^{hd}$) and foreign currency ($\tilde{L}_t^{hf}S_t$), invest in government bonds ($\tilde{B}_t$), and hold a certain required reserve ratio (rr) of FX liabilities at the central bank. The required reserve ratio is assumed to be constant.

On the one hand, banks pay interest rates i_t^{dd} on deposits in domestic currency and i_t^{df} on deposits in foreign currency, where the latter is a function of foreign interest rates (i_t^*) and the country risk premium ($Prem_t$). On the other hand, they receive interest rates i_t^{bd} on loans in domestic currency and government bonds, and i_t^{bf} on loans in foreign currency. $fc\left(\frac{\tilde{L}_t+\tilde{B}_t}{\tilde{D}_t}\right)$ is a penalty cost function that implies lower profits when a bank does not maintain the exogenously set loan-to-deposit target. The exact specification of the cost function is not needed, as only its derivatives are necessary for the interest rate dynamics (Roger and Vlček, 2011). These costs are paid to the government and are thus assumed to be private, not social.

After rearranging, the optimality conditions constitute the UIP equations adjusted by the derivative of the penalty function[22] and the reserve requirements

$$(1+i_t^{dd}) = (1+i_t^{*})(1+Prem_t)\dot{S}_{t+1}\frac{1}{1-rr} - \frac{rr}{1-rr}\dot{S}_{t+1} + \left(\frac{\tilde{L}_t^d+\tilde{L}_t^f+\tilde{B}_t}{\tilde{D}_t} - \eta\right)\omega_1, \qquad (30)$$

and

$$(1+i_t^{bd}) = (1+i_t^{*})(1+Prem_t)\dot{S}_{t+1}\frac{1}{1-rr} - \frac{rr}{1-rr}\dot{S}_{t+1} + \left(\frac{\tilde{L}_t^d+\tilde{L}_t^f+\tilde{B}_t}{\tilde{D}_t} - \eta\right)\omega_{2,} \qquad (31)$$

where η is the loan-to-deposit ratio requirement and ω_1 and ω_2 are scaling parameters. Thus, the penalty costs provide a closing mechanism for the model by balancing deposits and loans in the economy. If banks face high demand for loans, increasing both lending rates will encourage saving and, at the same time, discourage borrowing. Such a mechanism is similar to the debt-elastic premium described in Schmitt-Grohe and Uribe (2001).[23] Note that if the reserve requirement rate goes to zero, and without a loan-to-deposit ratio requirement, (30) and (31) transform into the pure UIP condition.

Rearranging the first-order conditions with respect to loans and deposits in foreign currency, and assuming that $fc'(\tilde{D}_t^f)$ and $fc'(\tilde{L}^f)$ equal zero, gives

$$\left(1+i_t^{df}\right) = (1+i_t^{*})(1+Prem_t), \qquad (32)$$

$$\left(1+i_t^{bf}\right) = (1+i_t^{*})(1+Prem_t)\frac{1}{1-rr} - \frac{rr}{1-rr}. \qquad (33)$$

Note that under the assumption that $Prem_t$, i_t^{*} and $\dot{S}_{t+1} - 1 = \Delta s_{t+1}$ are reasonably small, so that their products are close to zero, (30)–(33) can be rewritten in more economically intuitive forms:

$$i_t^{dd} = \frac{i_t^{*} + Prem_t}{1-rr} + \Delta s_{t+1} + \left(\frac{\tilde{L}_t^d+\tilde{L}_t^f+\tilde{B}_t}{\tilde{D}_t} - \eta\right)\omega_1, \qquad (34)$$

---

[22] The derivative of the cost function with respect to domestic deposits is $\left[fc\left(\frac{L+B}{D}\right)\right]' = -fc'(D)$ and the derivative with respect to foreign deposits is assumed to be zero, i.e. $\left[fc\left(\frac{L+B}{D}\right)\right]' = 0$.

[23] In the model equation, we modify the UIP condition to introduce some inertia into the reaction of the nominal exchange rate to shocks.

$$i_t^{df} = \frac{i_t^* + Prem_t}{1 - rr} + \Delta s_{t+1} + \left(\frac{\tilde{L}_t^d + \tilde{L}_t^f + \tilde{B}_t}{\widetilde{D}_t} - \eta\right)\omega_2, \tag{35}$$

$$i_t^{df} = i_t^* + Prem_t, \tag{36}$$

$$i_t^{bf} = \frac{i_t^* + Prem_t}{1 - rr}. \tag{37}$$

It is important to stress that (34) gives a UIP relationship, i.e. how the spread between domestic and FX deposit rates affects the exchange rate dynamics. The RSD deposit rate itself is basically determined by the NBS key policy rate:

$$i_t^{dd} = i_t^{mp} - Prem_t^{dd} \tag{38}$$

where $Prem_t^{dd}$ refers to the spread between the policy rate and the interest rate on deposits in the local currency. The strong relationship between the two rates is obvious from Figure 4.

3.4.2 Retail Banks

There is a continuum k of monopolistically competitive retail banks which purchase loans from wholesale banks, differentiate them, and extend them to households and firms. The assumption of monopolistic competition introduces stickiness into lending rates. The intuition behind this is similar to the price stickiness described in Section 3.3.2. Following Benes and Lees (2007), there is a probability to re-optimize the lending rate $(1 - \xi_n)$ where $n \in d, f$ refers to loans denominated in the domestic and foreign currencies. Retail banks are endowed with a linear production function. The profit maximization is

$$\max_{i_t^{ln}(k)} \sum_{t=0}^{\infty} \xi_n^t [(1 + i_t^{ln}(k)) L_t^n(k) - (1 + i^{bn}) L_t^n(k)], \tag{39}$$

subject to the downward-sloping demand curve

$$L_t^n(k) = \left(\frac{1 + i_t^{ln}(k)}{1 + i_t^{ln}}\right)^{\frac{\theta}{1-\theta}} \tilde{L}_t^n, \tag{40}$$

where θ determines the interest rate elasticity of demand.

The first-order conditions imply a Phillips curve-type of equation for both interest rates

$$i_t^{ld} = \frac{1}{1+\beta} i_{t-1}^{ld} + \frac{\beta}{1+\beta} i_{t+1}^{ld} + \frac{(1-\xi_d)(1-\xi_d\beta)}{\xi_d(1+\beta)} \left(\frac{i_t^{bd} + Sprd^d}{i_t^{ld}}\right), \tag{41}$$

$$i_t^{lf} = \frac{1}{1+\beta} i_{t-1}^{lf} + \frac{\beta}{1+\beta} i_{t+1}^{lf} + \frac{(1-\xi_f)(1-\xi_f\beta)}{\xi_f(1+\beta)} \left(\frac{i_t^* + Prem_t + Sprd^f}{i_t^{lf}}\right). \tag{42}$$

where $Sprd^d$ and $Sprd^f$ denote the spreads (markup) within the banking sector on domestic and foreign loans, respectively.

Finally, the total supply of loans and deposits equals demand. The following identities therefore hold:

$$\tilde{L}_t^n = L_t^n \tag{43}$$

$$\tilde{D}_t^m = D_t^m \tag{44}$$

$$\tilde{F}_t^B = F_t^B \tag{45}$$

(46)

where $n \in \{hd, hf, fd, ff\}$ and $m \in \{d, f\}$.

3.5 Central Bank

The central bank targets year-on-year headline inflation ($\log\dot{P}4$) four periods ahead using the policy rate (1-week BELIBOR).

The central bank's reaction function has the form

$$i_t^{mp} = \phi_i i_{t-1}^{mp} + (1-\phi_i)\left(\bar{r}_t^{mp} + \log\dot{P}_t^{target} + \phi_p\left(\log\dot{P}4_{t+4} - \log\dot{P}_{t+4}^{target}\right)\right) + \varepsilon_t^{mp}, \tag{47}$$

where i_t^{mp} is the key policy rate of the NBS, $\bar{r}_t^{mp}$ is the equilibrium real interest rate, $\log\dot{P}_t^{target}$ is the inflation target, and the term in parentheses denotes the deviation of inflation from the target four periods ahead. The parameter ϕ_i denotes interest rate smoothing and ϕ_p is the weight on the deviation from the inflation target (the reactiveness of the monetary authority). The monetary policy shock is ε_t^{mp}.

3.6 Government

The government collects lump-sum taxes, receives bank penalty payments, issues bonds, and consumes final (government) goods. Its budget constraint is

$$P_t G_t + (1 + i_{t-1}^{bd}) B_{t-1} = B_t + \Gamma_t \tag{48}$$

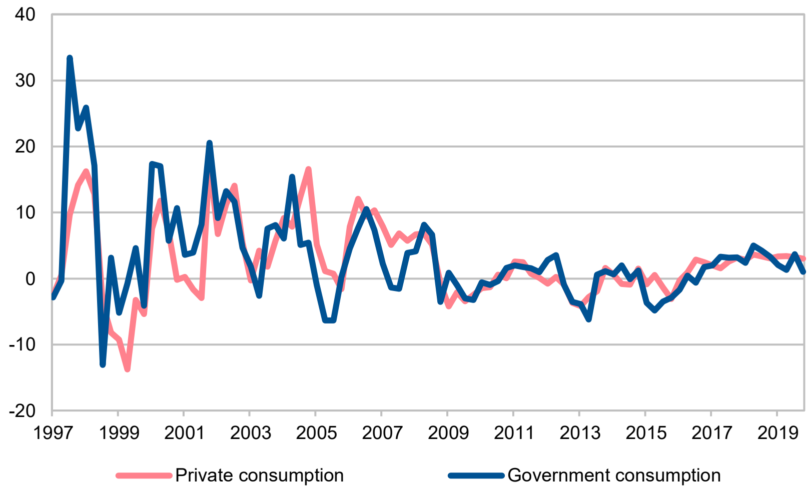
where Γ_t denotes lump-sum taxes adjusted for penalty payments from banks. The government budget constraint does not constitute a fiscal policy rule (see Andrle et al., 2009). The ratio of government bonds to government spending is assumed to be constant.

The government adjusts its spending to keep the ratio of its nominal expenditures to the nominal consumption of households constant. This simple rule is in line with the empirically observed constant ratio of private to public consumption.

$$\frac{P_t G_t}{P_t C_t} = \left(\frac{P_{t-1} G_{t-1}}{P_{t-1} C_{t-1}}\right)^{\rho_g} \left(\frac{PG}{PC}\right)^{1-\rho_g} \exp(\varepsilon_t^G). \tag{49}$$

This simple fiscal rule is consistent with the empirically observed strong correlation between private and government consumption growth (see Figure 6).

Figure 6. **YoY Growth of Private and Government Consumption**
(in %)



Source: NBS ans SORS.

3.7 Rest of the Model

3.7.1 Exports

The real export function is standard, relating real exports to foreign demand and the real exchange rate, approximating the price competitiveness of exporting firms:

$$X_t = X_{t-1}^{\rho_x} \left(\frac{P_t^*}{P_t^X / S_t} \right)^{\omega_x} (Y_t^*)^{\omega_{Y^*}} \exp(\varepsilon_t^X) \tag{50}$$

where parameter ρ_x determines the persistence of demand for exports, ω_x is the sensitivity of demand for exports to relative price changes, and ω_{Y^*} is the scaling parameter for foreign demand.

The price of exports, P_t^X, is independent of domestic business cycles and is determined in international markets.[24] The price of exports is derived from the terms of trade.

The terms of trade follow the autoregressive process

$$TOT_t = TOT_{t-1}^{\rho_{tot}} \exp(\varepsilon_t^{tot}) \tag{51}$$

implying that they are linked with the price of exports

[24] This feature enables us to capture the counter-cyclical behavior of the trade balance in developing economies (Aguiar and Gopinath, 2004), as model exports are exogenous and model imports are pro-cyclical.

$$TOT_t = \frac{P_t^X}{P_t^* S_t} \tag{52}$$

Foreign demand is set as an exogenous AR(1) process.

$$Y_t^* = Y_{t-1}^{*\,\rho_{Y^*}} \exp\left(\varepsilon_t^{Y^*}\right). \tag{53}$$

3.7.2 Net Foreign Asset Accumulation

From the household budget constraint, the government budget constraint, and the profits of monopolistically competitive firms and banks, it is possible to acquire an equation for net foreign assets

$$F_t^B = F_{t-1}^B (1 + i_{t-1}^*) \dot{S}_t (1 + Prem_{t-1}) - P_t^X X_t + P_t^* S_t M_t, \tag{54}$$

where a positive value of net foreign assets F_t^B implies that the country is a net debtor.

3.7.3 Risk Premium

The risk premium adjusts to the level of, and the change in, the net foreign asset position of the country. The underlying intuition is that the net foreign asset position affects the country's level of foreign borrowing, which further drives up the risk premium and therefore the country's interest expenses. The equation describing the law of motion for the risk premium is given in the following form:

$$Prem_t = \rho_{Prem} Prem_{t-1} + (1 - \rho_{Prem})\left(\omega_p (F_t^B - F_{t-1}^B) + \omega_{pl} F_t^B\right) + \varepsilon_t^{premia}, \tag{55}$$

where ω_p determines the sensitivity of the country risk premium to changes in foreign borrowing and ω_{pl} drives the response of the country risk premium to the level of foreign borrowing.

4 Model Properties

An examination of the model properties allows us to assess to what extent the model captures the aggregate business cycle features (stylized facts) of the Serbian economy. During this process, we tried to replicate and further supplement the features of the existing QPM model of the NBS while also taking into account the empirical evidence and underlying economic logic.

The model-consistent filtration, resulting in estimated gaps and trends in the historical data, was carried out using the Kalman filter. This is presented in Section 4.3. Particular attention was paid to parameter calibration and estimation, which are shown in Section 4.1. In order to assess the model's dynamic properties, we first carried out impulse response analyses, the results of which are presented in Section 4.2.

4.1 Calibration and Sensitivity Analysis

The parameters of the model were calibrated.[25] During the calibration phase, several criteria were used to specify the parameter values, e.g. sensitivity and impulse response analysis, expert judgment, conformity with the literature, and other countries' experiences.

We can divide the model parameters into four groups: 1) steady-state parameters, which determine the properties of the model in the long run i.e. along the balanced growth path, 2) steady-state shares, 3) transient parameters determining the short-run dynamics of the model, and 4) persistence.

The steady-state parameters were mostly calibrated by taking into account recent trends while also employing expert judgment. Their values are given in Table A1.

The calibration of the transient parameters is carried out by inspecting the model's impulse responses and performing a sensitivity analysis to establish theoretically grounded behavior of the model and reflect specific features of the Serbian economy. The values for these parameters are given in Table A2.

Due to macroeconomic volatility experienced during the 1990s, the Serbian economy is highly euroized. This implies a weaker interest rate channel, resulting in a reduced ability of the central bank to affect interest rates on FX deposits and loans. In our case, the level of euroization is 78% for household deposits and 70% for loans to firms and households. However, since we did not explicitly differentiate the two within the model, we chose to set the value for the level of euroization at 73%, taking into account the fact that the total volume of loans is significantly higher than the total volume of deposits. Since we assume a constant euroization level in the model, relatively stable observed values over time are consistent with the model specification (see Figure 2).

As shown in Fuhrer (2000), in order to successfully account for the hump-shaped response of consumption to various shocks, it is necessary to introduce habit formation in consumption.

[25] A small proportion of the parameters were estimated by OLS regression. This was used for the persistence of exogenous variables and some long-run trends.

Its introduction reflects the fact that consumers, while reacting to shocks hitting the economy, wish to smooth the level of consumption, thus contributing to more sluggish responses of consumption. We set different values of habit formation in the model for the two types of households. Net savers are less likely to change their consumption pattern, so their habit formation parameter is set to 0.8, compared to 0.3 for borrowers.

Price rigidities are introduced following Calvo's approach, with the probability of an arbitrary firm receiving a signal to re-optimize the price calibrated at 0.35. This implies that firms change their prices approximately every nine months, which is more frequent in comparison with the whole euro area, as documented in Smets and Wouters (2003), and some countries of the eurozone (see Levy and Smets, 2010).[26]

Since the Serbian economy is a small open economy with significant dependence on inflows of foreign capital, the sensitivity of the country risk premium to the share of foreign borrowing in total consumption has an important role in explaining its dynamic response to various shocks. We set this parameter to 0.005, which is the value that implies desirable effects and behavior of the model.

Proper calibration of the parameters representing the penalty costs for the banking sector in the extended uncovered interest rate parity equation turns out to be very important for obtaining plausible properties of the model. Higher penalty parameters put more weight on the loan-to-deposit requirement in determining deposit and lending rates. Depending on the spreads between those rates, this affects the real marginal costs of firms and therefore overall inflation. By employing sensitivity analysis, we determined the values of these parameters to be 0.03 in the equation for domestic deposit rates and 0.02 for wholesale rates on loans.

The estimated parameters include persistence for trend growth of the real variables in the model. These are shown in Table A3 along with the persistence which appears in the structural equations.

The ratios in the model are calculated mostly by taking into account the recent data, but also the structural breaks that occurred in the aftermath of the global financial crisis of 2008. They are presented in Table A4.

4.2 Impulse Responses

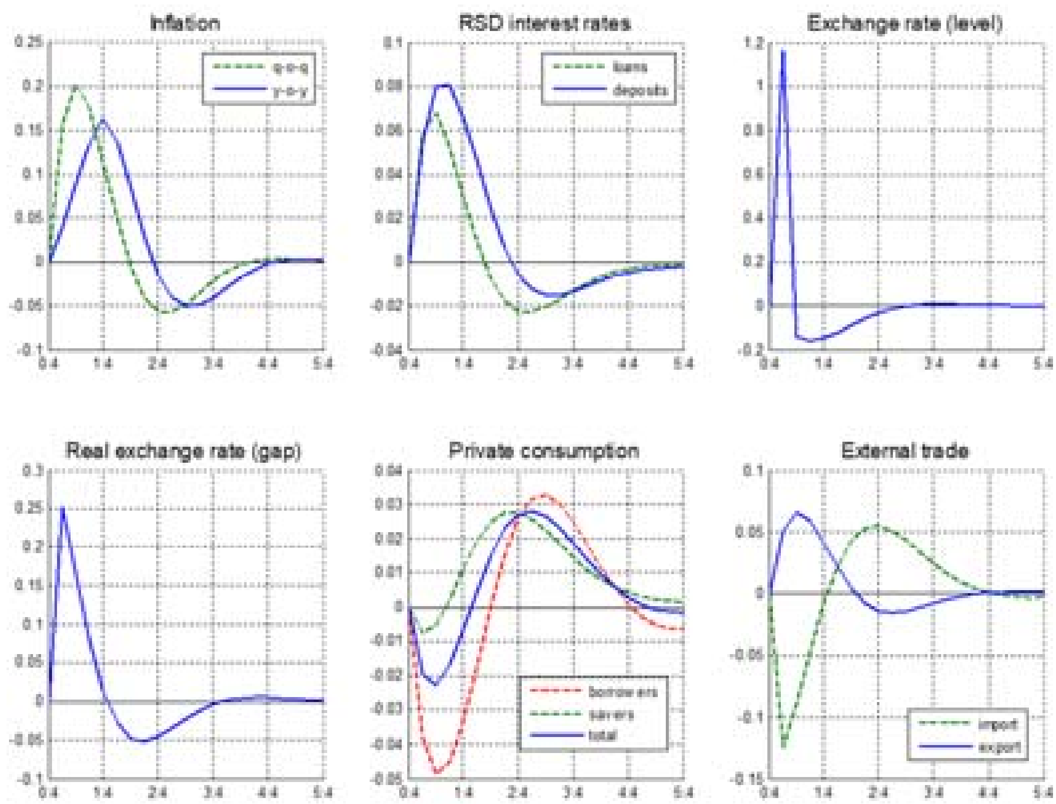
4.2.1 Unexpected Depreciation

An unexpected nominal depreciation has strong real consequences in a euroized/dollarized economy. Given foreign currency borrowing, any depreciation raises the repayment costs for borrowers. As a result of the nominal depreciation, borrowers reduce their consumption.

[26] For instance, Italian and Portuguese firms have a tendency to change their prices every ten and twelve months, respectively. On the other hand, Slovakia and Romania exhibit more frequent price changes, at intervals of approximately four and five months, respectively. For the Czech Republic, Murarik (2011) calculates the implicit length of periods without price changes (due, for example, to a change of store or a change in the specific type of product) for all selected items at ten to eleven months.

This behavior is demonstrated by an unexpected exchange rate depreciation shock which depreciates the nominal exchange rate by 1 percent initially. The one-off depreciation shock leads to a pickup in the inflation rate, triggering monetary policy tightening by the central bank. Both the deposit rate and the lending rate in local currency go up. While savers keep their consumption almost unchanged initially, borrowers reduce their spending. The reduction is driven by higher domestic lending rates in the case of domestic borrowers and also by the higher repayment costs of those who borrowed in foreign currency. Besides households, foreign currency-borrowing firms also face higher repayment costs due to the nominal depreciation. This creates additional upward pressures on domestic prices. Overall consumption improves in the subsequent periods, as inflation is expected to rise, reducing real deposit and lending rates.

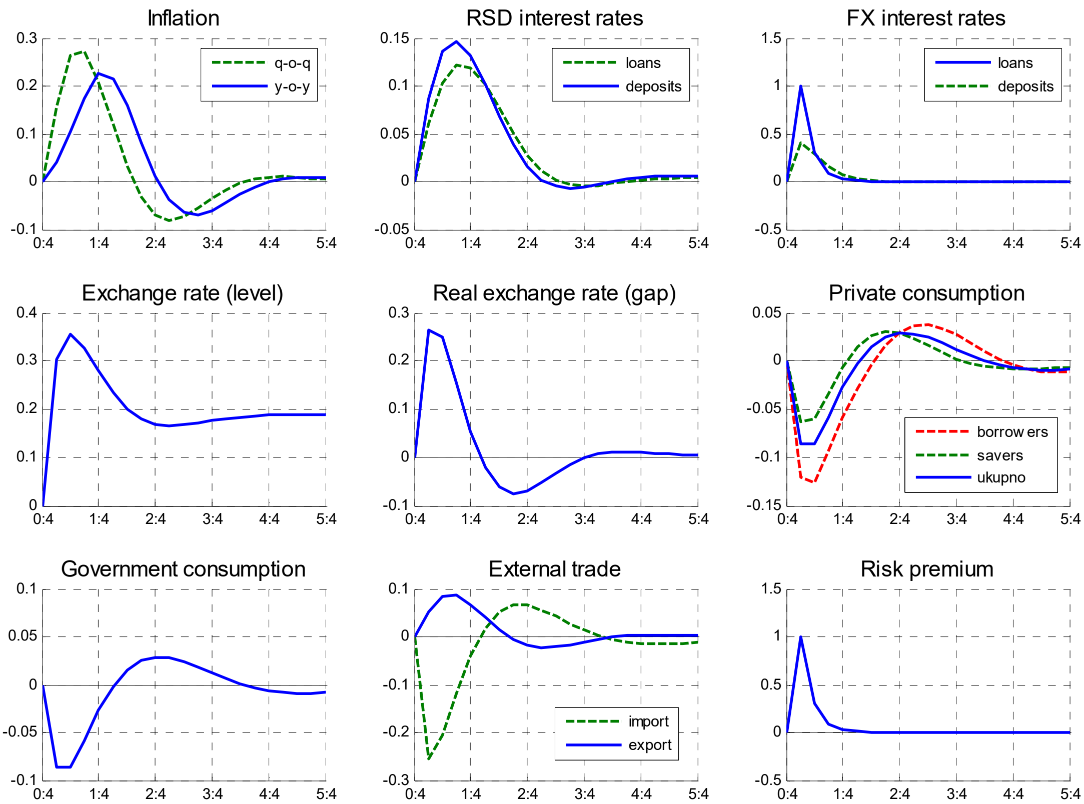
Figure 7 **Nominal Exchange Rate Shock (Deviations in percent/p.p. from Steady State)**



4.2.2 Risk Premium Shock

High capital mobility makes economies vulnerable to risk premium shocks. As can be seen from Figure 8, as a direct result of a rise in the country risk premium of 1 percentage point, FX interest rates in Serbia go up and the domestic currency consequently depreciates on impact. The depreciation leads to a higher inflation rate, triggering monetary policy tightening. The higher interest rates (both domestic and FX) have a negative impact on domestic consumption (mitigating the inflation pressures stemming from the depreciation), but, on the other hand, this is offset by an improvement in net exports caused by real depreciation of the dinar. The shock eventually fades away as the risk premium converges back to the steady state.

Figure 8 **Risk Premium Shock (Deviations in percent/p.p. from Steady State)**

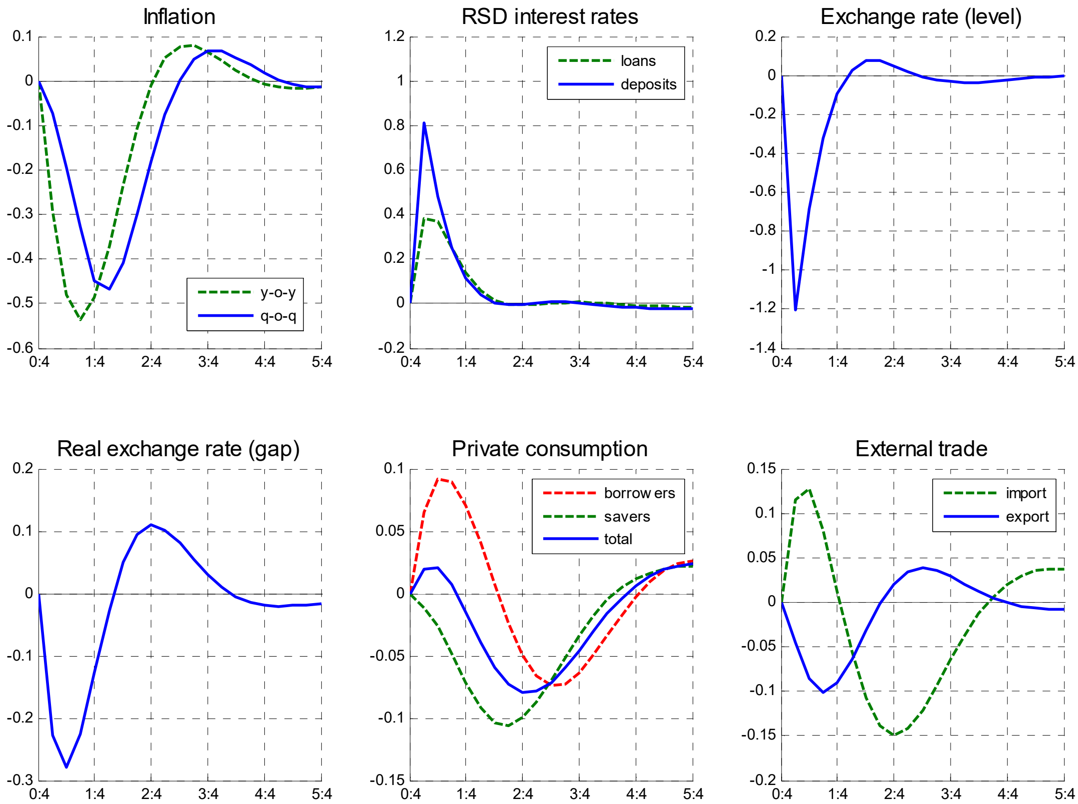


4.2.3 Increase in Policy Rate

Here we assume an unexpected shock to the policy rate of 100 bp. Changes in the policy rate pass through two main channels– the FX channel and the interest rate channel. First, raising the policy rate leads to a nominal appreciation of the dinar through the UIP. This implies real appreciation, thus creating disinflationary pressures through lower real marginal costs, but also negatively affecting GDP due to impaired external competitiveness. At the same time, the high real interest rate is a disincentive to consumption. As a result, there are negative effects on GDP along with the disinflationary pressures. Note that in the case of borrowers' consumption, the initial effect of the shock is positive, as their installment payments become lower due to the nominal appreciation (see Figure 9).

Because of euroization, however, the effect of the interest rate channel on inflation is significantly weaker than the effect of the FX channel.

Figure 9 **Increase in Policy Rate (Deviations in percent/p.p. from Steady State)**

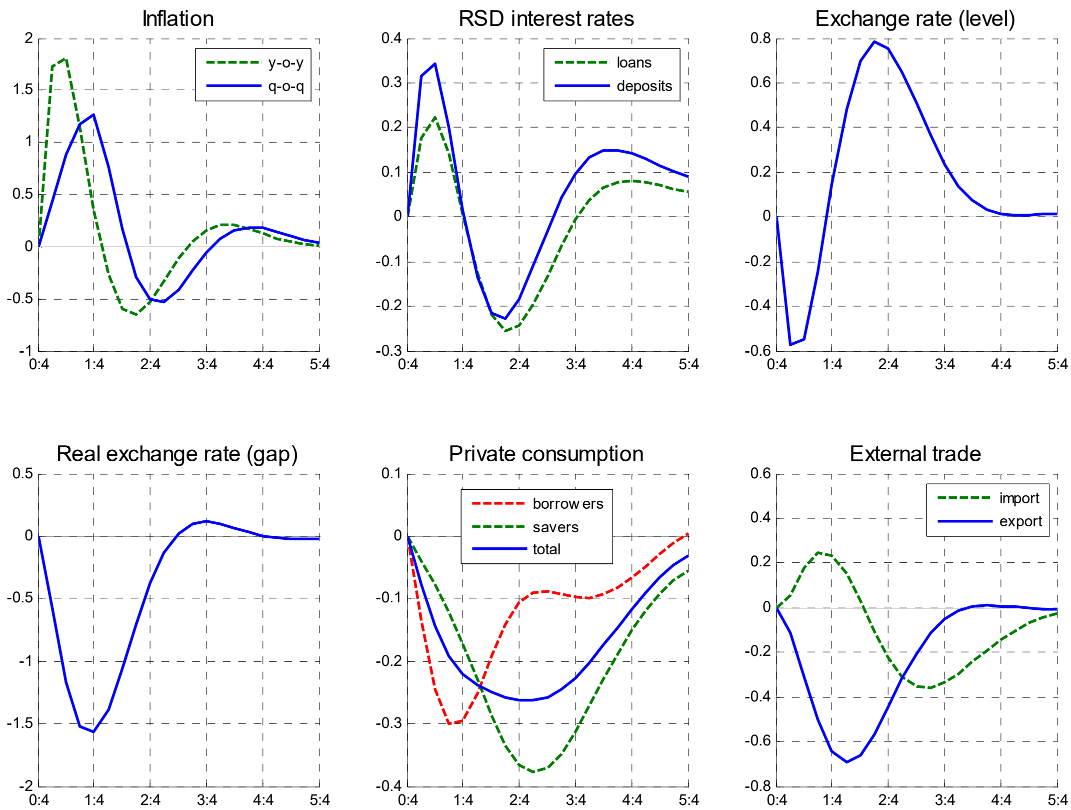


4.2.4 Inflation (Cost-push) Shock

Unexpected shocks to inflation are frequent and are caused by factors of both external and internal origin. In many cases, these shocks are related to a single component of the consumer basket.

A 1 p.p. shock to quarterly inflation (see Figure 10) leads to an increase in y-o-y inflation which, due to its persistence, reaches its peak three quarters afterwards. The central bank gradually raises its key policy rate, reacting to y-o-y inflation four periods ahead. This reaction leads to nominal exchange rate appreciation, which, together with rising inflation, results in real exchange rate appreciation. The subsequent fall in price competitiveness is the main driving force behind worsening net exports, which, in turn, are the main reason why real GDP reaches a trough one year after the initial shock. At the same time, the real exchange rate appreciation lowers the real marginal costs of intermediate goods producers, contributing further to disinflation. Initially higher real deposit rates make spending less favorable, implying a gradual fall in consumption as well as prices.

Figure 10 **Inflation Shock (Deviations in percent/p.p. from Steady State)**



4.3 Putting the Model to the Serbian Data

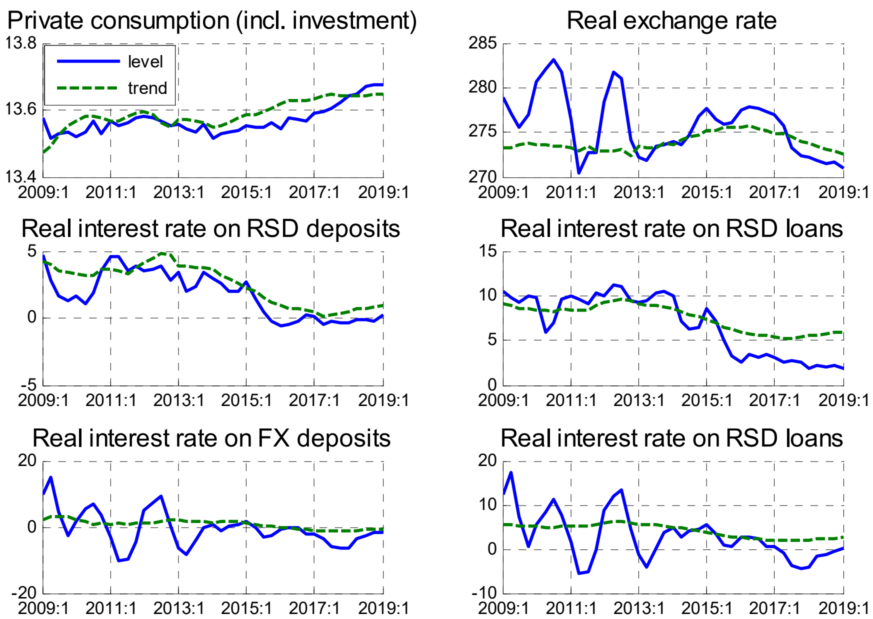
4.3.1 Filtering the Data

The model presented in the previous sections is stationary, showing no growth in the steady state. In order to match the data, the model is enriched by simple equations for trends. These state that the trend in a variable converges to its steady state with some persistence. There are also links among different trends, such as the real UIP equation, which links the real Serbian interest rate (IR) to the real eurozone IR, the risk premium, and the RER trend. We are not going to go into the details of these equations here. Instead, we will present the filtration of some of the variables in the following charts.

For the purpose of matching the model with the data, we use the standard two-sided Kalman filter (Harvey, 1990; Hamilton, 1994). The plausibility of the model is assessed by inspecting the filtration results rather than by using formal likelihood-based estimation methods. More specifically, parameters of the model which are outputs of the calibration and sensitivity analysis are considered as already given in the Kalman filtering process. The standard deviations of the residuals and the measurement errors are calibrated to match the data moments and the macroeconomic story. To this end, a shock decomposition is conducted and checked for consistency with economic intuition.

The filtration is done in a model-consistent way, meaning that the estimated gaps (consumption, interest rates, ...) and trends (premium, interest rate equilibrium, ...) of the variables are related in the way we described in the equations in Section 3.

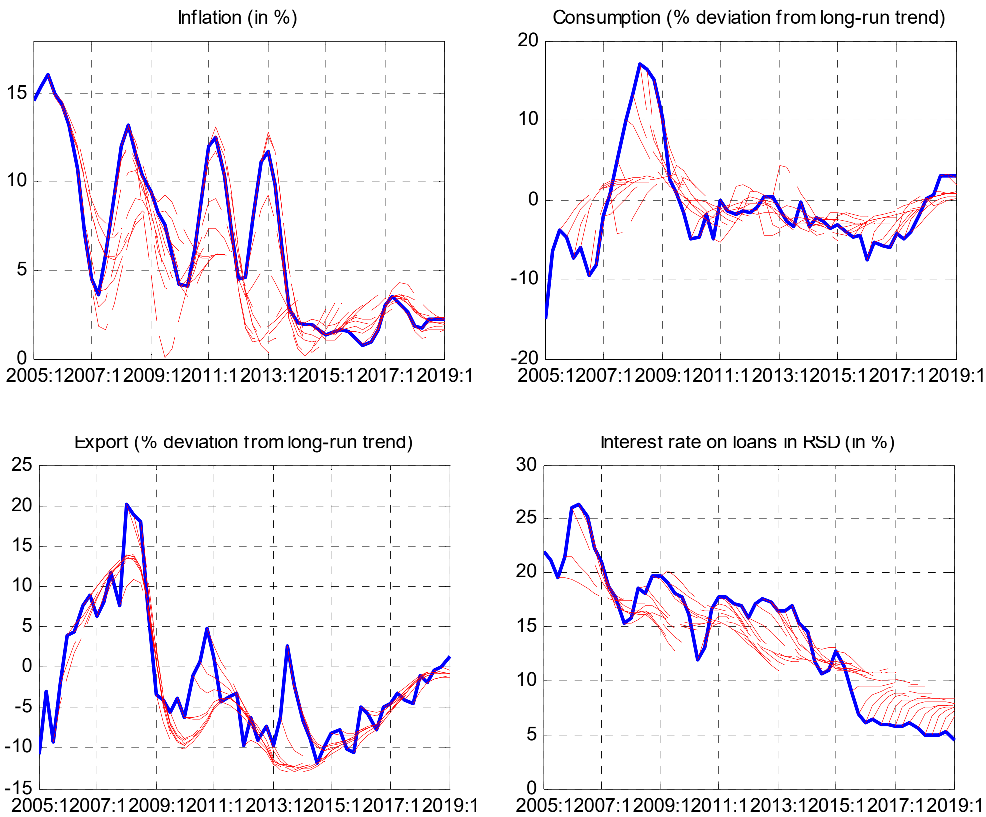
Figure 11 **Filtered Data**



4.3.2 Historical Simulations

The purpose of the historical simulations is to check how the model would have predicted the variables in certain past periods. In the case of non-stationary series, we focus on business cycle movements, i.e. gaps. Here, we present how the model explains inflation, the consumption gap, the export gap, and the domestic lending rate. As models are just a simplification of reality, historical simulations are far from perfect, but we can say that the variables are reasonably well predicted. We have to add at this point that this model is intended to be used not for forecasting, but rather for simulation (impulse response) purposes, so in calibrating the model we put more emphasis on the latter.

Figure 12 **Historical In-sample Simulations**



5 Conclusion

The main goal of this paper was to present the results of a technical cooperation project between the Czech National Bank and the National Bank of Serbia (NBS) focusing on the development of a DSGE model incorporating financial euroization. The model was developed to serve as complementary analytical tool in the NBS's Monetary Analyses and Statistics Department. Some of the modeling choices (constant steady-state shares of financial euroization, exogenous exports, etc.) reflect the preference of the authors to keep the model tractable. The model captures the main features of the transmission of monetary policy and various shocks in a highly euroized economy, such as the Serbian one, with inflation targeting as the strategy of the central bank. The financial euroization in the model means that banks collect deposits from saver households (not firms, for reasons of simplicity) in the domestic currency (RSD) and foreign currency (EUR) and extend loans to borrower households and firms, also in both RSD and EUR.

In the final derivation of the model, the FX deposit rate and FX lending rate are functions of the interest rate on foreign (EUR) bonds and a country risk premium, as well as a required reserve ratio (RRR). The central bank uses its policy rate (set by the Taylor rule) as its main monetary policy instrument. The policy rate has a one-to-one impact on the RSD deposit rate and an indirect effect on interest rates on RSD loans.

The properties of the model, i.e. the impulse responses, are broadly in line with the theory for dollarized economies and the observed relationships between variables in Serbia. A negative risk premium shock leads to an increase in FX-indexed rates, inducing depreciation of the dinar and higher inflation pressures and therefore triggering monetary policy tightening. A depreciation, on the one hand, leads to an improvement in net exports, but, on the other hand, together with higher interest rates, dampens domestic demand.

The fact that 70% of loans are FX-indexed obviously makes the interest rate channel weaker, while the bulk of the transmission goes through the exchange rate. However, even the latter is mitigated by the reaction of borrowing households. In the case of a depreciation of the domestic currency, this is due to a reduction in their consumption, as their current interest payment burden on EUR-indexed loans expressed in RSD goes up.

Note that continuation of the declining trend of the euroization in Serbia would have implications on the features in the model. In that case, interest rate channel of the monetary policy transmission mechanism would become stronger, while the impact of the exchange rate on the financial position of firms and households would be smaller.

The coefficients of the model were calibrated so as to get reasonable impulse responses and to try to explain the data behavior for Serbia as much as possible. Estimating the model using the Bayesian estimation technique is one of the tasks that lie ahead as longer time series become available (some of the data are only available since 2010).

As the model is in the gap form, putting it to the data was quite challenging. Filtering the data in order to extract gaps was done in a model-consistent way. We admit, however, that the explanation of the data requires further work. Nevertheless, we consider the model to be a useful tool for analyzing the relationships between the variables for a highly dollarized economy, especially having in mind the complexity of the channels in such an economy. In this regard, the model complements the existing QPM model, which covers inflation in much more detail but is less detailed when it comes to monetary policy channels and euroization.

To conclude, we see this stage as just the beginning of the process of developing a DSGE model for Serbia. Further development of equations, recalibration to better fit the data, and/or Bayesian estimation of the coefficients are the main future tasks relating to the model presented in this paper.

Appendix A: Calibrated Values of Parameters

Table A1 **Calibrated Values of Steady-state Parameters**

Parameter	Value
Private consumption growth	2%
Government consumption growth	1.5%
Export growth	9%
Import growth	6%
Terms of trade growth	0%
Real exchange rate growth	0%
Real interest rate on domestic deposits	2.5%
Real interest rate on foreign deposits	1.9%
Real interest rate on domestic loans	8.9%
Real interest rate on foreign loans	5.9%
Foreign real interest rate	0%
Domestic inflation target	4%
Foreign inflation target	2%
Risk premia	2.5%

Table A2 **Calibrated Values of Transient Parameters**

Parameter	Value
Monetary policy persistence	0.65
Reaction coefficient to deviation from target	2
Share of imports in domestic intermediate production	0.65
Discount factor	0.99
Habit formation (savers)	0.8
Habit formation (borrowers)	0.3
Share of foreign currency deposits in total deposits	0.76
Sensitivity of country risk premium to growth in ratio of foreign borrowing to consumption	0
Sensitivity of country risk premium to level in ratio of foreign borrowing to consumption	0.005
Penalty parameter of banking sector for i_t^{dd}	0.05
Penalty parameter of banking sector for i_t^{bd}	0.02
Elasticity of exports with respect to real exchange	0.1
Elasticity of exports with respect to foreign demand	1.5
Lending rate stickiness – domestic currency	0.4
Lending rate stickiness – foreign currency	0.4
Price stickiness	0.65
Wage stickiness	0.8

Table A3 **Estimated Parameters – Autoregressive Coefficients for Trend Growth in Real Variables in the Model**

Parameter	Value
Autoregressive coefficient for trend growth in private consumption	0.65
Autoregressive coefficient for trend growth in government consumption	0.46
Autoregressive coefficient for trend growth in exports	0.7
Autoregressive coefficient for trend growth in imports	0.73
Autoregressive coefficient for trend growth in real exchange rate	0.5
Autoregressive coefficient for trend growth in terms of trade	0.83
Autoregressive coefficient for trend growth in real interest rate on domestic deposits	0.76
Autoregressive coefficient for trend growth in real interest rate on foreign deposits	0.76
Autoregressive coefficient for trend growth in real interest rate on domestic loans	0.63
Autoregressive coefficient for trend growth in real interest rate on foreign loans	0.69

Table A4 **Ratios in the Model**

Parameter	Value
Ratio of private consumption to total consumption	82%
Ratio of government consumption to total consumption	18%
Ratio of nominal exports to nominal foreign borrowing	46%
Ratio of nominal imports to nominal foreign borrowing	76.5%
Ratio of loans to total loans	87%
Loan-to-deposit ratio	1.7
Government-bonds-to-deposits ratio	25.4%
Ratio of private consumption and investment to GDP	97%
Ratio of government consumption to GDP	22%
Ratio of imports to GDP	49%
Ratio of exports to GDP	30%
Ratio of loans to total consumption	28%
Ratio of bonds to total consumption	4%

References

Abel, A. B. (1990): “Asset Prices under Habit Formation and Catching Up with the Joneses.” American Economic Review, 80(2):38-42.

Aguiar, M, and G. Gopinath (2004): “Emerging Market Business Cycles: The Cycle is the Trend.” NBER Working Papers 10734, National Bureau of Economic Research, Inc.

Andrle, M., T. Hledik, O. Kamenik, and J. Vlcek (2009): “Implementing the New Structural Model of the Czech National Bank.” Working Papers 2009/2, Czech National Bank, Research Department.

Benes, J. and K. Lees (2007): ”Monopolistic Banks and Fixed Rate Contracts: Implications for Open Economy Inflation Targeting.”

Berg, A., P. D. Karam, and D. Laxton (2006): “Practical Model-based Monetary Policy Analisys: A How-To Guide.”

Bruha, J., T. Hledik, T. Holub, J. Polansky, and J. Tonner (2013): “Incorporating Judgements and Dealing with Data Uncertainty in Forecasting at the Czech National Bank.” Research and Policy Notes 2013/02, Czech National Bank, Research Department.

Calvo, G. A. (1983): “Staggered prices in a utility-maximizing framework.” Jounal of Monetary Economics, 12(3):383-398.

Castillo, P., C. Montoro, and V. Tuesta (2006): “An estimated stochastic general equilibrium model with partial dollarization: a Bayesian approach.” Documents de Trabajo (Banco Central de Chile), (381):1.

Christiano, L.J., M. Eichenbaum, and C. L. Evans (2005): “Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy.” Journal of Political Economy, 113(1): 1-45.

Copaciu, M., V. Nalban, and C. Bulete (2016): “REM 2.0 An estimated DSGE model for Romania.” In 11th Dynare Conference, Brussels, National Bank of Belgium.

Dixit, A. K. and J. E. Stiglitz (1977): “Monopolistic Competition and Optimum Product Diversity.” American Economic Review, 67(3):297-308.

Đukić, M., J. Momčilović, and Lj. Trajčev (2011): “Structure and use of the medium term projection model in the National Bank of Serbia.” Economic Annals, 56(188):32-61.

Erceg, C. J., D. W. Henderson, and A. T. Levin (2000): “Optimal monetary policy with staggered wage and price contracts.” Journal of Monetary Economics, 46(2):281-313.

Gerali, A., S. Neri, L. Sessa, and F. M. Signoretti (2010): “Credit and Banking in a DSGE model of the Euro Area.” Journal of Money, Credit and Banking, 42(s1):107-141.

Goodhart C.A.E, Osorio C., Tsomocos D (2009): „Analysis of Monetary Policy and Financial Stability: A New Paradigm“, CESifo Working Paper No. 2885, Munich: CESifo Group Munich

Hamilton, J. D. (1994): Time series analysis, volume 2 of Princeton university press Princeton.

Harvey, A. C. (1990): Forecasting, structural time series models and the Kalman filter. Cambridge university press.

Iacoviello, M. (2005): “House prices, borrowing constraints, and monetary policy in the business cycle.” American economic review, 739-764.

Ize, A. and E. Rerrado (2006): Real Dollarization, Financial Dollarization and Monetary Policy. Central Bank of Chile Santiago de Chile.

Levy, D. and F. Smets (2010): “Price setting and price adjustment in some European Union countries: Introdiction to the special issue.” Managerial and Decision Economics, 31(2-3):63-66.

Martinez J-S. and Tsomocos D. (2012) „Liquidity Effects on Asset Prices, Financial Stability and Economic Resilience“, mimeo, Oxford: University of Oxford

Murarik, R. (2011): “Price Setting Behaviour in the Czech Republic, Micro Data Evidence.” Statistica: Statistics and Economy Journal.

Rennhack, R. and M. Noyaki (2006): Financial Dollarization in Latin America. International Monetary Fund.

Roger, S. and J. Vlcek (2011): “Putting financial and external linkages into a small macro model for emerging market and low income countries.”

Schmitt-Grohe, S. and M. Uribe (2001): “Closing Small Open Economy Models.” Departmental Working Papers 200115, Rutgers University, Department of Economics.

Smets, F. and R. Wouters (2003): “An Estimated Dynamic Stohastic General Equilibrium Model of the Euro Area.” Journal of the European Economic Association, 1(5):1126-1175.

Smets, F. and R. Wouters (2007): “Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach.” CEPR Discussion Papers 6112, C.E.P.R. Discussion Papers.

UroševiĆ, B. And N. Grga (2014): “Application of dynamic stochastic general equilibrium models to the case of the Serbian economy.” Economic annals, 59(201):35:68.

Yeyati, E. L. (2006): “Financial dollarization: evaluating the consequences.” Economic Policy, 21(45):62-118.

Yun, T. (1996): “Nominal price rigidity, money supply endogenity, and business cycles.” Journal of Monetary Economics, 37(2-3):345-370.

ASSESSMENT OF THE REPUBLIC OF SERBIA'S SYSTEMIC RISK AND THE LIKELIHOOD OF A SYSTEMIC CRISIS

Darko Kovačević

The views expressed in the papers constituting this series are those of the author, and do not necessarily represent the official view of the National Bank of Serbia.

Financial Stability Department

THE NATIONAL BANK OF SERBIA

Belgrade, 12 Kralja Petra Street,

Telephone: (+381 11) 3027 100

Belgrade, 17 Nemanjina Street,

Telephone: (+381 11) 333 8000

www.nbs.rs

Assessment of the Republic of Serbia's Systemic Risk and the Likelihood of a Systemic Crisis

Darko Kovačević

Abstract: The purpose of this paper is to set up the Systemic Stress Indicator (SSI) of the Republic of Serbia’s financial system based on the proposed modification of the systemic stress testing approach that allows for a more appropriate aggregation of the observed indicators within the financial system segment. It also weighs up the advantages of the proposed approach compared to the aggregation methods most frequently used in literature. It proposes a mathematical formulation of the systemic risk level of the financial system and an analytical framework of the early warning system based on an assessment of the likelihood of a systemic crisis occurrence in case of an arbitrary number of regimes. The SSI has demonstrated the ability to correctly identify crisis periods and the systemic risk level of the Republic of Serbia’s financial system. It is suggested that probabilities of a systemic crisis occurrence in a given period are in perfect sync with the dynamics of undetected periods. An optimal period in the case of used indicators and relatively short time series is six months, which may provide timely signals to policy makers to mitigate negative effects on financial and macroeconomic stability.

Key words: Financial stability, Systemic risk, Financial crisis, Macro-financial linkages, Markov-switching model, early warning model
JEL Code: C32, G01, E44

Non-Technical Summary

An important lesson learned from the global financial crisis of 2008 was that national regulators and other financial system supervisory authorities lacked adequate tools to facilitate the analysis and measurement of systemic risk in real time. The preconditions for the development of analytical tools for systemic risk identification implied, on the one hand, the understanding of economic processes leading to the build-up of these risks, and, on the other hand, the analysis of events that may generate material costs in the real economy. Financial instability may become systemically significant through an exogenous and/or endogenous shock, if a larger part of the financial system is hit at the same time. The start of a crisis at a systemic level is usually accompanied with a trigger event, while recovery is a long-standing process, and the moment of exiting the crisis is often unknown. Moreover, although each financial crisis is unique in terms of causes and the channels of transmission to different market segments, it is important to compare different systemic events by using the indicators to measure the level of systemic stress.

The aim of this paper is to construct the Systemic Stress Indicator (SSI) of the Republic of Serbia's financial system, based on the modification of approach to systemic stress evaluation proposed by Hollo and a group of authors, which enables the aggregation of the observed indicators within financial system segments. As the composite indicator is based on movements in different financial market segments, it enables a solid evaluation of linkages between these segments, i.e. it allows for the possibility to evaluate the systemic component and individual risk factors. We have presented a detailed analytical framework for the calculation of the composite indicator and the systemic risk component, based on Markov switching models with dynamic transitive probabilities. We have identified the advantages and shortcomings of different approaches to the aggregation of sub-indices of financial system segments, based on which we have proposed an approach that is consistent with the composite indicator methodology and enables exact mathematical formulation of the systemic risk level. We have also validated the number of systemic crisis regimes by applying the Gaussian component method. Both approaches assessed in the same way the number of regimes and moments of SSI distributions, which suggests the uniqueness of the obtained results.

The paper also contains an analytical framework for early warning signals, which is based on the methodology used for SSI calculation and which is unique not only for the assessment of the current stress level in the system, but also for the assessment of probability of a systemic crisis in future in case of an arbitrary number of regimes. The quality of classification of the proposed model was validated by assessing the measures of the quality of classification in case of an arbitrary number of regimes, and based on public statements of economic policy makers and experts.

It has been demonstrated that the SSI can accurately identify crisis periods, the level of systemic risk of the financial system, and that it can assess the probability of a systemic crisis occurrence, providing significant information about the degree of risk accumulation in financial markets and potential implications for financial and macroeconomic stability.

Contents

1 Introduction **78**
2 Creation of a composite indicator of systemic stress **80**
2.1 Financial system segments and selected indicators 82
2.2 Assessment of a systemic stress period and the number of regimes 87
2.3 Analysis of SSI movements 90
2.4 Analysis of risk factors and stress level in the observed period 93
3 Assessment of the likelihood of a systemic crisis **95**
3.1 Critical levels of the likelihood of a systemic crisis occurrence and optimal forecast horizon 98
3.2 Assessment of the likelihood of a systemic crisis occurrence 99
3.3 Verification of an early warning signal of a systemic crisis occurrence 102
4 Conclusion **104**
Appendix 1 Sub-index aggregation method analysis **105**
Appendix 2 List of indicators for SSI calculation **112**
Appendix 3 Statistics of estimated values of model parameters **113**
Appendix 4 Models for assessing the probability of a systemic crisis occurrence **114**
Appendix 5 News history of the 2008–2010 crisis **116**
References **117**

1 Introduction

The global financial and economic crisis of 2008 revealed weaknesses in the regulation of financial systems internationally. One of the important lessons learnt during the crisis is that national regulators and other authorities responsible for supervising the financial system did not have appropriate tools to facilitate real-time risk perception and assessment. Another problem was that even when regulators were aware of the risk and when unfavourable trends that could affect the overall financial system were assessed, they did not have appropriate regulatory mechanisms to take emergency intervention. The financial crisis showed that so-called macroprudential regulation and supervision should be added to microprudential supervision of financial markets and intermediaries (Borio 2003) in order to identify imbalances and vulnerabilities of the overall financial system. Therefore, a large number of initiatives were introduced with the aim of reviewing existing regulatory frameworks. At the same time, intensive research work was initiated in order to develop new analytical tools, which the authorities responsible for macroprudential supervision would use for timely decision-making.

The prerequisites for the development of analytical tools for identifying systemic risks include the understanding of processes leading to the accumulation of these risks, on the one hand, and the analyses of events that may result in material costs on the other (an occurrence of a financial crisis with considerable costs in the real economy (Crotty 2009)). The three main sources of systemic risk are mentioned in scientific literature: the accumulation of financial imbalances (Obstfeld and Rogoff 2009), e.g. robust credit growth is linked to a sharp rise in the price of assets, exogenous external and/or internal shocks affecting financial market actors (Houben et al. 2004), and the financial contagion effect (Schwarcz 2008).

International experience shows that financial crises with high costs for the real economy are often preceded by persistent and excessive asset growth. During these periods, growth in consumption and investments, as well as expansion of lending for financing further growth, may become a self-perpetuating process, accompanied by risk accumulation. These periods are usually characterised by the weakening of banks' standards during the loan approval process, which results in riskier categories of borrowers being able to access banks' funds. This seemingly sustainable credit growth, consumption and investments may be interrupted by even minor financial shocks if they affect several segments of the financial system at the same time.

Financial instability may become systemically insignificant, through an exogenous external and/or internal shock, if a larger segment of the financial system is concurrently affected. As a result, depending on the risk accumulation and balance sheet imbalances, many financial institutions may face a lower capital quality and business problems, which are often called the primary effect. A financial contagion can occur endogenously regardless of the above processes. If one or more banks adversely affect the business sustainability of other financial institutions or even financial markets as a whole, as a result of negative external and/or internal shocks, a contagion effect called the secondary effect is present.

It most often occurs in case of unexpected problems with the operation of systemically important financial institutions, which may reflect on other financial institutions and financial markets. It is important to emphasise that in this case the effects do not necessarily result from the macroeconomic environment, but from the inadequate business policy, operating model and risk management function of a single, systemically important, financial institution. The onset of any crisis is usually accompanied by a triggering event, whereas the recovery is a long-lasting process with an often unknown ending. In addition, although each financial crisis is unique in terms of its cause and the channel of transmission to various market segments, it is important to compare various systemic events with certain indicators used to measure the systemic stress level.

The financial stress level largely depends on the size of the shock that hit the system, the risk accumulation level and financial system imbalances and responses of decision makers responsible for preserving macrofinancial stability, as well as market expectations in terms of these responses. The first step in establishing a uniform methodology for calculating indicators used to assess financial sector soundness, which are implemented by a large number of national regulators, are financial soundness indicators proposed by the International Monetary Fund for the purpose of a cross-country comparison and establishment of a macroprudential policy framework at national levels (IMF 2008). Unlike numerous indicators that are related to the economic activity and can be expressed in a money equivalent, the financial stress level does not have its unique form in the real economy. Since it cannot be directly measured, the idea behind the creation of the SSI, presented in this paper, is to measure different stress manifestations. The SSI aims to aggregate different financial segments into a single composite indicator. Since the composite indicator focuses on financial market movements, it provides a good estimate of linkages between these segments, i.e. the possibility to assess a systemic component and movement of individual risk factors.

The main objective of this paper is the construction of the SSI of the Republic of Serbia. The main or general objective of the SSI and other financial soundness indicators is an assessment of the current financial system stress level or its lack. The paper provides a detailed analytical framework for the creation of indicators, even in case of the lack of high-frequency data, simultaneously considering the characteristics of the local financial system that are related to a relatively high level of euroisation and significant foreign ownership in the banking sector, where one of the primary monetary policy transmission mechanisms is the exchange rate channel. The aggregation of financial system segment indicators is discussed in detail, and the strengths and weaknesses of each approach are elaborated upon. The paper also provides a framework for assessing the likelihood of a systemic crisis occurrence based on the same framework, which is a natural upgrade of the said methodology.

The paper is divided into four parts. After the introduction, we present an analytical framework for the aggregation of different financial system segments and creation of a composite indicator in the second part by using dynamic Markov-switching regimes. In addition, the indicators of different segments of the Republic of Serbia's financial system are discussed in detail and historical episodes of increased stress are identified.

The third part deals with the development of a methodology and mathematical model for calculating the likelihood of a systemic crisis occurrence in an arbitrary time horizon based on the presented SSI methodology. We select the best early warning model based on classification quality measures in case of an arbitrary number of regimes. We also validate the obtained probabilities of a systemic crisis occurrence based on public statements of the Republic of

Serbia's economic policy makers about the condition of the financial system from 2008 until early 2021.

Finally, the fourth part contains a conclusion and the main results of the conducted analysis and contributes to scientific literature in the form of the overall methodology not only for assessing individual risks in the financial system and systemic risk as a whole, but also for providing systemic crisis warning signals.

2 Creation of a composite indicator of systemic stress

In order to answer some of the above questions, the present analysis will introduce a financial stress index called the Systemic Stress Indicator (SSI), based on the Composite Indicator of Systemic Stress (CISS), proposed in the paper by Hollo et al. (2012). The main or general objective of the CISS and other financial soundness indicators is an assessment of the current financial system instability/stress level. Therefore, a financial system can be defined as a set of financial markets, intermediaries and infrastructure. A separate sub-index of each financial system segment is calculated after a proper transformation of individual indicators. The main methodological innovation during the development of this type of indicator, presented in the paper by Hollo et al. (2012), is the use of the portfolio theory in the sub-index aggregation into a composite indicator. The portfolio aggregation takes into account concordance measures between different sub-indices. As a result, an indicator assigns a larger weight when stress prevails in several market segments simultaneously, which indicates that systemic risk/stress is higher if financial instability is widespread in most of the financial system.

Another aggregation element that characterises a systemic risk is the fact that portfolio weighting functions are calibrated for each sub-index based on their impact on the industrial production index as the economic activity measure of a country.

The first step in calculating the SSI is the creation of sub-indices for each of the selected financial system segments. The most frequently used aggregation approach is the equal variance method, which assigns the same weights to all observed indicators. It is desirable for the methodology to include a common variability factor that occurs in the data. For the effects of individual data series to be comparable, input data need to be standardised. We use the empirical cumulative distribution function to reduce each indicator to an interval between zero and one, using the following formula:

$$z_t = F_T(x_t) = \begin{cases} r/T & for\ x_{[r]} \leq x_t < x_{[r+1]}, r = 1,2,\dots,T-1 \\ 1 & for\ x_t \geq x_{[T]} \end{cases} \tag{1}$$

Where $x_{[r]}$, $r = 1,2,\dots,T-1$, $t = 1,2,\dots,T$ and T is the sample length. By doing so, each data point obtains its rank that corresponds to the quantile of the cumulative probability distribution function ranging from 0 to 1.

According to the portfolio aggregation of the composite indicator (Hollo et al. 2012), a high correlation between individual risks results in an increase in the overall portfolio risk. On the other hand, if the correlation between individual portfolio segments is low, the overall portfolio risk is reduced, i.e. individual risks are diversified.

Unlike the original approach, in which the authors obtain the sub-index value using the equal variance method, i.e. a weighted sum of indicators within a segment, we suggest a different creation of these indicators that would comply with the composite indicator creation approach and enable an exact mathematical formulation of the systemic risk level Sys . An analysis of the existing and proposed aggregation approach is discussed in detail in Appendix 1. If the overall risk of an individual segment is observed as one sub-index, the value of this indicator, based on the portfolio risk aggregation theory, can be obtained as follows:

$$SI_i^t = \sqrt{\left(\boldsymbol{W}_i \circ \boldsymbol{Z}_i^t\right)\mathbb{C}_t\left(\boldsymbol{W}_i \circ \boldsymbol{Z}_i^t\right)^T}, \tag{2}$$

Where $z_{i,j}^t$ represents transformed indicators within the segment i, w_i represents a time-invariant weighting function of equal variances, and $\mathbb{C}_t$ represents a correlation matrix, whereas $\circ$ is the Hadamard product:

$$\begin{aligned} \boldsymbol{Z}_i^t &= \left(z_{i,1}^t, z_{i,2}^t, \dots, z_{i,j_i}^t\right) \\ \boldsymbol{W}_i &= \left(w_{i,1}, w_{i,2}, \dots, w_{i,j_i}\right), \end{aligned} \tag{3}$$

where j_i is the number of financial system segment indicators i. The following are weighting functions: $w_{i,j_m} = w_{i,j_n}, \forall m,n \in J_i$.

Out of individual SI_i^t sub-indices, the SSI is obtained based on the following equation:

$$SSI^t = \sum_{j=1}^{q}\left(\boldsymbol{W_k} \circ \boldsymbol{SI}^t\right)\boldsymbol{I}\left(\boldsymbol{W_k} \circ \boldsymbol{SI}^t\right)^T + \left(\boldsymbol{W} \circ \boldsymbol{SI}^t\right)\left(\mathbb{C}_t - \boldsymbol{I}\right)\left(\boldsymbol{W} \circ \boldsymbol{SI}^t\right)^T, \tag{4}$$

Where $W_k = [w_k]$, $w_k = \begin{cases} 1 & \mathrm{k} = \mathrm{j} \\ 0 & \mathrm{k} \neq \mathrm{j} \end{cases}$, q represents the number of the observed segments, I is a unit matrix, $SI^t = \left(SI_1^t, SI_2^t, \dots, SI_q^t\right)$ is an appropriate vector of the sub-index value and $\mathrm{W} = \left(w_1, w_2, \dots, w_q\right)$ represents an appropriate vector of sub-index weights.

The first part of the equation (4) concerns the impact of individual segments, i.e. sub-indices, while the second part of the equation (4) exactly determines the systemic risk level Sys at the moment t:

$$Sys^t = \left(\boldsymbol{W} \circ \boldsymbol{SI}^t\right)\left(\mathbb{C}_t - \boldsymbol{I}\right)\left(\boldsymbol{W} \circ \boldsymbol{SI}^t\right)^T. \tag{5}$$

Typically, in case of negative shocks in a financial system, many financial system segments are concurrently affected, which results in a high correlation between segments (Lo Duca and Peltonen 2011). In addition, an adequate sum of members can be shown in a more compact form that is used in the paper by Hollo et al. (2012):

$$SSI^t = \left(\boldsymbol{W} \circ \boldsymbol{SI}^t\right)\mathbb{C}_t\left(\boldsymbol{W} \circ \boldsymbol{SI}^t\right)^T \tag{6}$$

Scientific literature provides numerous definitions of systemic risk. It is defined mainly based on the final effect on the real economy (De Bandt et al. 2009). Here we see a correlation between a financial system and the real economy that we discussed at the beginning of this paper. The sub-index weighting vector $\boldsymbol{W}$ is consistent over time and is estimated using the maximum reliability method by means of an analysis of sub-index effects on the y-o-y industrial production growth rate – IP:

$$\hat{l}(\boldsymbol{W}|\boldsymbol{SI}) = \frac{1}{\mathrm{T}}\sum_{t=1}^{\mathrm{T}} ln\big(f(\boldsymbol{X}|\boldsymbol{SI})\big), \text{ where}$$
$$f(\boldsymbol{SI}|\boldsymbol{W}) = \frac{1}{2\pi^{T/2}\sqrt{\det|\boldsymbol{\Sigma_t}|}} exp\left(-\frac{1}{2}\big((\boldsymbol{W}^T\boldsymbol{SI} - \boldsymbol{IP})\boldsymbol{\Sigma_t}(\boldsymbol{W}^T\boldsymbol{SI} - \boldsymbol{IP})^T\big)\right) \quad (7)$$

With restrictions:

$$\sum_{i=1}^{q} w_i = 1 \text{ и } w_i > 0$$

In case of a sub-index calculation and the SSI, $\mathbb{C}_t$ represents a matrix of time-varying correlation coefficients, which is defined as:

$$\mathbb{C}_t = \begin{bmatrix} 1 & \rho_{12,t} & \cdots & \rho_{1q,t} \\ \rho_{21,t} & 1 & \cdots & \rho_{2q,t} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{q1,t} & \rho_{q2,t} & \dots & 1 \end{bmatrix} \quad (8)$$

The general member of this matrix is the correlation coefficient between the variables i and j at the moment t:

$$\rho_{ij,t} = \frac{\sigma_{ij,t}}{\sigma_{i,t}\sigma_{j,t}} \quad (9)$$

where $\mathrm{i} = 1, \dots, \mathrm{q}$ and $j = 1, \dots, q$, $\sigma_{ij,t}$ is a covariance between the variables *i* and *j* at the moment t, and $\sigma_{\mathrm{i,t}}$ and $\sigma_{j,t}\sigma_{\mathrm{ij,t}}$ are standard deviations of the variables *i* and *j* at the moment *t*, which are simultaneously estimated using the exponentially weighted moving average (EWMA) based on the following formulas:

$$\sigma_{ij,t} = \lambda\sigma_{ij,t-1} + (1-\lambda)(s_{i,t-1} - \mu_{i,t-1})(s_{j,t-1} - \mu_{j,t-1}) \quad (10)$$
$$\mu_{i,t} = \lambda\mu_{i,t-1} + (1-\lambda)s_{i,t-1}$$

The parameter λ refers to a persistence of the covariance. High values $\lambda \to 1$ indicate high inertia and low reaction of the process to previous values and vice versa (Alexander 2009). Here $i = 1, \dots, q$ and $j = 1, \dots, q$, $\mu_{i,t}$ represents the expected variable value, whereas the parameter λ takes on a constant value of 0.93 (Louzis and Vouldis 2013).

2.1 Financial system segments and selected indicators

The main difference between the SSI and other financial soundness indicators is that the SSI is focused on the systemic dimension of financial stress. The SSI includes 25 indicators that reflect the financial stress level in six crucial segments of the Republic of Serbia's financial system: foreign exchange market, public finances, money market, capital market, banking sector and foreign environment.

One of the objectives during the selection of variables for the SSI construction is the coverage of the local financial system segments and characteristics of the foreign environment, which can be seen in Appendix 2. The segments are named as follows: foreign exchange market – FX, public finances – GOV, money market – MON, capital market – EQU, banking sector – BANK and foreign environment – FOR. The range of the local financial system indicators largely reflects the characteristics of the Republic of Serbia's financial system. The data cover the period from January 2008 to March 2021. Even though the time series includes a period of only 13 years, we deem that the systemic stress indicator efficiently defined the periods that occurred during the said time interval. The variables are taken on a monthly basis even though the methodology itself offers the option of using higher-frequency data. Due to a small number of financial institutions in the capital market and the lack of market indicators, monthly data from regulatory reports of financial institutions were used. Thus, some of the indicators used in developed economies (Hollo et al. 2012) could not be applied in the case of the Republic of Serbia.

The first analysed segment of the financial system was the foreign exchange market. The used indicators take into consideration the characteristics of the local financial system that are related to relatively high euroisation and significant foreign ownership in the banking sector, where one of the key monetary policy transmission channels, apart from the interest rate channel, is the exchange rate channel. Apart from the fact that euroisation reduces the efficiency of monetary and fiscal policies, exchange rate fluctuations are highly important in circumstances of excessive foreign-currency lending, as it is the case in Serbia, and the currency-induced credit risk. The exchange rate is a result of economic developments, the interest rate and specificities of the economy and its environment. Apart from exchange rate fluctuations, the volatility of the dinar against the euro is used to calculate the SSI. Volatility indicators are generally observed as a reflection of uncertainty, i.e. the higher the volatility the higher the values of the indicators used for measuring the stress level. The FX market may respond to new information on expected exchange rate fluctuations. However, it turns out that not all information will always have a critical impact on the exchange rate. After some time, it can be ascertained that some information was not relevant due to asymmetric information available to the participants. Therefore, exchange rate fluctuations observed over a longer period are expected to be smaller than those observed over a short period (Dornbusch 1976). Another FX market indicator used to create the SSI is the average daily difference between the buying and selling exchange rate of the dinar against the euro, based on indicative quotations in the interbank FX market, which is related to the liquidity of the currency pair. In general, the liquidity of assets is related to the speed and ease at which they can be traded. The difference between buying and selling exchange rates contains an incorporated liquidity risk premium and thus, a difference increase contributes to an increase in the SSI value. The last indicator refers to the net value of FX interventions of the National Bank of Serbia in the domestic FX market. All exchange rate features explained so far are related to the rate freely formed in the domestic foreign exchange market on the basis of the supply and demand of currencies. The National Bank of Serbia pursues a managed float exchange rate regime and intervenes in the interbank market to ease excessive short-term volatility of the dinar against the euro, with no intention to influence the exchange rate level or its trend, and to preserve price and financial system stability and maintain an adequate level of FX reserves. This rule of the National Bank of Serbia complies with its Memorandum on Inflation Targeting as a Monetary Strategy, which has been officially applied since January 2009, and with its monetary policy programmes. When intervening in the interbank market, the National Bank of Serbia absorbs the stress that would otherwise reflect on the dinar/euro exchange rate. With this in mind, the value of FX interventions in the domestic FX market reflects pressure on the local currency and thus impacts the SSI.

The second financial system segment used to assess the SSI is the money market, which is a financial market segment where short-term financial instruments are traded. In general, assets whose maturity is shorter than one year are traded. Transactions with giro money, short-term loans, discount and Lombard transactions and short-term securities transactions are mainly conducted in this market. Since the money market enables an applicant to trade certain assets for cash within the shortest possible time, liquidity is the main feature of this market. The first indicator of the money market condition is BEONIA (Belgrade OverNight Index Average), which is an interest rate formed as a weighted average of interest rates on overnight interbank loans. Pursuant to the Decision on Submitting Data on Overnight Interbank Money Market Loans to the National Bank of Serbia, an overnight loan means lending dinar funds from one bank to another with a repayment deadline being the end of the following business day from the lending day. By including BEONIA, we assume that in case of severe turbulences in the banking sector, banks might start borrowing excessively, which would result in an increase in the average interbank interest rate, leading to the increase of the SSI. However, if we look at historical fluctuations of this rate, we can see that BEONIA is usually in sync with the repo rate. Apart from the BEONIA value, we also use the standard deviation of the rates at which individual transactions are concluded in the interbank money market against the BEONIA rate. A significant deviation of the rates at which interbank market transactions are conducted from the BEONIA rate may be the signal of money market uncertainty. An increase in the deviation leads to an increase in the SSI value. Apart from the BEONIA rate, we also use the difference between the BEONIA rate and the key policy rate. The key policy rate is the main instrument of the National Bank of Serbia's monetary policy. A negative value of this indicator shows a sufficiently high liquidity level of the banking system, which reduces the SSI value as expected. On the other hand, if banks are forced to borrow at an interest rate higher than the key policy rate, it should signalise that they are facing a liquidity shortage. Apart from the above indicators, we also included the difference between an interest rate ceiling on overnight interbank market loans and the interest rate on credit facilities. According to the Decision on Interest Rates applied by the National Bank of Serbia in the Implementation of Monetary Policy, the interest rate on an overnight loan for maintaining banks' daily liquidity (credit facility) is determined at the level of the key policy rate plus 0.9 pp. As a pledge for this loan, banks need to deposit a certain amount of dinar securities. If a bank failed to take such loan from the National Bank of Serbia and was forced to borrow in the interbank market at an unfavourable interest rate, it means that it was in urgent need for additional liquidity, which can be interpreted as a sign of stress, and therefore the positive value of this indicator should increase the SSI value. Another liquidity ratio in the money market is the required reserve allocation. The average required reserve allocation for a given period on the last day of the maintenance period should be equal to or higher than 100% of calculated required reserves. Even though it is best that the allocated required reserve be exactly 100% of the calculated reserves, a higher amount is not particularly worrying as it implies that banks are liquid more than necessary. An amount lower than 100% implies a liquidity shortage, which could imply that banks are not able to allocate sufficient funds for the required reserve, i.e. they are not liquid enough. Relatively high euroisation of our economy makes us dependent on euro fluctuations in international markets. Therefore, the SSI includes the difference between three-month EURIBOR and the interest rate on three-month German Bunds (government bonds). EURIBOR is an average interbank interest rate offered among major European banks. In addition, the difference between three-month EURIBOR and an EONIA (Euro OverNight Index Average) interest rate was used. The EONIA interest rate is important as it does not only contain market expectations about the European Central Bank's monetary policy, but it also limits fluctuations of longer-maturity interest rates. The dynamics of difference fluctuations was the focus of empirical studies on the effects of financial crises on the money market (Tamakoshi and Hamori 2015), which indicate that certain fluctuations of this indicator may be an early warning signal of potential money market turbulences.

The capital market, as the third financial segment used to calculate the SSI, is a market where cash and cash equivalents are supplied and demanded long-term. It is an institutionalised market with clear rules of conduct for its actors and product standardisation. Trading in the capital market provides its participants with investment diversification and optimum maturity asset transformation. The global financial crisis showed that its effects can be considerable in spite of opinions about its limited effects on shallow and insufficiently developed markets, such as the Serbian capital market. The first representative capital market indicator is the CMAX transformation (Illing and Liu 2006) of BELEX15 index of the most liquid stocks on the Belgrade Stock Exchange, whose purpose is to identify a sudden or an extended drop in the equity price, which represents a capital market stress symptom. BELEX15 quantifies the performance of the most liquid segment of the domestic capital market. A stock index basket is comprised of common shares in the regulated market, which had minimum 80% of trading with concluded transactions over the past two quarters. Those stocks that meet this criterion are ranked against market capitalisation in free circulation. The CMAX transformation measures the maximum cumulative loss in a given period. The maximum cumulative loss over the past 12 months is observed for the purposes of this paper. Higher values of this indicator mean a bigger cumulative loss and consequently an increase in the SSI value and vice versa. If the value of company's shares is high, the market estimates that such company is valuable and stable. Money invested in it will not be lost as the risk of going out of business is low. Investors assess all relevant information when trading and their risk perception is recorded in the stock price. Hence, a value increase in this indicator implies a lower risk. The reason why stock exchange data are interpreted in a simple manner is because a stock exchange is one of the most liquid markets. All stock exchange transactions are conducted in real time and therefore the information is quickly incorporated in the price. Thus, a price or yield formed on the stock exchange is an anticipating indicator. In addition to the price index, the turnover index that is often linked to information asymmetry of market participants is also of particular interest (Chae 2005). Market participants cannot be ideally informed about business operations of entities and therefore they trade considerably less out of caution during financial stress periods. Lower trading values of the BELEX15 index indicate a higher SSI level. The indicators that we will use are realised volatility of the BELEX15 index stated through the closing price and realised volatility of the BELEX15 turnover. If price/trading fluctuations are significant, investors cannot assess with certainty the company they invest in. A high uncertainty level is interpreted as a stress signal and results in an increase in the SSI value.

The effects of financial crises on public finances are reflected in a decrease in public revenue due to economic downturn. At the same time, there is an increase in expenditures and deficit, which is covered with higher government borrowing in domestic and foreign markets, which becomes more expensive in case of financial imbalances. In addition, the necessity for government interventions in order to preserve stability of the economy becomes more pronounced. The global financial crisis faced countries with a challenge to maintain a consumption level that supports future economic growth. In case of longer crisis periods, there is a need to support aggregate supply as it will ease the pressure on public consumption, which is further aggravated if the deficit was high before the crisis. This can result in a sovereign debt crisis that is defined as economic and financial effects triggered by the perception of the government's inability to repay its public debt. It usually occurs when a country reaches a critical level of external debt and low economic growth at the same time. The first indicator of public finances is the difference between the yield to maturity of the domestic 10Y government eurobond and the 10Y German government bond, which implies the risk perception of the Republic of Serbia's debt against Germany's low risk debt. When the yield difference increases, the value of the SSI also increases since a higher potential yield is related to a higher risk. Another indicator is the difference between the quotes of the yield to maturity of the 10Y government eurobond. The difference between yield quotes has an analogous meaning, as explained in the paper. A substantial difference implies that investment in this bond is perceived as risky and results in an increase in the SSI value. High-risk premia incorporated in the difference suggest liquidity risk. Volatility indicators point to investors' uncertainty about the measurement of assets, i.e. the higher the volatility the higher the values of the indicators used for measuring the stress level, which indicates increased market turbulences. Therefore, the SSI calculation includes the realised volatility of the yield to maturity of the 10Y government eurobond. The Emerging Markets Bond Index Global Serbia (EMBI Global Serbia) is one of the sub-indices based on which EMBI Global is formed (Cavanagh and Long 1999) and it has an analogous meaning. Increased risk perception leads to an increased EMBI and consequently a higher SSI. High values of the consolidated fiscal deficit (IMF 2001) in relation to the gross domestic product imply unsustainable public debt of a country and its vulnerability as a debtor, which is for investors often a signal of an elevated fiscal policy sustainability risk. Attinasi et al. (2009) showed that countries facing the highest increase in yield margins on government securities are precisely the countries that had a high deficit-to-GDP ratio in the pre-crisis period. Thus, a constantly high deficit-to-GDP ratio is the indicator of the accumulation of fiscal risks and results in a higher SSI.

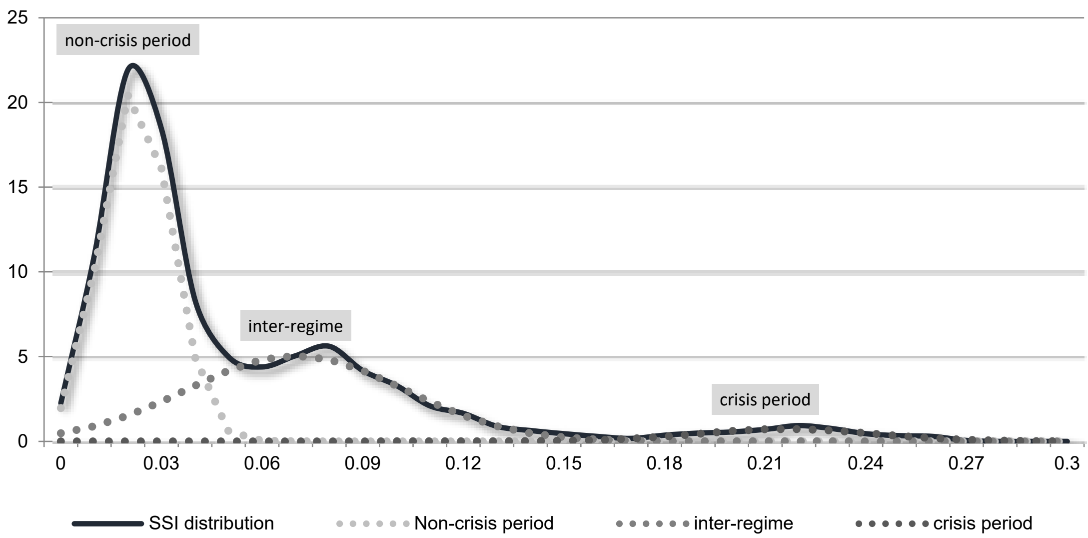
The banking sector constitutes a particularly important segment of the financial system. Taking into consideration that the Serbian financial system is bank-centric, with the banking system's share of more than 90% of financial sector assets, a significant impact on the SSI level is related to this segment. Since only a small number of banks are listed on the Belgrade Stock Exchange, no market indicators, but only regulatory and reporting data were used. The representative indicators used are the deviation or the total deposit gap that represents a cyclical component of deposits, assessed as a one-sided Hodrick-Prescott filter where the lambda parameter takes value 14400 (Louzis and Vouldis 2013). The advantage of applying gaps concerns the fact that they contain information on the cumulative effect of imbalances. In that regard, gaps are a more accurate imbalance measure in comparison to growth rates. If deposits are below their long-term trend at a certain moment, the gap is negative, and banks might face reduced funding sources. On the other hand, a positive deposit gap has a favourable effect on funding sources and liquidity of banks and consequently on the SSI. The next indicator is a credit gap that is also determined as a cyclical component of credit fluctuations using the one-sided Hodrick-Prescott filter. A negative gap implies that banks approve less loans than usual. It may occur due to the tightening of credit standards as a result of an increased risk perception or, on the other hand, reduced demand for loans that is related to dented consumption. In both cases it is an unfavourable signal that increases the SSI value. Taking into consideration a significant foreign ownership of the banking sector, the credit default swap (CDS) of parent banks was used, in which a weighting factor is proportionate to the share in total balance sheet assets of the domestic sector (Dumičić 2014). An assessment of the CDS includes an evaluation of the country's credit rating, fiscal and macroeconomic stability and the credit rating of the observed financial institution that covers operating indicators and a development strategy. The CDS is a leading indicator that captures the dynamics of crisis periods in real time and it therefore may be used to create early warning signals (Podpiera and Ötker 2010). This is supported by the fact that the values of the credit default swap of some countries increased several times in 2007, while financial soundness indicators did not show any risk accumulation signs until 2008.

The last segment used to calculate the SSI is the external environment, which is becoming more important due to market regionalisation and the effects of global trends on small, open economies, such as Serbian. The first indicator that was used to assess the external environment risk is the EMBI Global (Cavanagh and Long 1999). It defines markets of developing countries using the per capita income data and public debt restructuring history. During the aggregation into a composite EMBI Global, instruments that are more relaxed than other indicators are introduced and developing countries are included, which makes the indicator's range wider. Weighting factors per individual countries during the aggregation comply with their market capitalisation. The openness of the Serbian economy and inclusion in international capital flows emphasises the impact of international factors on domestic markets. Thus, it is particularly important to monitor capital markets in neighbouring countries taking into consideration a high correlation with the Serbian capital market, and to include an indicator related to the average value of stock exchange indicators in neighbouring countries as it may point to stress accumulation in these markets and a potential negative impact on domestic markets.

2.2 Assessment of a systemic stress period and the number of regimes

The objective of an analytical framework is to separate high financial stress periods from moderate or low stress periods. The framework assumes that the characteristics of the SSI depend on a space-state representation. For instance, financial stress is assumed to have the tendency to cluster around different local attractor levels in different regimes, showing persistence in different intervals, and only a transition between different regimes is sudden and unpredictable (i.e. stochastic). The SSI characteristics may be identified based on the empirical probability distribution function shown in Figure 1.

Figure 1 **Systemic Stress Indicator expressed as a sum of Gaussian distributions**



The SSI probability distribution is visibly multi-modal and largely asymmetrical towards the right-skewed tail. This feature of the probability distribution function suggests that the empirical probability function may be presented as a sum of appropriate distributions, whereby each of them characterises a separate switching regime. The Gaussian mixture model (Press 2007) in Figure 1 shows moments of parametric distributions presented in Table 1. Gaussian mixture models are parametric functions of the probability density obtained by weighted sums of Gaussian distributions.

Table 1 **Main statistics of Gaussian mixture distributions depending on the number of regimes**

	2	3	4	5
AIC	-655.38	**-688.76**	-685.42	-682.60
BIC	-640.03	**-664.21**	-651.66	-639.64
Likelihood	-332.69	-352.28	-353.71	**-355.31**

The model below provides estimates of Gaussian distribution moments in case of three different switching regimes based on the values of information criteria AIC and BIC, presented in Table 1.

Table 2 **Gaussian distribution moments in case of three regimes**

	Expected value	Standard regime deviation	Weighting function
Non-crisis period	0.0223	5.04E-05	56.2%
Inter-regime	0.0696	5.82E-04	36.2%
Crisis period	0.1929	1.60E-03	7.6%

As it can be seen from a preliminary regime analysis, in addition to an increased expected value, the indicator volatility rises in riskier periods.

Markov-switching models (Hamilton 1989) are one of the most popular non-linear techniques for analysing time series in scientific literature. These models use multidimensional structural equations that explain the behaviour of a space-state representation in case of several switching regimes. These models can also analyse complex dynamic behaviour patterns in time series by allowing switchings between structural equations. The improvement of Markov-switching models is a mechanism that enables switchings between different states with support of an undetected state variable – X^S. A significant advantage of introducing the state variable is that it enables dependence of a model structure change and probabilities of switching from one regime to another.

Starting from the definition of the generalised Markov-switching model:

$$y_t = \sum_{i=1}^{N_{nS}} \beta_i \, x_{i,j}^{N_{nS}} + \sum_{i=1}^{N_S} \Phi_j \, S_t x_{i,j}^{N_S} + \varepsilon, \tag{11}$$

Where N_S denotes switching parameters and N_{nS} non-switching parameters.

A model with only switching parameters included is obtained for $x_{i,j}^{N_{nS}}=0$, i.e.

$$y_t = \sum_{i=1}^{N_S} \Phi_j \, S_t x_{i,j}^{N_S} + \varepsilon \ , \tag{12}$$

where $x_{i,j}^{N_S} = [1 \quad 1 \quad \cdots \quad 1], i = 1,2, \dots , N_S$ and $|S_t|$ is the number of switching regimes. The number of regimes is not *a priori* known. Thus, the distribution of the dependent variable y_t must be assumed $\overline{y_t}$.. Under the assumption of the local normality of a dependent function $(y_i|S_t = i; \theta_i) \sim N(\mu_i^F, \sigma_i^F)$, where $i = 1,2,3$ and θ_i^F represents a set of parameters of an unrestricted or a full regime model, i the density of the conditional probability of the dependent variable is y_i obtained as follows:

$$f(y_t|S_t = i, \theta_i^F) = \frac{1}{2\pi^{n/2}\sigma_i^F} \exp\left(- \frac{\left((y_t - \mu_i^F)\right)}{2\sigma_i^{F^2}} \right) \tag{13}$$

The conditional regime probability j can be expressed through the following equation:

$$P_r(S_t = j|\psi_{t-1}, \theta_j^F) = \sum_{i=1}^{|S_t|} P_r(S_t = j, S_{t-1} = i|\psi_{t-1}^S, \psi_{t-1}', \theta_j^F) =$$

$$= \sum_{i=1}^{|S_t|} P_r(S_t = j|S_{t-1} = i, \psi_{t-1}^S, \theta_j^F) \, P_r(S_{t-1} = i| \, \psi_{t-1}', \theta_i^F), \tag{14}$$

Where $P_r(S_t = j|S_{t-1} = i, \psi_{t-1}^S, \theta_i^F)$ is a conditional probability of a transition from one regime to another based on historical information ψ^S, and θ_i^F represents an assessed vector of transition probabilities parameters in relation to the state variable X^S. A conditional transition probability can be defined by means of probit transformation $\Phi(\cdot)$:

$$P_r(S_t = j|S_{t-1} = i, \psi_{t-1}^S, \theta_j) = \Phi(\boldsymbol{X}^S, \boldsymbol{\psi}^S) \tag{15}$$

When new information on the probabilities of being in a regime j is obtained, the equation (13) can be updated:

$$\mathrm{P}_r(\mathrm{S_t} = j|\psi_t, \theta_j^F, \mathrm{y_t}) = \frac{f(\mathrm{S_t} = j, \mathrm{y_t}|\psi_{t-1}, \theta_j^F)}{f(\mathrm{y_t}|\psi_{t-1}, \theta_i^F)} = \\ = \frac{f(\mathrm{y_t}|\mathrm{S_t} = j|\psi_{t-1}, \theta_i^F) P_r(\mathrm{S_t} = j|\psi_{t-1}, \theta_i^F)}{\sum_{i=1}^{|\mathrm{S_t}|} f(\mathrm{y_t}|\mathrm{S_t} = i|\psi_{t-1}, \theta_i^F) P_r(\mathrm{S_t} = i|\psi_{t-1}, \theta_i^F)} \tag{16}$$

based on which a log-likelihood function is calculated:

$$logL = \sum_{j=1}^{|S_t|} \sum_{i=1}^{|S_t|} f(\mathrm{y_t}|\mathrm{S_t} = j, \mathrm{S_{t-1}} = i, \psi_{t-1}^S, \theta_j^F) \cdot \mathrm{P}_r(\mathrm{S_t} = j, \mathrm{S_{t-1}} = i|\psi_{t-1}^S, \theta_j^F) \tag{17}$$

A set of final model parameters is obtained by maximising the log-likelihood function $\max_{\theta_j^F}(\log L)$.

In a generalised case, a set of switching regime equations is obtained:

$$\begin{aligned} y_1 &= \mu_1 + \varepsilon_{1t}; \varepsilon_{1t} \sim \mathrm{N}(0, \sigma_1) \\ y_2 &= \mu_2 + \varepsilon_{2t}; \varepsilon_{1t} \sim \mathrm{N}(0, \sigma_2) \\ &\vdots \\ y_{|\mathrm{S}_t|} &= \mu_{|\mathrm{S}_t|} + \varepsilon_{|\mathrm{S}_t|t}; \varepsilon_{1t} \sim \mathrm{N}(0, \sigma_{|\mathrm{S}_t|}) \end{aligned} \tag{18}$$

whereas a transition regime matrix at the moment t is defined as:

$$P_t = \begin{bmatrix} p_{(1|1)}^t(\mathbf{X}^\mathrm{S}, \boldsymbol{\psi}^S) & p_{(1|2)}^t(\mathbf{X}^\mathrm{S}, \boldsymbol{\psi}^S) & \cdots & p_{(1|\mathrm{S}_t)}^t(\mathbf{X}^\mathrm{S}, \boldsymbol{\psi}^S) \\ p_{(2|1)}^t(\mathbf{X}^\mathrm{S}, \boldsymbol{\psi}^S) & p_{(2|2)}^t(\mathbf{X}^\mathrm{S}, \boldsymbol{\psi}^S) & \cdots & p_{(2|\mathrm{S}_t)}^t(\mathbf{X}^\mathrm{S}, \boldsymbol{\psi}^S) \\ \vdots & \vdots & & \vdots \\ p_{(\mathrm{S}_t|1)}^t(\mathbf{X}^\mathrm{S}, \boldsymbol{\psi}^S) & p_{(\mathrm{S}_t|2)}^t(\mathbf{X}^\mathrm{S}, \boldsymbol{\psi}^S) & & p_{(\mathrm{S}_t|\mathrm{S}_t)}^t(\mathbf{X}^\mathrm{S}, \boldsymbol{\psi}^S) \end{bmatrix} \tag{19}$$

Where X^S denotes a state variable vector and ψ^S denotes a system parameter vector. Detailed information on the expectation-maximisation algorithm used to assess model parameters can be found in Dempster et al. (1977)..

2.3 Analysis of SSI movements

Different specifications using a lagged dependent variable $\psi_t^\mathrm{S} = \mathrm{y_{t\text{-}h}}$ as a state variable for different number of regimes i are contained in Appendix 3. An auto-regression model with a unit lag follows more adequately the dynamics of the SSI movement during a crisis period, but caution is advised when including a dependent variable in the form of auto-regression terms in the equation (11). Inclusion of an auto-regression term and enabling its coefficients to have a switching property allow the smallest indicator variations, even at extremely low values, i.e. when an indicator is in the least risky regime, to be interpreted as stress episodes. As a rule, these models have better values of the information criteria AIC and BIC and the values of maximum likelihood. Numerous transitions between opposite regimes (e.g. from a non-crisis regime to the crisis one) indicate model's instability. Moreover, transition probability matrices show the tendency of not having maximum transition probabilities diagonalwise. This problem is mainly related to the inability of an optimisation algorithm to converge to a global optimum in case of a multi-modal problem or a large number of parameters being optimised. Therefore, during an assessment of the likelihood function in equation (17) the approach (Kovačević et al. 2014) was used, enabling the finding of a global optimum, even in case of a large number of parameters. Based on equations (4) to (7), as well as equation (17), a weighting vector W, per defined financial system segments, was obtained as:

$$\begin{aligned} w_{FX} &= 0.202 \\ w_{GOV} &= 0.125 \\ w_{MON} &= 0.229 \\ w_{EQU} &= 0.125 \\ w_{BAN} &= 0.194 \end{aligned} \tag{20}$$

Based on the weighting vector shown in (20), it is evident that the money market, FX market and the banking sector segments have the strongest estimated impact on the SSI, whereas the public finance segment, capital market and external environment have slightly lower values.

The specification of the preferred model is represented by an auto-regression term with a unit-lag of state variable $\text{SAR}(1)$ with three switching regimes in which a free member and the residual variance allow for switching parameters. We will name this model a full dynamic Markov model with three switching regimes, with a unit-lag of state variable F-DMS(3)-SAR(1). The preferred model has the best statistics related to information criteria values, as seen in Appendix 3. Additionally, the presented results also include a regime classification measure, proposed in the paper by Ang and Bekaert (2002) and Baele (2005), in case there are more than two regimes, which is obtained based on the following equation:

$$\text{RCM}(|S_t|) = 100\left(1 - \frac{|S_t|}{|S_t|-1}\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{|S_t|}\left(p_{i,t} - \frac{1}{|S_t|}\right)^2\right) \tag{21}$$

Lower RCM values denote a better regime separation ability. As it can be seen in Appendix 3, a full model F-DMS(3)-SAR(1) had the best regime classification results. In case of three switching regimes, a generalised Markov-switching model is obtained, based on the equations listed in Chapter 2.2.

$y_1 = 2.31\text{E-}02 + \varepsilon_{1t};\ \varepsilon_{1t} \sim N(0, 6.30\text{E-}05)$ for non-crisis period
$y_2 = 7.40\text{E-}02 + \varepsilon_{2t};\ \varepsilon_{2t} \sim N(0, 3.95\text{E-}04)$ for inter-regime
$y_3 = 1.87\text{E-}01 + \varepsilon_{3t};\ \varepsilon_{3t} \sim N(0, 1.86\text{E-}03)$ for crisis period

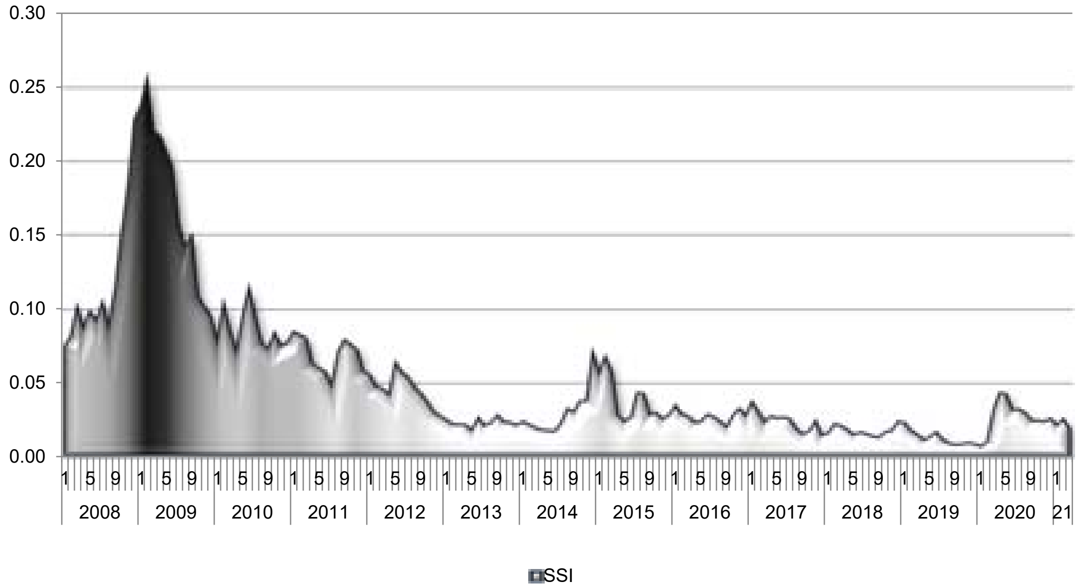
Since it is a full model F-DMS(3)-SAR(1), we will designate the model parameters as ***F***:

$$\mu_1^F = 2.31\text{E} - 02, \delta_1^F = 6.30E - 05$$

$$\mu_2^F = 7.40E - 02, \delta_2^F = 3.95E - 04$$

$$\mu_3^F = 1.87E - 01, \delta_3^F = 1.86E - 03$$

Figure 2 **Movement of SSI**



The assessed expected value and volatility differ statistically in different regimes. The values also comply with a preliminary analysis that was obtained using the Gaussian mixture model shown in Table 2. The expected value level is the lowest in case of a non-crisis period: $\mu_1^F = 2.31\text{E-}02$. In case of a crisis period, it is $\mu_3^F = 1.87\text{E-}01$, which complies with the expected behaviour of the SSI, i.e. the higher the regime risk the higher the SSI values. Correspondingly, volatility justifies economic fundamentals, where higher values of indicator fluctuations are expected, reflecting increased uncertainty in crisis periods. Volatility in a crisis period is $\delta_3^F = 1{,}86\text{E-}03$ and it is considerably higher compared to a non-crisis or an inter-regime period. In order to look into regime dynamics more closely, we have to identify the main characteristics of each period.

Figure 3 **Conditional regime probabilities using the F-DMS(3) SAR (1) model**

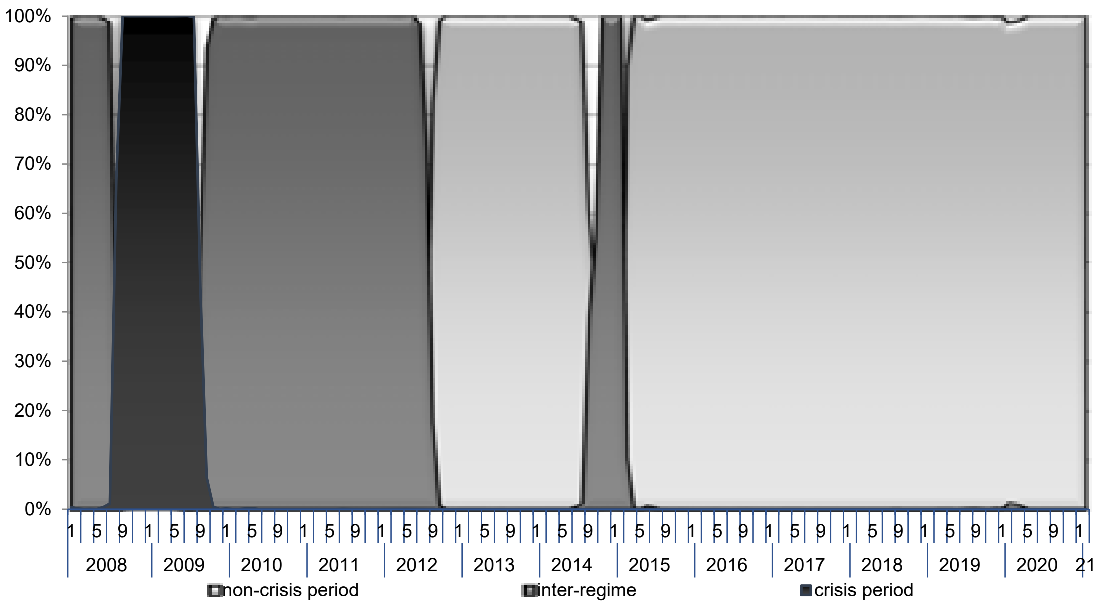
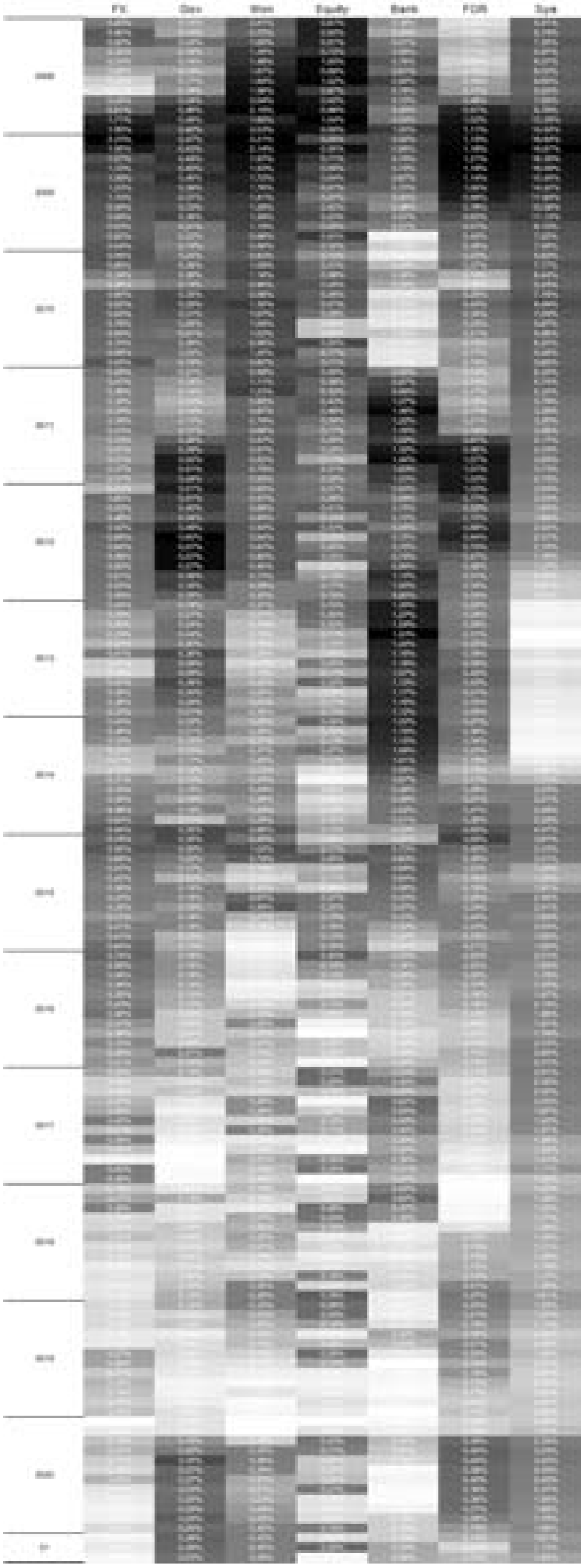


Figure 4 **Contribution of segments to the composite SSI**



2.4 Analysis of risk factors and stress level in the observed period

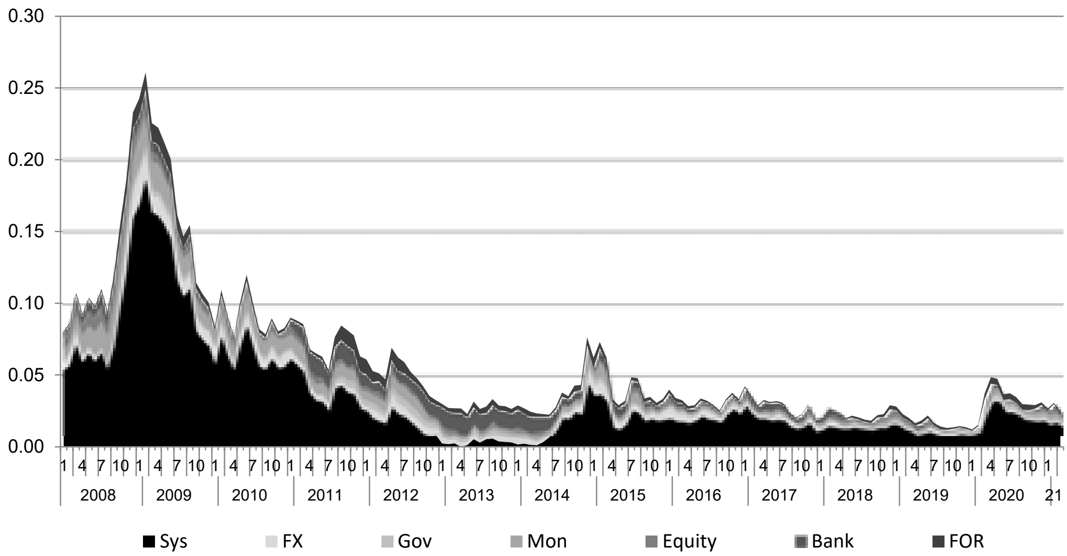
The SSI defines a pre-crisis period from January 2008 to September 2008. The first shocks in the domestic financial system first appeared in money and capital markets. The average monthly BEONIA value increased from 9.4% in January 2008 to over 17% in November and December 2008. Simultaneously, there was a reduction of excess dinar liquidity of banks, whereas the BELEX15 index lost over 70% of its value at the time.

The period between October 2008 and October 2009 was marked by the escalation of the global financial crisis. In the last months of 2008 we witnessed the worst financial crisis since the Great Depression of the 1930s. In January 2009 the prices of shares in all major global markets unexpectedly collapsed. A large number of US and European banks recorded enormous losses in their financial statements. Lehman Brothers went bankrupt in September 2008, while Goldman Sacks and Morgan Stanley sought state aid in order not to declare bankruptcy. The Federal Reserve System injected hundreds of billions of dollars as state aid into the banking system. The financial crisis spilled over into the European financial system for the first time since the introduction of the euro in 1999 as the official currency. Facing panic, regulators across Europe adopted additional measures of monetary and fiscal policies in order to ease the effects of the coming recession. At the same time, most central banks in Europe reduced their key policy rates.

The initial shocks in the domestic financial system in the first half of 2008 continued to deepen and intensify. The local currency weakened against the euro in the FX market by over 10% y-o-y, while the spread between the buying and selling rates increased, reflecting FX market uncertainty.

Public finance also suffered significant shocks. In December 2008 Serbia recorded historically the highest value of an emerging market bond risk premium of over 1,300 bp. At the same time, the consolidated fiscal deficit widened, equalling almost 4.2% of gross domestic product at the end of 2009. The average monthly BEONIA value reached its maximum of 18.2% in January 2009. Stock market indices saw a sharp drop. Over a period of eleven months, BELEX 15 dropped from over 1800 (June 2008) to 400 in April 2009. The banking sector saw a liquidity crisis as 6.2% of total deposits or RSD 1.3 billion was withdrawn from domestic banks. Thanks to the support of parent banks to their subsidiaries operating in the Republic of Serbia, the liquidity position was preserved, and major disruptions were prevented. The neighbouring countries also faced considerable shocks. Stock market indices in the neighbouring countries lost on average 55% of their value compared to January 2008, while the Emerging Market Bond Index Global for Serbia reached its maximum value in December 2008. Figures 2 and Figures 4 show that the SSI and the systemic component reached their historical maximum in February 2009 – the SSI peaked to 0.26, when the crisis culminated in the Republic of Serbia.

Figure 5 **Movement of the SSI by component and the breakdown of key risk factors**



The global economic crisis gradually subsided between October 2009 and August 2012 and assumed the local character of a sovereign debt crisis. A eurozone crisis broke out in late 2009. A few eurozone member states (Greece, Portugal, Ireland, Spain and Cyprus) were not able to repay or restructure their government debts or to inject sizeable funds into their financial systems without the support of other eurozone member states or international financial institutions. The European Central Bank responded by reducing interest rates and providing favourable loans in order to ensure the eurozone's financial system liquidity. In September 2012 the European Central Bank calmed financial markets by announcing it would provide unrestricted support to eurozone member states. The domestic financial system was now not as affected as it had been at the beginning of 2009. The SSI level visibly increased in June 2010, September 2011 and May 2012. The volatility in the domestic FX market increased in June 2010, which resulted in robust FX interventions by the National Bank of Serbia. The banking sector saw a decline in the credit activity and an increase in Greek-owned parent banks' risk due to an increase in Greek sovereign risk.

The next significant increase in the SSI took place in September 2011 mainly due to factors in the international environment – sovereign risk, particularly Greek and Italian, went up, spilling over to a increase in risk of banking groups in these countries. Moreover, there was a jump in the global emerging market bond index, indicating an increased investor risk perception of emerging markets. There was another mid-stress period in June 2012, when the domestic currency significantly dropped by 11.5% y-o-y. Simultaneously, for the first time since the crisis onset, the consolidated fiscal deficit exceeded 6%, while the average credit risk insurance premium of parent banks, particularly Greek and Cyprus banks, was at its peak, reflecting the sovereign credit risk.

The period between December 2012 and the beginning of 2020 was marked by a low risk with stable system risk values, except for a mid-stress period from September 2014 to April 2015. It was the period of a stable exchange rate and low and stable inflation. The BELEX15 index also stabilised. The banking sector was adequately capitalised and highly liquid, with lower credit activity during 2013 and early 2014. Individual episodes of increased stress occurred in some segments but did not significantly affect the SSI. The consolidated fiscal deficit rate reached 6.2% at the end of 2014 and started dropping afterwards, while EMBI Serbia reached somewhat higher values in comparison to the previous period.

The pandemic triggered by the coronavirus outbreak in early 2020 resulted in an unprecedented health and economic global crisis. To protect people’s lives, measures were adopted, accompanied with disruptions in international financial and commodity markets, elevated uncertainty and flight to safe-haven assets, which all together resulted in a sharp drop in the global economic activity. Due to increased uncertainty in the second quarter of 2020, an increase in the SSI was mainly affected by foreign market developments. The EMBI risk premium returned to a downward path at the end of April 2020, when it peaked at 312 bp (its highest value in 2020) and amounted to 137 bp at the end of the second quarter. A relatively stable exchange rate was maintained throughout 2020 due to timely and adequate FX interventions. Moderate depreciation pressures, in place since March as a result of the pandemic outbreak, gradually weakened during the year. In November and December 2020 appreciation pressures prevailed. An extensive set of economic measures introduced to mitigate negative effects of the pandemic resulted in an increase in the fiscal deficit and public debt, which was still among the smallest in Europe. The public debt-to-GDP ratio increased from 52.0% at the end of 2019 to 57.4% at the end of 2020. The recovery of the global economy that started in mid-2020 was slowed down by the second coronavirus wave in October 2020, which required the re-instatement of restrictive measures in many countries. Solid macroeconomic fundamentals and large-scale monetary and fiscal policy measures enabled the Republic of Serbia to remedy the consequences of the coronavirus epidemic more efficiently than other European countries.

3 Assessment of the likelihood of a systemic crisis

A large number of scientific papers have recently dealt with how to improve the early warning models, mainly by means of developing new analytical tools. Apart from traditional binary models that apply Logit or probit transformations (Berg and Pattillo 1998), literature offers multidimensional tools that use a large number of indicators when developing early warning indicators (Rose and Spiegel 2012, Bussiere and Fratzscher 2006, Frankel and

Saravelos 2010), by applying machine learning (Kou et al. 2019), while the application of Markov switching models is described in detail in Abiad (2003). A paper by Kliesen et al. (2012) emphasises the fact that financial crises are almost impossible to predict. They are instigated by various triggers and occur in various forms and thus new crises are difficult to predict by relying on traditional warning models only. An additional problem is that financial stress cannot be directly measured as it does not have a money equivalent.

We will apply the so-called univariate signalling approach that is based on an assessment of a signal that appeared during previous crisis periods, based on which critical values, above which the indicator points to a strong possibility of a crisis occurrence, are identified. Such an approach enables a simpler interpretation of the achieved results. A more detailed analysis of the types of approaches to creating early warning signals can be found in Kaminsky (1999).

In this paper, we propose the early warning signal approach that is based on the likelihood of a systemic crisis on the basis of the methodology described in the previous chapter. Unlike binary models, this approach enables an analysis of a larger number of regimes that can be classified, according to their intensity, as more or less risky. Although an ex-post identification using the SSI is a good choice, in a general case of undetected crisis periods, i.e. periods that can be estimated to be crisis periods only over time and during the identification of key crisis factors in the past and in drawing conclusions about periods of higher turbulences, early warning signals, on the other hand, raise awareness among decision makers of a potential systemic crisis occurrence and offer sufficient time for a timely response.

We will define a restricted model based on the unrestricted/full model described in the previous chapter. The main characteristic of dynamic Markov-switching models are conditional state transition matrices. Unlike static Markov models, probabilities of making transitions from one regime to another are time-dependent on the endogenous state variable X^S. Core building blocks for obtaining dynamic probabilities are presented in the paper by Diebold et al. (1994). Let us assume that the dependent variable can be defined as:

$$(y_i|S_t = i; \theta_i) \sim N(\mu_i^F, \sigma_i^F), \tag{22}$$

where $i = 1,2,3$ based on the results shown in Table 1 and Table 8, and θ_i represents a set of parameters of the regime i.

A restricted model is obtained using the assessed parameters of the full model that refer to distribution moments of the dependent variable y_t. Model parameters are divided into two subsets. The first subset is related to the parameters of distribution moments of the dependent variable - α_i^F of the regime i, while the second subset refers to the state variable parameters that are used when estimating conditional transition probabilities β_i^F, i.e. $\theta_i^F = (\alpha_i^F, \beta_i^F)$.

A restricted model is obtained by maximising log-likelihood functions based on lagged state variable h:

$$logL = \sum_{j=1}^{|S_t|} \sum_{i=1}^{|S_t|} f\left(y_t|S_t = j, S_{t-1} = i, \psi_{t-h}^S, (\alpha_i^F, \beta_i^R)\right) \cdot P_r\left(S_t = j, S_{t-1} = i|\psi_{t-h}^S, (\alpha_i^F, \beta_i^R)\right), \tag{23}$$

where $|\cdot|$ denotes the cardinality of the set.

The maximisation of the log-likelihood function results in a set of parameters β_i^R used in the restricted model.

A conditional probability of the regime j based on historical information ψ_{t-h}, on the basis of full model parameters $\left(\alpha_i^F, \beta_i^R\right)$ is obtained as follows:

$$P_r\left(S_{t+1}=j|\psi_{t-h},\left(\alpha_i^F,\beta_i^R\right)\right)=\sum_{i=1}^{|S_t|}P_r\left(S_{t+1}=j,S_t=i|\psi_{t-h}^S,\psi_{t-h}',\left(\alpha_i^F,\beta_i^R\right)\right)= \tag{24}$$

$$=\sum_{i=1}^{|S_t|}P_r\left(S_{t+1}=j|S_t=i,\psi_{t-h}^S,\left(\beta_i^R\right)\right)P_r\left(S_t=i|\,\psi_{t-h}',\left(\alpha_i^F\right)\right)$$

We use recursion to obtain a conditional probability of the regime j at the moment $t+h$ based on historical information ψ_t :

$$P_r\left(S_{t+h}=j|\psi_t\;,\left(\alpha_i^F,\beta_i^R\right)\right)=\sum_{i=1}^{|S_t|}P_r\left(S_{t+h}=j,S_{t+h-1}=i|\psi_t^S,\psi_t',\left(\alpha_i^F,\beta_i^R\right)\right)= \tag{25}$$

$$=\sum_{i=1}^{|S_t|}P_r\left(S_{t+h}=j|S_{t+h-1}=i,\psi_t^S,\left(\beta_i^R\right)\right)P_r\left(S_{t+h-1}=i|\,\psi_t',\left(\alpha_i^F\right)\right),$$

Where $P_r\left(S_{t+1}=j|S_t=i,\psi_t^S,\left(\alpha_i^F,\beta_i^R\right)\right)$ represents a conditional probability of transition from one regime to another based on historical information and state variable parameters ψ^S, β_i^R is an assessed vector of transition probabilities in relation to state variables X^S.

In a general case, a conditional transition probability can be defined using a probit transformation, as in the equation (15):

$$P_r\left(S_{t+1}=j|S_t=i,\psi_{t-h}^S,\left(\beta_i^R\right)\right)=\Phi\left(\mathbf{X}^S,\boldsymbol{\beta}^R\right). \tag{26}$$

If there are more than two regimes, we can use clustering to define a set of non-crisis regimes $\{nc\}$ and a set of crisis regimes $\{c\}$. Let us observe a transition matrix as follows:

$$P_r(\cdot\,|\,\cdot)=\begin{bmatrix}P_r(S_{t+1}=\{nc\}|S_t=\{nc\},\cdot) & P_r(S_{t+1}=\{nc\}|S_t=\{c\},\cdot)\\ P_r(S_{t+1}=\{c\}|S_t=\{c\},\cdot) & P_r(S_{t+1}=\{c\}|S_t=\{c\},\cdot)\end{bmatrix} \tag{27}$$

A conditional transition probability of staying in non-crisis regimes $\{nc\}$ at the time $t+1$ can be calculated as:

$$P_r(S_{t+1}=\{nc\}|S_t=\{nc\},\cdot)=\frac{\sum_{j=1}^{|S_{nc}|}\sum_{i=1}^{|S_{nc}|}P_r\,(S_t=i|\,\cdot)P_r(S_{t+1}=j|S_t=i,\cdot)}{\sum_{i=1}^{|S_{nc}|}P_r\,(S_t=i|\,\cdot)} \tag{28}$$

At the same time, a conditional transition probability of staying in crisis regimes $\{c\}$ at the time $t+1$ can be denoted as:

$$P_r(S_{t+1} = \{c\}|S_t = \{c\},\cdot) = \frac{\sum_{\mathrm{j}=1}^{|S_c|}\sum_{i=1}^{|S_c|} P_r\,(\mathrm{S_t} = \mathrm{i}|\,\cdot)P_r(S_{t+1} = j|S_t = i,\cdot)}{\sum_{i=1}^{|S_c|} P_r\,(\mathrm{S_t} = \mathrm{i}|\,\cdot)} \tag{29}$$

A conditional probability of transitions from non-crisis regimes $\{\mathrm{nc}\}$ to crisis ones $\{\mathrm{c}\}$ at the time $\mathrm{t}+1$ can be calculated as:

$$P_r(S_{t+1} = \{c\}|S_t = \{nc\},\cdot) = \frac{\sum_{\mathrm{j}=1}^{|S_c|}\sum_{i=1}^{|S_{nc}|} P_r\,(\mathrm{S_t} = \mathrm{i}|\,\cdot)P_r(S_{t+1} = j|S_t = i,\cdot)}{\sum_{i=1}^{|S_{nc}|} P_r\,(\mathrm{S_t} = \mathrm{i}|\,\cdot)} \tag{30}$$

A conditional probability of a transition from crisis regimes $\{\mathrm{c}\}$ to non-crisis ones $\{\mathrm{nc}\}$ at the time $\mathrm{t}+1$ is obtained trivially based on the equation (30).

We will name the probability that at least one crisis period will occur at the time interval $t+\gamma+1,\ t+\gamma+2,\ldots,\ t+\gamma+hor$ a probability of the occurrence of a systemic crisis PSC which can unambiguously be calculated as:

$$\mathrm{PSC}_{t+\gamma} = 1\text{-}\Big(\mathrm{P_r}\big(\mathrm{S}_{t+\gamma} = \{\mathrm{c}\}|\mathrm{S}_{t+\gamma\text{-}1} = \{\mathrm{nc}\},\cdot\big)\mathrm{P_r}\big(\mathrm{S}_{t+\gamma} = \{\mathrm{c}\},\cdot\big)\mathrm{P_r}\big(\mathrm{S}_{t+\gamma} = \{\mathrm{nc}\}|\mathrm{S}_{t+\gamma\text{-}1} = \{\mathrm{nc}\},\cdot\big)^{(\mathrm{hor}\text{-}1)} + \mathrm{P_r}\big(\mathrm{S}_{t+\gamma} = \{\mathrm{nc}\},\cdot\big)\mathrm{P_r}\big(\mathrm{S}_{t+\gamma} = \{\mathrm{nc}\}|\mathrm{S}_{t+\gamma\text{-}1} = \{\mathrm{nc}\},\cdot\big)^{\mathrm{hor}}\Big) \tag{31}$$

where $\gamma = 1,2,\ldots,\mathrm{h}$ and $\mathrm{P_r}\big(\mathrm{S}_{t+\gamma} = \{\mathrm{c}\},\cdot\big)$ is calculated recursively using the equation (25).

The equation (31) indicates that the probability of occurrence of a systemic crisis $\mathrm{PSC}_{t+\gamma}$ at the time $\mathrm{t}+\gamma$ represents an assessment of conditional probabilities that are based on the uncertainty of the regime in which we are at the previous moment $\mathrm{t}+\gamma\text{-}1$.

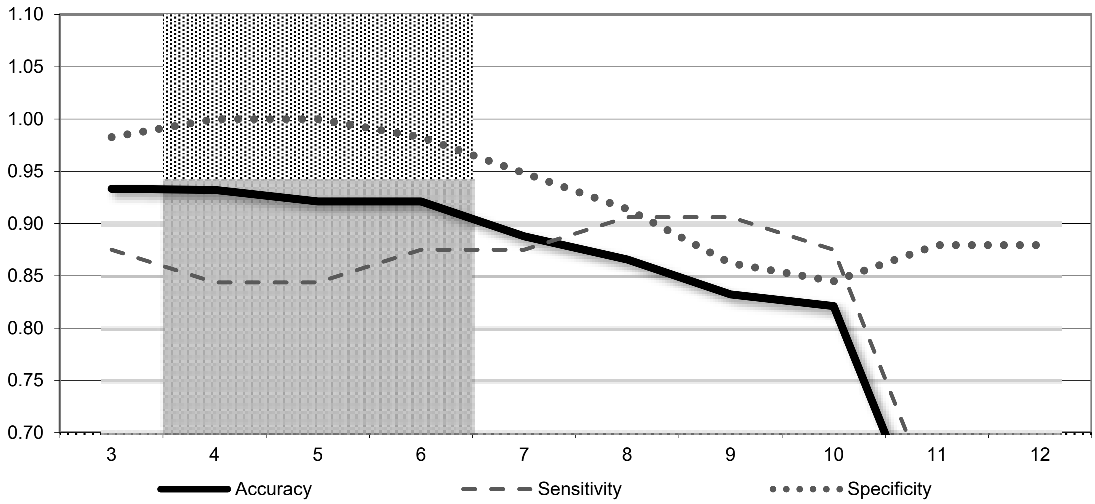
3.1 Critical levels of the likelihood of a systemic crisis occurrence and optimal forecast horizon

Appendix 4 details restricted models in which a lag h in a dependent state variable changes at an interval ranging between 3 and 12 months in case there are three regimes according to the analysis conducted in previous chapters. Prediction values below three months were not taken into consideration due to a short prediction horizon, which diminishes the usability of an early warning system. As it could be seen in the case of shorter lags, the values of the estimated likelihood are higher. Hence, in case of a three-month lag, the likelihood is 281.5, and 250.46 in case of a twelve-month lag. In addition, a measure of the regime classification quality – RCM decreases as the dependent variable lag increases, which reaches a minimum of 0.0194 in case of a ten-month lag. The values of the information criteria AIC and BIC decrease if the lag increases. In order to select the best model specification, the quality of each model classification needs to be assessed.

In three-regimes model, the equation (31) is used to calculate the probability that at least one crisis period will occur at the interval $t+\gamma+1, t+\gamma+2,\ldots, t+\gamma+$ hor for the time horizon hor = 6 months. We will name this probability of restricted models $\mathrm{PSC}^{\mathrm{R}}$. Table 11 in Table11 shows the upper and lower critical values of the probability of occurrence of the systemic crisis $\mathrm{PSC}^{\mathrm{R}}$, whereby we reach the maximum classification accuracy of the regimes shown in Figure 3, which were obtained based on full model probabilities.

As it can be seen in Table 11 in Table11, lower critical limits b_l in the case of all models are similar to the zero value and range at the interval of $[0{,}01-0{,}1]$. On the other hand, upper critical limits b_u, which are used to separate a non-crisis period from an inter-regime period, have a considerably wider interval $[0{,}15-0{,}8]$. Based on the accuracy of a multi-class classification (Sokolova and Lapalme 2009), we can see that the maximum classification accuracy significantly drops in case of state variable lags that exceed 10 months. In case of lags of 11 and 12 months, relevant maximum classification accuracies are around 0.57. Moreover, in case of these lags, the classification sensitivity is reduced to low values, which leads to the conclusion that an optimal prediction horizon must be strictly below 11 months. The classification specificity in the case of all models ranges from 0.84 up to the theoretical maximum value of 1.

Figure 6 **Assessment of multi-class classification parameters depending on a state variable lag**



A compromise between the classification accuracy and prediction horizon should be made. Figure 6 shows the accuracy, sensitivity and specificity of a classification depending on a prediction horizon and/or a state variable lag. It can be seen that in case of lags of up to six months, classification accuracy values of above 90% are obtained. In addition, in case of a six-month lag, classification sensitivity and specificity have high values, which justifies the use of the six-month lag and critical regime classification levels of 0.005 for the lower critical classification level and 0.760 for the upper one. Figure 16 in Appendix 4 shows families of classification accuracy curves in relation to the movement of the upper critical value in case of a six-month lag for the lower critical values set at 0.005, 0.1 and 0.2 respectively.

3.2 Assessment of the likelihood of a systemic crisis occurrence

Table 3 shows the values of transition probabilities based on the full model F-DMS(3) SAR (1) for the last moment in time t of the sample, i.e. March 2021, based on which, using the equations from (26) to (31), a transition matrix forecast was obtained at the interval of $t+1, t+2, \dots, t+6$ and shown in Table 4.

Table 3 **A transition matrix of the model F-DMS(3) SAR (1) in March 2021**

	Non-crisis	Inter-regime	Crisis
Non-crisis	98.9%	24.0%	16.8%
Inter-regime	1.1%	55.8%	34.1%
Crisis	0.0%	20.2%	49.1%

Transition probabilities behave in accordance with the expectation that the highest values are positioned diagonally, i.e. the probability of staying in the same regime is the highest. The probability of a transition from a crisis period to a non-crisis one is 16.8%, whereas the probability of a transition from a non-crisis period to a crisis one is close to 0%. Remote regimes have lower transition probabilities than the neighbouring regimes, which is in line with the expected behaviour of transition matrices. The forecasts of transition probabilities on the basis of the restricted three-regimes model with a six-month state variable lag R-DMS(3) SAR (6) show the same characteristics, which are presented in Table 4.

Figure 7 **Early warning signal of the full model F-DMS(3) SAR (1) and the restricted model R-DMS(3) SAR (6)**

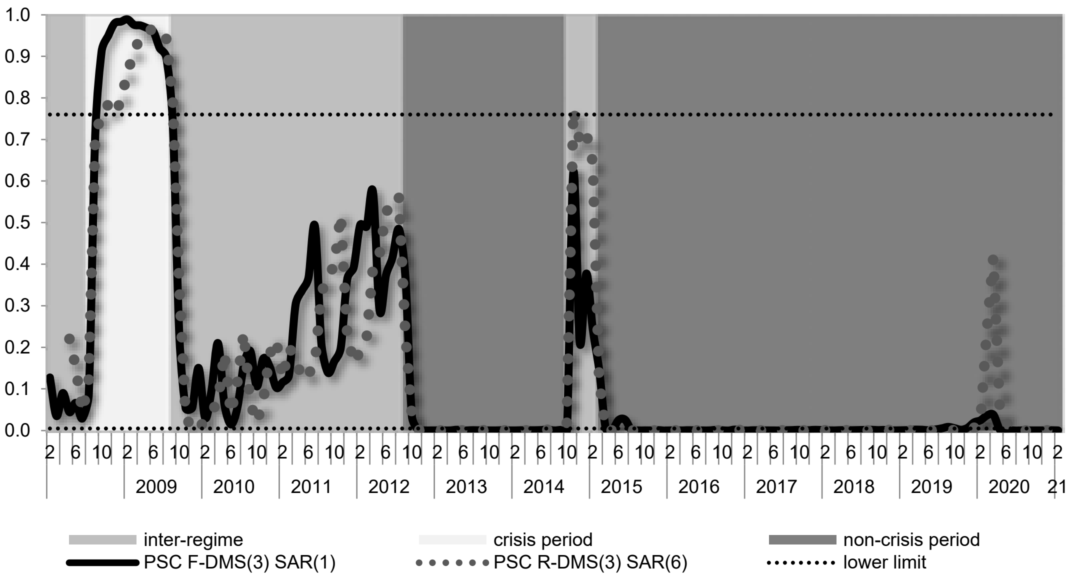


Figure 7 shows the movement of the likelihood of a systemic crisis occurrence based on the full model F-DMS(3) SAR (1) and the restricted model R-DMS(3) SAR (6). As it can be seen, the R-DMS(3) SAR (6) probabilities are in sync with the full model dynamics. However, probability inconsistencies can be detected in the inter-regime between November 2009 and August 2012, where a four-month lag in the restricted model probability is identified. Immediately after the crisis period, the restricted model probability dropped to low values in November 2009. However, this month was not incorrectly classified as a non-crisis one due to a low value of the lower critical probability level of 0.005. A lack of high probability values between October 2008 and February 2009 is also evident, but the level is above the upper critical value, which is why this period was correctly classified as a crisis period. There was also a slightly higher probability of a systemic crisis occurrence in December 2014, when the restricted model classified this month under a crisis one, while the full model probability had lower values that indicated an inter-regime. There was an apparent increase in the early warning signal of the reduced model in April and May 2020 to 0.241 and 0.416 respectively, but its levels were significantly lower than the upper critical limit values. Thereafter, the signal was on a continuous decline until April 2020. Both models classified non-crisis periods identically.

Table 4 **Forecast values of conditional transition probabilities from April 2021 to September 2021**

F-DMS(3) SAR (6)		Non-crisis	Inter-regime	Crisis
April 2021	Non-crisis	99.7%	25.1%	17.9%
	Inter-regime	0.3%	53.7%	35.4%
	Crisis	0.0%	21.2%	46.7%
May 2021	Non-crisis	99.6%	25.4%	18.2%
	Inter-regime	0.4%	53.3%	35.3%
	Crisis	0.0%	21.3%	46.5%
June 2021	Non-crisis	99.8%	23.7%	16.4%
	Inter-regime	0.2%	55.6%	35.6%
	Crisis	0.0%	20.6%	48.0%
July 2021	Non-crisis	99.2%	27.6%	20.8%
	Inter-regime	0.8%	50.3%	34.8%
	Crisis	0.0%	22.1%	44.5%
August 2021	Non-crisis	99.8%	24.0%	16.7%
	Inter-regime	0.2%	55.2%	35.6%
	Crisis	0.0%	20.8%	47.7%
September 2021	Non-crisis	98.9%	28.7%	22.1%
	Inter-regime	1.1%	48.8%	34.4%
	Crisis	0.0%	22.5%	43.4%

Figure 8 shows a regime probability forecast $\mathrm{P_r}(\mathrm{S_t} = \mathrm{j}|\cdot)$ in an observed six-month horizon, i.e. the period from April 2021 to September 2021, projected based on the equation (25). As it can be seen, an increase in the probability of being in a crisis period $\mathrm{P_r}(\mathrm{S_t} = 3|\cdot)$ was projected during the forecast period. A drop in a non-crisis regime probability $\mathrm{P_r}(\mathrm{S_t} = 1|\cdot)$ from 99.5% in the first forecast month to 98% in the sixth month was also projected in this period. At the same time, an inter-crisis regime probability $\mathrm{P_r}(\mathrm{S_t} = 2|\cdot)$ increased.

Figure 9 shows historical movement of the assessed Probability of a Systemic Crisis occurrence at any time over the next six months – PSC based on the full model PSC F-DMS(3) SAR (1) between January 2016 and March 2021. The probability forecast of a systemic crisis occurrence at any time between April and September 2021 on the basis of the restricted model PSC R-DMS(3) SAR(6) was obtained based on the equation (31). A maximum value was obtained in October 2021 with a 2.3% probability of a systemic crisis at any time between September 2021 and March 2022. These are low values that do not indicate a significant crisis occurrence probability in the observed period.

Figure 8 **Forecast regime probability movement in the observed six-month time horizon**

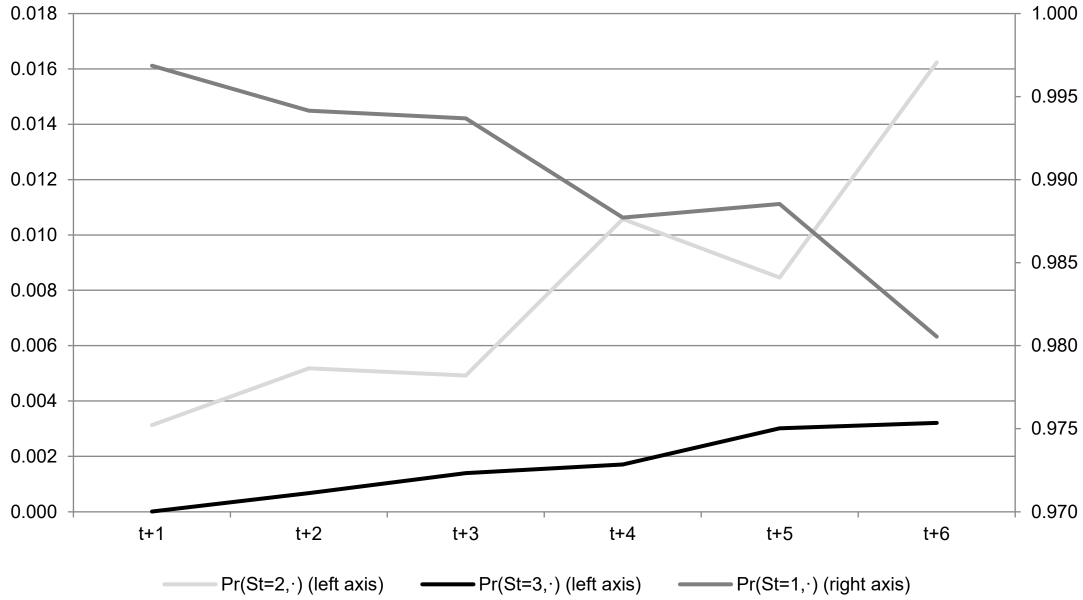
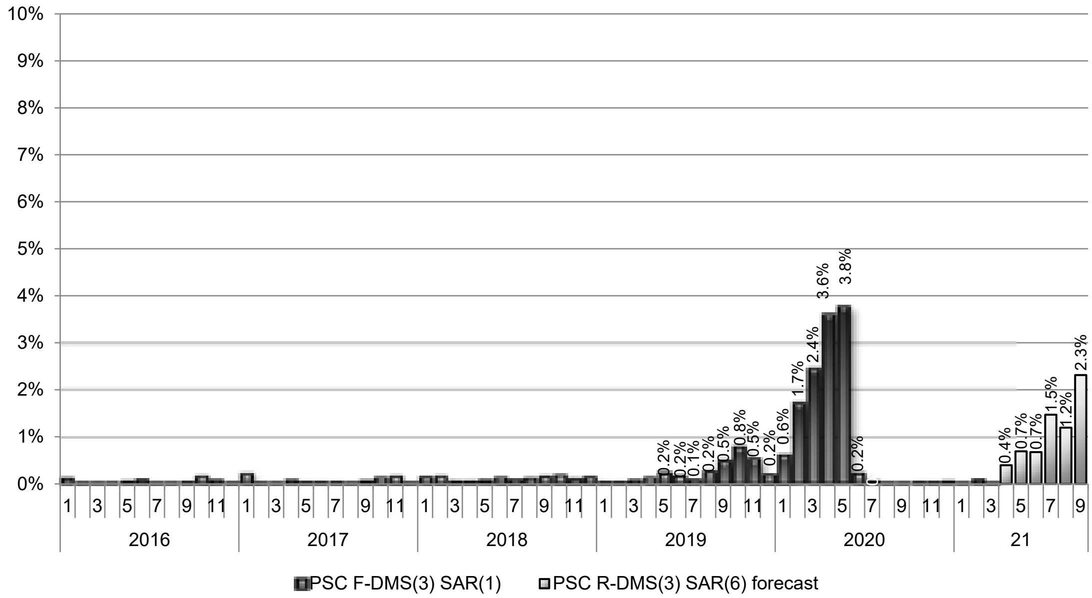


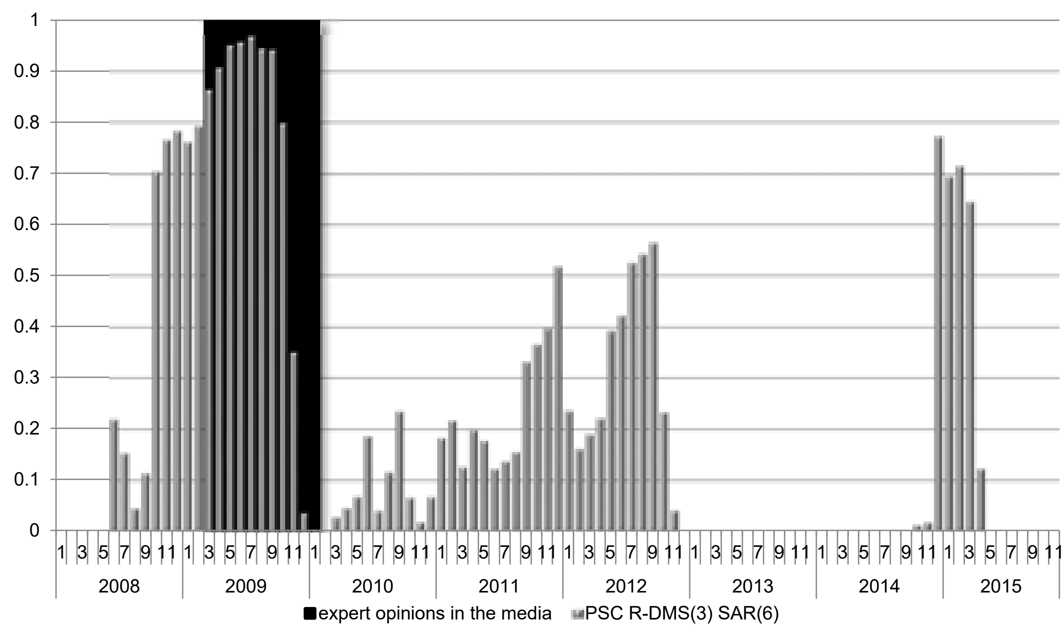
Figure 9 **Early warning signal movements of the full model PSC F-DMS(3) SAR (1) between January 2016 and February 2021 and a restricted model forecast PSC R-DMS(3) SAR(6) between March and September 2021**



3.3 Verification of an early warning signal of a systemic crisis occurrence

Figure 10 shows the movement of a systemic crisis probability, whereas the black rectangle indicates media statements on risk perception made by economic policy makers. Detailed expert statements and their dates are provided in Appendix 5. After optimistic expert announcements at the end of 2008, the first hint at a potential crisis appeared in the media in November, which suggested a possible crisis in the forthcoming period. The first public statement about the fact that the Republic of Serbia was in a crisis period appeared in March 2009. As it is evident in the figure, the systemic crisis occurrence probability approaches the upper critical level in October 2008. The systemic crisis probability appeared four months before the first official confirmation. The probability also returned to the normal levels in October 2009, i.e. three months before the first expert confirmation that the Republic of Serbia came out of the crisis, which appeared for the first time at the end of January 2010. According to the model specifications, an effective forecast period based on the probability of a systemic crisis is six months. However, in case of an empirical analysis on the basis of public expert statements, the said period is four months. There is an evident asymmetry between the crisis occurrence and end periods. It is apparent that the signalling period before the crisis occurrence is somewhat longer than the period after the crisis abates. This difference can be explained by the need for confirmation in case of bad news in order not to deepen market instabilities. On the other hand, there is apparent optimism coming from positive statements in order to ease the market pressure, which is reflected in a shorter period between early warning signals and the first official statements about the crisis end.

Figure 10 **Systemic crisis probability and media news headlines**



4 Conclusion

The Systemic Stress Indicator of the Republic of Serbia (SSI) is discussed in detail in this paper. The SSI is calculated based on the aggregation of six segments of the financial system into a composite index, which dynamically takes into account their mutual concordance measures. The paper provides a detailed analytical framework for the calculation of the composite indicator and systemic risk component, relying on the Markov-switching model with dynamic transition probabilities. Strengths and weaknesses of various approaches to the aggregation of the financial system segment sub-indices, which are included in the SSI calculation, were identified. In addition, switching regimes were validated using the Gaussian mixture model. Both approaches assessed the number of regimes and SSI distribution moments identically, which indicates the uniformity of the obtained results. The second part of the paper presents an analytical framework for early warning signals based on the methodology used for SSI calculation, providing a unique framework for assessing not only the current system stress level, but also the probability of a systemic crisis in advance. An analytical framework for assessing the probability of a systemic crisis occurrence in case of an arbitrary number of regimes is also provided. The classification quality of the proposed model is validated by means of assessments of classification quality measures in case of an arbitrary number of regimes, but it is also verified qualitatively based on public statements made by economic policy makers and experts.

It has been demonstrated that the SSI can properly identify crisis periods and the systemic risk level and assess the systemic stress component, providing significant information on the risk accumulation level in financial markets, and possible implications for financial and macroeconomic stability. The paper also contains a detailed analysis of the key instability factors in high stress periods between January 2008 and March 2021 based on SSI dynamics. The approach stemming from the composite index aggregation model was proposed to be used for aggregating indices of individual financial system segments into segment sub-indices. It also elaborates on the advantages of the proposed approach compared to the most frequently used methods in literature.

Unlike the *ex-post* identification of generally undetected crisis periods and key risk factors that triggered them in the past, early warning signals enable timely indication of a potential systemic crisis and factors that could trigger it, as well as the possibility to respond in a timely manner. The paper demonstrates that early warning signals properly monitor the dynamics of undetected periods and provide an assessment of the likelihood of a systemic crisis in case of an arbitrary time horizon. The optimal period in the case of the Republic of Serbia and relatively short time series is six months. It has been demonstrated that an effective prediction period is somewhat shorter and lasts between four and six months based on a qualitative analysis of public statements made by Republic of Serbia's economic policy makers.

The SSI discussed in the paper can be an excellent supplement to the existing analytical tools used to assess financial stability, such as macroprudential stress tests, a financial stress index and a banking sector stability index (NBS 2021), and provides additional information on a systemic risk measure in the Republic of Serbia's financial system.

Appendix 1 Sub-index aggregation method analysis

The Appendix contains the description of index movements using different data aggregation models and an explanation of the reasons for applying the proposed aggregation method. The first approach is the equal variance method which uses the same weights during the aggregation. The equal variance method generates an index that assigns equal importance to each variable. This is the most frequently used aggregation method in literature (Illing and Liu 2003). The main assumption of this model is the variable normality, which is at the same time its greatest weakness as it does not take into account an occurrence of a fat-tailed distribution, heteroskedasticity and volatility clustering (McNeil and Frey 2000). A strength of this approach lies in its simplicity and the fact that we do not know a priori aggregation weighting functions in a large number of cases. Therefore, the assumptions of this method are valid.

The second analysed method is a principal component analysis. This method is a statistical procedure that uses an orthogonal transformation in order to reduce data dimensionality taking into consideration a linear dependence between time series (Jolliffe 2002). Orthogonal components are called principal or main components. The number of principal components is less than or equal to the input space dimension. Principal components are ranked based on an explained variability of input space. The first principal component is defined as a component that explains the most of input data variations. The first principal component is used for each segment of the financial system and we will name this approach "the first PCA component". We will see that the key weakness of this approach is that the first principal component often explains a small amount of variation and therefore does not reflect data dynamics to a sufficient extent.

The next approach uses all principal components. Apart from the first component, each subsequent component explains the input data variability in descending order. This approach (Louzis and Vouldis 2013, Hatzius et al. 2010) is based on a weighted sum of principal component values proportional to the explained variability. To avoid including all principal components, only those that cumulatively explain more than the pre-determined amount of variation are included. The orthogonality of principal components helps to solve the multicollinearity problem in aggregating these components using a weighted sum.

Using data, defined at the beginning of this paper, we obtained standardised values for all aggregation methods shown in Figure 14. All indices have similar SSI dynamics. This is supported by high correlation between indices obtained through various aggregation methods shown in Table 5. The values of a linear correlation range from 0.79 to 0.94. The highest index inter-correlation was identified between the equal variance approach and all principal component method, whereas the lowest value was identified between the equal variance approach and the approach in which only the first principal component is used.

Table 5 **Inter-correlation values of various sub-index aggregation approaches**

	Equal variances	First PCA component	All PCA components	Portfolio aggregation
Equal variances	1.00	0.79	0.97	0.88
First PCA component	0.79	1.00	0.85	0.94
All PCA components	0.97	0.85	1.00	0.90
Portfolio aggregation	0.88	0.94	0.90	1.00

The reason is the fact that the first component is insufficient to capture data variability. This is particularly evident in segments of the FX market – FX and public finances – GOV, where the first principal components explain only 32.9%, i.e. 43.9% of the variability, respectively, as it can be seen in Table 6.

Although with similar dynamics, a significant deviation between different indices occurs in the period from May 2012 to January 2013. The indices that are based on the equal variance method and the principal components method are considerably higher compared to other two methods. When selecting a method that best describes the composite indicator of systemic stress of the Republic of Serbia's financial system, we will analyse this period in more detail in order to assess differences in the approaches. We will focus on two methods: the equal variance method and the co-variance method.

Table 6 **Data variance percentage according to principal components**

	1st component	2nd component	3rd component	4th component	5th component	6th component	7th component
FX	32.9%	32.3%	21.2%	13.5%			
GOV	43.9%	30.3%	16.7%	7.2%	2.0%		
MON	51.6%	16.4%	14.9%	9.5%	4.2%	2.8%	0.5%
EQU	51.1%	30.0%	16.3%	2.6%			
BANK	49.8%	33.5%	16.7%				
FOR	72.5%	27.5%					

Sub-index Figure 15 of each segment of the Republic of Serbia's financial system is shown in Figure 15. Conditional correlations are also in Figure 15 using the exponentially weighted moving average method (EWMA) of the segment's time series. Inter-correlations are aggregated by averaging the values of inter-correlations using the formula:

$$\overline{R} = \frac{1}{N}\sum R'_{i,j} \text{ in which } R'_{i,j} = \begin{cases} R_{i,j}, \forall i > j \\ 0 \quad \text{else} \end{cases}; \qquad (32)$$

A further analysis of asymptotic characteristics of multidimensional concordance measures is presented in the paper by Joe (1990).

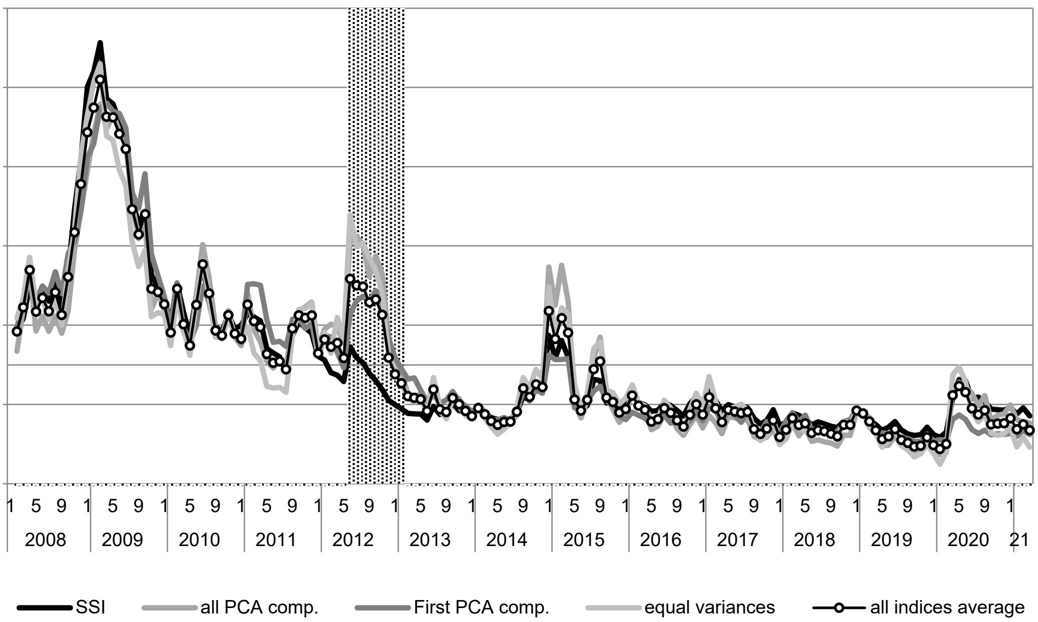
We can see a difference in the SSI level between May 2012 and January 2013, resulting from differences in individual segments, particularly in the case of public finances, money market and banking sector. Differences between sub-indices are the largest in the case of the pronounced difference between non-conditional correlation values, shown in Table 7, and conditional time-varying correlation obtained through the equations (8) to (10), and (32).

Table 7 **Values of non-conditional correlation between segment indicators**

	Non-conditional correlation
FX	0.1060
GOV	0.1426
MON	0.2593
EQU	0.3445
BANK	0.1770
FOR	0.4803

Based on an analysis of the difference of sub-indices for each individual segment and the difference between the non-conditional and conditional correlation values that are shown in Figure 15, a strong linear conditionality is evident in all cases, except in the case of the capital market, between the non-conditional correlation values and conditional correlation values in relation to a deviation of these two sub-indices. Another characteristic of the linear approximation shown in Figure 15 is that in case the non-conditional correlation and conditional correlation match, sub-indices will also match. The cross markers in Figure 15 are used to mark the period between May 2012 and January 2013, which saw the highest deviation between these two indices. As it can be seen, these are extreme values of a difference between correlations. It suggests that the principal component aggregation cannot properly follow the changing concordance measures between variables even though it offers the option of taking into account non-conditional correlations of a multi-dimensional input space. Although it is put to good use in case of input space reduction, it does not properly follow the dynamics, particularly when there is a sudden change in the input space co-variance, which occurs in case of regime changes, which is especially important when developing the SSI.

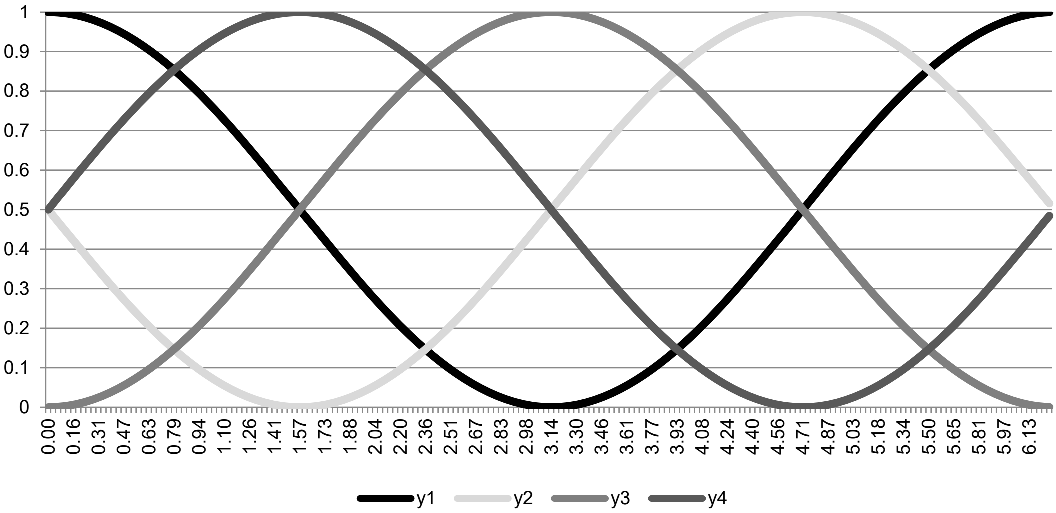
Figure 11 **Movement of the SSI using various aggregation methods**



This characteristic of the principal component methods can be seen in the following theoretical example. Assuming the input space shown in Figure 12 is in sync with the sinusoidal functions:

$$y_i = \frac{sin\left(\left(x + (\pi/_2)\right) i + 1\right)}{2}, i = 1,2,3,4 \tag{33}$$

Figure 12 **Example of comonotonic and anti-comonotonic signalsExample of comonotonic and anti-comonotonic signals**



It is evident that the sum of these signals in time is constant. There are also two pairs of signals that are anti-comonotonic (McNeil et al. 2015): $\{y_1, y_3\}$ and $\{y_2, y_4\}$, and they cancel each other out. After applying the all-principal method that is shown in Figure 1, it is evident that the first two components explain nearly 100% of the variability. These components are shifted in time, whereby the first component has a wider range in comparison to the second one. The third and fourth components have values that are close to zero. The index marked with a full black line is obtained by aggregating the first two components. The standardised index value ranges in the interval of $[-0{,}5 - 0{,}5]$. Normalisation that is used to reduce the standardised index to a unit interval results in both limit values in the interval, which indicates a lack of the method application. Namely, the aggregation of principal components covers the entire interval and therefore, they amplify the impact of extreme values. On the other hand, a portfolio aggregation has a lot slower dynamic. This index moves in the interval of $[0{,}1 - 0{,}25]$, which complies with the assumption about the pair index constancy of anti-comonotonic variables and the movement of a non-conditional correlation. Another advantage of this aggregation method is that it does not require additional normalisation and does not consequently generate extreme values of the unit interval and it enables a more accurate range of the index dynamics.

Figure 1 **Example of index movements based on all principal components approach and the portfolio aggregation method**

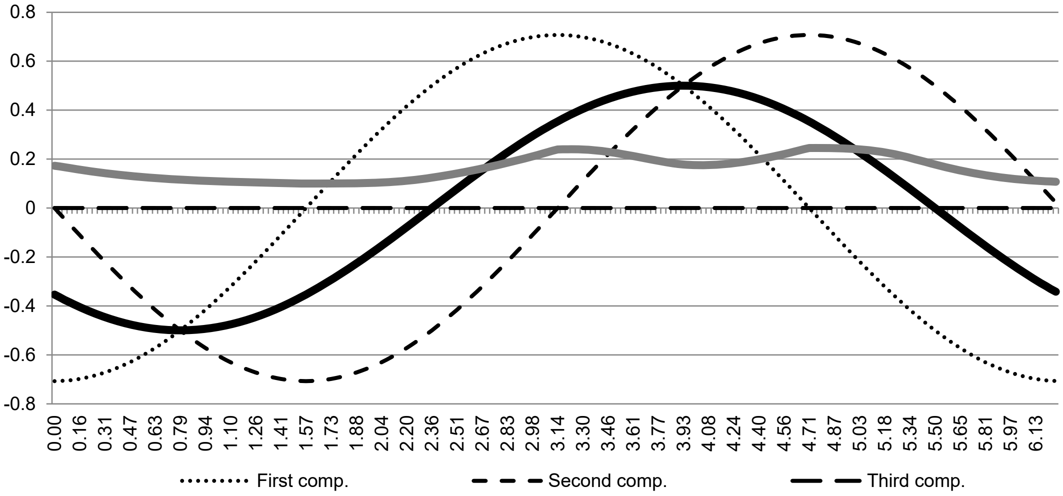


Figure 14 **Sub-index movement by applying the all principal components approach and the portfolio aggregation method by segments and conditional correlation values of sub-index series**

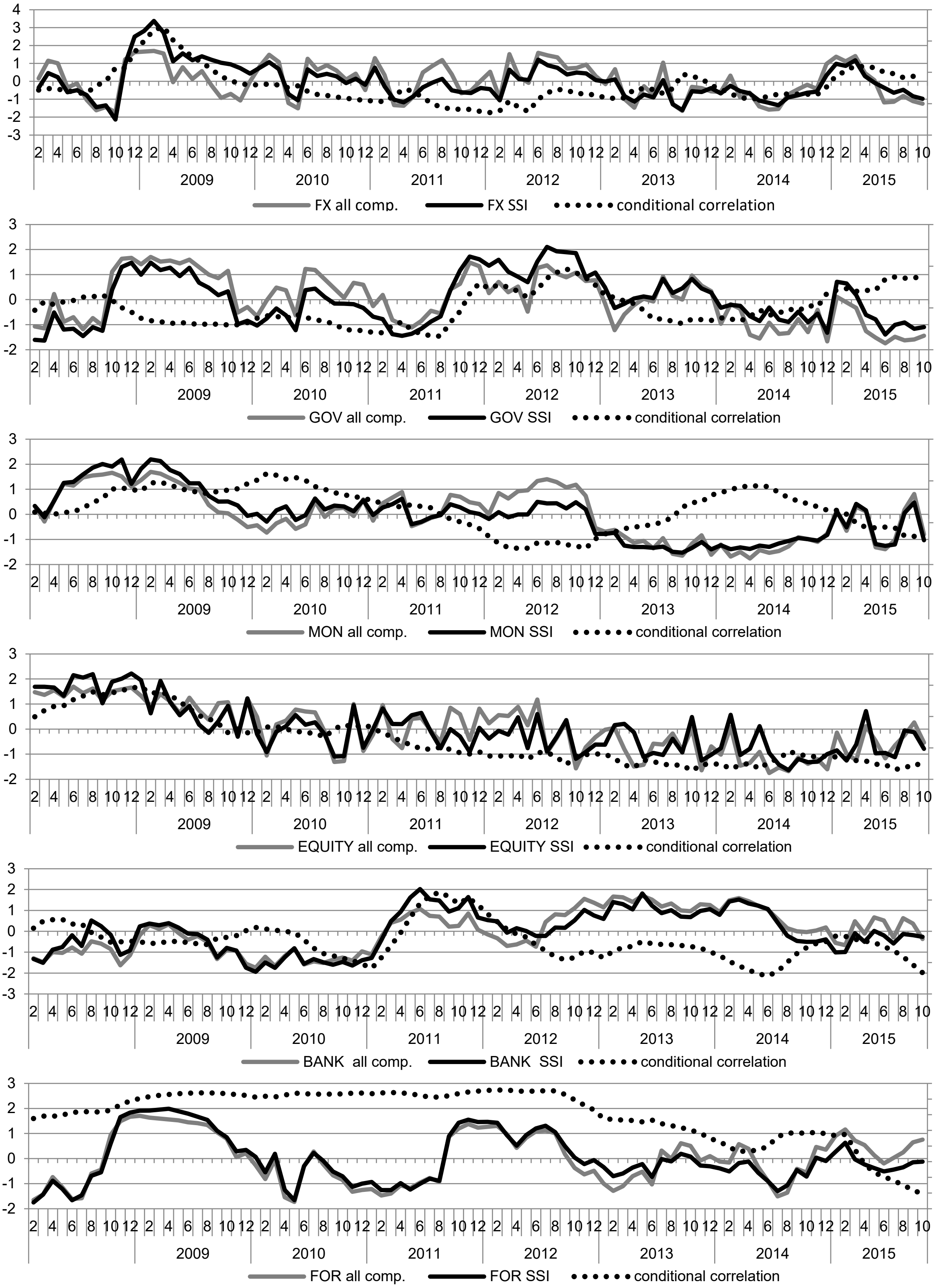
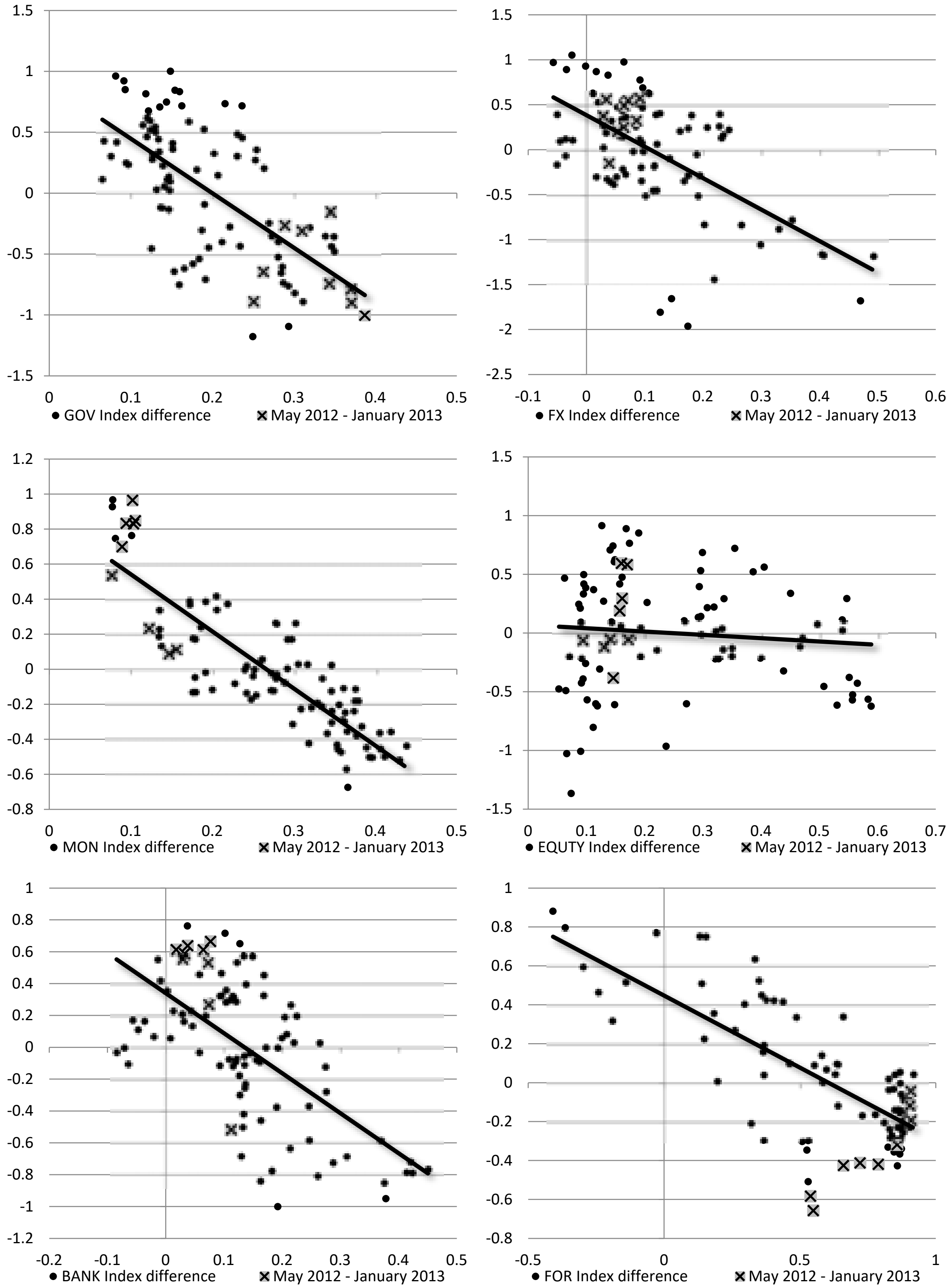


Figure 15 **Difference in sub-index values obtained by using the all principal and the portfolio aggregation method by segments depending on conditional correlation values of sub-index series**



Appendix 2 List of indicators for SSI calculation

Segment	Indicator	Designation
Foreign exchange market	Y-o-y growth of the RSD to EUR exchange rate	FX.EUR_RSD_MG
	Realised volatility of y-o-y growth of the RSD to EUR exchange rate	FX.RV
	Absolute value of the National Bank of Serbia's interventions in the FX market	FX.NBS_Interventions
	Difference between the selling and buying RSD to EUR exchange rates	FX.B_A_spread
Public finances	Difference between the yield to maturity of the 10Y government bond and the 10Y German government bond	Gov.YTM_RS_minus_BUND
	Difference between the selling and buying prices of the 10Y government bond	Gov.B_A_spread_RS
	Realised volatility of the yield to maturity of the 10Y government bond	Gov.YTM_RS_RV
	EMBI Global Serbia	Gov.EMBI_G_Serbia
	Consolidated fiscal result (% of GDP)	Gov.Deficit
Money market	Weighted average interest rate on overnight interbank market loans – BEONIA	Money.BEONIA
	Difference between BEONIA and the key policy rate	Money.BEONIA_minus_REFRATE
	Standard deviation – BEONIA	Money.BEONIA_StDev
	Difference between the maximum rate on overnight loans and interest rate on credit facilities	Money.DEPOSIT_CREDIT_Rates
	Average RSD required reserve allocation on 17th of each month	Money.Allocated_RR
	Difference between a three-month interest rate on EURIBOR and the yield to maturity of the German government bond	Money.EURIBOR_MINUS_BUND
	Difference between a three-month interest rate on EURIBOR and the overnight EONIA interest rate	Money.EURIBOR_MINUS_EONIA
Capital market	CMAX transformation of the stock market index BELEX15	Equity.CMAX_ BELEX15
	BELEX15 index turnover	Equity. BELEX15_turn
	Realised volatility of the stock market index BELEX15	Equity. BELEX15_RV
	Realised volatility of the BELEX15 index turnover	Equity.RV_ BELEX15_turn
Banking sector	Credit gap using a one-sided HP filter	Banking.Loan_Gap
	Deposit gap using a one-sided HP filter	Banking.Deposits_gap
	Weighted sum of CDS parent banks in line with a share in the balance-sheet total	Banking.weight_CDS
Foreign environment	Composite EMBI GLOBAL	FOR.EMBI_G_comp
	Average values of stock market indices in neighbouring countries	FOR.Stock_indexes

Source: NBS, MoF, Belgrade Stock Exchange.

Appendix 3 Statistics of estimated values of model parameters

Table 8 **Main statistics of different values of a dependent state variable lag and the number of regimes**

	F-DMS(2) SAR (0)	F-DMS(3) SAR (0)	F-DMS(2) SAR (1)	F-DMS(3) SAR (1)	F-DMS(2) SAR (2)	F-DMS(3) SAR (2)
Likelihood	225.93	286.29	227.84	**289.3**	226.69	285.97
Number of parameters	6	12	6	12	6	12
AIC	-439.86	-548.58	-443.68	**-554.6**	-441.37	-547.93
BIC	-424.66	-518.19	-428.48	**-524.21**	-426,24	-517.67
RCM	3.17%	2.26%	2.91%	**2.11%**	2.78%	2.64%
Normality (p value)	**3.95E-01**	1.00E-03	3.11E-01	1.00E-03	2.93E-01	1.00E-03
Autocorrelation	0.00E+00	1.13E-08	0.00E+00	2.97E-09	0.00E+00	8.47E-11
ARCH	4.47E-08	5.15E-03	2.27E-08	1.66E-02	4.23E-07	2.43E-03

Table 9 **Statistics of estimated parameter values of the F-DMS(3) SAR (1) model**

F-DMS(3) SAR (1)	Coefficient	Standard error	p-value
Variance (1)	6.30E-05	1.00E-05	0.00E+00
Variance (2)	3.95E-04	1.09E-04	4.00E-04
Variance (3)	1.86E-03	1.23E-03	1.33E-02
Expected value (1)	2.31E-02	8.61E-04	0.00E+00
Expected value (2)	7.40E-02	3.05E-03	0.00E+00
Expected value (3)	1.87E-01	1.31E-02	0.00E+00
$Pa_{(1\|1)}$	9.11E+01	8.88E+00	0.00E+00
$Pa_{(1\|2)}$	-3.81E+01	9.05E+03	9.97E-01
$Pa_{(1\|3)}$	3.61E+02	1.96E+06	1.00E+00
$Pa_{(2\|1)}$	2.47E+01	5.18E+00	0.00E+00
$Pa_{(2\|2)}$	-8.98E+00	7.55E+00	2.36E-01
$Pa_{(2\|3)}$	9.11E+01	8.88E+00	0.00E+00

Other parameter values of transition probabilities $\mathrm{Pa}_{(3|1)}$, $\mathrm{Pa}_{(3|2)}$ and $Pa_{(3|3)}$ are obtained on the basis of the presented transition probabilities.

Appendix 4 Models for assessing the probability of a systemic crisis occurrence

Table 10 **Statistics of reduced models depending on a state variable lag**

Horizon	AIC	BIC	Normality	ARCH	RCM	Likelihood
R-DMS(3) SAR(3)	-561.83	-542.53	1.00E-03	2.42E-04	0.0252	285.85
R-DMS(3) SAR(4)	-551.08	-536.01	1.00E-03	2.43E-04	0.0263	281.54
R-DMS(3) SAR(5)	-540.71	-525.71	1.00E-03	9.94E-06	0.0245	276.36
R-DMS(3) SAR(6)	-538.98	-524.05	1.00E-03	2.22E-06	0.0250	275.49
R-DMS(3) SAR(7)	-535.91	-521.05	1.00E-03	3.83E-07	0.0233	273.96
R-DMS(3) SAR(8)	-526.81	-512.01	1.00E-03	9.80E-08	0.0221	269.40
R-DMS(3) SAR(9)	-518.62	-503.89	1.00E-03	1.07E-07	0.0215	265.31
R-DMS(3) SAR(10)	-520.79	-506.14	1.00E-03	3.24E-09	0.0194	266.40
R-DMS(3) SAR(11)	-515.76	-501.18	1.00E-03	5.95E-10	0.0209	263.88
R-DMS(3) SAR(12)	-511.05	-496.54	1.00E-03	5.26E-11	0.0211	261.53

Table11 **Assessment of the reduced models classification quality depending on state variable lag**

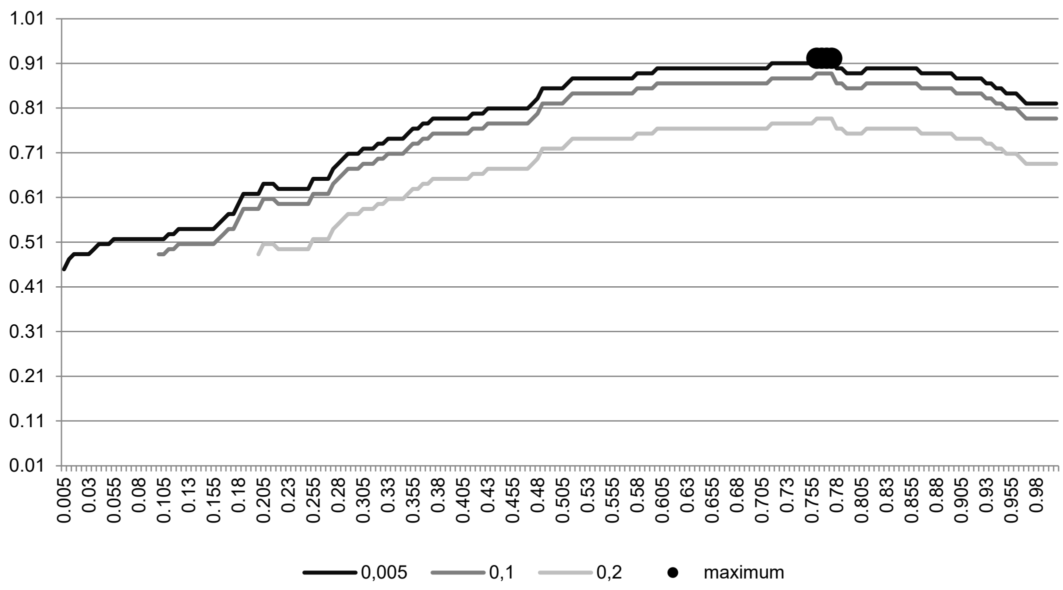
	b_l	b_u	Classification accuracy	Classification sensitivity	Classification specificity
R-DMS(3) SAR(3)	0.010	0.805	0.9333	0.875	0.983
R-DMS(3) SAR(4)	0.005	0.720	0.9322	0.844	1.000
R-DMS(3) SAR(5)	0.005	0.715	0.9211	0.844	1.000
R-DMS(3) SAR(6)	**0.005**	**0.760**	**0.9211**	**0.875**	**0.983**
R-DMS(3) SAR(7)	0.005	0.740	0.8878	0.875	0.948
R-DMS(3) SAR(8)	0.005	0.620	0.8656	0.906	0.914
R-DMS(3) SAR(9)	0.005	0.460	0.8322	0.906	0.862
R-DMS(3) SAR(10)	0.005	0.460	0.8211	0.875	0.845
R-DMS(3) SAR(11)	0.001	0.260	0.5767	0.532	0.879
R-DMS(3) SAR(12)	0.001	0.015	0.5767	0.551	0.879

Table12 **Statistics of estimated parameter values of the R-DMS(3) SAR (6) model**

R-DMS(3) SAR (6)	Coefficient	Standard error	p value
Variance (1)	6.30E-05	-	-
Variance (2)	3.95E-04	-	-
Variance (3)	1.86E-03	-	-
Expected value (1)	2.31E-02	-	-
Expected value (2)	7.40E-02	-	-
Expected value (3)	1.87E-01	-	-
$Pa_{(1\|1)}$	1.14E+02	1.94E+01	0.00E+00
$Pa_{(1\|2)}$	-2.79E+01	1.15E+01	1.65E-02
$Pa_{(1\|3)}$	-3.82E+01	5.19E+03	9.94E-01
$Pa_{(2\|1)}$	3.61E+02	1.30E+05	9.98E-01
$Pa_{(2\|2)}$	2.38E+01	4.54E+00	0.00E+00
$Pa_{(2\|3)}$	-7.25E+00	3.99E+00	7.11E-02

Other parameter values of transition probabilities $Pa_{(3|1)}$, $Pa_{(3|2)}$ and $Pa_{(3|3)}$ are obtained based on the presented transition probabilities.

Figure 16 **Family of curves of classification accuracy in case of three regimes depending on the lower and upper critical classification values**



Appendix 5 News history of the 2008–2010 crisis

“I expect that the crisis will be mitigated before its negative effects are reflected on our country.” **(Dnevnik, 3 October 2008)**

“Serbia will come out of this great global financial crisis intact…. It turned out that our restrictive monetary policy represents steady protection against crises such as the current one.” **(Dnevnik, 16 October 2008)**

“Global crisis threatens Serbia in 2009“ **(Stratfor, 4 November 2008)**

“Crisis came to Serbia earlier than expected and the responsible ones are ‘domestic economic actors’.” **(B92, 3 March 2009)**

”Crisis has not ended but it is slowing down.“ **(RTV Pink, 12 May 2009)**

“Serbia has emerged from the crisis according to economic indicators.” **(BETA, 20 January 2010)**

“Serbia has both formally and statistically come out of the crisis.“ **(Danas, 17 May 2010)**

References

Abiad, M. A. (2003) Early warning systems: A survey and a regime-switching approach, International Monetary Fund.

Alexander, C. (2009) Market Risk Analysis, Value at Risk Models, John Wiley & Sons.

Ang, A. and Bekaert, G. (2002) ‘‘Regime switches in interest rates’, Journal of Business & Economic Statistics, 20(2), 163-182.

Attinasi, M. G., Checherita-Westphal, C. D. and Nickel, C. (2009) ‘‘What explains the surge in euro area sovereign spreads during the financial crisis of 2007-09?’’.

Baele, L. (2005) ‘‘Volatility spillover effects in European equity markets’, Journal of Financial and Quantitative Analysis, 40(02), 373-401.

Berg, A. and Pattillo, C. A. (1998) ‘‘Are currency crises predictable? A test’’.

Borio, C. (2003) ‘‘Towards a macroprudential framework for financial supervision and regulation?’’, CESifo Economic Studies, 49(2), 181-215.

Bussiere, M. and Fratzscher, M. (2006) ‘‘Towards a new early warning system of financial crises’, journal of International Money and Finance, 25(6), 953-973.

Cavanagh, J. and Long, R. (1999) ‘‘Introducing the JP Morgan Emerging Markets Bond Index Global (EMBI Global)’’, Emerging markets research, JP Morgan Securities Inc.

Chae, J. (2005) ‘‘Trading volume, information asymmetry, and timing information’’, The journal of finance, 60(1), 413-442.

Crotty, J. (2009) ‘Structural causes of the global financial crisis: a critical assessment of the ‘new financial architecture’’, Cambridge Journal of Economics, 33(4), 563-580.

De Bandt, O., Hartmann, P. and Peydro, J. (2009) ‘Systemic risk in banking: An update’, Oxford Handbook of Banking, Oxford University Press, Oxford.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977) ‘Maximum likelihood from incomplete data via the EM algorithm’, Journal of the royal statistical society. Series B (methodological), 1-38.

Diebold, F. X., Lee, J.-H. and Weinbach, G. C. (1994) ‘Regime switching with time-varying transition probabilities’, Business Cycles: Durations, Dynamics, and Forecasting, 144-165.

Dornbusch R. (1976), ‘Expectations and Exchange Rate Dynamics’, The Journal of Political Economy, Vol. 84, pp. 1161-76

Dumičić, M. (2014) ‘Pokazatelji financijskog stresa za male otvorene visokoeuroizirane zemlje–primjer Hrvatske’, HNB Istraživanja, 43.

Frankel, J. A. and Saravelos, G. (2010) Are leading indicators of financial crises useful for assessing country vulnerability? Evidence from the 2008-09 global crisis, National bureau of economic research.

Hamilton, J. D. (1989) ‘A new approach to the economic analysis of nonstationary time series and the business cycle’, Econometrica: Journal of the Econometric Society, 357-384.

Hatzius, J., Hooper, P., Mishkin, F. S., Schoenholtz, K. L. and Watson, M. W. (2010) Financial conditions indexes: A fresh look after the financial crisis, National Bureau of Economic Research.

Hollo, D., Kremer, M. and Lo Duca, M. (2012) ‘CISS-a composite indicator of systemic stress in the financial system’.

Houben, A., Kakes, J. and Schinasi, G. J. (2004) Toward a framework for safeguarding financial stability, International Monetary Fund.

Illing, M. and Liu, Y. (2003) An index of financial stress for Canada, Bank of Canada.

Illing, M. and Liu, Y. (2006) ‘Measuring financial stress in a developed country: An application to Canada’, Journal of Financial Stability, 2(3), 243-265.

IMF (2001) Government finance statistics manual 2001, Statistics Department.

IMF (2008) Financial Soundness Indicators: Compilation Guide, International Monetary Fund.

Joe, H. (1990) ‘Multivariate concordance’, Journal of multivariate analysis, 35(1), 12-30.

Jolliffe, I. (2002) Principal component analysis, Wiley Online Library.

Kaminsky, G. L. (1999) Currency and banking crises-the early warnings of distress, International Monetary Fund.

Kliesen, K. L., Owyang, M. T. and Vermann, E. K. (2012) ‘Disentangling diverse measures: A survey of financial stress indexes’, Federal Reserve Bank of St. Louis Review, 94(5), 369-398.

Kovačević, D., Mladenović, N., Petrović, B. and Milošević, P. (2014) ‘DE-VNS: Self-adaptive Differential Evolution with crossover neighborhood search for continuous global optimization’, Computers & Operations Research, 52, 157-169.

Kou, G., Chao, X., Peng, Y., Alsaadi, F. E., & Herrera-Viedma, E. (2019). Machine learning methods for systemic risk analysis in financial sectors. Technological and Economic Development of Economy, 25(5), 716-742.

Lo Duca, M. and Peltonen, T. A. (2011) ‘Macro-financial vulnerabilities and future financial stress-Assessing systemic risks and predicting systemic events’.

Louzis, D. P. and Vouldis, A. T. (2013) ‘A financial systemic stress index for Greece’.

McNeil, A. J. and Frey, R. (2000) ‘Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach’, Journal of empirical finance, 7(3), 271-300.

McNeil, A. J., Frey, R. and Embrechts, P. (2015) Quantitative Risk Management: Concepts, Techniques and Tools: Concepts, Techniques and Tools, Princeton university press.

NBS (2021) ‘Annual Financial Stability Report 2020’, National bank of Serbia publications.

Obstfeld, M. and Rogoff, K. (2009) ‘Global imbalances and the financial crisis: products of common causes’.

Podpiera, J. and Ötker, M. I. (2010) The fundamental determinants of credit default risk for European large complex financial institutions, International Monetary Fund.

Press, W. H. (2007) Numerical recipes 3rd edition: The art of scientific computing, Cambridge university press.

Rose, A. K. and Spiegel, M. M. (2012) ‘Cross-country causes and consequences of the 2008 crisis: early warning’, Japan and the World Economy, 24(1), 1-16.

Schwarcz, S. L. (2008) ‘Systemic risk’, Georgetown Law Journal, 97(1).

Sokolova, M. and Lapalme, G. (2009) ‘A systematic analysis of performance measures for classification tasks’, Information Processing & Management, 45(4), 427-437.

Tamakoshi, G. and Hamori, S. (2015) The European Sovereign Debt Crisis and Its Impacts on Financial Markets, Routledge.

GLOBAL DEVELOPMENT TRENDS IN PAYMENT CARD INDUSTRY

Aleksandar Lagator

Available at www.nbs.rs

The views expressed in the papers constituting this series are those of the author, and do not necessarily represent the official view of the National Bank of Serbia.

Payment System Department

NATIONAL BANK OF SERBIA

Belgrade, 12 Kralja Petra Street

Telephone: (+381 11) 3027 100

Belgrade, 17 Nemanjina Street

Telephone: (+381 11) 333 8000

www.nbs.rs

Global development trends in payment card industry

Aleksandar Lagator

Abstract: This paper presents the most important global development trends in payment card industry. The payment cards industry is specific because it has a high level of multidisciplinarity since, besides the technological, it also includes various other aspects such as the legal, market, financial, social and even political. Hence, development factors of payment cards are not always only technological in their nature but are often also of market, legal, financial, social and political kind. The aim of this paper is to present these very aspects of development in a simple and concise manner, as much as this is possible, given that because of the multidisciplinary nature, the topics associated with payment cards are often extremely complex and voluminous. At the end of the paper, we detail the situation in Serbia which boasts a very high level of monitoring and realisation of technological and legal solutions available in the world. In addition to presenting development trends in payment card industry, the paper also offers some of the author's critical analysis of certain development trends in the sense of their advantages, shortcomings, problems in realisation and "lessons learned" from previous numerous experiences in the world.

Key words: payment cards, development trends, chip cards, EMV, CPA, contactless payments, e-commerce, PCI DSS, IFR, PSD2, NFC, HCE, strong authentication, 3-D Secure, SRC.
[JEL Code]: C32, G01, E44.

Non-Technical Summary

Because of its pronounced multidisciplinary nature, global development trends in payment card industry are affected by many factors that are not solely technological. These are often market, legal, financial, social and even political factors. Because of the strong competition in the global payment cards market, there are constant challenges in this area both in terms of new business and technological services and solutions, and in terms of problems in the market, such as unsettling healthy competition. This is why two paths for resolving the existing challenges have been dominant in recent times: new technological and new legal solutions.

The milestone in the development of technological solutions in payment card industry took place with the invention of chip payment cards in the 1990s. Chip technology provided the starting basis for later new technical solutions, such as contactless cards and mobile phone payments, and the protection principles of this technology are becoming increasingly present with the latest solutions for the security of online payments.

Besides technological development, the increasingly frequent challenges in the payment card industry are market-related, which is why the implementation of legal regulations in solving these challenges has become quite frequent recently. The aim of the new legal regulations is to correct disruptions in the market, while maintaining the principle of free market, which is not easy to fulfil, therefore each new regulation was usually preceded by years of analyses, consultations, public discussions and even court procedures.

Contents:

1 Introduction 124
2 A brief overview of the history of payment cards 124
3 Chip payment cards as the beginning of the recent history of payment cards 125
4 Contactless payment cards 128
5 Payments by mobile phone 128
6 New solutions for online payment security 129
7 Payment card industry data security standard – PCI DSS 132
8 Expansion of debit payment card use as a consequence of the 2007–2008 global economic crisis 133
9 National card systems as a growing trend 133
10 Countries as regulators and factors of payment cards development 135
11 Payment card development trends in Serbia 137
12 Conclusion 139
References 140

1 Introduction

Due to the persistently strong competition in the global payment cards market, challenges in this area are constantly present both in terms of new business and technological services and solutions, and in terms of problems in the market, such as unsettling healthy competition. This is why two paths for resolving the existing challenges have been dominant in recent times: new technological and new legal solutions, which is the topic of this paper.

The target group for the topics presented in this paper are primarily professionals whose scope of work includes payment cards or operations more or less associated with payment cards; next, professionals who are beginning or considering to begin to engage in these matters; as well as all those who find this field interesting for one reason or another.

In addition to presenting payment card development trends, the paper also offers some of the author's critical analysis of certain development trends in the sense of their advantages, shortcomings, problems in realisation and "lessons learned" from previous numerous experiences in the world.

2 A brief overview of the history of payment cards

American writer and journalist Edward Bellamy (1850–1898) is considered the originator of the idea of payment cards, as he was the first to use the phrase "credit card" in his Sci-Fi novel *Looking Backward*. In the book, this is a card which each citizen receives from the state, together with certain monetary assets, which he can use for purchasing using the card as a payment instrument.

The first kinds of payment cards are thought to be various types of coins made of plastic, copper, aluminium or steel, differing in shape and size, with embossed account number, merchant's name and symbol, and often with a small hole for attachment to pendants. Such coins were used in the late 19^{th} and early 20^{th} century for identification of accounts for collection in hotels and stores, by recording an imprint of the information on the coin instead of writing it down by hand. This sped up the collection process, though it still was not good enough for client identification, hence problems and fraud were frequent.

The first true payment cards are the so-called Charga-Plate cards that were used in the USA between 1930 and late 1950s. The idea for and the manufacturing of these cards are attributed to Farrington Manufacturing Co. These cards were made from a rectangular piece of metal, with the holder's name, town and state embossed on the metal, and were issued by large department stores to their customers. The information on the card is quickly taken by making an imprint with a piece of white and indigo paper, instead of writing the information down by hand, which accelerated the purchase, and since the store had identification data about the customer, the problems and fraud were limited and acceptable. In addition to being kept by the customer, these cards were often kept at points of sale.

The 1930s saw the emergence of payment cards, first by American and later by other international airline operators (Air Travel Card). These cards were based on a numbering system of card issuers (airline operators) and client's accounts. Payments by these cards were deferred (the customer was billed later) and they often implied certain discounts. These cards are considered the first international payment cards as they were accepted by all members of the International Air Transport Association.

The Diners Club card, created in 1950, is thought to be the first general use payment card (used with several merchants). It was a credit-type card and the entire debt was charged at the end of the month (charge card). Not long after, there appeared cards with a similar purpose Carte Blanche and American Express, with the latter soon becoming an international card.

The first true credit card (with a credit line based on the monthly collection of a certain percentage of the total remaining debt, i.e. revolving credit) is considered to be the BankAmericard that appeared in 1958, created by Bank of America and gradually licensed to other banks, initially in the USA and later in other countries of the world. Then, in 1976, a new common name for this card was adopted – Visa. Concurrently, as a response from market competition, a group of banks founded the Master Charge in 1966, which would eventually grow into MasterCard.

Over time, Visa and MasterCard became dominant payment cards in the international market, and remain so today. On the one hand, the contribution of these brands to the global development of the payment card industry is unquestionable. However, in time, the activity of the two brands became identified as a sort of monopoly or a duopoly (an increasingly used term), which is supported by the fact that they function very similarly, hold a dominant market share in most countries of the world, have been through similar stages of the transformation of ownership, and recently they are having more and more common owners in their increasingly complex ownership structure. In such conditions, over time we see a growing number of local and national card systems emerge in response to the existing situation, with the goal of demonopolizing and better regulating their markets, as well as lowering costs and the dependence on these two global card systems, which will be looked at more closely in this paper.

3 Chip payment cards as the beginning of the recent history of payment cards

From the 1970s until the 1990s, magnetic stripe payment cards were the dominant technology in the payment card industry. This technology enabled the digitisation of payment transactions via POS readers (point-of-sale reader) and ATMs (automated teller machines), and data transmission networks. However, the magnetic stripe was solely a passive digital medium for reading identification data, with no option of entering data or any other functionalities.

In the 1990s the first payment cards with a chip appeared, first in France (the first payment application on the chip called B4/B0), and then beyond, including the global card systems such as Visa and MasterCard. The main reasons for introducing chip payment cards were security enhancements (fraud prevention) and the possibilities for additional functionalities of payment cards.

As for security enhancements, chip payment cards enabled advanced cryptographic functionalities for highly reliable user authentication (encrypted PIN), identification (digital signature and dynamic authentication) of the card issuer and the card itself, as well as the prevention of payment card skimming. The use of chip payment cards significantly improved the security of transactions.

Regarding additional functionalities of chip payment cards, in addition to the main function of making payments at POS and ATMs, this technology also enabled the use of the existing chip payment card for reliable authentication for online payments, employee identification using the card for access to work premises, identification of citizens for electronic administration services, control of customers' age when purchasing alcohol, various commercial discount programmes for loyal customers (loyalty programmes), buying tickets in public transport, etc. Such additional functionalities were enabled by some card systems, e.g. by the German Girocard (authentication online, identification of citizens for electronic administration services, control of customers' age when purchasing alcohol) and the Japanese JCB (authentication online, identification of citizens for electronic administration services, employee identification using the card for access to work premises, buying tickets in public transport).

To lower the bank costs associated with the manufacturing of chip payment cards for various card systems, in 1994 Visa, MasterCard and EuroPay began a project called the EMV, which implied a common technical specification for chip payment cards for all three card systems, enabling the use of the same chip for all three card systems. In 1996, the first official version of the EMV technical specification was made public, and three years later, in 1999, the EMVCo consortium was founded, with the aim of further maintaining and developing technical specifications for support to chip payment cards. Until today, other card systems joined the consortium – JCB (2004), AMEX (2009), and UnionPay and Discover (2013). In 2005, the first version of a common payment application (CPA) on the chip was issued, and somewhat later, in 2008, the first version of the CPA protocol for contactless payments.

Nowadays, chip payment cards have become common and omnipresent, while the use of the magnetic stripe on payment cards is increasingly rare. However, though the use of the chip led to visible progress in card payments, the progress achieved was not at the level that had been expected and announced. Below we elaborate on this fact in more detail.

One of the main problems which the chip payment card technology brought along was its exceptional complexity and the high price of its implementation. That is why the migration from magnetic stripe to chip payment cards, as well as the implementation itself, was very slow to take place in almost all markets, with few exceptions – these included markets where this technology originated. Therefore, card systems such as Visa and MasterCard were forced to give some encouragements for the migration, and when this did not produce the desired results, they introduced rules shifting the liability for fraud (the so-called liability shift) to the side that did not introduce the chip technology. Lastly, they issued a strict obligation and a deadline for migration, gradually for different markets. After more than one decade, this finally yielded the planned results at the global level.

Though the chip technology enabled a high level of security and additional functionalities, chip cards were not used as much as initially expected, primarily because of the complexity and implementation costs. For instance, for the majority of optional functionalities (such as reliable authentication with online payments, employee identification using the card for access to work premises, identification of citizens for electronic administration services, buying tickets in public transport, etc.) there has to be an additional application on the chip, as well as more memory, which requires complex implementation and consequently additional costs, as well as the considerably higher final price of a chip payment card. For these reasons, the ideas about additional functionalities on the same chip (on the same chip card) were abandoned over time and instead, additional functionalities were implemented on separate chip cards intended for certain functionalities (e.g. personal ID card with a chip for electronic administration services, a special chip card for payments in public transport, such as the Oyster Card in London, special proximity cards with a chip for employee access control to work premises, etc.).

Interestingly, the most complex part of chip card functionalities is associated with cryptographic operations with offline payments (where payments are executed between a POS terminal and a chip card, without the use of the data transmission network). This pertains to Offline Data Authentication (ODA) process to authenticate the card issuer and the card itself, which requires complex cryptographic operations, as well as complex support on the part of the issuing bank in relation to the cryptographic pairs of keys and certificates. However, this functionality is important in terms of security only with offline payments. As for transactions via ATMs, this functionality is not used, and it is not necessary for online payments on POS terminals given that in this case (as well as with ATMs) a much simpler cryptographic method of card issuer authentication is used – the exchange of ARQC/ARPC cryptograms on a data transmission network, which does not require complex support on the part of the issuing bank in terms of cryptographic pairs of keys and certificates. As offline payments are being gradually abandoned across the globe, so does the need for complex ODA functionality on the chip card cease, without which bank procedures for chip technology would be incomparably simpler and implementation of chip cards much cheaper as well. It is not impossible that this very useful simplification may take place in the near future.

Regardless of the existence of the CPA specification, which offers the possibility of using the same chip application for multiple card systems, in practice Visa and MasterCard still use their special specifications, the so-called VSDC (Visa Smart Debit/Credit) and MasterCard M/Chip, primarily because of the specificities of the M/Chip specification which, as time goes by, deviates more and more from the common CPA specification. Still, given that it is free for use, the CPA specification has become a major basis for the development of many local and national card systems.

Chip payment cards have significantly contributed to the improvement of the security of payment card transactions. However, this only pertains to card payments (card present), and not to card not present payments, such as online payments, manual entries at POS terminals or phone orders, which remain at the bottom security level. This has led to the so-called fraud migration from previous payment manners to dominantly online payments which are less protected. Due to these reasons, a new problem was tackled – increasing the security of online payments, which will be discussed in more detail in the following chapters.

4 Contactless payment cards

The first contactless payment card is thought to be the South Korean Upass, which entered into use in 1995 for payments in public transport. Since 2008, Visa, MasterCard and American Express card systems almost simultaneously began offering contactless payment cards. Today, contactless payments have become very widespread and even dominant in some markets. The most often used are the so-called hybrid chip payment cards which have both contact and contactless payment interfaces.

Initially, contactless cards were intended primarily for specific types of payments implying lower amounts and a faster execution of payments for a large number of buyers at a sales point (fast-food stores, kiosks, public transport, etc). However, at some point, global card systems such as Visa and MasterCard launched a very intensive campaign to introduce this payment manner at all sales points, first through incentives for implementation, and then through a strict obligation to implement this at all sales points.

Some believe that the reason for such campaign is the advantage of these global card systems relative to local and national card systems, as the latter did not have a technical solution for contactless payments available at the time. Though the EMVCo consortium created the CPA specification for contactless payments as well, it is still of a general type, without sufficient technical details that would enable its easier implementation in practice. This fact contributed to the large delay in the implementation of contactless payments with local and national card systems.

In response to the problem of the lack of an open technical specification for contactless payments, Gemalto implemented the PURE specification, as well as the corresponding CPA payment cards available to all card systems; not long after, IDEMIA began offering a similar solution. In March this year, IDEMIA, G&D and NXP established the White Label Alliance (WLA) with the aim of further developing open technical solutions for contactless payments that would be available to all card systems.

Currently, several contactless payment solutions have been implemented by using the said open standards. For now, most implementations use the PURE solution (e.g. Bcard in Bulgaria and EFTPOS in Australia), and there are more plans for using new, announced open solutions as well.

Regardless of the existing problems and initially modest plans for contactless payments, this type of payments will most likely become dominant in the future. This was facilitated by a number of unexpected market requirements, such as simpler and faster payments without the PIN (up to a certain amount), and the outbreak of the coronavirus pandemic where contactless payment was recognized as a desirable measure in combating the pandemic.

5 Payments by mobile phone

The first concept of payments by mobile phone was the so-called m-commerce, which was first mentioned in 1997 in the sense of transferring functionalities of the existing e-commerce (online card payments) to mobile phones. Initially, this new form of payment brought an advantage only in terms of the greater mobility of buyers who are no longer bound to the computer. Advantages multiplied over time, notably in terms of sending push notifications to the mobile phone, and then in terms of greater security which the mobile phone offers.

In time, the use of mobile phones evolved into various new forms of payment associated with different payment channels (omnichannel).

First, with the emergence of Near-field Communication (NFC) of mobile phones, there also emerged the possibility of contactless payments by mobile phones at a POS terminal. Initially, to implement this payment type, the participation of the mobile phone operator and the SIM card manufacturer was needed, as well as the use of the so-called Single Wire Protocol (SWP), which implied connecting the SIM card and the NFC chip on the mobile phone with a wire through which they could communicate with each other via the SWP. However, because of a large number of users and the complexity of implementation, this solution was not widely accepted in practice.

With the introduction of a new concept called the Host Card Emulation (HCE), there was no longer a need for the above-described complex SWP solution, or for the participation of the mobile phone operator and SIM card manufacturer. The new concept, created in 2012 and more widely implemented as of 2014, implies a software simulation of the chip payment card using advanced cryptographic techniques, which soon gained a much wider application, including the replacement of contactless chip cards with the mobile phone, e-wallets on mobile phones (such as Google Pay and Apple Pay), and a new concept of security for online and mobile phone payments called Secure Remote Commerce (SRC), which will be discussed more in the next chapter.

Interestingly, soon after the emergence of first solutions for payments by mobile phone, mass predictions appeared stating that these solutions were revolutionary and would entirely replace the standard form of payment cards in a very short period. However, although there were many arguments in support of such conclusions, this has not yet happened. The reason is that there are still many limiting factors in practice, such as the following: obligatory NFC functionality which is still a feature of only a small percentage of mobile phones, more expensive ones as a rule; the still relatively complicated settings and use of mobile phones for payments; younger people, with greater technical knowledge as well as regular income, are primarily interested in payment card usage.

Regardless of the existing problems and limitations of mobile phone payments, this type of payment can indeed be considered as one with the best prospects in future, since the mobile phone has become a device we always carry with us. Still, we should not underestimate the continued need of the market for different solutions and payment channels for various purposes and groups of users, hence it is most likely unrealistic to expect that only one type of payment should become dominant over the majority of others.

6 New solutions for online payment security

The first card-based online payments appeared soon after the appearance of the Internet and the first websites, specifically in 1992 on the book sales site www.books.com.

In the beginning of card-based online payments, until 2000, there was no special additional protection for this type of payment, which gradually sparked a number of frauds. As mentioned earlier, with the appearance of chip payment cards, frauds further migrated from transactional payments where payment cards are present (which have become much more secure due to the appearance of chip technology) to online transactions which were less protected. The market responded rapidly with the emergence of new solutions for additional protection in this type of payment, and later with new binding rules, first by card systems, and then by regulators, primarily in the European Union.

The first advanced technological solutions for online payment security were introduced by Visa in early 2000, under the name Verified by Visa, and then under the name 3-D Secure, which was soon accepted by other card associations like MasterCard (under the name SecureCode), Discover (called ProtectBuy), JCB (called J/Secure) and American Express (called SafeKey). All these solutions are based on the 3-D Secure concept, which involves additional protection in 3 domains: card acceptor domain (in practice MPI-Merchant Plug In software module on the merchant’s website), card issuer domain (in practice ACS-Access Control Server for cardholder authentication) and the interoperability domain (in practice DS-Directory Server for control and connection of the other two domains).

These 3-D Secure solutions for online transaction security were first optional for payment participants, and then the card system rules introduced the concept of shifting liability to the party that did not implement them (liability shift). The full obligation to use them was imposed only for certain markets that were considered riskier, but not for all markets globally. The method of implementation was not strictly defined in terms of type and degree of protection, but there was a free choice of the method of protection by the participants (e.g. permanent password, one-time password, additional devices such as token devices, etc.).

For a long time, the 3-D Secure concept was owned by Visa (known as 3-D Secure Version 1). Consequently, other card systems had to either develop their own variants of this concept or pay the licence costs to Visa.

Given the numerous mentioned shortcomings of the existing solutions in the market related to the online payment security (absence of mandatory implementation, too much freedom in the choice of the protection degree, lack of an open standard for free use), in 2015 the European Union adopted the so-called PSD2 Directive (Revised Directive on Payment Services), which, among other things, introduced more detailed and stricter obligations for secure online payments. This directive introduced the mandatory implementation of the so-called strong customer authentication, which means the simultaneous application of at least two of the following three authentication mechanisms – something the user knows (e.g. permanent password), something the user has (e.g. a token or a mobile phone) and something the user is (e.g. a fingerprint, a face photo, an iris scan etc.).

In response to the technical requirements defined in the PSD2 directive, in 2016 the EMVCo association published a new, open version of the 3-D Secure specification that can be used freely by all card systems, known as 3-D Secure Version 2. In addition to free use, this directive has brought much wider possibilities for using innovative methods of user authentication including biometrics. This new version of the 3-D Secure specification has already been implemented in most of the European Union, which significantly reduces card payment frauds in this region and will most likely be disseminated by gradual implementation in the rest of the world, given the similar experiences with chip payment card implementation.

Market reactions to the PSD2 Directive and mandatory two-factor authentication for online payments were initially predominantly negative, primarily due to the necessary investment in implementation and short deadlines. However, after the first implementations, it was recognized that the new version of the 3-D Secure specification offered solutions that can be both more comfortable and simpler for users, and thus more competitive in the market. As an example, the earlier implementation of strong authentication with the complicated use of tokens has been replaced by a simpler use of SMS messages on a mobile phone, and biometric techniques are increasingly being offered which make it even easier for users to authenticate (face or iris recognition via a mobile phone, fingerprint read on a mobile phone, etc.).

Despite the great potential and new possibilities of biometric authentication online, currently the dominant solution in practice is the simultaneous use of a permanent password (what the user knows) and a one-time password sent as an SMS to a mobile phone, which can be considered somewhat burdensome for users. Using biometrics (what the user is) instead of e.g. remembering a permanent password would greatly facilitate authentication. However, when it comes to biometrics as one of the factors of authentication, this area is still in the development phase where there are still many open issues, predominantly related to reliability. Namely, it turned out that it is not enough to use only the so-called static physical characteristics of the user (e.g. facial features, an iris, fingerprints) because they can be copied. That is why the so-called behavioural biometrics have been introduced recently (such as voice recognition, handwritten signature dynamics, typing dynamics, gestures, etc.), which would significantly contribute to reliability and easier user authentication, but on the other hand the implementation of such solutions is extremely complex and often expensive. However, according to previous experiences, if in time there are better and more efficient authentication solutions and they become massified, it would inevitably lead to new accepted standards and cheaper implementation.

Given the great challenges in the implementation of secure online payment, primarily in terms of protection but also easier use, in recent years, global card brands have developed the concept of SRC (Secure Remote Commerce) for which the term “Click to Pay” is often used. On the one hand, this concept offers a simplified and unified use of various secure payment channels by users (from any computer, laptop, tablet or mobile phone), but on the other hand, its implementation is quite complex. The implementation of this concept involves several complex components (SRC system that connects and coordinates other components, DPA-Digital Payment Application through which the user communicates with the system, DCF-Digital Card Facilitator that stores and sends user and payment card data, SRC Initiator that exchanges data between the merchant and the DCF component, and the SRC Participating Issuer which automatically registers users in the SRC system), as well as complex cryptographic solutions based on the 3-D Secure protocol and tokenisation protocols.

For the sake of easier and broader implementation of the new SRC concept (Secure Remote Commerce) for safer and easier online payments, in June this year the EMVCo association published its technical specification “EMV Secure Remote Commerce Specifications, v1.0”, which will certainly facilitate the implementation of this new concept.

7 Payment card industry data security standard – PCI DSS

With the appearance of the first advanced security solutions in using payment cards, such as chip technology and online authentication, new components and new participants in card systems appeared, which increased the complexity of implementation, and thus the complexity of data protection. For easier and more efficient implementation of the data protection segment, in 2006 the global card systems Visa, MasterCard, American Express, Discover and JCB formed the so-called Payment Card Industry Security Standards Council, an association that in the same year published the first version of the standard called PCI DSS – Payment Card Industry Data Security Standard.

The PCI DSS standard first defined the most important security aspects only for individual participants in the card system, to eventually include all participants (acceptors, issuers, processors, merchants, and all other service providers) and security aspects (PIN, payment card identification data, data transmission channels, secure data storage, secure data access, etc.).

At present, the global card systems prescribe the mandatory implementation of the PCI DSS standard for all participants in the system without exception. However, the method of compliance testing depends on the size of the participant, in terms of the transactions volume of the participant – for participants with low volumes, a self-assessment questionnaire (SAQ) is sufficient, for participants with medium volumes, a Qualified Security Assessor (QSA) is required, while for participants with high volumes, an appropriate Internal Security Assessor (ISA) is required. Compliance testing is performed regularly, annually or quarterly, depending on the type of request.

The technical requirements of the PCI DSS standard are divided into six groups:

1. Security of data and system transfer network
2. Security of payment card and cardholder data
3. System vulnerability management
4. Strong access control implementation
5. Regular monitoring and testing of the security of data transfer network
6. Maintaining information security policy

Each of these groups of requirements is divided into subgroups that cover different aspects of data protection for different participants in the system.

PCI DSS standards definitely provide a very high level of data security for all card system participants. However, they also require significant additional costs related to implementation and regular compliance assessments, which is probably a major obstacle for many potential system participants. However, given such a broad obligation of compliance for all participants in the system without exception, it is only logical that massification could lead to a reduction in costs per participant. Unfortunately, the costs of implementation and regular compliance checks are still high, which for now favours larger participants over smaller ones.

8 Expansion of debit payment card use as a consequence of the 2007–2008 global economic crisis

Until the outbreak of the global economic crisis of 2007–2008, credit cards were predominantly used in developed countries, primarily due to the numerous benefits they offered. In addition to favourable credit lines, credit cards were used without restrictions in all types of payments, which was not the case with debit cards that were linked to current accounts and limited by the current account balance, which resulted in additional restrictions in certain types of payments.

It is interesting that, until 2007–2008, card not present payments such as online payments, telephone or email orders were very often not possible for debit payment cards. The reason for this was the very complicated and time-consuming procedure in case of fraud, which was not problematic for credit cards (because the user still had enough funds available due to high credit limits), but it was problematic for debit cards (because the limit is linked to the current account balance, which is much lower than the credit limit, so the user has significantly less funds at his disposal, until the end of the complaint procedure due to fraud). For these reasons, credit cards have long been used for online payments without any additional security (authentication), since in this type of payment, according to the card systems rules, responsibility predominantly rests with merchants. Therefore, in cases of fraud, after the completion of the complaint procedure, as a rule, the cardholder gets back his funds on the credit account.

It is known that the cause behind the global economic crisis in 2007–2008 was primarily the granting of insufficiently controlled loans and credits, which over time resulted in the so-called “bursting of the bubble” first in the United States, and then consequently in the rest of the world. Although this was primarily related to the real estate market, the practice of insufficiently controlled loans was also present in credit lines linked to credit cards. Namely, unlike the previous period when the monthly repayment of the remaining debt on credit cards was usually at 10–15%, that level fell to only 2–3% in 2007–2008, which drastically extended the repayment of debt (from 1–2 years to as many as 5–7 years), and thus drastically increased cardholders’ total debt under interests (from 15–20% to as much as 100–150%).

The described bad experiences of credit card users, as well as the deteriorating global economic situation, led to a massive shift from credit cards to debit cards, which was induced by both card users and issuers. However, in these new circumstances, the previous restrictions on the use of debit cards (such as disabled online payments) were no longer acceptable, which led to the accelerated development of solutions to those restrictions. It can be considered that this situation also contributed to the accelerated development of solutions for more secure online payments, along with the expansion of the use of debit cards.

9 National card systems as a growing trend

The first national card systems emerged in parallel with today’s global card systems such as Visa and MasterCard. The first ever was French Carte Bleue established in 1967, with the European Eurocheque following soon in the same year, with its own variant of a payment card

– the Eurocheque guarantee card. However, the Eurocheque system turned out too expensive for merchants, offering lower quality technical solutions compared to the ever-growing Visa and MasterCard, so Eurocheque cards were less and less accepted by merchants, which finally led to the system closing in 2001. The Eurocheque cards had been predominantly used in Germany, so they continued to be used in that country under a new name EC (Electronic Cash), while being replaced by MasterCard-owned Maestro cards in the rest of Europe. This situation led to the emergence of the German national payment card Girocard in 2007.

Hence, the first national cards systems were established mainly in developed countries, initially with the intention to keep up with the existing systems or even take the lead in payment card development (e.g. French Carte Bleue), and in some cases also in response to the weaknesses of existing card systems and the need to become independent from them (e.g. German Girocard). It can be said that the first national card systems reflected the aspiration for faster development and competitiveness, rather than mere necessity.

The new national card systems emerging with time have been more driven by necessity than by faster development and competitiveness. Namely, the rapid spreading of global card systems Visa and MasterCard strengthened their domination, which gradually resulted in substantial cost increase for system participants and full dependence on those systems in the markets which had no alternative solutions. This problem was first identified in developed economies, only to soon become widely recognised. Such situation was a trigger for the emergence of more and more national card systems worldwide, which remains the current trend.

The fact that generally around 95% of all card transactions are realised in the country and only 5% as international transactions speaks further in favour of national card systems development. Hence, there is a growing number of countries in which national payment cards are being more used that those of global brands.

Table 1 **Overview of national card systems dominant in local markets**

Country	Name of the card system	Characteristics
GERMANY	Girocard	90 mn cards issued
FRANCE	Cartes Bancaires	64.5 mn cards issued, 83% of all cards
DENMARK	Dankort	84% share in the turnover
SPAIN	ServiRed, Sistema 4B, Euro 6000	Over 70 mn cards issued
ITALY	Carta Si, PagoBancomat	-
NORWAY	Bank Axept	7 mn cards issued
BELGIUM	Bancontact	15.7 mn cards issued
BELARUS	BelCard	5 mn cards issued
JAPAN	JCB	77 mn cards issued
CHINA	UnionPay	6 bn cards issued, 100% share in the turnover in the country
SOUTH KOREA	BC Card	52 mn cards issued
SAUDI ARABIA	SPAN	100% share in the turnover in the country
AUSTRALIA	eftpos	70% debit card transactions in the country
NEW ZEALAND	eftpos	60% transactions on POS terminals
CANADA	Interac	-

Source: Websites of selected central banks and national card systems.

Table 2 **Overview of national card systems established in the last ten years**

Country	Name of the card system	Characteristics
BRASIL	ELO	Launched in 2011
INDIA	RuPay	Launched in 2012
RUSSIA	Мир	Launched in 2014
TURKEY	Troy	Launched in 2016

Source: Websites of selected central banks and national card systems.

The number of new national card systems is evidently on the rise worldwide and this trend is likely to continue.

10 Countries as regulators and factors of payment cards development

It is well known that one of the cornerstones of the modern global economy is the free market principle, which should also be the dominant factor of development. However, markets occasionally experience smaller or lesser disturbances, of shorter or longer duration. In situations of extended disturbances, it is only natural and logical that the state should take a more active role in addressing the problem, primarily via the legislation. Such examples can also be found in the payment cards area.

The first payment card-related problems in the market date back to early 1970s, mostly in connection to litigations between merchants and global card systems Visa, MasterCard and American Express.

Already at their onset, card schemes Visa and MasterCard imposed a disputable rule that issuing banks must chose whose cards they are going to issue. After banks' appeals to the U.S. Department of Justice, which were upheld based on the existing antitrust legislation, Visa and MasterCard card schemes were forced to lift this rule and allow banks to simultaneously issue both types of cards.

In the 1990s, Visa and MasterCard card schemes set a new disputable rule preventing the banks issuers of their cards from simultaneous issuing of other cards such as Discover and American Express. The case ended up in the US court in 2001 and the rule was found to have seriously jeopardised the principles of healthy competition and ordered to be abolished.

Simultaneously with the above-described case, in the 1990s large US merchants initiated another court case related to the Visa and MasterCard rule on obligation to accept all types of cards, since credit card acceptance was much costlier for merchants than the acceptance of debit cards. The outcome of this court process was the negotiation with the card schemes to abolish the rule, so that the merchants are free to choose which types of cards they are going to accept.

The period since 2000 has seen dozens of various litigations between merchants and global cards schemes Visa, MasterCard and American Express, primarily related to high multilateral interchange fees which negatively affected merchant costs and final sale prices. These court litigations were conducted also outside the US (e.g. in the European Union and UK), but in most cases without final rulings, as almost all rulings were either cancelled at higher courts or returned for retrial. The reason for this situation most probably lies in extreme complexity of these issues and a large number of separate cases without a single and consolidated approach.

After years of analyses, consultations, discussions and court cases related to high multilateral interchange fees, in 2011 the US Fed adopted the so-called Regulation II (Debit Card Interchange Fees and Routing). This regulation pertains to debit cards only and caps the debit card interchange fee at 21 cents plus 0.05% of the transaction. Exempt from this rule are “small merchants” who are separately defined in the Regulation.

Several years later, in 2015, the European Union (EU) adopted the Interchange Fee Regulation solving this problem in the EU, by strictly defining caps on interchange fees for debit (0.2%) and credit (0.3%) cards. The Regulation further allowed merchants to choose which cards to accept and forbade the equalisation of final merchant fees in cases where different interchange fees are applied for different cards (unblending).

A similar, though somewhat different approach to regulating interchange fees was applied by Australia, which as early as 2003 introduced periodical, the so called benchmark analyses of the current levels of interchange fees, based on which it initially issued suggestions for their independent updating by card brands and later introduced mandatory caps for interchange fees, updated on as needed basis informed by periodical benchmark analyses, as defined by the Interchange Standard for the EFTPOS System, a part of the Payment Systems Act (Section 18). The current maximum interchange fee for credit cards is 0.8%, and for debit cards – 15 cents if defined as a fixed amount or 0.2% if expressed as a percentage.

For the sake of better regulation of payment systems, in 2007 the EU adopted the Payment Services Directive (PSD) which was transposed by end-2009 into laws of all EU countries. In this Directive, payment cards are mentioned as one among numerous payment instruments.

A few years later, in 2015, the EU issued the Revised Directive on Payment Services (PSD2) with a view to regulating safer and more innovative payment systems. Additions related mostly to better regulation of Account Information Services (AIS), and safer payment card use, by introducing mandatory Strong Customer Authentication (SCA). The concept of strong customer authentication assumes the so called two-factor authentication i.e. simultaneous application of minimum two out of three authentication mechanisms: something the user knows (e.g. a permanent password), something the user has (e.g. a token or a mobile phone) and something the user is (e.g. a fingerprint, a face photo, an iris scan etc.). The strong authentication concept is typically associated with online card payments, though in the Directive it actually pertains to all types of card payments (with few exceptions for smaller transaction amounts), which indirectly also means the obligation to use chip cards (as something the user has). Market responses to the PSD2 and mandatory two-factor authentication were initially predominantly negative, primarily as it assumed investments and short implementation deadlines, but soon thereafter the market witnessed novel solutions which, apart from meeting the new requirements, also offered more comfortable and simpler technical solutions for users so it may be said that after initial problems, the Directive did after all bring higher security and more innovative and comfortable solutions for users.

11 Payment card development trends in Serbia

As in many other countries, the use of payment cards in former Yugoslavia begun in 1980s, with payment cards of global card systems. Initially, those were predominantly Visa and Diners cards. After the imposing of sanctions toward FR Yugoslavia in 1992, one of the direct consequences were the problems in using the cards of global card systems. Seeking to overcome this problem, domestic banks launched a domestic YUBA card in 1996.

Following the breakdown of Yugoslavia and the lifting of sanctions in 2000, payment cards of the global card systems were reintroduced in parallel with the use of the domestic YUBA card. However, at that time payment cards were underdeveloped and insufficiently used and the situation was much worse even compared to the period before sanctions. According to some data, the number of payment card users was twice higher in 1992 than after 2000.

Given a rather unfavourable situation with payment cards at that time, the lack of viable technical solutions of the domestic YUBA card and absence of any development in the area, in 2003 the National Bank of Serbia launched the national DinaCard project aiming to speed up the development of this payment area.

Within merely a few years, the national DinaCard project gave a significant boost to development of payment cards in the country, not only DinaCard, but also Visa, MasterCard, Diners and American Express. The total number of cards rose from few hundreds to several million, the number of POS terminals from several thousand to tens of thousands, while ATMs which numbered in dozens grew to several thousand. Under the influence of a new competitive DinaCard and new regulations of the National Bank of Serbia, the costs borne by the participants of the payment card system were significantly reduced, primarily merchant fees and costs of card users.

After the initial success, the national DinaCard project encountered many challenges, typical for all national card systems. In the first place, it is the extremely strong competition of global card systems, mainly in terms of pretty aggressive marketing and pressure on system participants (primarily banks and then merchants as well), and later also in terms of the quick introduction of novel technological solutions which bring progress as well as advantages over the domestic system, where the implementation of new technological solutions is typically slower.

New problems facing the national DinaCard project are of a completely different nature compared to the initial ones, so they may be viewed as additional drivers of further development of card payments in the country. In order to ensure further competitive influence on the domestic payment card market, the National Bank of Serbia and DinaCard system keep abreast of global card development trends, seeking to replicate in Serbia both new technological solutions and legal regulations.

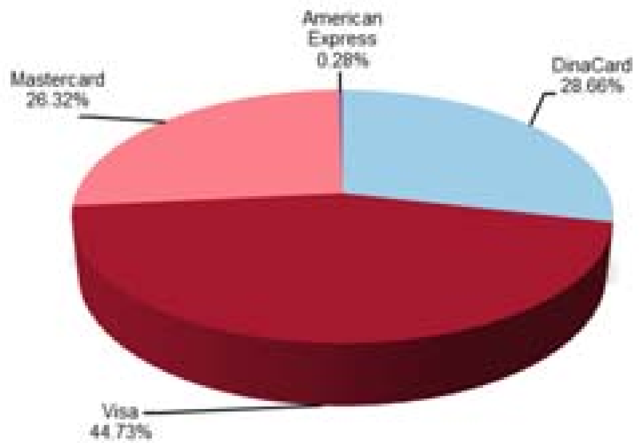
In order to transpose PSD provisions into the Serbian legislation, in 2014 the National Bank of Serbia adopted the Law on Payment Services. In relation to the EU Interchange Fee Regulation, in 2018 the National Bank of Serbia adopted the Law on Multilateral Interchange Fees and Special Operating Rules for Card-Based Payment Transactions. These laws consistently implement the provisions of EU directives into the Serbian legislation, as part of Serbia's EU accession.

Speaking of the latest trends of technological development of payment cards in Serbia, it can be said that all global trends are present Serbia as well. Expectedly, regarding new technological solutions, global card systems are dominant, mainly Visa and MasterCard, and they may be even stronger here with their implementation compared to other markets, owing to the strong competition from the domestic DinaCard. On the other hand, the DinaCard system continuously keeps pace with the new technological trends in order to maintain healthy competition and positively impact the development.

In 2010, the DinaCard initiated cooperation with the Discover system resulting in a joint DinaCard-Discover payment card, which can be used abroad in global Discover, DCI and Pulse networks. In 2018 the DinaCard established cooperation with the Union Pay International, leading to the acceptance of China Union Pay cards in Serbia, with a joint card soon to follow – DinaCard-UPI acceptable internationally in the global Union Pay International network. As of 2019 all newly issued DinaCards have been chip-based and DinaCard contactless cards will soon be issued. Compared to several years ago, the number of online card transactions has also increased multiple times.

While in many countries the national card system is dominant compared to other card systems, this is not the case in Serbia. The goal of the National Bank of Serbia, in introducing the DinaCard system and additional legislation, was not to suppress global card systems in the country, but to restore a healthy competition and remove the existing problems in the market. That this goal was accomplished is demonstrated also by a rather equitable distribution of market shares of large card systems in Serbia, presented below.

Chart 1 **Distribution of the shares of big card systems in Serbia in 2020**



Source: NBS.

12 Conclusion

Due to the multidisciplinary nature of this area, global payment card development trends are affected by a number of factors which are not exclusively technological in nature. Very often these include market, legal, financial, social and even political factors. Due to the unrelenting strong competition in the global payment card market, there are constant challenges in the area, both in terms of new business and technological services and solutions and issues undermining healthy competition. Two dominant courses of action in addressing the existing global challenges are new technological and legislative solutions.

The key word for addressing many existing and future problems in this area is – finding the right balance. The cause to the majority of problems is a disturbed market balance, which sooner or later forces the regulator to take action. It is very important that the regulatory action is also adequately calibrated, in order to preserve the free market principles, while at the same time correcting negative phenomena, which in most cases is not an easy task.

Speaking of smaller participants in the payment card market, primarily local and national card systems, they suffer from a chronic inferiority compared to the global card systems. Global card systems are, on the one hand, the main drivers of market development, but due to their dominance they tend, intentionally or unintentionally, to impose technological solutions, costs and rules. Therefore, the competition in the form of local and national card systems and occasional legal interventions are of great importance for regulating the payment card market.

References

EMVCo LLC, 2014, “A Guide to EMV Chip Technology, version 2.0”, https://www.emvco.com/
Thales Digital Identity and Security, 2021, https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/inspired/biometrics
EMVCo LLC, Jun 2021, EMV Secure Remote Commerce Specifications, v1.0, https://www.emvco.com/emv-technologies/src/
PCI DSS Compliance Guide, 2021, https://www.pcicomplianceguide.org
PCI DSS DOCUMENT LIBRARY, 2021, https://www.pcisecuritystandards.org/document_library
Interchange Fee Regulation (IFR), 2015, Regulation (EU) 2015/751 of the European Parliament and of the Council
Revised Directive on Payment Services (PSD2), 2015, DIRECTIVE (EU) 2015/2366
Reserve Bank of Australia, 2012, Interchange Fees in the EFTPOS System
Law on Payment Services, 2018, “RS Official Gazette”, Nos 139/2014 and 44/2018
Law on Multilateral Interchange Fees and Special Operating Rules for Card-based Payment Transactions, 2018, “RS Official Gazette”, No 44/2018

CIP - Каталогизација у публикацији
Народна библиотека Србије, Београд

336.71

WORKING Papers Bulletin / editor-in-chief Jorgovanka
Tabaković. - [Engl. ed.]. - 2021, No. 1- . - Belgrade : National Bank
of Serbia, 2021- (Belgrade : Institute for Manufacturing Banknotes
and Coins - Topčider). - 30 cm

Ima izdanje na drugom jeziku: Зборник радова (Народна банка Србије) = ISSN 2787-3226
ISSN 2787-3315 = Working Papers Bulletin
COBISS.SR-ID 47831561