CONTENTS

Inflation and Population Age Structure:
The Case of Emerging Economies
   Darya Antonova, Yulia Vymyatnina

Text Mining-based Economic Activity Estimation
   Ksenia Yakovleva

Inflation Forecasting Using Machine Learning Methods
   Ivan Baybuza

Review of the Bank of Russia – IMF Workshop
Recent Developments in Macroprudential Stress Testing
   Elizaveta Danilova, Evgeny Rumyantsev, Ivan Shevchuk

Fear of Forward Guidance
   Alex Isakov, Petr Grishin, Oleg Gorlinsky
Russian Journal of
MONEY AND FINANCE
Founded and published by the Bank of Russia

Editor
Ksenia Yudaeva, Bank of Russia

Deputy Editors
Konstantin Styris, New Economic School, Bank of Russia
Nadezhda Ivanova, Bank of Russia

Co-Editors

Alex Boulatov, International College of Economics and Finance
Philipp Kartavov, Lomonosov Moscow State University
Iikka Korhonen, Bank of Finland Institute for Economics in Transition (BOFIT)
Dmitry Makarov, International College of Economics and Finance
Alexander Muravyev, Higher School of Economics — St.Petersburg
Maksim Nikitin, International College of Economics and Finance

Managing Editor
Olga Kuvshinova, Bank of Russia

Deputy Managing Editor
Nadezhda Petrova, Bank of Russia

Assistant Editor
Nina Bodrova, Bank of Russia

Editorial contacts:
Address: 12 Neglinnaya Street, Moscow, 107016 Russia
E-mail: journal@mail.cbr.ru

© The Central Bank of the Russian Federation, 2018

The articles with additional materials provided by authors are freely available at: www.cbr.ru/eng/money-and-finance

The opinions and views set out in the articles are those of the authors and do not necessarily reflect the official opinion of the Bank of Russia.

ISSN 2618-6799
CONTENTS

Monetary Policy

Inflation and Population Age Structure: The Case of Emerging Economies
Darya Antonova, Yulia Vymyatnina ................................................................. 3

Big Data

Text Mining-based Economic Activity Estimation
Ksenia Yakovleva ............................................................................................ 26

Inflation Forecasting Using Machine Learning Methods
Ivan Baybuza ..................................................................................................... 42

Conference Review

Review of the Bank of Russia – IMF Workshop ‘Recent Developments in Macroprudential Stress Testing’
Elizaveta Danilova, Evgeny Rumyantsev, Ivan Shevchuk ............................... 60

Discussion Paper

Fear of Forward Guidance
Alex Isakov, Petr Grishin, Oleg Gorlinsky ..................................................... 84

All the articles are freely available at: www.cbr.ru/eng/money-and-finance
The Russian Journal of Money and Finance is a scholarly peer-reviewed journal published by the Bank of Russia. Since 2018 it is issued quarterly both in Russian and in English.

The Journal was established in 1938 (from the middle of 1941 to the end of 1945 was not published).

The Journal is included in the list of leading scientific journals of the Higher Attestation Commission and in the Web of Science — RSCI.

Inflation and Population Age Structure: The Case of Emerging Economies*

Darya Antonova, European University at St. Petersburg
daryaad@gmail.com

Yulia Vymyatnina, European University at St. Petersburg
yv@eu.spb.ru

This paper examines the relationship between inflation and population age structure for emerging market economies. We form an unbalanced panel of data for 21 countries for the period 1950 – 2017 and include a number of additional control variables – terms of trade, exchange rate regime, debt-to-GDP ratio, broad money supply growth rates, and PPP-adjusted GDP per capita index. After estimating a variety of model specifications and robustness checks we conclude that the elderly group (65+) in these sample of countries is deflationary, the young group (0 – 19) shows weak signs of being deflationary, and the working group (20 – 64) is found to be inflationary. The deflationary effect of the elderly has been found in some studies for OECD countries, but the findings regarding the young group being deflationary and the working group being inflationary are new. Therefore, the question about the general empirical relation between inflation and the population age structure remains unsettled, and it is probable that the relation between population age structure and other macroeconomic variables is different for emerging market economies and for advanced countries.

Keywords: inflation, population age structure, ageing, emerging market economies

JEL Codes: E31, E40, O11, P52

doi: 10.31477/rjmf.201804.03

1. Motivation and literature review

Since inflation targeting remains the main monetary policy regime among central banks – 40 of them have adopted it,1 – understanding of how and why various factors influence inflation as a key concern for monetary policy

---

* The authors are very grateful to the two anonymous referees of the Russian Journal of Money and Finance and to the participants of the First International Research Conference 'Inflation: New Insight for Central Banks', hosted by the Bank of Russia, for their valuable comments and suggestions.

authorities. Recent attention to unconventional monetary policies and their results has prompted a search for unconventional explanations of inflation dynamics, especially in developed countries. One of the previously unexamined factors that might influence inflation is demographic trends.

The fact that population dynamics, or, more precisely, changes in the labour force, have a direct influence on the economy’s output is widely recognized in growth theory. It is also well understood in macroeconomics that, if a factor of production is suddenly depleted, it will affect its relative price and, through changing production costs, the total output of the economy. The effect of such an event on inflation was mostly disregarded except for the discussion occurring in the early years of World War II (Galbraith, 1941; Hansen, 1941; Keynes, 1940). After that and until the late 1990s there were no studies addressing the link between demographic trends and macroeconomic indicators. A general analysis of how demographic trends influence macroeconomic indicators, including savings rates, inflation, and economic growth was conducted by Lindh (1999), who also addressed the potential of demographic trends for forecasting purposes. Another interesting example suggesting a potential use of demographic trends for macroeconomic forecasting is a very recent study by Buckles et al. (2018), which demonstrates, using the US data, that fertility behaviour might be a leading indicator of economic recessions and not a reaction to the negative economic situation, as was thought earlier. The long-term nature of demographic changes (excluding war periods, natural disasters and similar events that have a striking impact on demographic structure) suggests that any influence of demographic trends on macroeconomic indicators will also be long-term.

The connection between inflation and the population age structure has been often discussed in the context of explaining Japan’s prolonged stagnation since the early 1990s through its ageing population, which leads to a deflationary environment and to expectations of decreasing economic growth in the future. Anderson et al. (2014) confirmed that Japan’s ageing population introduced deflationary pressures on the economy through lower economic growth expected in the future, falling land prices, and dis-saving behaviour. The latter might be surprising, but in the case of Japan most of the savings were held abroad, and their repatriation into the country resulted in a real exchange rate appreciation, the deflationary effect of which was larger than the inflationary effect of additional spending. Yoon et al. (2014), after considering a group of 30 OECD countries in the period 1960 – 2013, also claimed that population ageing and reduced population growth might exert a sizeable deflationary effect on the economy. These empirical findings have been corroborated by the political economy argument advanced by Bullard et al. (2012), i.e. that – elderly people living on their savings prefer a lower inflation to economic growth and unemployment reduction, and thus consistently vote for political parties that pay more attention to inflation. However, the findings of Juselius and Takáts (2015, 2016a, 2016b) do not support the political economy
view, and favour instead a purely economic explanation of the link between inflation and demographic indicators.

Several studies have identified inflationary pressures from older generations: retired people tend to dis-save and stop contributing goods and services to the economy, but continue to demand goods and services. Such effects were confirmed based on the data for the OECD economies by Lindh and Malmberg (1998, 2000), Juselius and Takáts (2015, 2016a, 2016b), and Andrews et al. (2018), as well as for specific countries such as Sweden (Lindh, 1999; Bruer, 2002), the US (McMillan and Baesel, 1990) and Australia (Lenehan, 1996). By the same reasoning, the very young should be inflationary, as was found by examining the OECD data in recent studies by Juselius and Takáts (2015, 2016a, 2016b) and Andrews et al. (2018).

The latter two studies use a similar approach and pay close attention to fine-tuning the age structure, so their results merit a closer look. Juselius and Takáts (2015, 2016a, 2016b) use a group of 22 OECD countries with data covering the period 1955 – 2010. Apart from using the usual dependency ratio and the division of the population into three different groups – young (0 – 19), working age (20 – 64) and elderly (65 and older) – they make use of population polynomials that allow for better capturing the impact of the population structure on inflation, but also economize on the number of estimated coefficients in case of very small age breakdowns (e.g. intervals of 5 years). Their results suggest that the very young (0 – 19) and the elderly (65+) are inflationary, though the very elderly (presumably 80+) might be deflationary. Since the latter group is rather small in each country, the results are to be interpreted with caution. The working age population is found to be deflationary.

Andrews et al. (2018) find similar results for the same group of countries and the same time period as the previous authors, but they employ an additional method – panel-data VAR – and also break down the group of elderly people into young old (65 – 79) and older old (80+), in order to study the potentially contradictory effects of the elderly group size on inflation. They conclude that the older cohort of elderly people is deflationary, while the younger old are inflationary, and that the issue in general merits further research. Both this and the previous paper find that the demographic structure explains a large share of inflation variation.

Therefore, existing studies of the relation between inflation and population age structure have found contradictory evidence as regards the influence of the elderly – inflationary or deflationary or both – depending on whether the elderly group is divided into further sub-groups. Besides, to the best of our knowledge, all existing empirical studies of this issue focused on OECD countries. While recognizing that demographic indicators, due to the advantage of slow changes, might be attractive for forecasting purposes, it is important to also note that this would only hold true in case of established and well-explained regularities. To be of use for monetary policy conduct, the empirical relation between inflation and population age structure have to be (a) established, and (b) explained using a suitable economic theory.
With this paper we aim to fill in the gap in the existing literature by examining whether the link between inflation and population age structure exists for emerging economies as well, whether it is the same or different for these countries in comparison with OECD economies, and how the differences, if any, might be explained. We adopt the methodology used in some of the previous studies for OECD countries to check if the results differ for emerging market economies. Addressing this issue would allow us to determine whether the factor of population age structure is relevant for monetary policy conduct in developing countries.

2. Data

We use data for 21 emerging market economies: Algeria, Argentina, Belarus, Brazil, Bulgaria, Chile, China, Croatia, Czech Republic, Hungary, India, Kazakhstan, Malaysia, Mexico, Peru, Philippines, Poland, Russia, Saudi Arabia, South Africa, and Ukraine. Some of these countries are usually considered to be more on the ‘developed’ spectrum – e.g. Poland or the Czech Republic. We consider countries from the former socialist block to be an important part of our sample, since their inflation experience is different from that of most OECD countries, and is closer to some emerging market economies; however, at the same time, their demographic dynamics are closer to those of OECD countries.

We have selected this particular group of countries on the basis of data availability. Other countries that we considered – Armenia, Azerbaijan, Cambodia, Columbia, Georgia, Indonesia, Kyrgyzstan, Moldova, Mongolia, Romania, Tajikistan, Thailand, Turkey, and Uzbekistan – either lack some of the data needed, most importantly on GDP and real and/or policy interest rate, or only have certain data used in our analysis available for a very short period of time. We considered that having less than 20 years of data for any country will make our results less reliable, and therefore we have opted for a smaller group of countries.

We use annual data on: inflation rate, population age structure data (grouped into age cohorts spanning five years each: 0 – 4, 5 – 9, 10 – 14 etc.), real GDP, real interest rate, policy interest rate, terms of trade, debt-to-GDP ratio, exchange rate regime dummies, commodity export dependence dummy, and GDP per capita (PPP). Most of the data cover the period of 1990 – 2017. For some countries (Chile, Argentina, India) we have longer periods of data. Details on the period for which the data are available for each country, as well as on the sources of the data, are provided in Appendix 1.

We used the data collected to calculate output gap (with the Hodrick-Prescott filter applied to the real GDP data), real policy interest rate (policy interest rate corrected for some indicator of inflation expectations – see the next section for details), and three major population groups: young (0 – 19), working (20 – 64) and old (65+). The sum of young and old divided by working gives us the dependency
ratio. Using our data index for PPP-adjusted GDP per capita, based on data from the World bank we calculate the ratio of a country’s GDP per capita to the USA GDP per capita (both PPP-adjusted). This index is used as one of our control variables.

At a first step we analyse the data for different countries and compare them with those for OECD countries from the studies of Juselius and Takáts (2015) and Andrews et al. (2018). In the advanced economies sample, inflation, though volatile, remains within the bounds from -5% to +25%; however, in our sample, inflation is much more volatile due to several episodes of hyperinflation followed by a period of high inflation (Figure 1). As Figure 1 suggests, the relation between the age structure of the population and inflation might be non-linear. The largest part of this non-linearity, however, seems to come from different age groups having different dynamics over time and being related to inflation in a different manner.

![Figure 1. Inflation variation and dependency ratio](image)

*Note: Hyperinflation episodes are excluded.*

The simplest indicator of population age structure is the dependency ratio. The typical dependency ratio pattern for OECD countries exhibits a peak around the 1960s – 1970s, with a decline afterwards and an increase in the last 10 – 15 years. In our full sample, a slight increase in the dependency ratio is seen only in the last few years (see Figure 2).

A closer analysis of dependency ratios for separate countries demonstrates that most countries in our sample have dependency ratio profiles seemingly close to those of OECD countries (see Figures 1A, 2A, 3A, 4A in Appendix 2).
These are countries of the former Soviet Union, former socialist countries, and a number of other countries. Several countries have rather different dependency ratio dynamics – e.g. China, Brazil, Croatia, and Malaysia (see Figure 5A in Appendix 2). There is no clear pattern for the dependency ratios seen in our sample of countries – some are closer to advanced economies with an ageing population, while others have an obviously younger population structure. It is interesting to note that India and China, the two most populated economies, have slightly different population dynamics at the end of the sample: in India the dependency ratio is still going down, while in China the ageing has only started to be seen recently. The only common pattern for all countries seems to be the fact that dependency ratio dynamics are mostly driven by the young age group.

Figure 2. Dependency ratio and the breakdown of population (ths.) into three age groups (sample mean)

We have omitted hyperinflation episodes from the graphs to make the potential link between inflation and demography more pronounced. However, it is important to stress that the inflation experience of the countries in our sample is very diverse: some countries, such as South Africa or Saudi Arabia had relatively low inflation throughout the whole sample, while a number of other countries had years with inflation of over 1000%, which makes our sample much less homogenous in comparison with the samples used by researchers concentrating on OECD studies.
3. Methodology

Our methodology is defined by two considerations: (1) comparability with previous OECD studies, and (2) data specificity. Our data, unlike those for OECD countries, present much more variation, especially in terms of inflation rates, which are, in almost all cases, clearly non-stationary due to hyperinflation episodes. This means that we have to give more careful consideration to the inflation indicator that we use. Both Juselius and Takáts (2015) and Andrews et al. (2018) mention that demographic indicators, being inertial in nature, should correspond to low-frequency inflation, and they consider annual rates of inflation as being representative of the latter.

It is not very obvious why they did not opt for the use of a frequency filter to extract the low frequency inflation component itself. In our case, with hyperinflation episodes present, a simple application of any filter does not remove all hyperinflation episodes, and in case of a band-pass, reduces the number of observations. Since for many countries we have only slightly over 20 years of data, we consider using a band-pass filter to be undesirable. The problem of hyperinflation episodes stays if we apply a filter on a 3-year moving average. We did not opt for a 5-year moving average for the same reason that we discarded the use of the band-pass filter. In the end, we first removed hyperinflation episodes from the series, and then applied a HP-filter to remove the more volatile part of the inflation series.

Due to the constraints imposed by the dataset size, we opted against using population polynomials as in Juselius and Takáts (2015) and Andrews et al. (2018), and decided instead to work with the three population groups of young, working and old. This setup allows us to compare our results with those of Yoon et al. (2014), and with some of the results of Juselius and Takáts (2015) and Andrews et al. (2018). Though we do not aim to replicate the results previously obtained for the OECD countries, we wished to make our results comparable with those of the previous studies, and therefore tried to follow the most common path in our baseline models.

We checked our data for stationarity using panel data stationarity tests (see Appendix 3 for details), and in general our data proved to be stationary. To avoid spurious regressions as much as possible, we have used a ‘White period’ specification in EViews 10 that computes standard errors that are robust to serial correlation (based on Arellano, 1987, and White, 1980). This allowed us to make inferences from the results that we get even if we have some autocorrelation in residuals. We tested for serial correlation in our results by directly checking the presence of autocorrelation in residuals. Most of our specifications do not have

---

2 We consider, somewhat arbitrarily, that hyperinflation starts from 100% a year. In this we follow the classification suggested, for example, in Acocella (1998, p. 130).

3 We regress residuals on their lagged values and check if the coefficient of such a regression is significantly different from 0 (following the procedure suggested in EViews 10 manual for dynamic panel data (http://www.eviews.com/help/helpintro.html#page/content/panel-Panel_Estimation_Examples.html)).
autocorrelated residuals. We checked for cross-section dependence by using the Pesaran CD test and found that it is absent in almost all cases at the 10% level.

We divided our work into two parts: building baseline models for the whole panel and then providing a robustness check by adding various control variables to account for the specifics of the selected groups of countries, or changing some of the indicators used. Our control variables include dummies for the exchange rate regime (1 if different from the free-floating, which is characteristic for OECD countries) and for commodity export dependence, broad money supply growth rate, terms of trade, debt-to-GDP ratio, and index of relative PPP-adjusted GDP per capita.

4. Baseline models

Our baseline models consider the relationship (if any) between inflation and various indicators of the population age structure: (1) dependency ratio, (2) dependency ratio split between the young and old population to check if the effects are different, (3) three population groups – young, working and old. During the next stage, we added a real interest rate and output gap as the two most important variables that are likely to affect inflation. At this stage, we used the real interest rate, defined as the lending interest rate (weighted average lending interest rate offered by the banking system to the private non-financial sector) adjusted for inflation, as measured by the GDP deflator. This indicator, supplied by the World Bank, reduces the potential of endogeneity in our models, since the dynamics of the GDP deflator and CPI-based inflation are never fully the same for our group of countries. Therefore, our baseline model has the form:

\[ \pi_{it} = c + \mu_i + \mu_t + \beta \text{dep}_{it} + \varepsilon_{it}; \]  

\[ \pi_{it} = c + \mu_i + \mu_t + \beta_1 \text{dep}_{ity} + \beta_2 \text{dep}_{ito} + \varepsilon_{it}; \]  

\[ \pi_{it} = \mu_i + \mu_t + \beta_1 s_{ity} + \beta_2 s_{ito} + \beta_3 s_{itw} + \varepsilon_{it}; \]  

where \( \pi_{it} \) is inflation in country \( i \) at time period \( t \), \( c \) is a constant (we drop it for the last specification since otherwise we have the three population shares summing up to unity, and with a constant included this results in perfect multicollinearity), \( \mu_i \) is a cross-section (country) fixed effect, \( \mu_t \) is a period (time) fixed effect, \( \text{dep}_{it} \) is the dependency ratio as defined above, \( \text{dep}_{ity} \) is the dependency ratio for the young dependents (young divided by working), \( \text{dep}_{ito} \) is the dependency ratio of the elderly dependents (old divided by working), \( s_{ity} \) is the share of young in total population, \( s_{ito} \) is the share of old in total population and \( s_{itw} \) is the share of working in the total population.

A closer study of model (3) reveals a problem, however: it is not clear how to interpret the results of such regression, since we cannot conclude that if a share of young increases this will increase or decrease inflation, since the share of the other two groups of population should decrease in that case. We decided to break
down model (3) into three different models with only two groups in each one. So model 3 became:

\[ \pi_{it} = \mu_i + \mu_t + \beta_1 s_{it}^y + \beta_2 s_{it}^o + \varepsilon_{it}; \]  
(3.1)

\[ \pi_{it} = \mu_i + \mu_t + \beta_1 s_{it}^y + \beta_3 s_{it}^w + \varepsilon_{it}; \]  
(3.2)

\[ \pi_{it} = \mu_i + \mu_t + \beta_2 s_{it}^o + \beta_3 s_{it}^w + \varepsilon_{it}. \]  
(3.3)

We opted for the fixed effects for the cross-section dimension since the countries in our sample are very different with regard to their political and economic situation. We decided to include time-fixed effects in our specification to account for any possible common events (e.g. the oil crises of 1973 and 1979, the Asian financial crisis of 1997 – 1998, the Great Recession of 2008 – 2009, etc.). To check our reasoning, we verified through the likelihood ratio test that we do need both cross-section and period fixed effects. Detailed results for models (1) – (3.3) are provided in Table 1.

We also added model (0) in which we regressed our inflation series on the constant, only keeping both country and time fixed effects present. As can be seen from Table 1, model (0) explains around 60% of variation of the dependent variable. We consider this result as supportive of the fact that we have extracted the part of inflation that might be termed low-frequency, and which can be meaningfully compared with the dynamics of demographic indicators, since adjusted R² for the regression of the original inflation series on a constant with both types of fixed effects present explained only 22% of the total variation.

From the results for models (1) – (3.3) in Table 1, we can see that the age structure of the population seems to have some effect on inflation, though results should be treated with caution due to the presence of serial correlation. Results from various regressions reinforce each other: dependents are deflationary, and if they are detached, the elderly group appears to be deflationary, while the younger group is modestly inflationary. If we check this against the results from models (3.1) – (3.3), then the young are inflationary if they increase at the expense of either the old or the working, and the old are deflationary if they increase at the expense of either the young or the working. The results for the working group depend on which group should decrease: they are deflationary if the old decrease and inflationary if the young decrease. For future reference, we use two baseline specifications taken from here – models (2) and (3.1). While they reinforce each other in terms of qualitative results, model (2) is less restrictive in terms of keeping the third group constant, while model (3.1) allows for further comparison with the previous studies.

---

4 For each of the models (0) – (21) described in this paper, we conducted standard tests on whether we should include country fixed effects, period (time) fixed effects, or, both (LR-test results). In all cases, the tests invariably confirmed that both types of fixed effects should be included. For the sake of brevity, we do not report the details of the results of these tests in the paper, but they are available upon request.
### Table 1. Estimation results for models (1) – (3.3). Dependent variable: $\pi_{it}$

<table>
<thead>
<tr>
<th></th>
<th>Model 1</th>
<th>Model 2</th>
<th>Model 3.1</th>
<th>Model 3.2</th>
<th>Model 3.3</th>
<th>Model 0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.2260***</td>
<td>0.3236***</td>
<td>0.3381***</td>
<td>-3.9219***</td>
<td>0.5290***</td>
<td>0.1036***</td>
</tr>
<tr>
<td>$dep_{it}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$dep_{it}$</td>
<td>-0.1325***</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$dep_{it}$</td>
<td>0.0576*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_{it}$</td>
<td></td>
<td>0.1910*</td>
<td>4.4509***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_{it}$</td>
<td>-4.2600***</td>
<td>-4.4509***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_{it}$</td>
<td>4.2600***</td>
<td>-0.1910*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country effect†</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes**</td>
<td>Yes**</td>
</tr>
<tr>
<td>Time effect‡</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes**</td>
<td>Yes**</td>
</tr>
<tr>
<td>$R^2$ adjusted</td>
<td>0.6060</td>
<td>0.6633</td>
<td>0.6944</td>
<td>0.6944</td>
<td>0.6944</td>
<td>0.5935</td>
</tr>
<tr>
<td>Number of</td>
<td>826</td>
<td>826</td>
<td>826</td>
<td>826</td>
<td>826</td>
<td>826</td>
</tr>
<tr>
<td>observations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Durbin-Watson statistic</td>
<td>0.0320</td>
<td>0.0370</td>
<td>0.0388</td>
<td>0.0388</td>
<td>0.0388</td>
<td>0.0321</td>
</tr>
<tr>
<td>F-statistic</td>
<td>15.4223</td>
<td>19.2634</td>
<td>22.0587</td>
<td>22.0587</td>
<td>22.0587</td>
<td>14.8446</td>
</tr>
<tr>
<td>Probability of</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>F-statistic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Probability of LR test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Probability of cross-section correlation test</td>
<td>0.0175</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0085</td>
</tr>
<tr>
<td>Probability of serial correlation test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

* denotes significance at 10% level, ** denotes significance at 5% level, *** denotes significance at 1% level.
† denotes that for this type of fixed effects it was verified that it should be present in the model, and that both types of fixed effects jointly should be present in the model.
‡ denotes that for this type of fixed effects it was verified that it should be present in the model, and that both types of fixed effects jointly should be present in the model.

Our results as regards the dependency ratio are at odds with the results of Juselius and Takáts (2015, 2016a, 2016b), who find for the group of 22 OECD countries that dependents are inflationary, not deflationary. The breakout of the dependency ratio for the young and the old suggest that the overall negative sign for the dependents is explained by the influence of the elderly. This is also at odds with the results of the studies by Juselius and Takáts – according to them it is the very old (a group of 80+) who are deflationary (a result, corroborated by Andrews et al., 2018). Both of these studies treat results for the very old age
group with care, since the group is comparatively small. For countries in our sample, due to lower longevity, this group is even smaller, so we do not attempt to check for this effect.

Even though we have used a specification that computes standard errors that are robust to serial correlation, we treat the results for models (1) – (3.3) with caution due to the clear presence of serial correlation in residuals. Therefore, we proceeded to check if these were spurious regressions by adding to the selected models (2) and (3.1) variables that are most likely to influence inflation – output gap ($y_{it}$) and real interest rate ($r_{it}$). Real interest rate stands as a proxy for monetary policy, and output gap might influence inflation in the course of the business cycle, especially if the monetary policy did not account correctly for the pressure stemming from the output gap. Such specifications are in line with the studies of Juselius and Takáts (2015, 2016a, 2016b) and Andrews et al. (2018), thus allowing for further comparison. While the output gap is less likely to influence the inflation indicator, corresponding to a more stable part of inflation, persistent mistakes when taking it into account in monetary policy decisions might influence inflation in the longer-term. Adding these two regressors – real interest rate and output gap – might be considered as a crude proxy of either the Phillips curve or of the Taylor rule (‘solved’ for inflation). Since our main concern in this paper is to check if the population age structure is related in any robust way to inflation, we consider that if population age structure indicators remain significant when all of the other, more conventional, factors of inflation are accounted for, then it may be the case that modelling inflation should account for changes in demography. Model (4) in Table 2 below is for the two dependency ratios, and model (6) – for the shares of young and old. Model (8) stands for specification with real interest rate and output gap, but without variables of the age structure.

Since models (4), (6), and (8) have all produced counter-intuitive positive signs of real interest rate coefficients, we have attempted lags of the real interest rate and of the output gap in case of potential endogeneity. Though, as we mentioned earlier, the real interest rate we use is calculated using GDP deflator inflation, thus reducing a potential endogeneity problem, it may still be present. Since monetary policy influences inflation with variable lags that are usually estimated to be about 2 – 2.5 years, but might be longer, we included initially six lags of real interest rate and four lags of output gap. Results for specification with the full number of lags are provided in Appendix 4 (models (4.1), (6.1), and (8.1)). In Table 2, below, we report specifications with only those lags of real interest rate that are statistically significant (model (5) for dependency ratios, model (7) for shares, and model (9) for specification without demographic variables).
Table 2. Estimation results for models (4) – (9). Dependent variable: $\pi_t$.

<table>
<thead>
<tr>
<th></th>
<th>Model 4</th>
<th>Model 5</th>
<th>Model 6</th>
<th>Model 7</th>
<th>Model 8</th>
<th>Model 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.2128***</td>
<td>0.2242***</td>
<td>0.1758***</td>
<td>0.2335***</td>
<td>0.0651***</td>
<td>0.0565***</td>
</tr>
<tr>
<td>$dep_{it}^c$</td>
<td>-0.0345</td>
<td>0.0971**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$dep_{it}^s$</td>
<td>-0.8318***</td>
<td>-0.7042***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_{it}^c$</td>
<td>0.3116**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_{it}^s$</td>
<td>-2.4446***</td>
<td>-1.9822***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$y_{it}$</td>
<td>0.0006</td>
<td>0.0008</td>
<td>0.0000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{it}$</td>
<td>0.1292***</td>
<td>0.1010***</td>
<td>0.1142***</td>
<td>0.0917***</td>
<td>0.1367***</td>
<td>0.1043***</td>
</tr>
<tr>
<td>$r_{it+1}$</td>
<td>0.0864***</td>
<td>0.0813***</td>
<td>0.0022***</td>
<td>0.0024***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{it+3}$</td>
<td>0.0024***</td>
<td>0.0022***</td>
<td>0.0024***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Country effect*</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
</tr>
<tr>
<td>Time effect*</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
</tr>
<tr>
<td>$R^2$ adjusted</td>
<td>0.6594</td>
<td>0.7265</td>
<td>0.6851</td>
<td>0.7371</td>
<td>0.6453</td>
<td>0.7094</td>
</tr>
<tr>
<td>Number of observations</td>
<td>494</td>
<td>476</td>
<td>494</td>
<td>476</td>
<td>494</td>
<td>476</td>
</tr>
<tr>
<td>Durbin-Watson statistic</td>
<td>0.2368</td>
<td>0.7351</td>
<td>0.2147</td>
<td>0.6873</td>
<td>0.2432</td>
<td>0.7830</td>
</tr>
<tr>
<td>Probability of F-statistic</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Probability of LR test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Probability of cross-section correlation test</td>
<td>0.5561</td>
<td>0.2071</td>
<td>0.4737</td>
<td>0.4184</td>
<td>0.0752</td>
<td>0.5305</td>
</tr>
<tr>
<td>Probability of serial correlation test</td>
<td>0.0000</td>
<td>0.0862</td>
<td>0.0000</td>
<td>0.0736</td>
<td>0.0000</td>
<td>0.0859</td>
</tr>
</tbody>
</table>

* denotes significance at 10% level, ** denotes significance at 5% level, *** denotes significance at 1% level.

As can be seen from Table 2, adding demographic structure indicators improves slightly the goodness of fit (compare models (4) – (7) with models (8) – (9)). It is also clear that the output gap is not statistically significant, and its removal improves the fit of the regression (compare model (4) against (5), and model (6) against (7)). This is not a very surprising result for our sample, since the countries in question went through several periods of political and economic instability, often accompanied by hyperinflation episodes, thus blurring the relation between output and inflation until the situation fully stabilizes. It might also be the case that a simple HP-filter does not really remove the cycles from the GDP data (Aguiar and Gopinath, 2007). However, we consider that the most conventional method of obtaining the output gap is suitable for our purpose, i.e. that of establishing the presence of certain empirical regularities. Yoon et al. (2014) mention that the very concept of an output gap depends on the method used to measure the potential output, and the latter clearly depends on the population age structure.
The real interest rate is significant in all cases, and it is positively related to inflation indicator. Adding lags of the real interest rate does not change the result; all significant lags have positive coefficients. While this is contrary to what is theoretically expected, it can be explained by the specifics of the countries in our sample. For countries that are dependent on natural resources exports (a third of countries in our sample are resource-dependent according to UNCTAD – Algeria, Argentina, Chile, Peru, Saudi Arabia, Kazakhstan, and Russia) and that use ‘dirty float’ policies to stabilize their exchange rates (which is the case for these countries for most of our sample period), an increase in the costs of domestic borrowing does not necessarily decrease output and inflation pressures. Indeed, it is often the result of an attempt by monetary authorities to reduce inflationary pressures and stabilize exchange rates at the same time. Appreciation pressure in situations of high-resource prices results in the central bank buying foreign currency to keep the nominal exchange rate at a certain level, and this contributes to inflation (Charemza et al., 2009; Sosunov and Zamulin, 2007). This hypothesis calls in for additional control variables that we use later on, e.g. exchange rate regime dummy, commodity export dependence dummy, and terms of trade index. Another explanation of this positive relation between real interest rate and inflation is that the cause and effect are reversed, i.e. that in periods of high inflation the nominal policy rate is high, thus pushing real interest rates up. There might be several rounds of such policy rate increases coinciding with periods of high inflation, accounting for the positive connection between inflation and real interest rates.

Comparing models (2) and (3.1) to models (4) – (5) and (6) – (7), respectively, we see that as regards the absolute size of the demographic indicators, coefficients becomes smaller when the real interest rate and output gap are included, however the sign stays the same. It is interesting to note that the deflationary effects of the old group are more pronounced and always highly statistically significant, regardless of whether we consider the share of this age group relative to the total population or relative to the working population. This is in contrast with the findings of both Juselius and Takáts (2015, 2016a, 2016b) and Andrews et al. (2018), who find that the young group is inflationary (as is weakly suggested in our case) and highly statistically significant, while the elderly are also inflationary, and only the very old are deflationary. However, our results concerning the deflationary effect of the old group are in line with the results of Yoon et al. (2014), whose main concern is that significant deflationary pressure from the elderly might make higher inflation targets unattainable.

5. Robustness checks

Before proceeding to add controls accounting for the specific features of the studied group of countries, we perform a robustness check of our results by using different indicators of the real interest rates. Since the relevance of the indicator that we use – the lending rate corrected for inflation using GDP deflator – for the
monetary policy and, hence, inflation, might be questioned, we use the nominal policy interest rate and correct it for expected inflation calculated in three different ways. The first two ways acknowledge the fact that inflation in developing and emerging economies is characterized by high inertia (Crowley, 2010; Edwards, 1998; Chopra, 1985). In the first case we mode individual inflation series as AR(1) processes and use one-step static forecasts from these AR(1) models to get the expected inflation (models (10) and (11)). In the second case we use a very crude rule for calculating expected inflation by assuming adaptive expectations with a simple rule of expecting 90% of today’s inflation for the next period (models (12) and (13)). In the third case we rely on the fact that inflation expectations are, effectively, based on our previous experience of inflation (Trehan, 2010), and we use Allais’ method for specifying inflation ‘perceptions’ – as he called them – following the algorithm outlined in Barthalon (2014) (models (14) and (15)). We are aware that none of these ways of extracting inflation expectations from the data are ideal. However, they are simple to produce and allow for various ways of defining the real policy interest rate and robustness check. The results of the robustness checks of models (5) and (7) with three different options for the real interest rate are provided in Table 3.

As can be seen from Table 3, our results presented in Table 2 are robust to the changing real interest rate indicator in the baseline models (5) and (7), which include the real interest rate and its lags. Mostly, regressors retain their relative importance. Except for coefficients for the share of the young or for the dependency ratio of the young, which becomes insignificant in some modifications, coefficients are about the same in absolute size and retain their signs and significance.

We proceed with adding variables that might be relevant for inflation dynamics in the countries under consideration: broad money growth rates ($\Delta m_t$), exchange rate regime dummy (with 1 corresponding to some sort of exchange rate manipulation and 0 otherwise – $\text{exr}_t$), terms of trade index ($\text{ttr}_t$), debt to GDP ratio ($\text{debt}_t$), and GDP PPP index ($\text{y}_{ppp}$). We also tried to include a dummy for the commodity export dependence, but this resulted in a near-singular matrix response regardless of the estimation method used. Since we could not get data on the share of commodities in exports for our sample of countries based on the same method of calculation and for a reasonable period of time (at least 20 years), we had to drop this control variable. Growth rates of monetary aggregates could partially explain the inflation dynamics during periods of hyperinflation and subsequent adjustment. Though we have removed hyperinflation episodes from our inflation indicators, money supply growth rates could still be important in explaining inflation dynamics. Yoon et al. (2014) use the growth rate of M2 as a control variable and find it statistically significant. In our case, across all countries, we obtained only broad money (broader money aggregate compared to M2) growth rates.

---

5 The authors are very grateful to the anonymous referees for suggesting a comprehensive list of control variables on top of our own ideas.
Table 3. Estimation results for models (10) – (15). Dependent variable: $\pi_{it}$

<table>
<thead>
<tr>
<th></th>
<th>Model 10</th>
<th>Model 11</th>
<th>Model 12</th>
<th>Model 13</th>
<th>Model 14</th>
<th>Model 15</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.1317***</td>
<td>0.1331***</td>
<td>0.1713***</td>
<td>0.1714***</td>
<td>0.2311***</td>
<td>0.2250***</td>
</tr>
<tr>
<td>$depay_{it}$</td>
<td>-0.0664*</td>
<td>-0.0688*</td>
<td>-0.0914*</td>
<td>0.0417</td>
<td>0.0696</td>
<td>0.0929</td>
</tr>
<tr>
<td>$depsy_{it}$</td>
<td>-0.2649*</td>
<td>-0.4826***</td>
<td>-0.6783***</td>
<td>-1.0705***</td>
<td>-1.5468***</td>
<td>-2.0614***</td>
</tr>
<tr>
<td>$s_{it}$</td>
<td></td>
<td>0.0417</td>
<td>0.0696</td>
<td>0.0929</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$s_{it}^*$</td>
<td>-1.0705***</td>
<td>-1.5468***</td>
<td>-2.0614***</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$r_{it}$</td>
<td>0.1368***</td>
<td>0.1304***</td>
<td>0.1222***</td>
<td>0.1133***</td>
<td>0.0391*</td>
<td>0.0322</td>
</tr>
<tr>
<td>$r_{it}^*$</td>
<td>0.1035***</td>
<td>0.0992***</td>
<td>0.0661***</td>
<td>0.0638***</td>
<td>0.0169</td>
<td>0.0156</td>
</tr>
<tr>
<td>$r_{it}^*$</td>
<td>0.0045***</td>
<td>0.0042***</td>
<td>0.0046***</td>
<td>0.0043***</td>
<td>0.0022***</td>
<td>0.0019***</td>
</tr>
<tr>
<td>Country effect</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes**</td>
<td>Yes**</td>
<td>Yes**</td>
</tr>
<tr>
<td>Time effect</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes**</td>
<td>Yes**</td>
<td>Yes**</td>
</tr>
<tr>
<td>$R^2$ adjusted</td>
<td>0.7848</td>
<td>0.7887</td>
<td>0.7434</td>
<td>0.7521</td>
<td>0.6907</td>
<td>0.7058</td>
</tr>
<tr>
<td>Number of observations</td>
<td>476</td>
<td>476</td>
<td>476</td>
<td>476</td>
<td>476</td>
<td>476</td>
</tr>
<tr>
<td>Durbin-Watson statistic</td>
<td>0.6329</td>
<td>0.5989</td>
<td>0.5543</td>
<td>0.5313</td>
<td>0.2766</td>
<td>0.2632</td>
</tr>
<tr>
<td>Probability of F-statistic</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Probability of LR test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Probability of cross-section correlation test</td>
<td>0.5105</td>
<td>0.4705</td>
<td>0.7125</td>
<td>0.6949</td>
<td>0.2274</td>
<td>0.1986</td>
</tr>
<tr>
<td>Probability of serial correlation test</td>
<td>0.0687</td>
<td>0.0613</td>
<td>0.0548</td>
<td>0.0512</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

* denotes significance at 10% level, ** denotes significance at 5% level, *** denotes significance at 1% level.

We added the exchange rate regime dummy, since a number of countries in our sample have been involved in managing exchange rates for a considerable part of the sample period, and such practices tend to affect inflation. To capture the fact that inflation might also be influenced by current account pressures, we include terms of trade as another control variable. Yoon et al. (2014) use this control variable for OECD countries, and find it to be negative and statistically significant. Debt-to-GDP ratio might be an important factor adding to inflation for the countries we considered, since in case of prolonged budget deficits and growing indebtedness of the government, central banks in these countries could have been called in for help, thus exerting direct inflation pressure. A similar control variable – budget balance change – was used for OECD countries in Yoon et al. (2014), who found it to have a statistically significant and positive effect on inflation.
We also added an index that relates the country's GDP per capita based on PPP to that of the USA. This allows us to capture the standards of living of the country in relation to an external criterion, and to verify that the relation between inflation and population age structure is not a spurious one. We opted for an index rather than PPP-adjusted GDP per capita to ensure that all variables in our regressions have similar units of measurement by being shares of a unit.

Table 4 shows the first results for the model with two dependency ratios with the full list of control variables and the most parsimonious specification with removed insignificant regressors (models (16) and (17)), and then results for the model with the two shares – young and old – also with the full list of controls and the most parsimonious specification (models (18) and (19)). We have also verified that our results of models (16) – (19) are robust to the choice of real interest rate indicator by re-estimating these models by using the same options of real interest rate as in Table 3. Results are available upon request.

As can be seen from Table 4, inclusion of a full control list resulted in a substantial decrease in the observations included, due to a reduced period of data availability. Besides, we could not obtain data on debt-to-GDP ratio for a third of our sample. Therefore, even though this control variable is significant, we have opted to drop it and concentrate on other control variables, since the excessively high level of debt might influence inflation through the interest rates that are included in our regressions. Broad money supply growth rates are completely unrelated to our inflation indicator. We consider this an additional confirmation that we have managed to extract the low-frequency inflation corresponding to long-term trends, since the latter should be related not so much to monetary factors as to more fundamental ones. It might be noted that the regime of control over the exchange rate resulted in lower inflation for the countries in our sample, meaning that the exchange rate served as a nominal anchor helping to stabilize the economy (Edwards, 2011). Terms of trade in the final specification are slightly inflationary, which is explained by the fact that at least a third of the countries included in our sample are dependent on commodity exports; and higher resource prices, increasing terms of trade index, exert inflationary pressure. This is in contrast with the findings of Yoon et al. (2014) for OECD countries, which is also to be expected, since OECD countries are mostly net importers.

The relative level of GDP per capita does play a role in inflation dynamics for the countries in our sample: the higher standard of living corresponds to lower inflation. This is a logical result, since lower inflation usually produces a more stable economic environment, which allows for higher economic growth and better standards of living. However, presence of the standard of living indicator does not make variables reflecting population age structure totally irrelevant – while the share of the young and the young dependency ratio both become insignificant, the share and dependency ratio of the old remain statistically significant and retain a negative sign.
Table 4. Estimation results for models (16) – (21). Dependent variable: $\pi_{it}$

<table>
<thead>
<tr>
<th></th>
<th>Model 16</th>
<th>Model 17</th>
<th>Model 18</th>
<th>Model 19</th>
<th>Model 20</th>
<th>Model 21</th>
</tr>
</thead>
<tbody>
<tr>
<td>$c$</td>
<td>0.2181***</td>
<td>0.3292***</td>
<td>0.3888***</td>
<td>0.3679***</td>
<td>0.0974***</td>
<td>-0.0283</td>
</tr>
<tr>
<td>$\text{dep}_{it}$</td>
<td>-0.1391**</td>
<td>-0.0368</td>
<td>0.2126***</td>
<td>-0.0283</td>
<td>-0.0283</td>
<td>-0.0283</td>
</tr>
<tr>
<td>$\text{dep}_{in}$</td>
<td>0.2126***</td>
<td>-0.0283</td>
<td>-0.0283</td>
<td>-0.0283</td>
<td>-0.0283</td>
<td>-0.0283</td>
</tr>
<tr>
<td>$\text{dep}_{in}$</td>
<td>-0.3635*</td>
<td>-0.7510***</td>
<td>-0.2720***</td>
<td>-0.2720***</td>
<td>-0.2720***</td>
<td>-0.2720***</td>
</tr>
<tr>
<td>$\text{dep}_{in}$</td>
<td>-0.5283***</td>
<td>-0.1762**</td>
<td>0.5961***</td>
<td>0.5961***</td>
<td>0.5961***</td>
<td>0.5961***</td>
</tr>
<tr>
<td>$\text{dep}_{in}$</td>
<td>-0.6964***</td>
<td>-0.0840</td>
<td>-1.6814***</td>
<td>-1.6814***</td>
<td>-1.6814***</td>
<td>-1.6814***</td>
</tr>
<tr>
<td>$\text{dep}_{in}$</td>
<td>0.5961***</td>
<td>-0.0357**</td>
<td>0.0349**</td>
<td>0.0349**</td>
<td>0.0349**</td>
<td>0.0349**</td>
</tr>
<tr>
<td>$\text{sy}_{it}$</td>
<td>-0.3635*</td>
<td>-0.7510***</td>
<td>-0.2720***</td>
<td>-0.2720***</td>
<td>-0.2720***</td>
<td>-0.2720***</td>
</tr>
<tr>
<td>$\text{sy}_{in}$</td>
<td>-0.0840</td>
<td>-0.1762**</td>
<td>0.5961***</td>
<td>0.5961***</td>
<td>0.5961***</td>
<td>0.5961***</td>
</tr>
<tr>
<td>$\text{sy}_{in}$</td>
<td>0.1356</td>
<td>0.5961***</td>
<td>-0.3635*</td>
<td>-0.3635*</td>
<td>-0.3635*</td>
<td>-0.3635*</td>
</tr>
<tr>
<td>$\text{so}_{it}$</td>
<td>0.0009</td>
<td>0.0009*</td>
<td>0.0049**</td>
<td>0.0049**</td>
<td>0.0049**</td>
<td>0.0049**</td>
</tr>
<tr>
<td>$\text{so}_{in}$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0049**</td>
<td>0.0049**</td>
<td>0.0049**</td>
<td>0.0049**</td>
</tr>
<tr>
<td>$\text{so}_{in}$</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0049**</td>
<td>0.0049**</td>
<td>0.0049**</td>
<td>0.0049**</td>
</tr>
<tr>
<td>$\text{y}_{it}$</td>
<td>0.0704**</td>
<td>0.0492**</td>
<td>0.0701**</td>
<td>0.0488**</td>
<td>0.1450***</td>
<td>0.1559***</td>
</tr>
<tr>
<td>$\text{r}_{it}$</td>
<td>0.0450*</td>
<td>0.0607***</td>
<td>0.0470**</td>
<td>0.0595***</td>
<td>0.1127***</td>
<td>0.1203***</td>
</tr>
<tr>
<td>$\text{r}_{in}$</td>
<td>0.0445***</td>
<td>0.0355***</td>
<td>0.0455***</td>
<td>0.0455***</td>
<td>0.0455***</td>
<td>0.0455***</td>
</tr>
<tr>
<td>$\text{r}_{in}$</td>
<td>-0.0382</td>
<td>-0.0379*</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\text{∆m}_{it}$</td>
<td>-0.0305***</td>
<td>-0.0004</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>$\text{ttr}_{it}$</td>
<td>-0.0077</td>
<td>0.0244***</td>
<td>0.0079</td>
<td>0.0235***</td>
<td>0.0128***</td>
<td>0.0108***</td>
</tr>
<tr>
<td>$\text{debt}_{it}$</td>
<td>0.0525***</td>
<td>0.0561***</td>
<td>0.0525***</td>
<td>0.0561***</td>
<td>0.0525***</td>
<td>0.0561***</td>
</tr>
<tr>
<td>$\text{y}<em>{ppp}</em>{it}$</td>
<td>-0.1127**</td>
<td>-0.4438***</td>
<td>-0.1201**</td>
<td>-0.4271***</td>
<td>-0.1201**</td>
<td>-0.4271***</td>
</tr>
<tr>
<td>$\text{y}<em>{ppp}</em>{it}$</td>
<td>-0.2396***</td>
<td>-0.1791***</td>
<td>-0.2396***</td>
<td>-0.1791***</td>
<td>-0.2396***</td>
<td>-0.1791***</td>
</tr>
<tr>
<td>Country effect</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
</tr>
<tr>
<td>Time effect</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
<td>Yes***</td>
</tr>
<tr>
<td>$R^2$ adjusted</td>
<td>0.7578</td>
<td>0.7578</td>
<td>0.7578</td>
<td>0.7578</td>
<td>0.7578</td>
<td>0.7578</td>
</tr>
<tr>
<td>Number of observations</td>
<td>149</td>
<td>149</td>
<td>149</td>
<td>149</td>
<td>149</td>
<td>149</td>
</tr>
<tr>
<td>Durbin-Watson statistic</td>
<td>0.7053</td>
<td>0.5764</td>
<td>0.7431</td>
<td>0.5644</td>
<td>0.6285</td>
<td>0.6229</td>
</tr>
<tr>
<td>F-statistic</td>
<td>58.8464</td>
<td>23.2489</td>
<td>62.4273</td>
<td>23.7899</td>
<td>70.1053</td>
<td>67.1899</td>
</tr>
<tr>
<td>Probability of F-statistic</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Probability of LR test</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>Probability of cross-section correlation test</td>
<td>N/A</td>
<td>0.0244</td>
<td>N/A</td>
<td>0.0250</td>
<td>0.0094</td>
<td>0.0175</td>
</tr>
<tr>
<td>Probability of serial correlation test</td>
<td>0.0841</td>
<td>0.0516</td>
<td>0.0732</td>
<td>0.0493</td>
<td>0.0231</td>
<td>0.0392</td>
</tr>
</tbody>
</table>

* denotes significance at 10% level, ** denotes significance at 5% level, *** denotes significance at 1% level.

* denotes that for this type of fixed effects it was verified that it should be present in the model, and that both types of fixed effects jointly should be present in the model.
Interestingly, in both cases of population age structure representation – dependency ratios or shares of young and old in the population – we end up with the same most parsimonious specification in terms of control regressors. This suggests similar dynamics of the share of young in the total population and the ratio of young to the working population, with the same applying to the group of old. It should be mentioned that, in comparison with previous results without additional controls, young dependents become not significant, and that the sign changes from children being inflationary to them being deflationary, though with a smaller absolute size of a coefficient in comparison with the old group. This is in contrast with studies of Juselius and Takáts (2015, 2016a, 2016b) and Andrews et al. (2018) who find the young group to be consistently inflationary. There are no direct counterparts to these results in other studies.

We consider the results obtained from model (17) and model (19) to be the most reliable in terms of their quality (though with some reservations about their cross-section and serial correlations), as they capture the influence of monetary policy on inflation (through the dummy for exchange rate regime, and real interest rate) and that of the external factors (through the terms of trade index), they account for countries’ specificity (through country fixed effects), and for any common temporary shocks that these countries might have had (through period fixed effects). The old group is deflationary – it is significant in all models, and the effect is always deflationary, regardless of which particular specification is chosen. This is in line with the results of Yoon et al. (2014) for OECD countries, and is partly corroborated by the evidence from Juselius and Takáts (2015, 2016a, 2016b) and Andrews et al. (2018), though only for the very old group and with some reservations. It is more difficult to make a definite statement about the young group, since in regressions without additional controls they demonstrate an inflationary effect, while in regression with additional controls there is weak evidence for them being deflationary. While we have not calculated the working group directly in our regressions, we can deduce that if this group’s relative size increases at the expense of the other two groups, the net result will be inflationary (due to a lower deflationary effect from the old and, possibly, the young).

The fact that the elderly in our group of countries are deflationary rather than inflationary can be explained by specific features of these countries. These countries are usually characterized by low savings, low pensions, and high levels of poverty among the elderly (see e.g. Barrientos et al., 2003). These factors explain why there is no inflationary effect from the elderly – they cannot spend their savings, and they usually receive many of the services they need from within their households. This is in line with the fact that emerging countries in our sample are characterized by higher levels of traditional values, and concentration on survival (Dülmer et al., 2015), including close family ties, with a much higher proportion of grown-up children living with parents (Szoltysek and Poniat, 2018). This makes it likely that children are raised with the help of family members
(grandparents baby-sitting, second-hand goods for children handed down, etc.), making children-related consumption relatively small and providing arguments in favour of the deflationary effects of the young group.

An inflationary effect from the working group might be in part explained through the ‘Duesenberry effect’. Davies mentions population inflation as one of the important factors influencing both inflation and monetary policy conduct in poorer countries of the 'South and East' as he calls them (Davies, 2002, pp. 6-7). He suggests that inflationary pressures of the relatively young population of the developing countries striving for consumption patterns of a higher social class ('Duesenberry effect') make the conduct of monetary policy much more difficult compared to the developed countries of the global North and West, where population pressures are less pronounced. Our results provide tentative support for the existence of the 'Duesenberry effect' which is in line with the results of the World Values Survey, suggesting that a higher share of young people in emerging economies considers being rich as important.6

The specifications that have been tried so far did not allow for studying the dynamic effects of demographic changes on inflation. While adding lags of the dependent variable is a desirable way to address the issue, this would require finding an instrument for inflation, which is a difficult task and might be better addressed in further studies. We have opted to check if the lags of our population age structure variables add something to the picture of inter-relations between inflation and demography. We have added to models (17) and (19) 10 lags of either shares or dependency ratios of both young and old groups. In Table 4 models (20) and (21) present the most parsimonious specifications excluding insignificant regressors. While the quality of these regressions is also compromised by the presence of serial and cross-section correlations, we rely on the use of standard errors that are robust to serial correlation. We have opted not to use more than 10 lags since this would decrease the size of the sample and might further compromise the validity of the results.

As can be seen from models (20) and (21), the old group is deflationary both at present, as well as 10 years ago, while the latter has a larger negative coefficient, implying that the deflationary effect of the elderly is consistent over time. In a sense, this might be viewed as confirming the tentative results of Juselius and Takáts (2015, 2016a, 2016b) and Andrews et al. (2018) that the very old group is deflationary. An interesting result from models (20) and (21) is that while the present young group seems to be deflationary, the young group of 10 years ago proves to be inflationary. Since within 10 years roughly half of this group would have entered the working group, this indirectly confirms the inflationary effect of the working group. While models (20) and (21) provide some hints about the dynamics between inflation and population age structure, a more thorough and better-structured investigation of the issue is needed, and constitutes a good topic for further research.

---

6 Results of wave 6 at http://www.worldvaluessurvey.org/WVSOntline.jsp
6. Concluding remarks

The purpose of this paper has been to investigate the existence and nature of a link between inflation and population age structure for a group of emerging market economies comprised of 21 countries, and to compare the results with those from the previous studies of the same issue for OECD countries. Our selection of countries was based mostly on the basis of data availability, since we tried to get at least 20 years of data for all countries in our sample.

Since in our sample a number of countries had hyperinflation episodes, we had to adjust the inflation indicator by excluding those episodes, and by filtering out the cyclical component to concentrate on the more stable component of inflation, which could be deemed as corresponding to low-frequency inflation. This indicator was proved to be stationary and, in our view, captures the more stable and long-term oriented part of inflation. This contrasts our study with those focused on OECD countries, where annual inflation was considered as corresponding to low-frequency inflation without additional adjustments.

Another difference of our study from those for OECD countries is the set of control variables used. They include terms of trade, exchange rate regime characteristics, debt-to-GDP ratio, broad money supply growth rates, and PPP-adjusted GDP per capita index. All these factors have a direct or indirect bearing on the monetary policy conduct of the countries in our sample, and have the potential to influence inflation.

We have found that variables describing population age structure are jointly significant in most settings, and this is in line with the previous OECD studies. Our main result is that the old group (65+) is deflationary, when considering both the share of this group in the total population, and its dependency ratio. The effect is consistent across various model specifications. This corresponds well to the conclusions of Yoon et al. (2014), but is not exactly the same as the results of Juselius and Takáts (2015, 2016a, 2016b) and Andrews et al. (2018). The explanation for the elderly being deflationary in our study involves such specific features of the group of the countries we studied as low savings, low pensions, higher levels of traditional values and concentration on survival. Unlike many OECD countries, elderly people in emerging markets are usually at the bottom of income distribution, so it is not surprising that they are not inflationary.

We have found weak signs that the young group is deflationary as well, though the effect is not so consistent across different specifications, and it is not always significant at 10% level. It is not surprising that the young are weakly deflationary in the selected group of countries, again due to higher levels of traditional values and a high level of family ties, implying that people tend to live in extended families, providing free childcare and elderly care.

An indirect conclusion stemming from our results concerning the working group of the population is that this group tends to be inflationary, which is in contrast with the results found for the OECD countries. This can be explained by the
fact that, in our sample, the working group composition is skewed towards younger people, for whom the 'Duesenberry effect' – the desire to imitate upper classes' consumption patterns – might be at work, as is confirmed by the results from the World Values Survey.

Since our results are different to some of the conclusions of those focusing on OECD countries, the issue of the empirical relation between inflation and the population age structure is far from being resolved. Further studies might concentrate on the relationship between the population age structure and other macroeconomic indicators, e.g. savings rate dynamics and economic growth rates, for emerging market economies, and compare the results with those obtained for advanced countries. Another important line for further inquiry is an examination of the dynamic effects of population age structure on inflation, both for emerging market economies and the OECD countries. An accumulation of evidence concerning empirical regularities of this sort should be helpful in developing a suitable macroeconomic theory that accounts for the composition of the population in a given country.

Finally, we would like to stress that the growing share of the elderly in most of the countries, including emerging economies, might become a concern for monetary policy conduct, since this group has a negative relation to inflation and thus might present problems when an expansionary monetary policy is needed. It is important to highlight that we consider our results as having qualitative rather than quantitative value, in showing the relative importance of various age groups for inflation dynamics.

Appendices are available at www.cbr.ru/eng/money-and-finance; dx.doi.org/10.31477/rjmf.201804.03

7. References


Text Mining-based Economic Activity Estimation

Ksenia Yakovleva, Bank of Russia
yakovlevakv@cbr.ru

This paper outlines a methodology for constructing a high-frequency indicator of economic activity in Russia. News stories from internet resources are used as data sources. News data is analysed using text mining and machine learning methods, which, although developed only relatively recently, have quickly found wide application in scientific research, including economic studies. This is because news is not only a key source of information but a way to gauge the sentiment of journalists and survey respondents about the current situation and convert it into quantitative data.

Keywords: economic activity estimates, nowcasting, text mining, machine learning, Big Data, data mining, topic modelling, sentiment analysis

JEL Codes: C51, C81, E37

doi: 10.31477/rjmf.201804.26

1. Introduction

Understanding the current economic situation and its future dynamics is extremely important for implementing timely and effective economic policy. Of special value to politicians, businesses, and other economic agents is the ability to track real-time economic and financial indicators to enable prompt decision-making.

Scientific literature has recently devoted increasing amounts of attention to the development of high-frequency internet-based indicators. This data includes virtually all publicly available information: online-shopping websites, job search engines, news sources, social media, etc. That said, the internet presents practically all of this information in an unstructured form, i.e., as text. This makes it impossible for the individuals to promptly process large information volumes on their own, so researchers in a variety of areas, economics among them, are developing new statistical methods to extract and analyse unstructured internet data.

Interest in this kind of information stems from its advantages over conventional statistical data. These advantages are provided by the sheer volume of information and the speed of its collection, as well as data variety and reliability. On top of that, it can be used to construct new, unique indicators that are not available in official statistical reports.
In this study, we constructed a high-frequency indicator (computed using a daily news flow), which enabled a country’s economic activity trends to be assessed. The PMI index was chosen as a target indicator reflecting economic activity. The results obtained suggest that text analysis allows for the monitoring of the current economic situation on a daily basis with fairly high accuracy.

Section 2 of the paper reviews the key literature dealing with the use of text analysis and Google Trends in the financial and economic areas. Section 3 provides a methodology for constructing the news index. The Conclusion sums up the results of the study and considers the ways of refining the news index.

2. Literature review

The foreign literature provides a fairly large number of papers studying text analysis in economics. These papers mainly deal with short-term forecasting – nowcasting – using text information. Although forecasting stock market movements is a challenging task because of the market’s high volatility and heavy data noise, quite a few papers explore this subject. Alsing and Bahceci (2015), for instance, attempt to forecast movements in share prices of three major companies (Walmart, Netflix, and Microsoft) using relevant publications in the Twitter social network. This social network was chosen because it provided the opportunity to extract user opinions and conduct their sentiment analysis in order to predict future share price movements. The study found an artificial neural network that predicted Walmart stock price movements with 80-percent accuracy to be the best model.

Economic research papers on text analysis usually deal with short-term forecasting (nowcasting) of GDP and industrial output growth, inflation, the unemployment rate, sales, and other indicators.

Ardia and Bluteau (2017) attempted to predict U.S. industrial output growth based on newspaper stories covering the period from 2001 to 2016. The model thus constructed provides evidence that text analysis dramatically improves the accuracy of industrial output growth forecasting over horizons of nine months and one year.

Another study (Nyman et al., 2014) suggests that text analysis substantially improves the accuracy of forecasting the performance of the Michigan Consumer Index performance. The study only used texts containing the words ‘anxiety’ and ‘excitement’. Two time series were constructed using the average number of occurrences of each word per story in each month, with the difference between them established. The indicator thus constructed was included in the regression, enhancing the model’s predictive power.

---

1 Purchasing Managers Index is a macroeconomic indicator of business activity in the manufacturing and services sectors calculated based on surveys of managers.
2 Sentiment analysis identifies and categorises the author’s emotional attitude to a particular object.
3 MCI is a monthly index of American consumers’ confidence constructed by Michigan University.
Entirely new, high-frequency indexes can be constructed based on text analysis. One example is the Uncertainty Index (Bloom and Baker, 2016). The idea behind constructing this index was to aggregate newspaper stories containing the three keywords related to uncertainty, the economy, and policy, so as to monitor the instability of economic policy. This index was constructed for various countries (including Russia) and various activity areas, such as health service and national security.

Also worth noting among the tools for nowcasting economic indicators is Google Trends4, which analysts have started using extensively for short-term forecasting of various indicators. Information obtained from Google Trends is used primarily to get an insight into the current economic activity and sentiment.

Most of the papers claim that using Google Trends as part of forecasting models improves the accuracy of forecasting results. Choi and Varian (2012) used Google search query statistics for nowcasting car sales. The study finds that the first-order autoregression model with the Google Trends time series has a better predictive power than that of the conventional first-order autoregression model.

Aside from car sales, Google Trends improves the accuracy of forecasting jobless claims, consumer confidence, and tourist statistics. The regression results have shown that autoregression models (AR-models) that include the Google Trends variables yield a 5 – 20% higher prediction accuracy than conventional AR-models (Choi and Varian, 2012).

Studies by Russian researchers have also started to use text analysis and the Google Trends service. Goloshchapova and Andreev (2017), for example, suggested a new approach to constructing inflation expectations indicators based on the algorithms of internet user opinions (comments) regarding inflation. It was found that the index measuring inflation expectations based on the internet may become a viable analogue of official survey data.

3. Main results

The key macroeconomic indicator reflecting a country’s economic activity is GDP growth, which all analysts, the business community, and various government agencies keep track of. However, GDP data is published on a quarterly basis, and with a significant lag, which prevents essential information from being obtained in a timely manner. A large number of papers concerned with GDP nowcasting have sought to address this drawback. With digital information and Big Data technology development, GDP nowcasting has also started to be gradually modified. (Kapetanios and Papailias, 2018). GDP nowcasting can rely on a great variety of data, ranging from textual information

---

4 Google website analysis of users' searches of a particular term relative to the total number of search queries.
to electronic payment system data on credit and debit card transactions (Galbraith and Tkacz, 2015).

Thorsrud (2016), for instance, only used textual information from daily business newspapers to forecast GDP growth. The author deconstructs news topics into time series using latent Dirichlet allocation and employs the dynamic factor model for forecasting GDP.

Thorsrud (2016) was taken as a basis for this study but with a number of differences. One of them is that it is not the quarterly GDP number that is forecast but the diffusion index of business activity⁵ (Purchasing Managers’ Index, PMI)⁶ calculated on a monthly basis. GDP was replaced because of its limited news database, as quarterly GDP numbers are insufficient for forecasting.

Diffusion indices are superior in that they have a strong predictive power and are leading indicators closely correlated with business cycles.

PMI, released on the first business day of the month⁷, is a key diffusion index, which fairly accurately tracks business cycle dynamics, as evidenced by its close correlation with GDP (Figure 1).

3.1. News data collection and processing

News data is the basis of this study, as the business activity index will be forecast based on them. An important problem to be addressed in dealing with text analysis is the choice of a news source, as this should reflect the current economic situation with maximum accuracy and cover a significant proportion of the population.

The following key criteria are to be met in choosing a news source. First, the news should be concerned with economic issues; second, a sufficiently long time series should be available on the internet (at least 3 – 4 years); and third, web scraping⁸ should be simple, i.e., information should be easy and fast to extract from the website.

A news resource solely devoted to economic developments both in Russia and abroad was found to meet these criteria. It should be noted from the start that using just one data source is most probably insufficient for such analysis to be carried out, as it may be affected by a number of factors, such as less than comprehensive coverage of economic topics, a specific target audience, etc. On the other hand, Thorsrud (2016) also uses just one business newspaper, the fourth

---

⁵ A diffusion index constructed based on survey data. Each respondent answers questions related to business conditions: new orders, the labour market, contract performance time, and so on.

⁶ The Composite PMI Index is a weighted average of the Manufacturing Output Index and the Services Business Activity Index. The Manufacturing PMI Index is based on five key indicators with the following weights: new orders – 0.3, output – 0.25, employment – 0.2, timing of raw materials and supplies deliveries – 0.15, raw materials and supplies inventories – 0.1. The Services PMI Index is calculated by weighing percentages of respondents’ answers with the following weights: improvement/growth – 1.0, unchanged conditions – 0.5, worsening/decline – 0.0.

⁷ https://www.markiteconomics.com

⁸ Web scraping – a technology for extracting data from web sites.
largest in terms of readership. Still, reliance on just one source may become a problem, so further research is needed to increase the number of sources used.

**Figure 1. GDP growth rate and PMI**

The resulting sample had just over 59,000 news items (Table 1). News items were collected over a time span of four and a half years – from 2014 to 2018 – and contained 54 month-long periods. The sample was from the outset split into training and testing parts, with the former used to construct the model and optimise its parameters, and the latter serving as a tool to assess the quality of the model constructed. The learning sample, as a rule, accounts for 75 – 80% of the raw data, although there are no strict rules in this regard.

<table>
<thead>
<tr>
<th>Table 1. Descriptive statistics of news data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Period</strong></td>
</tr>
<tr>
<td>Number of news stories mined, total</td>
</tr>
<tr>
<td>Number of news stories per month (on average)</td>
</tr>
<tr>
<td>Learning sample</td>
</tr>
<tr>
<td>Test sample</td>
</tr>
</tbody>
</table>

As the text is unstructured data, it needs to be converted into a structured form, allowing for the identification of unknown patterns and construction of topics.

---

Before textual data is converted into a quantitative structured vector, it needs to be processed. The processing of news data is an important stage of text analysis: first, it allows for the reduction of data dimensionality, thereby drastically accelerating information processing; and second, a better text preparation from the outset helps obtain final results of higher quality that can be more easily interpreted.

Processing involves several steps. The first step is stemming: reducing words to their stems using the free MyStem\textsuperscript{10} software developed by Yandex in 1997 to conduct morphological analysis of the Russian language. It’s operating principles are outlined in an article by one of Yandex’s founders, Ilya Segalovich (Segalovich, 2003). The second step involves text processing: removing punctuation, numbers, unnecessary spaces, and ‘stop words’\textsuperscript{11}. Filtering news texts in this way significantly reduces the raw data without losing its semantic component. Table 2 provides examples of raw and processed texts. The processed words in the filtered texts are called terms and used as the basis for a document-term matrix (dtm). Each row of the matrix denotes an individual term, and each column is a separate document.

<table>
<thead>
<tr>
<th>Table 2. Input and processed news texts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Raw news text</td>
</tr>
<tr>
<td>U.S. initial jobless claims rose last week (July 21–28) to 365 thousand, according to U.S. Labor Department report, up 8 thousand from last week’s revised number of 357 thousand (versus an initial estimate of 353 thousand). Experts surveyed by Bloomberg expected the jobless claims to increase by 17 thousand over the week of July 21-28 from the earlier announced level to reach 370 thousand. Thus preliminary data suggest that jobless claims rise was twice as low as forecast.</td>
</tr>
</tbody>
</table>

3.2. Construction of a quantitative indicator

Two indicators were constructed as part of this study. The first indicator reflects the quantitative component of the topics, i.e., how frequently the topic is mentioned in news data. The second represents the emotional component, i.e., the tone of the news.

Since we are interested in all news items rather than a specific topic, this study analyses all news data collected by web-scraping.

To identify which topic a particular news story represents, we use topic modeling enabling data to be automatically sorted out by topic.

\textsuperscript{10} https://tech.yandex.ru/mystem/ [in Russian]

\textsuperscript{11} ‘Stop words’ are connective words of minor semantic importance that connect and make logical transitions between sentences, paragraphs, etc. They include conjunctions, prepositions, interjections, particles, parenthetic words, demonstrative pronouns, as well as some nouns, verbs, and adverbs.
The most popular topic models can be divided into two groups: algebraic and probabilistic. Among algebraic models are the Vector Space Model (VSM) and Latent Semantic Analysis (LSA). Algebraic models allow for the identification of the weights of words and the assessment of similarities between the texts under consideration.

The base probabilistic models are Probabilistic Latent Semantic Analysis (PLSA) and Latent Dirichlet Allocation (LDA). The PLSA develops the LSA model and allows a better identification of possible topics of the document. But the number of parameters increases with an increase in the number of documents, necessitating the overfitting of the model (Vorontsov and Potapenko, 2013). The LDA model was designed to eliminate this overfitting problem.

In exploring the key topic model, algebraic models were found to be unsuitable for handling the data mining task, as they largely aim to extract the key words from the text and compare them. The best among the probabilistic models is LDA, because it eliminates the main PLSA drawbacks, and it is the most widespread probabilistic model – used, among other places, in Thorsrud (2016).

The LDA model, however, also has some drawbacks, one of them being the absence of linguistic notions: it only takes into account the frequency of word occurrence, rather than the order and meaning. Another drawback is that it assumes that words and documents follow a normal distribution, whereas the Poisson distribution is closer to reality.

LDA is a three-layer hierarchical Bayesian model in which each document in the corpus is modelled as a collection of nonobservable, latent topics. According to LDA, each word in a textual document belongs to an unknown topic, and each topic is modelled from the initially specified probabilities of the topics:

\[
p(\theta,z,w|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^{N} p(z_n|\theta) p(w_n|z_n,\beta),
\]

where \(\alpha\) and \(\beta\) are given vectors – the Dirichlet distribution parameters; \(\theta \sim \text{Dir}(\alpha)\) is the distribution of topics in each document; \(z_n \sim \text{Dir}(\beta)\) is the distribution of words in each document; \(w_n\) is the conformity of the words in the document to the topics.

The LDA model requires the initial parameters, or the number of topics here, to be set. The rest of the parameters were used by default. The analysis that we conducted showed that the optimal number of topics for all news data collected was 50. When there are more topics, they tend to overlap and to be duplicated, and when there are fewer topics, several tend to be merged into one. We therefore apply the LDA model to the pre-processed corpus, taking the number of topics as 50.

---

12 A corpus is a set of textual documents. A pre-processed corpus means that all words that do not make sense from the perspective of machine learning, such as numbers, punctuation marks and other characters are eliminated to filter out noise in the documents.
The LDA model produced a list of words (unigrams) relevant to all of the 50 topics. Figures 2 and 3 present 100 unigrams for two topics in the form of a word cloud. To identify the topics of the first and second word clouds, one needn’t scan all the words. According to Lau and Newman (2010), the first 10 words contain 30% of information about the topic, which is adequate for full understanding. The most conspicuous words suggest that documents related to the first word cloud are concerned with oil, while in the second example they deal with fiscal policy.

Figure 2. Unigrams for oil

The most conspicuous words: oil, price, oil product, petrol, demand, oil inventories, volume, grow up, growth, extraction, OPEC, U.S., Saudi, Brent, market

Source: author's estimates

Figure 3. Unigrams for fiscal policy

The most conspicuous words: deficit, finance ministry, pension, revenue, reserve, current, outflows, fund, ruble, first, level, GDP, balance, budget, export

Source: author's estimates
The examples of 15 major topics constructed using LDA are given in Table 3, which provides five words for each topic. The words are then used to sort out the topics. It was decided to select five words rather than ten, in order not to overload the model, and to accelerate information processing; and also because words relevant to economic issues are informative enough, so their automatic classification should not present any difficulties.

<table>
<thead>
<tr>
<th>Topic</th>
<th>Key words (unigrams)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Topic 1</td>
<td>Bond, credit, yield, securities, volume</td>
</tr>
<tr>
<td>Topic 2</td>
<td>Dollar, ruble, exchange rate, euro, currency</td>
</tr>
<tr>
<td>Topic 3</td>
<td>Ukraine, Ukrainian, Crimea, Kiev, hryvna</td>
</tr>
<tr>
<td>Topic 4</td>
<td>Fund, investor, asset, financial, investment</td>
</tr>
<tr>
<td>Topic 5</td>
<td>Exports, commodity item, products, imports, ton</td>
</tr>
<tr>
<td>Topic 6</td>
<td>Bank, banking, Sberbank, capital, VTB</td>
</tr>
<tr>
<td>Topic 7</td>
<td>Debt, IMF, creditor, support, default</td>
</tr>
<tr>
<td>Topic 8</td>
<td>Natural gas, Gazprom, delivery, cubic meter, stream</td>
</tr>
<tr>
<td>Topic 9</td>
<td>Oil field, oil, project, Rosneft, oil production</td>
</tr>
<tr>
<td>Topic 10</td>
<td>Development, project, investment, business, establishment</td>
</tr>
<tr>
<td>Topic 11</td>
<td>China, Chinese, People’s Republic of China, yuan, Asia</td>
</tr>
<tr>
<td>Topic 12</td>
<td>Source, energy, coal, electricity</td>
</tr>
<tr>
<td>Topic 13</td>
<td>USA, Trump, American, Obama, state</td>
</tr>
<tr>
<td>Topic 14</td>
<td>Oil, price, barrel, oil production, OPEC</td>
</tr>
<tr>
<td>Topic 15</td>
<td>Finance Ministry, budget, revenue, expenditure, deficit</td>
</tr>
</tbody>
</table>

The quality of topics thus produced was assessed by comparing them with actual events. Figures 4 and 5 present fluctuations in the intensity of topics from January 2014 to July 2018, declining in some cases and rising in others. Figure 4 shows the Ukrainian topic, the intensity of which surged in March 2014, when the Crimean referendum was held. In subsequent periods, the focus on the Ukrainian topic started tapering off, and its intensity in economic media is now at its lowest since early 2014, based on our estimates.

A surge in the number of news stories is also evident in the American topic (Figure 5). Their number more than doubled because of the November 2016 presidential election in the U.S.

Quality assessment shows that the first five words identified using the LDA model help capture key peaks of developments, forming quantitative indicators.
3.3. Construction of the emotional indicator

To recognise the emotional colouring of the text, the problem of its classification needs to be addressed. In our case, classification should seek to determine the tone of the text as ‘positive’ or ‘negative’.

The automated analysis of the text tone often uses the following approaches:
1. rule-based approaches;
2. dictionary-based approaches;
3. supervised machine learning;
4. unsupervised machine learning.
An analysis of each approach provided in Voronina and Goncharov (2015) suggests that the rule-based approach is the most accurate but also the most difficult to implement, as well as the most labor- and time-intensive. The dictionary-based approach has substantive constraints, and currently there is no Russian-language dictionary of economic terminology. The unsupervised approach offers the lowest prediction accuracy. This study therefore used the supervised learning approach to determine the tone of the text. This approach, as a rule, provides for a relatively high quality of classification. The chosen method was the support vector machine (SVM), which marks samples as belonging to two categories using a separating hyperplane, so that the distance from it to the nearest data points of the set is maximised.

A number of empirical studies suggest that the SVM method is well suited to text classification, offering advantages over other methods (see, for example, Joachims, 1998). Basu and Walters (2003), exploring automated classification of news texts, found the SVM method to be superior to neural networks as regards classification quality. Sassano (2003) used the Reuters database to show how the SVM method could improve text classification accuracy.

To train the classifier using the SVM method, relevant tones were identified in the sample. The symbols ‘1’ and ‘-1’, corresponding to positive and negative tones, were assigned manually to individual news stories. A total of 3,438 news stories were used in constructing the model, of which 2,600 (76%) were chosen as ‘training’ and 838 – as ‘test’ ones (24%).

**Table 4.** Tones in the sample

<table>
<thead>
<tr>
<th>Actual value</th>
<th>Model</th>
<th>Positive</th>
<th>Negative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>TP</td>
<td>350 (TP)</td>
<td>96 (FN)</td>
</tr>
<tr>
<td>Negative</td>
<td>FP</td>
<td>206 (FP)</td>
<td>186 (TN)</td>
</tr>
</tbody>
</table>

Using values from Table 4, a metric can be computed showing the quality of the classifier that we have constructed.

\[
\text{Accuracy} = \frac{TP + FP}{TP + FP + FN + TN} = 64\%
\]

The accuracy of the classifier is computed as the share of accurately predicted values in the total number of values in the test sample. The tone was correctly predicted for 536 out of 838 news stories (64%).

The result of the SVM method is the distribution of tone estimates (‘1’, ‘-1’) and their probability. If the probability is less than 60%, these texts are excluded from the sample, because they identify the tone with low accuracy and may produce biased results. For the remaining texts, the tone was multiplied by probability in
order to assign greater weights to news stories which are classified with the highest accuracy.

3.4. Construction of the forecast model

The quantitative series with topics are multiplied by the relevant tone, forming new time series which will be subsequently used to construct the forecast model for the PMI index.

To eliminate the daily noise in the time series of each topic, all the data is smoothed using the three-month (87 days) moving average. The three-month moving average was chosen because the PMI index itself has a fairly strong predictive power, since the managers surveyed take into account, among other things, the economic situation anticipated in the future. So it is not only the current month’s values that need to be taken into account (conducting data smoothing for 30 days) but also data for previous months. The results suggest that the model using the three-month moving average has the best predictive power.

To compare daily topics with the monthly PMI index, the topics are converted into monthly series by identifying the average monthly value. As a result, 50 monthly time series regressors were obtained, each characterising a particular topic.

When the number of regressors (50 regressors) is larger than that of observations (34 observations), the usual linear regression cannot be used. One way of addressing this problem is to employ machine learning methods enabling a larger number of regressors to be used. They help reduce data dimensionality, minimising information losses. Among them are models with regularisation and factor models.

We tested LASSO and Ridge regressions as regularised models. The models’ key parameters, lambda\(^{13}\) and alpha\(^{14}\), were chosen in such a way as to minimise the mean square errors of the equation. Lambda equalled 0.23, alpha – 0.1. As a result, out of the 50 topics initially given, the regularised model left 24, which is also a fairly large number. A large number of regressors may result in the problem of model overfitting, so the regularised models had to be rejected.

Another way of reducing dimensionality is factor analysis. Principal Component Analysis is one of the most frequently used factor analysis methods. It uses an algorithm to move from the initial data to new groups where data have similar relationships (Orlov and Lutsenko, 2016).

\(^{13}\) Lambda is a regularisation parameter introducing a complexity penalty: at \(\lambda = 0\), LASSO regression reduces to the conventional method of least squares; and as \(\lambda\) increases, the number of variables declines until it becomes zero.

\(^{14}\) Alpha determines which model type is the most suitable: at \(\alpha = 0\), Ridge regression is suitable, at \(\alpha = 1 – \text{LASSO regression}\).
Figure 6. Share of principal component dispersion

Source: author’s estimates

The first five principal components usually account for the main share of dispersion (Figure 6). The regression equation of the PMI index on the first five principal components shows that the first principal component is nonsignificant at the 10% significance level. The following four principal components are significant and explain 85% of the regression (Table 5).

Table 5. Regression components

<table>
<thead>
<tr>
<th>Variable</th>
<th>B-coefficient</th>
<th>t-statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Principal component</td>
<td>0.23</td>
<td>10.28 ***</td>
</tr>
<tr>
<td>(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal component</td>
<td>-0.11</td>
<td>-3.87 ***</td>
</tr>
<tr>
<td>(3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal component</td>
<td>-0.13</td>
<td>-3.60 **</td>
</tr>
<tr>
<td>(4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal component</td>
<td>-0.24</td>
<td>-5.00 ***</td>
</tr>
<tr>
<td>(5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-statistic</td>
<td>39.67 ***</td>
<td></td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>0.85</td>
<td></td>
</tr>
</tbody>
</table>

Note: ***, **, and * denote the significance of coefficient estimates at 0.1, 1, and 5% respectively.

The model constructed was tested on a test sample covering the period from February 2017 to August 2018. The comparison of predicted results with the actual PMI values demonstrates the fairly strong predictive power of the model, and its adequately chosen specification (Figure 7). To assess the quality of the model constructed, mean absolute error (MAE) was used. The MAE value for this forecast equaled 1.0 percentage point, while in using the first-order autoregression model AR(1), it was 2.7 percentage points.
4. Conclusion

This paper presents a model estimating economic performance based on news data. The calculations presented in the paper show that the use of unstructured information such as news is as important a component of economic activity forecasting as the use of conventional statistical indicators.

The methodology developed addresses the task of forecasting economic performance fairly accurately, as evinced by the model quality metrics obtained. This suggests that news data has a fairly strong predictive power.

Moreover, this model has potential for further refinement in a variety of directions: first, expanding the available news database through the use of both other news sources and social media; and second, different topic models need to be used so as to identify the best of them. The latent Dirichlet allocation method produces fairly good results but fails to capture some relationships. Therefore, the use of bigrams (two-word combinations) instead of unigrams could be considered along with other topic methods. This also applies to the identification of the text tone, the accuracy of which needs to be brought to 85% – 90%.

The news index obtained can be used not only to monitor economic performance on a daily basis but also to develop other indicators, allowing faster responses to the current economic situation and prompt decision-making.
5. References


Inflation Forecasting Using Machine Learning Methods

Ivan Baybuza, Ludwig Maximilian University of Munich*
ibaybuza@nes.ru

Inflation forecasting is an important practical problem. This paper proposes a solution to this problem for Russia using several basic machine learning methods: LASSO, Ridge, Elastic Net, Random Forest, and Boosting. Despite the fact that these methods already existed in the early 2000s, for a long time they remained almost unnoticed in the professional literature related to the forecasting of inflation in general, and Russian inflation in particular. This paper is one of the first attempts to apply machine learning methods to the forecasting of inflation in Russia. The present empirical study demonstrates that the Random Forest model and the Boosting model are at least as good at inflation forecasting as more traditional models, such as Random Walk and autoregression. The main result of this paper is the confirmation of the possibility of more accurate forecasting of inflation in Russia using machine learning methods.

Keywords: inflation forecast, machine learning, boosting, random forest

JEL Codes: C53, E37


1. Introduction

It is difficult to overestimate the importance of inflation forecasting for rationally thinking and acting economic agents: numerous economic obligations, including wages and interest rates, are usually expressed in nominal prices. In practice, central banks implement monetary policy guided mainly by their expectations of how inflation will behave in the short or medium term, rather than by its current values, since the rate of inflation does not react immediately to the tightening or easing of monetary policy, but rather with a certain lag. Therefore, price forecasting is important both for households and businesses, and for official authorities.

* The author is a graduate of the HSE/NES Joint Bachelor’s Program in Economics (BAE 2018) and a student on the Master’s Programme in Economics of Ludwig Maximilian University of Munich, Germany.
One of the key tasks of central banks is to maintain price stability. Over the past three decades, many central banks have adopted inflation targeting policies to solve this problem.

Inflation targeting policy is based primarily on public confidence in monetary authorities. Central banks introduced the practice of publishing their own forecasts for inflation and other key macroeconomic variables in order to establish and maintain this confidence. As a result, the problem of the quality of these forecasts became more acute. The Bank of Russia completed the transition to an inflation targeting regime in December 2014, and in these circumstances, the forecasting of inflation in Russia is now more important than ever.

The task of inflation forecasting can reasonably be divided into two parts: short- and long-term forecasting. It is not easy to explicitly divide up forecast horizons, but it is intuitively clear that forecasting inflation for one quarter and for five years are different tasks. While it may be sufficient to use the idea of the neutrality of money (Romer, 2012, p. 514–515) for predicting inflation in the long term, inflation forecasting in the short term is a more challenging problem.

According to current research in this area (for example, Stock and Watson, 2007 and 2008; also Faust and Wright, 2013), it is known that the behaviour of inflation in the short term can be closely approximated using fairly simple models based only on a time series of inflation. At the same time, predictions involving other macroeconomic indicators are worse than single-factor models.

Inflation forecasting using other macroeconomic variables as predictors has two serious limitations: on the one hand, a rather large number of potentially informative predictors, and on the other hand, the limited duration of the time series available, which leads to the so-called ‘curse of dimensionality’ (Stock and Watson, 2011). The dynamics of inflation are influenced, to varying degrees, by many different macroeconomic factors, and the maximum length of the time series that can be used is about 700 time observations (US, monthly data).

The above limitations can potentially lead to the so-called ‘overfitting problem’, the result of a small number of time observations relative to the number of explanatory variables. Here, overfitting means adjusting the model to random patterns in the ‘learning’ sample which are absent in the general population. In other words, the model achieves a very low prediction error in sample, but gives a very inaccurate forecast when predicting out of sample. This is why, in practice, multivariate models often give less accurate inflation forecasts than univariate ones.

The overfitting problem can be solved by pre-selection of explanatory variables on the basis of theoretical concepts, for instance, by selecting rates of real activity. However, this approach has a number of potential vulnerabilities. First, the predictive power of explanatory variables can change over time, as well as depending on the forecast horizon. Quite often, variables that predict well one or two months ahead can give a very inaccurate inflation forecast six months or
a year ahead, and vice versa. Second, overfitting of explanatory variables for a particular data sample remains possible. The selection of explanatory variables itself can be viewed as a kind of initial hyperparameter, on which the quality of the forecast primarily depends.

For these reasons, pre-selection of parameters should be carried out in a sub-sample of data, but not solely on the basis of theoretical concepts, in order to improve the prediction results.

Approaches to solving the overfitting problem have long been a preoccupation of the field of computer science known as machine learning (ML). Over the past few decades, many different ML models have been created: LASSO regression (Least Absolute Shrinkage and Selection Operator), Ridge regression, Principal Components Analysis, Decision Trees, Random Forest (RF), Boosting, Neural Networks, etc. Their application has led to significant breakthroughs in areas such as text categorisation (Sebastiani, 2002) and image recognition (Simonyan and Zisserman, 2014). However, many powerful ML methods such as the Random Forest model have started to be used to predict macroeconomic variables only relatively recently. In particular, works have begun to appear in which the authors use ML methods to predict inflation (Chakraborty and Joseph, 2017). The author of this paper has not come across any published works on this topic regarding the Russian economy.

In the light of the above, the main purpose of this paper is to test ML methods for forecasting inflation in Russia. For this task, the following popular ML methods were selected: LASSO and Ridge regressions, Elastic Net model, Random Forest model, and Boosting. Forecasts obtained using ML methods are compared with results obtained using traditional econometric methods: Random Walk, Autoregressive model of order 1 (AR(1)), and Autoregressive model of order p (AR(p)).

This paper has the following structure: Section 2 contains a review of literature on the subject; Section 3 describes data and methods used in the study; Section 4 presents the models used in this paper; Section 5 shows the results of the study; and Section 6 contains the conclusion to the study.

2. Literature review

This review presents the works that influenced the choice of methods used in this paper. Works by Stock and Watson (2008) and Faust and Wright (2013) are frequently cited in the field of inflation forecasting. The basic models used in this study (random walk, autoregressions of orders 1 and p) were chosen based on the results obtained in these works. In the next two works by Chakraborty and Joseph (2017) and Garcia et al. (2017), ML methods are used to predict inflation. Finally, the work by Andreev (2016) is important for understanding the actual methods used by the Bank of Russia to forecast Russian inflation.

A more detailed review of each of the works is provided below.
Stock and Watson (2008) compare many different models, dividing them into four main groups. The first group includes models based only on the use of a time series of inflation: Autoregression of Moving Average (ARMA), RW, as well as the authors’ own model with an Unobserved Components and a Stochastic Volatility (UC-SV). In the second group, the authors include models in which the explanatory variables are indicators of economic activity, first of all, and then the unemployment rate and the output gap. The third group includes models in which predictions are based on expected inflation or forecasts such as surveys of professional forecasters (SPFs). The fourth group includes models in which the explanatory variables are variables other than indicators of economic activity, i.e. variables which are not used in the second group models. The authors model future inflation for four quarters, obtaining forecasts in pseudo-real time in a rolling window of 10 years. The quality criterion of the model is the Root Mean Squared Error (RMSE) relative to the error of the UC-SV model as a reference. The main conclusion of the work is that, as far as the quality of the forecast is concerned, models based on indicators of economic activity do not systematically improve on univariate models that take into account only the dynamics of inflation itself.

The work by Faust and Wright (2013) is a review and comparison of the best models for inflation forecasting at that time. The authors compared 17 models, including the AR model, the UC-SV model, the RW model in two variations (RW and its modification, RW-AO2), the Phillips Curve-based model, Structural Vector Autoregression (SVAR), Bayesian Model Averaging (BMA), the Dynamic Stochastic General Equilibrium model (DSGE), and more. The authors nowcast inflation in the current quarter (h=0) and forecast it up to eight quarters ahead, obtaining forecasts in pseudo-real time in an expanding window. The quality criterion for the model is the Root Mean Squared Error (RMSE) of the model relative to the error of the AR(1) model, which is selected as a reference. The authors conclude that the models based on other predictions (primarily surveys of professional forecasters) have the best predictive capacities. Moreover, the authors conclude that very simple methods, such as random walk, predict inflation surprisingly well. In general, the findings of their review correlate well with those of the previous work: both studies support the hypothesis that the best multi-factor models, both with activity rates and with other variables as predictors, systematically fail to surpass the best univariate models that use only a time series of inflation.

The work by Chakraborty and Joseph (2017) is a review of the best ML methods in terms of their practical application for solving several important problems faced by central banks. For our purpose, the forecasting of inflation in the UK, as carried out by the authors of the article, is of particular interest. The authors forecast inflation for the medium-term horizon of two years, using data

---

2 A variant of the random walk model was proposed by Atkeson and Ohanian (2001).
for various macroeconomic variables (money supply, unemployment rate, bank rates, GDP, and other indicators) between the first quarter of 1988 and the fourth quarter of 2015. The models are trained on an initial 15-year window (the first quarter of 1990 to the fourth quarter of 2004) with further quarterly expansion to the fourth quarter of 2013; the quality measure is Mean Absolute Error (MAE) for the entire forecasting period, with the period before and after the crisis of 2008 taken separately. The authors apply the following models: Nearest Neighbours, Decision Tree, RF, Neural Networks, Support Vectors, Support Vectors combined with Neural Networks, Ridge regression, VAR(1), AR(p), and AR(1). AR(1) is used as the benchmark. As a result, the authors come to the conclusion that the model of Support Vectors combined with Neural Networks gives the highest-quality forecasts for the entire forecasting horizon, while the Random Forest model performs best in the post-crisis period. However, all other ML methods also show better results than traditional benchmarks in the form of VAR(1), AR(p), and AR(1).

In Garcia et al. (2017), inflation in Brazil is predicted by ML methods in 12 different periods. The first forecast period is five days before the release of inflation statistics, and the twelfth forecast period is 11 months and five days before the release of inflation statistics. In their work, the authors use various macroeconomic variables (a total of 59 series), reflecting the state of the financial market, the labour market, the balance of payments, and the public debt of the country. They also use professional forecasts of Brazilian inflation collected during the FOCUS survey conducted by the Central Bank of Brazil. In the work, a time series from January 2003 to December 2015 is used. Models are trained on a nine-year rolling window; the RMSE and the MAE serve as criteria for the quality of the forecast. The authors apply the following models: RW, AR(p), where the number of lags is chosen based on the Bayesian information criterion, the dynamic factor model using principal component analysis, three varieties of LASSO regression, RF, Complete Subset Regression (CSR), and two models based only on surveys of professional forecasters. As a result, the authors come to the conclusion that inflation for five days and for one month and five days is best predicted by the LASSO model and the model that uses forecasts obtained from the surveys. However, the authors note that the LASSO model actually cuts off all variables except survey forecasts; if survey forecasts are excluded from the model, the LASSO result is significantly worse. From the third forecast horizon (two months and five days) to the last (11 months and five days), the best results are given by the CSR model. In addition, the RF model and the dynamic factor model perform better than the AR and RW models over all forecasting horizons.

Andreev (2016) describes a combined forecast method for forecasting inflation in Russia. The main goal of the algorithm presented in the work is to combine different models without preliminary screening of potential predictors. The following models are included in the algorithm: Vector Autoregression,
Bayesian Vector Autoregression (BVAR), Ordinary Least Squares regression (OLS regression), RW, Linear Threshold Autoregression (LTAR), and the Unobserved Components (UC) model. Multivariate methods, such as VAR, BVAR, and OLS, are trained not on all variables, but only on some combination of them. In other words, $n$ subsets of the initial data array are created, with only some of the variables included. After training, they are aggregated on each sub-sample of the base model. Thus, combined forecasts using the VAR model, the BVAR model, and the OLS model are constructed. Finally, multivariate and univariate models are aggregated in a combined forecast. This algorithm makes use of the different strengths of particular models, using all variables.

In our study, we used the RF and boosting methods, which belong to the class of so-called ‘ensemble’ methods. The methods of this class are based on the general idea of using multiple training algorithms and subsequently combining them so that the final forecast is more accurate than any individual forecast. We expect that results of our study may be of practical value to the Bank of Russia, by expanding the range of potential methods for inflation forecasting.

3. Data description and methodology

The main sample consists of 92 macroeconomic series, excluding the time series of price level (93 series if this is included). The data reflect the state of business activity, industrial production, the financial market, the employment rate, the balance of payments, and the prices of the main export goods of the Russian economy. The consumer price index (CPI) was chosen as a measure of inflation. Explanatory variables are taken for the period from February 2002 to June 2016 (173 observations). CPIs are used for the period from February 2002 to January 2018 (192 observations). The data represent either the monthly rate of change of the relevant variables or their monthly values in levels. All series are adjusted for seasonal and calendar factors, rendered stationary and standardised. Table 1 of the Appendix gives all the variables and the means by which they were transformed.3

Researchers working with this data may come up against the so-called ‘ragged edge problem’ associated with uneven disclosure of data on various indicators by statistical agencies. Inflation data are published by the Federal State Statistics Service (Rosstat) at the beginning of the month following the reporting month, while other key macroeconomic variables are published at the end of the month. This problem may prevent real-time forecasting. However, when forecasting in pseudo-real time, when all data are known, the problem is not so pressing. Therefore, no adjustments are made for the ragged edge problem in this work.

3 The author of this paper expresses his deep gratitude to Professor Konstantin Styrin of the NES for providing the data.
After adjusting for seasonal and calendar factors, the first difference in the logarithms of price levels was taken to obtain the monthly inflation:

$$\pi_t = \log(CPI_t) - \log(CPI_{t-1}).$$

Figure 2 (see Appendix) shows trends in monthly inflation over time; in principle, it shows that the first difference of logarithms may be sufficient to render the series stationary. We conducted the extended Dickey-Fuller test, which confirmed the hypothesis that the price level in Russia is an integrated process of order 1.

The quality of forecasts is estimated using the root mean squared error (RMSE) of the forecast:

$$RMSE = \sqrt{\frac{1}{T - T_0 - 1} \sum_{t=T_0}^{T} (\pi_t - \hat{\pi}_t)^2}.$$

The forecasting is performed in pseudo-real time out of sample, i.e., for dates outside the limits of the estimation sample, with a rolling 10-year window. Predictions are obtained out of sample, because when training with a large number of predictors, the high accuracy of forecasting in the training sample likely indicates not the high quality of the model, but its overfitting. Inflation is predicted 1, 2, 3, ..., 24 months ahead, following the last available monthly inflation value in pseudo-real time. In addition, average inflation values are calculated on the horizon of one month, two months, one quarter, half a year, one year, a year and a half, and two years. The models are compared in two ways: according to individual monthly inflation forecasts and according to average values on the above mentioned time horizons.

All RMSE values are considered in relation to the corresponding value of the AR(1) benchmark model: if the RMSE value of any model on a certain horizon is less (more) than one, then this model predicts inflation in such a month or over such a horizon better (worse) than the benchmark model.

It is important to note that no multivariate models, except for the AR(1) and LASSO combination, employ inflation lags in training or predict inflation using them. This is done deliberately to determine the strength of ML models in comparison with single-factor standards.

4. Models

This section describes the models used in the work. To begin, two traditional econometric models based only on past trends, the random walk model and the AR model, are briefly described. The following is a description of the ML models (LASSO, Ridge, Elastic Net, Random Forest, and Boosting) used in this paper, as well as their various specifications.
4.1. Random walk (RW)

This is the simplest model considered in this paper, but it has a good predictive capacity, which in practice is not so easy to improve on by including additional predictors. Mathematically, the model has the following form:

$$\pi_{t+h} = \pi_t + \varepsilon_{t+h},$$

where $h = 1, 2, \ldots, 24$, $\varepsilon_i$ is an unexpected fluctuation in inflation.

4.2. Autoregression (AR)

This paper uses a recursive version of the autoregression model. This means that, in order to predict for $h$ periods in advance, the missing inflation values between the time $t$ and the time $t + h$ are consistently predicted. According to Faust and Wright (2013), the iteration method produces more accurate forecasts than the direct method (when the inflation forecast is immediately constructed at the time $t + h$). The number of lags $p$ in the AR($p$) model is selected using the Bayesian Information Criterion (BIC). Mathematically, the autoregressive model has the following form:

$$\pi_t = \alpha_0 + \sum_{j=1}^{p} \alpha_j \pi_{t-j} + \varepsilon_t,$$

where $\varepsilon_i$ is an unexpected fluctuation in inflation.

4.3. Models with regularisation

A typical feature of the so-called ‘overfit’ model (i.e. where the model is adjusted for random patterns in the ‘training’ sample which are absent in the general population) is that attributes (explanatory variables in terms of the regression analysis, or predictors) can be assigned large coefficients in the resulting solution. If the number of attributes is larger than the number of time observations, or if there are correlated attributes, the problem of minimisation of the root mean squared error can have an infinite number of solutions. In such a situation, attempting to achieve a perfect approximation of noisy data can lead to a dramatic change in the coefficient values and to their abnormal increase. That is why it is reasonable to control the size of the coefficient vector, taking it into account when constructing the quality function. For this purpose, an additional term, a regulariser or penalty term based on the norm of the coefficient vector, is added to the basic function in the minimisation problem.

Mathematically, the regression with regularisation has the following form:

$$Q_R(x) = Q(x) + \alpha R(x),$$

where $Q(x)$ is a certain quality function (in this paper, RMSE), $R(x)$ is the penalty for the norm of the regression coefficient vector, and $\alpha$ is the hyperparameter responsible for the ratio of accuracy to the size of the parameter vector.
The hyperparameter $\alpha$ is determined out-of-sample (i.e. outside the sample used for the estimation of model parameters) through cross-validation: an optimal value of the hyperparameter $\alpha$ is a value which allows us to obtain the most accurate forecast out-of-sample, i.e. which minimises $Q(x)$ (see for example Murphy, 2012, p. 206).

There are many different regularisers, but the most common one looks like this:

$$R(x) = \gamma \sum_{i=1}^{q} |x_i| + (1 - \gamma) \sum_{i=1}^{q} x_i^2,$$

where $\gamma \in [0; 1]$, $q$ is the dimensionality of the parameter vector $x$.

A regulariser with $\gamma$ equal to 1 is called a $L_1$-regulariser, and the OLS model together with this regulariser is called a LASSO model (or LASSO regression). A regulariser with $\gamma$ equal to 0 is called a $L_2$-regulariser, and the OLS model together with this regulariser is called a ridge regression model. When $\gamma \in (0; 1)$, a mixed regulariser is obtained, and the OLS model including it is called an elastic net model.

Each of the regularisation methods has its advantages and disadvantages. An important advantage of $L_2$-regularisation is the presence of a clear analytical solution when used together with a quality function that is differentiable in a closed form. For the root-mean-square quality function, the analytical solution looks like this:

$$x = (X^TX + \alpha I)^{-1}X^T y,$$

where $X$ is a matrix, each column of which is a time series of observations of an attribute (predictor), $I$ is a unity matrix of size $q \times q$, and $y$ is a vector of responses (values of the indicator, which the model seeks to explain using a combination of attributes).

Moreover, as can be seen from the solution, the addition of $L_2$-regularisation ensures a positive determinant of the matrix, making it invertible. Thanks to this, the problem will always have a single solution, which is very convenient in practice.

The $L_1$-regulariser, in contrast to $L_2$, does not provide a unique solution, but has another important property. Thanks to its mathematical conditionality, the use of this regulariser leads to the zeroing of some coefficients in the final solution. When minimising the convex RMSE function, the unconditional minimisation problem $Q(x) + \alpha ||x||_1$ can be reduced to a conditional minimisation problem in the form:

$$\begin{cases}
\hat{x} = \arg\min_x Q(x) \\
||\hat{x}||_1 < S
\end{cases}, \text{ for some } S$$

The solution of this system is the intersection of the admissible set $||\hat{x}||_1 < S$ with the line of the level nearest to the absolute minimum.
In a two-dimensional case, it will look like this:

![Figure 1. Regularisation](image)

*Note: $L_1$-regularisation (left), $L_2$-regularisation (right)*

Quite often, the solution of such a system will be the intersection of the admissible set with the level line at the vertex of the admissible set, which will mean the zeroing of one of the components (see the graph on the left of Fig. 1). Thus, the $L_1$-regulariser screens certain attributes, which allows the most important attributes of the object to be focused on. Because of this property, the LASSO model is well suited for pre-screening of variables before using other models. This paper provides an example of such a combination: pre-selection of parameters using the LASSO for later use in the OLS regression together with the first inflation lag. In other words, attributes pre-selected by the LASSO are incorporated into the AR(1) model.

Both regularisers have their advantages and disadvantages, so in practice a mixed regulariser known as elastic net regression is often used. In this model, the parameter $\gamma$ is the hyperparameter, the value of which needs to be determined. Often it is selected using cross-validation, but this optimisation problem already has a hyperparameter ($\alpha$), which is found using cross-validation. In this case, it is necessary either to allocate a part of the sample for the cross-validation, which is inadvisable due to the limited sample size of available time series, or to use the same learning sample, which can also lead to overfitting of the model. In this paper, the parameter is selected ad hoc and is defined as 0.5. This value is chosen because it is equidistant from the extreme cases ($L_1$- and $L_2$-regularisers). Moreover, this value for $\gamma$ is also used in the work by Chakraborty and Joseph (2017).

### 4.4. Random forest (RF)

The Random Forest model is based on bootstrap union of so-called decision trees and was first proposed by Breiman (2001). A binary decision tree is used as the basic algorithm for the Random Forest. A binary tree is a graph consisting of
‘parent’ or ‘root’ nodes (interim nodes) and ‘leaf’ terminal nodes. A decision tree is constructed in stages. The first stage is the optimal division of the entire sample $X$ into two sub-samples: $X_1(i,p) = \{x|x_i \leq p\}$ and $X_2(i,p) = \{x|x_i > p\}$ according to the specified quality function $Q(X, i, p)$. Next, each of the sub-samples is iteratively broken down using the same principle. The breakdown stops when a stopping criterion is fulfilled.

After that, $n$ leaf nodes are created, each of which corresponds to a certain preserved sub-sample (which could contain only a single point). If the regression analysis problem is solved (as in the case of modelling and forecasting of inflation), each leaf node is assigned the average value of the explanatory variable across the points (observations) in the corresponding sub-sample. The resulting tree is a connected graph of root nodes, each of which contains a threshold predicate that breaks down the sub-sample into two parts, and leaf nodes, each of which contains the predicted values of the explanatory variable. Usually, the quality function is specified in the following form:

$$Q(X, i, p) = H(X) - \frac{|X_1|}{|X|} H(X_1) - \frac{|X_2|}{|X|} H(X_2),$$

where $H(X)$ is an informativeness criterion.

The informativeness criterion shows how homogeneous the objects (observations) in the sub-sample are in terms of the explanatory variable. The main idea here is to maximise this homogeneity, and to break the sample down into two parts, in each of which the spread of values of the explanatory variable is minimal. Therefore, for the regression analysis problem, the quadratic deviation is used as a loss function and the following informativeness criterion is minimised:

$$H(X) = \min_c \frac{1}{|X|} \sum (y_i - c)^2.$$

As we know, the minimum value of this kind of function is achieved when $c$ is equal to the average value of the target (explained) variable.

$$H(X) = \frac{1}{|X|} \sum (y_i - \frac{1}{|X|} \sum y_j)^2.$$

In other words, the main goal of the algorithm at each stage is to minimise the sum of the weighted average variance within each of the two sub-samples resulting from the breakdown. Using the constructed tree, we can get predictions for values of the target variable with the new values of the explanatory variables.

The main advantage of the decision tree model is that the trees allow us to simply create effective (in terms of minimising the variance of the target variable) nonlinear dependencies. However, there is a serious drawback: the overfitting problem. For any given sample, it is possible to create a tree of such depth that
it will make no error at all. The RF model is designed to compensate for this drawback and reduce the variance of the base model. To do this, on the basis of the actual sample $X$, $N$ artificial sub-samples of the original sample length are generated using the bootstrap. Also, an artificial sub-sample does not include all the attributes, only a random set. Randomisation therefore occurs in two directions.

Next, for each resulting artificial sample $\tilde{X}_n$, a decision tree $t_n(x)$ is built. The tree is built in such a way that in each leaf node there are at least $l$ observations (in this paper the value $l$ is equal to five). The final output of algorithm is the average across all constructed individual decision trees:

$$T_N(x) = \frac{1}{N} \sum_{i=1}^{N} t_i(x).$$

The number of decision trees in a RF is an important hyperparameter: the more trees there are in the forest, the more reliable the result is. At the same time, the more trees there are in the forest, the longer the operation time of the algorithm. Usually, the number of trees is chosen so that the resulting output of the RF stops changing. In this paper, 200 decision trees in each RF sufficed.

When working with time series, the drawback to using a regular bootstrap is loss of information extracted from data. This problem can be solved in two ways: by abandoning the bootstrap entirely and randomising samples only by attributes, or by using the block bootstrap (see for example Efron and Tibshirani, 1994), in which not individual points (observations) but entire blocks of a certain length (in this paper, blocks of length 10 are used) are randomly selected. Therefore, in this paper we consider two specifications of the model: without any bootstrap and with the block bootstrap.

As described in Section 3, in the current study of inflation forecasting using ML methods, data were made stationary. In traditional econometric models, this is done to avoid the so-called ‘spurious’ regression problem. However, the results of many ML algorithms, including the RF model and Boosting, are not subject to this problem. This paper therefore presents the results of the RF model using both data in a stationary form and data which have not been transformed. It is important to note that, in all specifications, the RF models were trained and used to obtain inflation forecasts (stationary series), but not forecasts of accumulated CPI values (non-stationary series). Accordingly, all RMSE calculations and their comparisons are made in the same way for all models.

4.5. Boosting

The gradient Boosting model was first proposed by Friedman (2000). The idea behind the gradient boosting algorithm is similar to the idea behind the RF model: both algorithms are ensemble methods. The base Boosting model can
represent any collection of models, but often, as with the RF model, a decision tree is selected. The main difference between the Boosting and the RF is that the base models are not trained independently, but rather taking into account the results of operation of the model on the previous iterations. The algorithm’s operation can be described as follows:

1) The first base model is trained on the whole sample:

\[ b_1(x) = arg\min_b \sum_{i=1}^l (b(x_i) - y_i)^2. \]

2) After the first step, the ensemble Boosting algorithm results in the first trained base model:

\[ B_1(x) = b_1(x). \]

3) Next, residuals are calculated that are equal to the difference between a true value and a predicted value based on the first Boosting model:

\[ e_1^1 = y_i - B_1(x_i). \]

4) The following model is trained on these residuals:

\[ b_2(x) = arg\min_b \sum_{i=1}^l (b(x_i) - e_1^1)^2. \]

5) We add a new model to the algorithm obtained in the previous step with a certain coefficient \( \gamma \in (0; 1] \). This technique is called ‘step reduction’. In this paper, the coefficient \( \gamma \) is equal to 0.2. This technique helps to improve the model’s operation and avoid overfitting. A new model is obtained:

\[ B_2(x) = B_1(x) + \gamma b_2(x). \]

6) Then, the algorithm is built up iteratively until the end. As a result of operation of the algorithm, the final model is:

\[ B_N(x) = \sum_{i=1}^N \gamma^{i-1} b_i(x). \]

The algorithm terminates once all training cycles are completed. The number of training cycles is an important hyperparameter. Usually, in the training sample, errors tend to vanish as the number of iterations is increased. However, out of sample, too many iterations can actually increase errors as that the model begins to adjust for noise. In this paper, the number of iterations is 100.

As already mentioned above, both untransformed (non-stationary) and transformed (stationary) data are used for the RF and Boosting models.
5. Results

5.1. Overall results

The main results of this study are presented in Tables 2–5 of the Appendix. The Tables show the relative RMSE values of all specifications of each model for all forecast horizons (Table 2 – from the 1st to the 6th month, Table 3 – from the 7th to the 12th month, Table 4 – from the 13th to 18th month, Table 5 – from the 19th to the 24th month). Also, Table 6 of the Appendix gives the relative RMSE values for the average inflation over the horizon of 1, 2, 3, 6, 12, 18, and 24 months. We can draw the following main conclusions from the results:

1) The use of ML methods can improve the quality of forecasting of Russian inflation compared to reference models (benchmarks) that use only lags of inflation as predictors. However, significant disadvantage of these models in comparison with classical econometric models is the loss of interpretability in the classical sense.

2) The ensemble methods (RF and Boosting) predict average inflation better than the base model from the second month onwards.

3) Among all three specifications of the RF model, the specification with untransformed data gave the best result when forecasting both inflation in individual months and average inflation over entire forecast horizon. Comparing the results of the two specifications of the Boosting model, with stationary and non-stationary data, leads to the same conclusion.

4) Relatively speaking, the regularised models provide less accurate forecasts over all forecasting horizons. The AR(1) model combined with LASSO gave results worse than the base AR(1) model, except for forecasts one month ahead.

5) The AR(1) model combined with LASSO gave the highest quality results when forecasting inflation over the horizon of one month. The models that use only lags of inflation as predictors (RW and AR) gave the same quality of forecasts over this horizon. The remaining methods were less accurate with respect to forecasts than the base model.

A more detailed analysis of the results is provided below.

5.2. Results of using RF and Boosting models

As mentioned above, the ensemble models that use untransformed data gave better results than similar models that use transformed data. This is not surprising, since the application of the RF and Boosting models does not require prior transformation of data to a stationary form. In addition, it is worth noting that the specification of the RF model that uses the block bootstrap gave slightly worse results than the model without the block bootstrap when forecasting both inflation for specific months and average inflation over entire forecast horizon. This result probably indicates that data within time series have a high degree of serial correlation and that their random mixing with replacements negatively
affects the result of the model’s operation, even when a special block bootstrap is applied.

The RF model, which uses untransformed data for forecasting inflation for individual months, performed better than the base model. With the exception of the inflation forecast one month ahead, the RF model generally predicts inflation for individual months more accurately. At the same time, the RF model predicts average inflation much better: when forecasting average inflation two years ahead, the error is 60% smaller than the error obtained using the base AR (1) model.

To understand this significant difference, let us compare monthly and accumulated inflation values over the same horizon, as predicted by both models. Figure 3 of the Appendix shows the dynamics of monthly and accumulated inflation (actual and forecast) over the forecast horizon of 12 months (the justification for taking this horizon being that RMSEs of monthly inflation forecasts 12 months ahead are almost identical: the value of the relevant RMSE is equal to 0.9827). The chart clearly shows a sharp short-term surge of inflation, associated, for the most part, with the fall of the rouble at the end of 2014. The AR(1) model represents this surge with some lag when forecasting each subsequent monthly figure for inflation over the entire horizon, as a result of which, around a year later, high average accumulated inflation appears, which is very different from the actual case. Here, the RF model, which is not trained on the inflation lag, shows higher resistance to such shocks. It is this very shock resistance which explains the differences between the model’s predictions for average inflation.

The Boosting model using untransformed data is similar to the corresponding RF model. Figure 4 of the Appendix shows changes in inflation over 13 months plus the average cumulative inflation (since the relative RMSE value at 13 months is also approximately equal to 1) for the Boosting model and the AR(1) model. The graph shows that the Boosting model forecast is also more resistant to shocks than the AR(1) model forecast. The explanation given above concerning the RF model also goes, more or less, for the difference. It is worth noting that the Boosting model reacts to shock more sensitively than the RF model. This is related to the fact that the Boosting model learns from past forecast errors, which makes it more adaptable to a specific training sample.

The RF and Boosting model algorithms have several common features: both algorithms are ensemble methods and are based on the decision tree model. Therefore, it is logical that the results of their application should be similar. This paper draws the following conclusion: we observe a similarity in forecasting using separate features of the RF model and the Boosting model. Such a correspondence may indicate the reliability of the results obtained.

### 5.3. Results for regularisation models

All regularisation models showed lower prediction accuracy than the base model. This was primarily because these models are, to some extent, unstable.
Figures 5–10 of the Appendix show a change in the number of explanatory variables left by the LASSO model when forecasting inflation 1, 2, 6, 12, 18 and 24 months ahead. The number of explanatory variables picked at each forecast horizon is quite volatile. We see that bursts occur at certain moments when the model selects an abnormally large number of explanatory variables. On the other hand, from the forecast horizon of one month onward, for some dates, all variables are cut off. At the same time, the average number of predictors remaining in the model plummets as the forecast horizon changes from one to two months. This instability may indicate that, among the initial set of variables, there are no explanatory variables that either individually or as part of a small group of variables could effectively predict the dynamics of inflation.

We consider that the combined LASSO and AR(1) model confirms this hypothesis. The LASSO enables preliminary selection of variables with subsequent addition of the first inflation lag. With regard to regressions, the coefficient vector is of particular interest. For each attribute, the average absolute value of the coefficient vector was calculated (for many variables, this value was 0, since the characteristic could have been eliminated by the LASSO before) when forecasting inflation for each forecast horizon. In other words, for example, a coefficient matrix was used for developing 53 one-month-ahead forecast options. Some of the values in this matrix are equal to 0 because the LASSO model had already eliminated some of the variables. This is followed by a review of the absolute values of non-zero coefficients. Then, the average absolute value of the coefficient is calculated for each explanatory variable. As a result of data standardization, the coefficient values show the relative strength with which the predictors account for future inflation. This operation is performed for each forecast horizon separately.

Table 7 of the Appendix presents, in descending order, the five largest average absolute values of the coefficient for each forecast horizon of 1, 2, 3, 6, 12, 18, and 24 months. The table does not present the results for all months, since these seven periods are sufficient to identify the general pattern. We see that for all forecasting months, on average, it is the first lag of inflation that remains the main explanatory factor; the coefficients of all the other variables turn out to be significantly less. At all forecasting horizons from the sixth month onward, the values for the indicators obtained are significantly less than the coefficient of the first inflation lag. It should be noted that expansion of the forecast horizon gradually identifies indicators that do in fact account for inflation. Thus, when inflation is forecast for the 24th month ahead, two explanatory variables survive: the first inflation lag and loans to individuals over one year, while all other variables are almost completely zeroed out.

The quality of the Ridge model is similar to that of the LASSO model, even slightly worse. This makes sense: overfitting may be a problem for the LASSO model, but it is even more of an issue for the Ridge model. It is worth recalling here that, due to the peculiarities of the regulariser, the LASSO model has a strong
ability to cut off some of the variables in the regression, something the Ridge model is not capable of. Therefore, the forecast quality of the Ridge model is even lower than that of the LASSO model. The results of the elastic net model are almost identical to the results of the LASSO.

6. Conclusion

This paper aimed to prove the viability of ML methods for forecasting Russian inflation, compared to traditional methods. As the results demonstrate, this conjecture has been confirmed. Not all methods performed equally well in solving this problem: the regularisation models showed lower forecasting quality compared to the base model.

Both ensemble methods (RF and Boosting) showed results comparable to the basic AR(1) model in predicting monthly inflation. At the same time, they showed significantly better results when forecasting average inflation over a horizon of more than two months. We can therefore conclude that the RF and Boosting models show promise when applied to the task of forecasting Russian inflation.

This paper also addressed the issue of data transformation. According to our results, the RF and Boosting models perform better with untransformed rather than transformed data. This conclusion could be of use in further research using ML methods, since data transformation is a standard preparatory element of almost any empirical macroeconomic research using time series.

In addition to the models used in this article, there exist a number of other nonlinear ML algorithms, such as neural networks. Researchers would be well advised to also test these algorithms in future work on forecasting Russian inflation.

Appendix is available at www.cbr.ru/eng/money-and-finance; dx.doi.org/10.31477/rjmf.201804.42

7. References


Review of the Bank of Russia – IMF Workshop 'Recent Developments in Macroprudential Stress Testing'

Elizaveta Danilova, Bank of Russia
danilovaeo@cbr.ru
Evgeny Rumyantsev, Bank of Russia
rumyancevel@cbr.ru
Ivan Shevchuk, Bank of Russia
shevchukiv@cbr.ru

In September, the Bank of Russia held a joint workshop with the International Monetary Fund in Moscow on macroprudential stress testing. IMF experts, members of the research community, staff members of central banks, and regulators from 16 countries shared their approaches to and methodologies of macroprudential stress testing and systemic risk analysis. This publication provides a brief review of the workshop and the key findings of the studies presented.

Keywords: macroprudential stress testing, systemic risk analysis, Bank of Russia, IMF
JEL Codes: C18, E37, E58, G17, G20

doi: 10.31477/rjmf.201804.60

1. Introduction

In the wake of the 2007 – 2009 global financial crisis, many central banks, in addition to their traditional objective of securing price stability, were given an official mandate to maintain financial stability. Unlike inflation targeting, which has been successfully implemented for several decades and the mechanism of which has been extensively explored in research studies, central banks’ policies regarding the maintenance of financial stability is an area where a complex of universal approaches has yet to be developed and where very extensive studies are underway. Central banks’ experience suggests that the goal of securing financial stability requires a system for identifying, monitoring and assessing

---

1 Workshop program and presentations (in English) are available at http://www.cbr.ru/analytics/fin_stab/
systemic risks to be developed along with an array of macroprudential policy instruments.

One instrument that has started to be widely used after the global crisis for analysing risks – of individual banks and financial organisations as well as those of the financial system in general – is stress testing (Danilova and Markov, 2017). Stress testing helps assess financial organisations’ stability, estimate the amount of capital required to cover losses should a crisis scenario materialise, compute the size of recapitalisation needed under such circumstances and conduct quantitative estimation of individual risk types. Macroprudential Stress Testing (MST) is distinguished by its ability to factor in the effects of financial organizations’ interconnection as well as the financial sector’s effects on the real economy. MST is an area that is being extensively developed in leading central banks, and a single, generally accepted methodology has yet to be developed.

The Bank of Russia has been extensively developing an MST concept of its own since 2017. The concept was published for discussion and consultation in October 2017 (Bank of Russia, 2017).

The joint Bank of Russia – IMF workshop consisted of eight sessions, each covered in a separate section of this review: 1) development of MST methodology and assessment of its effectiveness; 2) macroeconomic scenarios for MST; 3) risk assessment using Big Data; 4) using stress testing to analyse nonfinancial companies’ risks; 5) assessment of nonbanking financial organisations’ risks as part of MST; 6) assessment of feedback effects between the financial sector and the real economy; 7) development of interconnection analysis in the financial sector; 8) policy issues: optimal disclosure of stress-test information and assessment of macroprudential policy effectiveness.

2. Development of MST methodology and assessment of its effectiveness

Four MST concepts were presented in this session: those of the European Central Bank (presenter – Christoffer Kok), the Bank of Japan (Wataru Hirata), the Bank of Russia (Elizaveta Danilova), and the National Bank of Poland (Oskar Krzesicki and Marcin Borsuk).

The European Central Bank (ECB) was one of the first regulators to start developing MST, for the benefit of, among others, the European Banking Administration (EBA) and the Single Supervisory Mechanism (SSM). The ECB’s

---

2 For example, the first stress testing exercise conducted by the Federal Reserve in 2009 became the central element of anti-crisis policy. This took place amid the market’s severe lack of confidence in banks, as their investments in structured products, derivatives, and other risk assets were unknown. Instead of preventive infusion of government funds into the capital of all major banks, stress testing was conducted in order to recapitalise only fragile banks as necessary. Stress testing showed the actual situation at the major banks: a capital shortfall was identified, but its size proved to be smaller than the market expected; confidence was restored; and almost all banks were able to get hold of the required capital in the market without resorting to government support.
A stress-test concept was published in February 2017 (Dees et al., 2017). The ECB only uses stress testing for banks. The ECB’s methodology presented at the seminar by Christofer Kok, includes the computation of credit and market risks, the estimation of net interest income and fee and commission income, the assessment of operating risk, and the forecast of banks’ other income and losses. The aggregated stress testing results and data on the distribution of losses across banks are published on a regular basis in the ECB’s financial stability review. MST results are compared with European banks’ estimates calculated as part of supervisory stress testing by the bottom-up method.

The key macroprudential element of the ECB stress test is assessing feedback effects between the real economy and the financial sector. This methodology factors in banks’ response to negative shocks, when a decline in capital in the stress environment brings about lending contraction, thereby further exacerbating recession, hitting companies’ creditworthiness and entailing more problems for banks (see Section 7 for details). In addition, the ECB’s MST takes into account the impact of macroprudential measures – in particular, the ratio of the countercyclical buffer size to banks’ capital. Banks’ behaviour as shocks are realised depends on their balance sheet structure and risk-return ratios. To assess banks’ responses to shocks and their impact on lending, the ECB uses a variety of model-based approaches: regression analysis, empirical models for adjusting bank balance sheets, vector autoregression (VAR) models for individual banks’ balance sheets, models of portfolio structure optimisation, modelling a relationship between the capital level and the cost of funds and agent-based models (ABM). Christofer Kok pointed out the need to model not only banks’ but also other economic agents’ (nonfinancial companies, households) responses. To assess various banking risks and to forecast profits, losses, and balance sheets, a complex of satellite models is employed. Christofer Kok believes that the effectiveness of the stress testing methodology needs to be verified via backtesting and via the analysis of case studies and the sensitivity of a model’s key assumptions.

As Wataru Hirata showed in his presentation, the assessment of relationships between the financial system and the real economy also plays a central role in the Bank of Japan’s stress testing concept. MST aims to assess not only the stability of the financial sector but also its efficiency in performing its financial intermediation function. Like the ECB, the Bank of Japan confines its stress testing to banks; but unlike the ECB, which only analyses the Eurozone’s major banks, the Bank of Japan also provides estimates for smaller, including regional, banks (a total of 370 banks). The inclusion of smaller banks is essential, since, based on Japan’s experience, it is above all the regional banks that suffer profitability problems and the most severe lending contraction in stress events.

The Bank of Japan’s approach to MST provides for three components: an econometric macrofinancial model, a model for estimating long-run profitability (mainly for regional banks), and a joint modeling project with major banks.
Aggregated results of the stress test are published twice a year in the Bank of Japan’s Financial System Report. Stress testing uses one unlikely crisis scenario (comparable to the 2007 – 2009 crisis in terms of severity) and complementary scenarios reflecting current market risks (in 2016, for example, Japanese banks’ resilience to an increase in the cost of dollar funding was checked; in 2017 their resilience to a crisis in the real estate market; in 2018, the model assumes an escalation of low-income borrowers’ problems).

Jointly with major banks, the Bank of Japan experts analyse detailed bank data, and assess credit risks through constructing models that use different data aggregation levels (by industry, by individual borrower, by asset size, or broken down into production sector and nonproduction sectors). Analysis suggests that the total loss level is best predicted by a model in which loans are aggregated on the level “40 industries and two asset classes by size”, whereas the levels of bad loans by industry are the most accurately identified by a model using data of individual borrowers. The Bank of Japan is currently implementing a joint stress testing project with banks to assess a potential impact of interest rate hikes on borrowers’ creditworthiness and the banking sector’s credit risks.

Elizaveta Danilova presented a MST concept implemented at the Bank of Russia. It aims to cover not only banks but also noncredit financial organisations (NFOs: insurance companies, nongovernment pension funds, and broker/dealer firms). The rationale for such a wide coverage (despite NFO’s small share in the financial system’s assets) is that, first, NFOs are extensively interconnected with the banking sector and, second, the Bank of Russia (unlike many other central banks) is a megaregulator responsible for, among other things, regulating and supervising the NFO sector. Stress testing applies to all companies belonging to financial groups owned by the same beneficiary, which allows for taking into account both financial support within the group, and contagion from other companies of the group. MST covers financial groups accounting for over 80% of the financial sector’s assets. The most detailed data (from the credit registry, trade repository, and banks’ reporting for securities investments) is used to analyse credit and market risks.

A two-stage approach is used to assess credit risk: the first stage evaluates the relationship between macroeconomic indicators and financial indicators of corporate borrowers; and in the second stage, the relationship between borrowers’ financial position and a bank’s provisions for individual loan types is examined. The Bank of Russia is currently fine-tuning its models. The construction of a default probability model is underway, based on information on defaults provided by major credit history bureaus.

The Bank of Russia uses MST not only for quantitative assessment of risks and their dynamics but also as a tool to implement financial stability measures. This policy can be regarded as addressing two goals: an anti-crisis one (measures to safeguard financial stability when risks are already realised) and a macroprudential one (regulatory measures to prevent the accumulation of systemic risks).
MST helps determine bank recapitalisation needs and plan anticrisis refinancing policy. Stress testing is performed on five time horizons (two days, one month, one quarter, one year, and two years) so as to identify the time when particular measures of support need to be stepped up.

Interconnection analysis is also conducted as part of stress testing along with assessment of the contagion effect (the amount of losses sustained by stress test participants as a result of their direct counterparties’ default in the previous stage). Early prevention of contagion helps restrain further shock propagation. The high ratio of the contagion effect to capital deficit may point to a bank’s high systemic importance.

Oskar Krzesicki and Marcin Borsuk dwelt on the National Bank of Poland’s approach to stress testing. Stress testing only involves banks (34 banks which account for over 80% of banking sector assets). Aggregated stress test results are published twice a year in the National Bank of Poland’s financial stability review. The central bank is not a supervisory authority in Poland, but it has been interacting with a relevant authority since 2013, comparing the results of macroprudential stress tests and those conducted by the supervisory authority. An important element of a macroprudential stress test is a macro model which is used to identify a shock scenario. Economic growth slowdown in Poland’s major trading partner countries (the EU and, accounted for separately, Germany and the UK, as well as the U.S.) and oil and gas price hikes are deemed to be the key stress factors. The National Bank of Poland uses historical shocks identified based on the Hodrick-Prescott filter, and takes into account analysts’ macroeconomic forecasts and dispersion of forecast values reflecting the level of uncertainty of the macroeconomic situation. To assess credit risk and the interest margin, satellite models are employed. Separate regressions are constructed for the corporate loan portfolio, unsecured consumer loans, and mortgage loans. A regression for forecasting the interest margin is also constructed in which the net-interest-income-to-assets ratio is used as the explained variable. Explanatory factors in the equations are macroeconomic variables, lagged explained variables, and individual banks’ variables.

Foreign currency loans account for a substantial 34-percent portion of the banking system’s mortgage loan portfolio in Poland. Mortgage lending risks are assessed using credit history bureau microdata on borrowers’ income and quality of loans. The test estimates to what extent a currency and interest rate shock changes the payment-to-borrower’s-income ratio and the borrower’s default likelihood.

Also, the National Bank of Poland assesses the impact of individual banks’ risks on their balance sheets and profit and loss accounts. If the capital adequacy ratio falls below the statutory minimum, the bank is assumed to be in a situation of default on interbank loans. Krzesicki and Borsuk note, however, that banks’ interconnection is low in Poland’s interbank loan market, so no domino effect occurs: a bank’s default does not result in a counterparty bank’s insolvency.
3. Macroeconomic scenarios for a macroprudential stress test

Session 2 of the workshop dealt with developing macroeconomic scenarios for MST. The authors indicated key criteria which stress scenario parameters should meet and methods that can be used to analyse macroeconomic and financial relationships. They also discussed their experience and practices of developing stress scenarios for implementing MST.

Mindaugas Leika of the International Monetary Fund started his presentation with describing challenges for stress scenario designers. Stress scenarios should meet the requirements of plausibility and sufficient severity of losses, while the results of stress tests conducted based on these scenarios should lend themselves to formulating conclusions regarding further economic policy improvement. These requirements make it necessary to employ certain econometric tools. GaR (growth-at-risk) models based on (Adrian et al., 2018) help estimate the likelihood of a stress event and establish its relationship with macrofinancial risks. The severity of scenarios (which is equivalent to the amount of forecasted losses) can be estimated using the CaR (capital-at-risk) model. To estimate economic agents’ operations aiming to reduce the overall risk level or to assess the effects of macroprudential policy, general equilibrium models (DSGE), structural vector autoregressions (SVAR), or ABM can be applied.

Economic models used for MST purposes should meet the following requirements: they should be based on clear methodology; factor in structural features specific to national economies as the impact of exogenous shocks is modelled; take into account country-specific economic policy rules; enable data on a large number of objects to be included in analysis; and allow for the tracking of macrofinancial relationships in the economy.

The stress testing horizon is usually two to three years, while a longer horizon may result in an overestimation of losses at the time of a shock. The results of stress testing may show larger losses on a long horizon, as this implies keeping a bank portfolio unchanged. In reality, however, banks tend to adapt to the stress environment so as to minimise losses.

The GaR model constructs probability distribution of real GDP growth rates in relation to financial conditions. The degree of investor risk tolerance is an endogenous variable of the model (the volatility level of financial conditions and economic growth are interdependent). The GaR model, thus, links information contained in financial variables to the expected GDP performance based on tail risk.

The GaR model allows for the identification of the measures of financial conditions which have the strongest impact on future real GDP performance and shows the relative importance of these variables. The model shows a relationship between the financial sector and the real economy by constructing the probability distribution of aggregate output growth in relation to the severity of financial market conditions. The model allows for the assessment, among other things,
the severity of scenarios and comparing them with the current financial market conditions.

The GaR methodology explores a wide range of financial indicators, which can be roughly split into three groups: 1) the cost of risk (risk spreads, return on assets, price volatility, inflation); 2) lending measures (leverage – the debt-to-equity ratio, lending growth rate); 3) external conditions (global investors’ risk appetite, commodity prices, exchange rates, regional growth rates).

The key findings reported by Leika based on the GaR model are as follows. Milder financial conditions help maintain higher economic growth rates and lower volatility levels in the short run (up to four quarters), but result in lower growth rates and high uncertainty in the medium term (more than one year). Mild financial market conditions heighten investor risk tolerance, jeopardising, however, economic growth in the medium term. Maintaining mild financial conditions at present is a drag on economic growth in the future.

Stress testing is especially important when developing recommendations for economic authorities. As part of this policy, the IMF, based on the results of stress testing, evaluates the effect of economic policy instruments on the GaR level. An analyst chooses the desired level of GDP growth at risk. Then the results of implementing microprudential and macroprudential instruments affecting the performance of explanatory variables in the GaR model (leverage, housing prices, lending growth rates) are estimated on a 12-month horizon. Comparing the implications of various economic policy measures, the analyst chooses the most efficient instrument – the one that secures GaR volatility reduction with the least decline of the mean output growth value.

To conduct structural economic analysis, the IMF relies on various versions of DSGE models, structural and global VAR and Flexible System of Global Models (FSGM) used for drawing up the World Economic Outlook. Models employed for MST purposes factor in an array of key variables characterising the condition of an economy and its financial system. Moreover, the model’s tools enable macroprudential authorities’ policy to be taken into account.

Iuri Lazier of the Central Bank of Brazil noted in his presentation that design of stress scenarios should proceed from the following assumptions: economic conditions are worsening, a negative impact of external conditions is observed or specific risks are realised.

The quantitative parameters of stress scenarios can be constructed based on market expectations, historical experience, expert estimates, statistical methods, or a combination of these. The Bank of Brazil studies four negative scenario types: baseline, worst-case, worse-than-worst-case (based on historical experience), and a structural scenario.

In conducting MST, the Bank of Brazil looks at the performance of major macroeconomic variables (GDP, interest rates, exchange rates, unemployment) and global factors (international interest rates, sovereign risk premia, and commodity prices).
The baseline, or the most likely, scenario is designed based on professional analysts’ forecasts, with a number of parameters (commodity prices, unemployment, sovereign risk premia) assumed to remain unchanged. The level of interest rates in the international financial market is set in line with the U.S. Federal Reserve forecast.

The magnitude of a negative shock of a variable under observation in the worst-case scenario is set on the border of a 95% confidence interval. Commodity prices decline linearly at an annual rate of 5%. The unemployment rate soars to 80% within 6 quarters. The cost of borrowing in the international financial market is formed in line with a Federal Reserve forecast. The values of other variables are computed using the VAR model based on a conditional forecast.

The parameters of the worse-than-worst-case historical and structural scenarios are chosen based on historical observations (the largest variable deviations). Further performance of the model’s variables is formed based on the proportions of relationships between the model’s parameters. Under the worse-than-the-worst-case scenario, the scale of the shock is determined using data on the largest actually observed drop of the variable in history. To implement a structural scenario, a two-year period that saw the negative path of the variable under observation is selected, followed by a year of stabilisation. Proportions of the change in variables are carried over to the future stress testing horizon.

Stress scenario design and evaluation procedures at the Bank of Russia were reviewed by Andrei Lipin. In conducting stress testing, the Bank of Russia relies on a macrofinancial model consisting of a complex of modules which comprise a base semi-structural model of economy functioning and a set of auxiliary (satellite) models estimating individual macrofinancial indicators. The base semi-structural model is constructed in neo-Keynesian logic on a quarterly basis and features more flexible assumptions than the classical DSGE model, thus providing a more flexible forecasting tool in the stress environment. The model specification is based on a system of behavioural equations characterising a relationship between inflation, business activity, exchange rate movements, interest rates, terms of trade, and external sector variables (see Bank of Russia, 2017).

The strength of this model is that it includes the function of monetary policy response to stress as part of the Bank of Russia’s inflation targeting mandate. Significant inflationary risks in the stress environment are therefore automatically alleviated in the model thanks to the Bank of Russia’s proactive response.

The Bank of Russia’s macroprudential stress test provides for stress scenarios on a horizon of up to two years. The scenario envisages a sharp drop in oil prices, and their staying at this level for two years. In addition to an oil price shock, the scenario assumes a deterioration of external conditions of the economy’s functioning, including in global financial markets, which boosts net capital outflows from the Russian economy.
4. Risk assessment using Big Data

Modern banks extensively use machine learning and Big Data analysis to construct risk assessment models and to promote their products. The sophistication of reporting data collection happens in parallel; the most advanced countries move from supervisory forms to the data-centric model, where supervisory authorities have virtually continuous access to financial organisation accounts. Therefore, information of increasingly high granularity becomes available to central banks conducting macroprudential testing.

This session heard two presentations on using Big Data for assessing financial system risks.

The presentation of Fabrizio Lopez Gallo Dey from the Bank of Mexico dealt with using payment system data to assess bank risks for loans to nonfinancial organisations. It became evident after the 1994 – 1995 crisis that information available about the financial system was insufficient for monitoring individual sectors, markets, financial institutions, and transmission channels. Steps have therefore been taken to arrange information exchange among the country’s regulatory authorities, with granular data collection on a daily basis currently being implemented.

The Bank of Mexico was one of the first central banks to use daily payment system data to analyse systemic risks. Based on this data, incoming and outgoing payments were evaluated for individual firms. The difference between incoming and outgoing payments was assumed to be indicative of a company’s income. Expert estimates established the threshold of income decline below which a company was assumed to be in default, ceasing payments on its liabilities to other companies and banks. This identified the network effect which helped evaluate banks’ sensitivity not only to corporate borrowers’ defaults, but also to that of companies associated with borrowers.

In his presentation, Evgeny Rumyantsev of the Bank of Russia reviewed risk assessment in the mortgage lending segment relying on Big Data. He presented a technique for assessing the risks of a mortgage loan portfolio using the migration matrix. Unlike conventional approaches, under which the migration matrix depends on how long the loan has been in default, the likelihood of transition is, under the technique in question, described by such characteristics of the loan as LTV (the loan-to-value ratio), PTI (the payment-to-income ratio), and the number of a borrower’s household members. Also, macroeconomic indicators such as unemployment, household income and inflation performance are used to model the likelihood of transition. All macroeconomic variables used are regional, enabling the specific features of the region where the loan was provided to be factored in.

Relationships between the probabilities of transition, loan parameters, and macroeconomic variables are identified using Dirichlet regression, which is widely employed for composite data. Regression coefficients are estimated by the Markov chain Monte Carlo method (Bayesian approach), which allows for
the computing of the joint distribution of model parameters rather than their pointwise estimation. This forecasting exercise provides data on the distribution of losses in the loan portfolio. The collection of a large number of loan portfolio parameters reveals differences in loan portfolio quality across banks, allowing for a more accurate risk assessment with the characteristics of borrowers and loans in the loan portfolio taken into account.

5. Using stress testing to assess nonfinancial companies' risks

Credit risk associated with lending to corporate borrowers is traditionally a key risk for various countries’ banking sectors: loan claims on nonfinancial organisations account for the largest share of assets in banks’ balance sheet structure (about 37% in Russia, for example). In a number of countries, nonfinancial companies not only borrow massively from the banking sector, but also raise a substantial amount of debt in financial markets, including from non-residents. In the event of a crisis, a large amount of companies’ external debt may become a source of instability even if the situation in the banking sector is stable. This heightens the systemic importance of nonfinancial companies in a stress environment, making it necessary to regard them as a direct object of stress testing (rather than as an element of credit risk assessment for banks). A sizable external debt of nonfinancial companies is seen, in particular, in Turkey and Indonesia.

Pelin Sumer from the Central Bank of Turkey and Linda Hakim from the Bank of Indonesia presented similar approaches to assessing company credit risks. A company’s financial stability is associated with the interest coverage ratio (ICR) calculated as the ratio of a company’s earnings before interest and taxes (EBIT) to net financial expenses. A company’s EBIT is calculated by multiplying the revenue of the previous period by the average EBIT profitability over the last three years. In modelling net financial expenses, potential interest rate hikes under a stress scenario and the movements of the national currency exchange rate are taken into account. Companies with the ICR value of less than 1.5 show an elevated default probability. To assess loan loss risks for such companies, a probability of default (PD) value of 15% is applied calculated by Moody’s model on 1970 – 2012 data for companies with the ICR of less than 1.5. The level of losses upon a borrower’s default (Loss Given Default, LGD) is computed using World Bank data for relevant regions and countries. The amount of expected loss is thus computed from the following formula:

$$EL = EAD \times PD \times LGD,$$

where $EAD$ is the amount of loan claims on the borrower.

An increase in expected losses is reflected by changes in the capital adequacy ratio and helps assess the effect of losses on the financial stability of individual banks.
This approach helps assess the sensitivity of bank portfolios’ credit quality to interest rate hikes and exchange rate fluctuations but can be complemented by companies’ other financials. This approach can be successfully employed in countries where foreign currency-denominated or floating interest rate loans account for a substantial share of companies’ debt. As regards other loan types, techniques for forecasting companies’ revenues depending on a macroeconomic scenario still need to be developed. This will allow changes in the debt burden of companies in stress to be predicted. Also, given that the default probability of companies with similar financials may vary between countries due to national financial and business specifics, it would be worthwhile to replicate the Bank of France’s approach and develop countries’ own models of default probability forecasting based on their default statistics.

Because of this, the Central Bank of Turkey also uses an alternative approach to risk assessment for companies, assigning them scores calculated using Altman’s models (Altman, 1968). Altman’s model coefficients are calculated on Turkish companies’ default data. After establishing a relationship between the assigned score and macro variables, an analysis of companies’ financial resilience to macroeconomic shocks can be conducted.

6. Assessing risks of nonbanking financial organisations in a macroprudential stress test

Most central banks only cover credit institutions in conducting MST (in line, for example, with the Bank of Japan and ECB concepts described above). This stems from banks’ domination of many countries’ financial systems, and this is where systemic risks first emerge in the event of a crisis. On the other hand, long-term institutional investors (insurance companies and pension funds), which are not subject to the risks of short-term asset withdrawal, help absorb shocks.

The situation has been changing over the past decade: as the Financial Stability Board’s review on the risks of the parallel (shadow) banking system suggests, nonbanking financial intermediaries’ asset growth has steadily outperformed that of the banking sector. Financial reforms in the wake of the 2007 – 2009 financial crisis mainly consisted of toughening banking sector regulation. The 10-year period of low interest rates has brought down institutional investor returns. This situation has enhanced the role of asset management institutions, and recent years have increasingly seen initiatives to assess the systemic risks of nonbanking financial organisations (systemic stress initiative, Bank of England technique for risk assessment of investment funds, etc.). Laurent Clerc from the Bank of France, Inro Lee from the Bank of Korea, and Alina Kuraeva from the Bank of Russia presented their findings and insights regarding this issue.

Laurent Clerc discussed the Bank of France’s current study of insurance companies specialising in life insurance and asset management. The study became
relevant as legislation (Sapin II ³) was passed in December 2016 giving the Supreme Financial Stability Council (a collegiate authority for France’s macroprudential policy) powers to apply macroprudential measures to these organisations (the right to impose a moratorium on early termination of life insurance contracts, and various measures to manage investment funds’ liquidity). The Bank of France study assesses two channels whereby these financial organisations can generate systemic risk, i.e., a risk of mass financing withdrawal and that of ‘fire sales’, entailing direct or indirect contagion. Life insurance companies may encounter a surge of early contract terminations in the event of a crisis, when people seek to invest funds in risk-free assets (with those who terminate policy contracts earlier than others enjoying the greatest benefits). Investment funds also encounter this risk; moreover, they are heavily interconnected with other financial organisations via the funding channel (insurance companies and pension funds take the lead, followed by banks, with households in third place).

The study analysed a sample of France’s major funds and insurance companies accounting for over 50% of total assets under management. The partial equilibrium model (Greenwood et al., 2015; Cetorelli et al., 2016) was employed to examine funds’ behaviour in the event of interest rate rises, which the model regards as an initial shock. The concept of an investor’s risk budget was used, showing the possible amount of losses an investor is prepared to tolerate. Heterogeneity is ensured by adding an individual stop loss limit (a minimum price triggering the sale of an asset). After the shock has happened, the investor’s risk budget dwindles, hence investors want to sell a part of their portfolios. The sales are assumed to start with the most liquid assets, followed by less liquid ones, which are sold at a larger discount. Further negative repricing also hurts other investors (an indirect effect). With assumed risk budget of 34%, asset selloffs stand at about 13% of assets under management, while the second-tier effects are limited. Contagion may, however, arise in the event of a substantial yield curve shift (more than 300 bps).

As Inro Lee reported, the Bank of Korea is currently working on integrating the model of the macroprudential stress test for the banking sector developed in 2012 with the model of stress testing nonbanking organisations completed in June 2018. Korea’s key nonbanking financial organisations include insurance and credit card companies. Inro Lee pointed out the increasing share of these institutions in the financial system and the significant degree of interconnectedness of this sector, with its regulation remaining relatively loose.

The Bank of Korea’s model of stress testing nonbanking financial organisations is made up of several modules: a module for calculating operating income/loss, modules for estimating losses from credit and market risks, modules for estimating net interesting income and capital adequacy, and a risk transmission module. An individual model is constructed for each nonbanking organisation

³ Sapin II Law (Law No. 2016-1691) is the bill on transparency, the fight against corruption and the modernisation of the economy.
type, while business income (fees and commissions, insurance premiums, etc.) is estimated using regressions with macroeconomic and financial variables. Risk transmission is estimated based on the mutual positions of stress test participants (via liquidity, credit, and market risks).

**Alina Kuraeva** presented a scheme for assessing nonfinancial organisations’ risk as illustrated by insurance companies. Insurance companies’ market risk is assessed based on the z-spread approach for securities and real property repricing under a single scenario. Counterparty risk is assessed using the LGD methodology allowing for changes in ratings. Insurance risk is determined as a scenario-based worsening of an insurance business’s operating results using standard Solvency II formula with simplifications. The sum of an insurance company’s risk is set against its available equity, the results are included in an overall assessment of the financial group’s stability; the insurance company’s stability is, among other things, assessed with support from its parent financial group factored in.

As the presenter noted, the parameters of this methodology are subject to calibration as the risk-oriented approach to regulation (Solvency II principles) for Russian insurance companies is introduced and quantitative studies of the Russian insurance market are conducted.

**7. Assessing feedback effects between the financial sector and the real economy**

The conventional approach to stress testing provides for assessing the immediate effect of a stress scenario on the indicators of financial sector stability. For microprudential stress testing, a stress scenario implies shocks to the real economy, such as negative GDP growth and a rise in unemployment; while for commodity countries shocks are represented by a drop in commodity prices, a fall in export revenues, and weakening of national currencies. The implications for the real economy of a bank’s potential problems revealed by a stress test are of course not assessed.

MST, by contrast, implies that a stress test assesses not only banks’ stability but also their ability to perform their key function of lending to the economy and transforming household savings into investment. Economic growth and the pace of the economy’s recovery after an initial shock in turn depend on lending and investment performance.

If the financial system is unstable, banks have to undertake deleveraging – curtailing lending and restructuring bad loans, which means they are unable to provide loans to efficient businesses and thus to help the economy’s recovery. One vivid example of the impact of financial stress on economic development is provided by Japan, where the burden of restructured loans weighed down on bank

---

lending, resulting in the ‘lost decade’ of the 1990s (Danilova and Markov, 2017). Therefore, after obtaining the results of stress testing on the macro level, a second estimation should be launched to find out how the economy’s further weakening would affect the financial sector.

Assessment of feedback effects between the financial sector and the real economy was discussed in the Session 6 of the IMF – Bank of Russia workshop. In their presentation, Stefan Schmitz from the National Bank of Austria and Mindaugas Leika from the IMF noted that ignoring feedback effects could result in a substantial underestimation of systemic risk, as stress not only brings about the financial system’s direct losses (direct effect) but also produces indirect negative effects. They are divided into so called front-book and post-book effects.

The impact of financial stress on the economy materialises through front-book effects under which the worsening conditions of the financial sector’s functioning make intermediaries curtail financial services or provide them at a higher price. The post-book effects act in the opposite direction: an economic activity decline stemming from a financial intermediation squeeze worsens economic agents’ solvency, which in turn causes an additional financial sector stress.

The issue of feedback between the financial sector and the real economy is closely related to financial intermediaries’ behaviour mode and their response to stress. After the emergence of stress, financial intermediaries have to decide on their countermeasures under the given financial and regulatory constraints. The extent to which the worsening of financial sector conditions affects the real economy, and the other way round, largely depends on how most financial intermediaries deal with the problem.

One of important triggers of banks’ response to stress is changes in the cost of funding. Schmitz and Leika pointed out in their presentation that the increasing cost of funding is an essential factor of business activity revision and provided examples of how much major banks’ cost of funding rose after the 2007 – 2009 global financial crisis. The estimation from a simultaneous equations system illustrated feedback between the cost of funding and capital adequacy. As the cost of funding increases, a bank’s stability declines on the back of rising interest expenses, but the opposite relationship is also significant: a decline in financial stability sends the cost of funding higher (Basel Committee on Banking Supervision, 2015).

In choosing assumptions for the construction of optimisation models, it is important to identify financial intermediaries’ business priorities in a stress environment. The recovery of financial stability may be driven by deleverage, changes in asset risk profile, and bank recapitalisation (or a combination of the three). A Basel Committee on Banking Supervision survey has shown that, following stress testing results, most banks prefer to improve their financial
stability primarily through recapitalisation. Similar conclusions are found in the ECB reports regarding plans for restoring European banks’ financial stability. That said, one should bear in mind that this option of financial stability improvement is best used under relatively favourable market conditions.

In a stress environment, the main channel of financial intermediaries’ adaptation to a shock is deleverage. In the models presented, deleverage is achieved by revising lending prices and selling a part of (primarily non-core) assets. In revising banking product prices, it is important to take into account the level of pass-through of funding costs to deposit and loan rates (Harimohan et al., 2016). This largely depends on the nature of a shock (idiosyncratic or systemic), an intermediary’s market power, and also its objective function in balance sheet optimisation.

Therefore, in an environment of stress and rising funding costs, financial intermediaries are to address the optimisation problem. On the one hand, a rise in expenses needs to be restrained (sources of finance need to be optimised), while on the other hand, the asset portfolio needs to be rebalanced. Several portfolio optimisation models were presented at the workshop. Meanwhile, the issue of choosing optimisation criteria remains open to discussion. In particular, models with various optimisation criteria, including Risk-adjusted Return on Capital (RAROC) and the multi-period optimisation based on Economic Value of Equity (EVE), have been developed.

One example of the deleverage optimisation model allowing for feedback effects between the financial sector and the real economy was provided by Laura Valderrama of the IMF. The model seeks to identify an inter-sector equilibrium under which various types of financial intermediaries interact, factoring in with their effect on macroeconomic performance. The model covers banks, with banking and trading books analysed separately, as well as nonbanking organisations – comprised in one part by borrowers receiving bank loans, and in the other part by bank traders and shareholders.

The key model components in the banking sector block include the current profit maximisation function, a balance sheet equation, and regulatory restrictions. The nonbanking sector block of the model defines a borrower’s default probability functions, which rise monotonically in proportion to lending growth but decline monotonically with economic growth. The securities market appears to be in equilibrium resulting from the operations of traders, who can buy securities from banks at low prices during fire sales. Investors play the role of additional capital suppliers and buy banks’ shares in proportion to their increase in return.

In addition to the financial sector, the model contains a macroeconomic block represented by the IS curve, the extended Phillips curve, the Taylor rule.

---

5 Basel Committee on Banking Supervision, 2018.
6 EU banks have to submit restructuring plans to authorities; the documents contain a detailed list of bank reactions to both idiosyncratic and macro-economic shocks (European Central Bank, 2018).
and interest rate parity. An illustrative numerical example shows that feedback effects are economically significant, thus aggravating the depth and length of a recession. Therefore, the incorporation of feedback effects in the model improves its accuracy and enhances its relevance to MST.

The importance of the feedback effects in estimating banking sector stability was pointed out by the Bank of Japan’s Wataru Hirata. As mentioned above, banking sector problems resulted in a ‘lost decade’ for Japan, hence this aspect acquires a special importance for Japan’s system of MST.

To assess the implications for economic growth of the weakening of banking sector financial stability, the Bank of Japan uses an econometric equation linking nominal GDP growth to a change in the real economy’s bank debt, which also depends on the banking sector’s profitability and capital adequacy. The incorporation of the feedback effect in the equation doubles the deviation of GDP growth rate from the baseline scenario. In modelling the deleverage process, banks’ profitability heterogeneity needs to be factored in, as banks with low or negative profitability are the most prone to lending contraction.

8. Development of financial sector contagion analysis

The 2007 – 2009 global financial crisis revealed structural imbalances in the global financial system and lack of information about interconnection among the financial sector’s participants. In this context, studies examining contagion risk, the channels of its occurrence, and implementation of measures to contain financial panic and restrain shock propagation acquire special relevance. Session 7 of the workshop was devoted to these issues.

The presentation of Fabrizio Lopez Gallo Dey from the Bank of Mexico examined the contagion mechanisms, their modelling methods and examples of regulatory practices. It also dwelt on solvency issues, constraints on funding sources and the role of payment systems in a stress environment.

The author noted a lack of consensus among economists regarding the concept of contagion, which King and Wadhwani (1990) regard as an increase in correlation among markets, while Kodres and Pritsker (2002) regard it as a price change in one segment in response to a situation in another segment, Hartmann et al. (2004) regard it as a rise in an overall default probability. Various contagion criteria are indicated: declaration of default (Allen and Gale, 2001; Eisenberg and Noe, 2001), a fall in profitability (Battiston et al., 2012; Tasca and Battiston, 2016), an asset price drop (Kiyotaki and Moore, 2002; Caballero and Simsek, 2013), and constraints on funding sources (Fourel et al., 2013; Acharya et Merrouche, 2013).

Fabrizio Lopez Gallo Dey presented the results of several contagion models which cover: 1) the financial sector with external shocks introduced (banks, insurance companies and brokerages, pension and investment funds); 2) the domestic market as a closed system; and 3) the banking sector. The variables
of the models included deposit and securities portfolio changes, money market operations, derivatives and repo transactions. In addition to individual losses, contagion effect modeling assumes the realisation of a macroeconomic shock with the following parameters: interest rates, exchange rates, GDP, inflation rate, and stock indexes.

To finance the liquidity deficit in a situation of limited access to the funding market, a bank may operate under the following four scenarios. A bank withdraws its short-term funds 1) from all positions in the money market on a pro rata basis (Hard, 2016), or 2) according to a certain structure of preferences with regard to counterparties. Possible alternatives include 3) withdrawing all funds from the interbank market and covering the balance through selling liquid assets, and 4) first selling liquid assets, then withdrawing funds from the interbank market.

It was emphasised that contagion effects are dangerous because they can amplify the initial shock manifold, triggering a number of further defaults. The presenter noted that confining analysis to just one round was inadequate, resulting in an underestimation of financial risks. The Bank of Mexico takes account of stress test results in developing its regulatory framework. Specifically, contagion risks in Mexico’s financial system have been reduced by imposing restrictions on transactions within a group.

Lopez Gallo Dey regards clustering analysis as a promising area of research, as world practices lack an optimal algorithm of splitting a cluster into homogenous subgroups that would serve as an intermediary link between an individual bank and the banking sector. The most popular model, presented in Craig and von Peter (2014), has limited applicability. The key issue to be addressed is the criteria and assessment of the extent of interconnection, which affects the propagation of contagion in the financial sector. One of interconnection determinants is regularity and volume of cash movement between market participants via payment systems.

Cluster elements need to be grouped to forecast contagion dynamics. It is highly inhomogeneous: seizing an individual bank, contagion is to a certain extent localised within its subgroup, acquiring a cumulative nature in the event of default. The closeness centrality indicator, the betweenness centrality indicator, and the eigenvector centrality indicator are used as additional analysis instruments.

Xiaobei He from China’s Tsinghua National Institute of Financial Research focused on the price channel of contagion and presented the relevant model. The price channel of contagion risks makes itself felt through the effect of asset price decline in response to asset selloffs by one or several major participants encountering liquidity risk. The asset price decline in turn affects the participants owning these assets (or assets highly price-correlated with them) in their portfolios and marking these assets to market, which produces a negative financial result and a negative effect on capital.

Modeling the impact of ‘fire sales’ on the market value of assets is conducted using the Cont and Schaanning (2017) model, which estimates elasticity of
an asset for sale price in relation to the market depth. According to estimates provided in the presentation, China’s securities market depth is more than 130 times lower than that in the U.S. for government bonds and more than 216 times lower for corporate bonds. Due to the low liquidity of secondary markets, the model presented showed high sensitivity of asset prices even with a relatively low volume of sales. Therefore, the price channel of contagion is, based on Xiaobei He’s estimate, of major significance.

Given the low liquidity of the national securities market, the model assumes successive asset sales, with the riskiest assets being the first, and the least risky ones the last, to be sold; and also selling first the most liquid and then the least liquid assets. The deleverage strategy for the trading portfolio was therefore modelled based on its impact on a bank’s financial position. Fire sales start with the most liquid commercial papers, then corporate bonds are sold, and the last to be sold are the least liquid government securities. This sequence allows the fastest sale of securities at the lowest discount and with a maximum positive impact on capital adequacy.

The initial shock assumed in the model is a twofold rise in the level of non-performing construction sector loans (NPL), because it is in this sector where banks’ highest credit risks are concentrated. The model’s NPL level doubles from 1.04% to 2.08%, bringing about a shortfall in a minor bank’s capital. In the next stage, the model estimates contagion triggered by this bank selling a part of its assets so as to recover its capital adequacy. The first round of contagion produces a capital shortfall at three banks. It is, thus, shown that even small banks can give rise to significant systemic effects that need to be taken into account in macroprudential stress testing.

The European Central Bank’s Christoffer Kok provided a review of contagion models and the use of network analysis for macroprudential regulation purposes at the ECB. Worth noting are, among other things, models based on accounting data, those analysing direct credit exposure and allowing for the structure of investment portfolios, as well as static and dynamic models, depending on the assumed nature of network effects. Contagion models may assume fire sales, retention of liquidity by banks, and overall macroeconomic and individual shocks.

Christoffer Kok made special mention of changes in central banks’ approaches to safeguarding financial stability under the influence of the global financial crisis. Prior to the crisis, regulation proceeded from the assumption that the market was self-sufficient and hence it was reasonable to limit government intervention. The key risks were attributed to fundamental factors, such as the profitability level, competitive positions, and an individual bank’s capital adequacy, whereas bank runs and liquidity shortages were regarded as possible consequences rather than root causes of these. On top of that, the scale and depth of potential network effects were underestimated. The financial market, with its ongoing sophistication of the structure of instruments, had therefore largely outpaced legislation and regulation development, paving the way for the crisis.
The global financial crisis has revealed the need to develop network analysis methodologies and early response systems to localise potential risk. The assessment of the degree of interconnection among market participants and ensuring efficient intermediation have come to be treated as macroprudential policy priorities. Banking sector stress testing, taking into account contagion effects, has gained wide acceptance (Henry and Kok, 2013; European Central Bank, 2016; Dees et al., 2017). It seeks to estimate the banking sector’s aggregate losses under the adverse scenarios of risk contagion among the participants. Meanwhile, methodologies currently employed across countries are, as Christoffer Kok pointed out, still in their early stage of development, and their effectiveness has yet to be explicitly evidenced in practice.

The degree of interconnection among banks varies across market segments. The absence of direct credit exposure does not mean that banks are isolated from the perspective of contagion. It is banks, insurance companies, brokerages and other firms that can act as intermediaries. An individual organisation’s position in a counterparty network – i.e., the number and volume of transactions, the number of businesses in a banking group, the degree of business diversification and its geography – plays an important role. This provides a basis for identifying a group of systemically important financial institutions whose default would have serious implications for the entire market. Therefore, these institutions are subject to especially close supervision by regulatory authorities.

Analysis of financial sector structure uses, among other approaches, the theory of graphs, which enables interaction among market participants to be reflected in a standardised form. This standardisation substantially widens the array of analytical instruments available, including the measures of centrality. These instruments are used to forecast potential implications of particular regulatory measures. The findings of Halaj and Kok (2013) suggest that a reduction of the limit of interbank loan exposure from 25% to 10% of a lending bank’s Tier 1 capital (Basel III) has proved highly effective in limiting contagion risk for major banks.

In Christoffer Kok’s view, a major problem in analysing network effects is limited access to data and low potential for its processing. A large number of contagion channels and their interconnection call for a large variety of statistical data, ranging from general market information to financial accounting access to which is limited. On the one hand, the network analysis of the banking sector requires unified standards for the comparability of results to be ensured, while on the other hand, the specifics of individual organisations need to be taken into account. Given these requirements, banks are themselves asked to take part in stress testing and provide expert evaluation of default probabilities to the regulator.

The inconsistency of network data presents an additional difficulty. One example is provided by Covi et al. (2018), which analysed the causes of discrepancy between the Page Rank indicators, centrality with regard to its own vector and
degree. These measures provide inconsistent assessment of the interconnection structure, and researchers lack consensus on what the optimum version is.

Kok also pointed out the ambiguity of regulation aiming to limit banks’ susceptibility to risk. On the one hand, a rise in credit exposure, the amount of transactions within one group, and the number of counterparties amplify contagion risk, but on the other hand, this brings about a positive synergy. Therefore, the toughening of regulatory requirements regarding, for example, the share of a maximum loan in a bank’s capital may produce negative effects. The need for a balanced approach which would both secure financial stability and create conditions for economic growth is emphasised.


Central banks conducting macroprudential stress tests, as a rule, disclose their results in an aggregated form. Detailed information disclosure may give rise to a market panic if the test results prove to be negative. At the same time, in some countries (in the U.S. for example) this instrument is extensively used as part of supervision, and, if the stress test results are negative, banks are required to recapitalise.

Dmitry Orlov from the University of Rochester presented the key findings of his paper (Orlov et al., 2018), which discusses optimal approaches to information disclosure regarding stress test results.

The disclosure of stress test results is clearly positive from the perspective of market discipline and provision of the right incentives to banks ex ante, also enhancing the credibility of central bank policy. If a stress test produced positive results, their disclosure is certainly worthwhile. If, however, the stress test results were negative, this may trigger a market panic and bank runs.

The author’s analysis suggests that gradual information disclosure may make sense. If a bank has failed a stress test, it would be best not to release the results at first but to require the bank to recapitalise. After the bank has recapitalised, information can be disclosed.

Marco Gross from the IMF discussed the EuraCe 2.0 ABM model, which the IMF constructed to look into endogenous causes of a phase change in a business cycle and to assess macroprudential policy potential to reduce the economy’s output volatility. The model incorporates households, firms, banks, a central bank, and government responsible for implementing fiscal policy. The model describes markets for real property, including rental property, mortgage loans, corporate loans, and investment and final consumption goods.

The impact on the banking sector of a central bank’s macroprudential policy, regarding, for example, the tightening of capital requirements, is modelled via

7 This is mostly relevant to supervisory rather than MST.
an increase in banks’ funding costs. The model also factors in the possibility of a regulator imposing lending restrictions, for example, on loans with a small down payment. Numerical experiments show that countercyclical macroprudential policy oriented to tightening capital requirements reduces the economy’s output volatility. That said, measures based on borrowers’ characteristics (the debt ratio, the size of down payment) directly affect the cycle and are more effective in reducing the amplitude of business cycle fluctuations.

Also, banks’ assessment of risks at a point in time, based, for example, on the current loss level, was found to be ineffective, since it fails to allow for a possible worsening of the borrower’s financial position should negative macroeconomic shocks occur. This problem can be partially addressed by assessing risks under the IFRS 9, where a borrower’s default probability is estimated based on a macroeconomic forecast.

10. Key takeaways from the workshop

The key findings of the workshop, including key recommendations for conducting stress testing, are as follows:

- Effective MST requires ensuring high-quality input data, sample completeness, and the use of the most relevant models. It is therefore best to choose granular data, to cover, among other things, small banks and nonbanking organisations, and to conduct backtesting of the models used. It is of major importance to work closely with financial organisations so as to test the adequacy of assumptions and results and to share experiences and practices of using risk assessment models.

- As stress scenarios are designed, they should be tested for plausibility, severity, and suggestiveness. The GaR model is an innovative method of stress scenario evaluation helping explicit assessment of these properties. In addition to modelling on historical data, stress scenarios should be designed with market expectations, expert estimates and statistical estimates taken into account. The regulator’s response as part of the monetary policy rule should also be borne in mind.

- Banks have made Big Data analysis an integral part of their operations, with many of them already relying on machine learning approaches. Central banks are also beginning to use these tools in MST. At the same time, such models should be used with caution in designing macroprudential policy measures, because models on which decision-making is supposed to be based should allow both the regulator and the banking community their clear and unambiguous interpretation.

- Techniques of nonfinancial organisations’ credit risk assessment continue to grow in sophistication. Most central banks are currently using simplified models, often based on just one measure of debt burden. Their further
development may move both towards a more complex analysis of bank borrowers’ financial operations and towards the sophistication of models identifying corporate borrowers’ default probability.

- The assessment of feedback effects between the financial sector and the real economy has become part and parcel of MST. Ignoring the feedback effects may result in a significant underestimation of systemic risk. The workshop discussed a number of approaches to feedback modelling, including the model of post-stress business optimisation. The choice of optimisation criteria, however, remains open to discussion. In particular, models have been developed using the risk-adjusted current return-on-capital (RAROC) and economic value of equity (EVE) measures as optimisation criteria.

- Modelling contagion effects is a relatively new area of systemic risk assessment. It requires highly granular data, which is often hard to obtain, and facilities for their processing. Contagion is dangerous in that it can amplify the initial shock manifold, triggering systemic risk. The difficulty of contagion analysis stems from the variety of channels through which it is realised: the channel of access to liquidity, those of asset prices and counteragent default, etc. Contagion effects are not always assessed as part of stress testing, special models are used to assess them, including network analysis methods.

- The disclosure of MST results requires a cautious and balanced approach. The transparency of results improves risk assessment by the participants and market discipline, but may trigger further market volatility. It is best to ration information disclosure (in terms of content and time) so as to prevent stigma effects and other negative implications of disclosing stress testing results.

11. References


Fear of Forward Guidance*

Alex Isakov, VTB Capital
alex.isakov@vtbcapital.com

Petr Grishin, VTB Capital
petr.grishin@vtbcapital.com

Oleg Gorlinsky, VTB
gorlinsky@vtb.ru

This article is a response to the review of Adrian et al. (2018) by Yudaeva (2018), which summarizes the case of the Bank of Russia against the publication of key rate forecasts, a communication strategy known as conventional forward guidance. We believe that the case in favour of publishing the Bank of Russia’s key rate forecast is at present not stated sufficiently coherently. Our note attempts to fill this gap. Extending the argument put forward by Adrian et al. (2018) we provide a comprehensive review of the working papers, staff notes, and leadership comments related to interest rate expectations and monetary policy communication by four central banks that have had practical experience with the application of conventional forward guidance. We conclude with an evaluation of the validity of the commonly voiced concerns regarding its adoption in Russia, based on the reviewed literature.

Keywords: inflation targeting, monetary policy, central bank communications

JEL Codes: E58


doi: 10.31477/rjmf.201804.84

1. Introduction

Monetary policy is a conservative area. New approaches and ideas undergo careful inspection, scrutiny and testing before being considered for adoption. The latest changes to the monetary policy framework in Russia were the transition to a freely floating exchange rate, and an inflation targeting regime in 2014. This setup is generally considered as preferable, as it allows for a smoother absorption of external shocks. However, such consensus is relatively recent – in the early 1990s the debate regarding the suitability of flexible (and, ultimately, floating) exchange rate arrangements for emerging markets was far from being settled (see

* We are grateful to Nadezhda Ivanova and Olga Kuvshinova (of the Bank of Russia) for their invaluable help and support while working on this paper, and also to Alexey Khazanov (of Boston College), Nikolay Frolov and Pavel Pikulev (of Sberbank) who kindly shared their ideas and comments.
the review by Frankel, 2003). Interestingly, even when the consensus shifted firmly in favour of floating exchange rates (e.g. see the discussion in Parrado, 2004) the monetary authorities released their grip on FX market only grudgingly – a phenomenon dubbed by Calvo and Reinhart (2000) as ‘fear of floating’. Central banks characterised by ‘fear of floating’ are those for which the declared flexibility of the exchange rate arrangement is believed to be higher than what is effectively practised.

Central banks’ transparency with regards to monetary policy and their adoption of forward guidance in their communications in particular, we believe, suffer from a similar pattern of reluctant adoption. ‘Fear of forward guidance’, then, is a situation in which the central bank would acknowledge and indeed promote transparency as a general principle of its communication policy, but in practice would limit the amount of disclosure regarding its intentions for future changes in the policy rate.

Before proceeding any further, we need to acknowledge that forward guidance is an umbrella term covering a diverse set of communication strategies, which can be broadly divided into three groups:1

1. ‘General policy comments’, as per Yudaeva (2018), are verbal qualitative descriptions of the future monetary policy stance;
2. ‘Conventional forward guidance’, as per Adrian et al. (2018), implies publication of the central bank’s forecast of the policy rate path or paths;
3. ‘Commitment’, as per Smith and Becker (2015), includes conditional and unconditional commitments to a published interest rate trajectory.

This article focuses exclusively on the second type of forward guidance and, more precisely, it is a response to the review of Adrian et al. (2018) by Yudaeva (2018). The latter summarises the case of the Central Bank of Russia (CBR) against the publication of key rate forecasts. Extending the argument put forward in Adrian et al. (2018), we provide a comprehensive review of the working papers, staff notes, and leadership comments related to interest rate expectations and monetary policy communication by four central banks that have had practical experience with the application of conventional forward guidance.

To reiterate, this is a wide review with a narrow goal. We will comprehensively review the working papers and other relevant bodies of work related to interest rate expectations and monetary policy communications by the central banks of New Zealand, Sweden, Norway and the Czech Republic. However, the primary question is narrow: did the private sector experience confusion between

---

1 We do not claim our classification is exhaustive or our naming convention universal. For instance, Gavin et al. (2013) calls interest rate path announcement with commitment ‘Odyssean forward guidance’, while Fileček and Matějů (2014) extend this Greek naming convention and calls the publication of the central bank’s interest rate forecast without commitment to this path ‘Delphic forward guidance’.
guidance and commitment, or did it demonstrate a decrease in independent efforts to understand economic developments?

The note is structured as follows: in the next section we will open the topic up for further discussion by examining evidence of mismatches between the CBR’s interest rate policy outlook and that of the market. This will be accomplished by looking at the evidence provided by public surveys. This exercise is challenging because the CBR’s interest rate projections are unavailable, which means we are limited to the outcomes of monetary policy meetings and CBR’s own communication on the subject. In the third and main section we review the literature generated by research conducted at the four central banks deemed to be in the ‘avant-garde on inflation targeting’ by Adrian et al. (2018): the Reserve Bank of New Zealand (RBNZ), Norges Bank (NB), Sverige Riksbank (SR) and the Czech National Bank (CNB). In order to ensure a comprehensive review is carried out, we have consulted all working papers, public staff notes and statements by the banks’ leadership available online, related to the interest rate expectations of the markets and to the management of such expectations, for monetary policy purposes, by the central bank. These archives typically span from the early 1980s to the present. The depth of the banks’ experience with conventional forward guidance is also substantial, ranging from 20 years for the original adopter, RBNZ, to 10 years for the CNB. In the fourth section we review the case against conventional forward guidance, as summarized in Yudaeva (2018), and comment on the salience of the concerns raised by the practising central banks. We also highlight the approaches that the aforementioned central banks used to avoid confusion between guidance and commitment. The last section concludes.

2. Motivation

Increasingly, clear communication on behalf of the central bank is seen as essential for an efficient monetary policy, a point highlighted both in Adrian et al. (2018) and Yudaeva (2018). In its official statements, the CBR (Bank of Russia, 2017) has stated that, in order to build confidence in its policy and reduce and anchor inflation expectations, it will increase its communication transparency and be more open about the measures it employs to achieve the goals of the monetary policy.

The costs generated by the bias in the interest rate expectations of the public and the markets are of the same nature as those arising from unanchored inflation expectations. Both are factors influencing the real interest rates, which in turn drive essential decisions regarding borrowing and saving, and capital allocation. This means that if inflation expectations are excessively high or interest rate expectations excessively low, then monetary conditions will be easier than what is consistent with the central bank’s goal and the bank would have to tighten its monetary policy further or for longer, than in the case of well-coordinated expectations. The reverse is true for excessively low inflation or high
interest rate expectations, which may lead to the bank repeatedly undershooting to the inflation target, resulting in less buoyant economic activity.\(^2\)

How relevant is this issue to Russia? For the purposes of this discussion, it is useful to rely on the suggestion of Blattner et al. (2008) to distinguish between short-term and long-term predictability when assessing the clarity and efficiency of a central bank’s communication. Central banks’ longer-term predictability is inherently difficult to measure, especially when the numerical guidance on the interest rate path is not available. Thus, we first review the evidence provided by the market’s ability to foresee the CBR’s decisions in the short term.

To assess the short-term predictability of the monetary policy, we rely on Bloomberg survey data of financial analysts. This data is most suitable for the purpose because the platform essentially allows for changes to the estimate submitted up to the moment of the announcement of the rate by the Board of Directors, and thus it excludes the possibility that the rate submitted rate is different because of events occurring between the time of submission and the publication of the decision.

**Figure 1a.** Share of correct estimates of the policy rate decision: New Zealand

![Figure 1a](image)

*Source: Bloomberg*

We review the data for January 2015 – September 2018, which is the period characterised by a floating exchange rate regime, excluding the initial shock

\(^2\) The ability of the central bank to influence longer term interest rates are implied by the “expectations theory” of the yield curve, e.g. see Cox et al. (2005). Empirical tests of this hypothesis are not uniformly supportive, i.e. Roush (2007) finds that they hold in case of monetary policy shocks, and are rejected when shocks stem from aggregate demand.
of Q4 2014. During this period, 29 Board of Directors’ meetings took place. According to the Bloomberg data, on average, 65% of the estimates submitted correctly predicted the Board’s decision regarding the key rate. In five cases (January 2015; April 2015; March 2017; April 2017; December 2017) the decision was expected only by 1/4 or less of those surveyed. In two cases (January 2015; December 2017), the decision was completely unexpected by the forecasters.

**Figure 1b. Share of correct estimates of the policy rate decision: Russia**

![Share of correct estimates chart](image)

*Source: Bloomberg*

To determine how typical these results are, it is useful to compare the average shares of correctly submitted estimates for 2015 – 2018 for Russia and for a country where the central bank had already adopted conventional forward guidance: 65% (Russia) vs 90% (New Zealand).

Longer-term predictability is more important from a monetary policy implementation perspective, and more costly – if the mismatch persists from a capital allocation perspective. However, it is also not possible to reliably evaluate it in the absence of published rate forecasts by the central bank. An exception would be when the central bank itself decides to comment on such divergences.

Such a case was registered in 2016, as documented in Yudaeva (2018), when the CBR identified an appreciable divergence between the CBR’s internal and market interest rate forecasts. This prompted the CBR to declare that it will be clearer in its communications and announce a commitment to not revise its interest rate for several months. Given the still relatively recent implementation of inflation
targeting, even this single episode of divergence between the expectations of the central bank and those of the market should serve as an early lesson.

We believe that the costs of such a mismatch deserve more in-depth study. The gravity of the issue regarding the management of interest rate expectations cannot be less serious than that of the management of inflation expectations, and is potentially even more consequential.

### 3. Interest rate expectations and their management: notes from the avant-garde

The focus of this review is exclusively central banks’ experience with conventional forward guidance (CFG). This means that it omits forward guidance as a commitment to the interest rate path. This strand of literature is concerned with examining the suitability of commitment at the zero lower bound (ZLB), which is not immediately relevant for Russia.

The scope of the review includes all public working papers, staff notes and statements made by the banks’ leadership and available online related to interest rate expectations of the markets and their management for monetary policy purposes by the four aforementioned central banks, which have practical experience with implementation of the regime.

<table>
<thead>
<tr>
<th>Central bank</th>
<th>IT since...</th>
<th>CFG since...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reserve Bank of New Zealand</td>
<td>1989</td>
<td>1997</td>
</tr>
<tr>
<td>Norges Bank</td>
<td>2001</td>
<td>2005</td>
</tr>
<tr>
<td>Sverige Riksbank</td>
<td>1995</td>
<td>2007</td>
</tr>
<tr>
<td>Czech National Bank</td>
<td>1997</td>
<td>2008</td>
</tr>
</tbody>
</table>

*Source: Clinton et al., 2017.*

Regarding the literature on conventional forward guidance, in the late 2000s Norges Bank (Qvigstad, 2009) called the publication of the projected paths for the central bank’s policy ‘the new frontier in central bank communication’ and ‘high-priority area for future research’, but lamented that, at the time, it had only been practiced in a few countries and for merely a few years. The next section takes stock of the evidence regarding the publication of projected paths that we possess one decade later.

Though the results of our reviews of the evolution of central banks’ thinking on influencing interest rate expectations will be detailed below, we believe it would be useful at this point to briefly outline our impression of a typical evolution pattern we have detected in such thinking:

1) Identification: the central bank issues working papers that document the methodology used to extract monetary policy expectations from available financial instruments and/or surveys (i.e. Söderlind and Svensson, 1997);
2) Evaluation 1: working papers that attempt to assess which degree the monetary policy expectations of the markets are aligned with those of the monetary authorities by looking at surveys or at the volatility of the money market rates around the time of the central bank’s decision announcements. Such studies usually find a relatively low predictability of the monetary policy and a scope for more transparent communication (i.e. Bulir et al., 2007);

3) Implementation: to address the challenge posed by monetary policy surprises, the central bank opts to adopt conventional forward guidance (i.e. Czech National Bank, 2008);

4) Evaluation 2: typically 2 – 3 years after implementation of forward guidance, working papers start to evaluate its effects on the short-term interest rate volatility, as well as its power to shape the yield curve (i.e. Drew and Karagedikli, 2008).

This is a highly stylised scheme. There are cases where two steps (usually, identification and evaluation) are packaged into one research note. However, we believe it is useful to organise the evolution of ideas on interest rate expectations and on the central bank’s communication into steps, in order to discern patterns such as the above.

3.1. Forward guidance research at the RBNZ

The year 1997 is commonly cited (e.g. in Clinton et al., 2017) as marking the start of conventional forward guidance in New Zealand. This, however, might be misleading for the modern reader, and we think that a fully modern monetary policy implementation system was not put in place by the RBNZ until 1999. The first review of interest rate expectations is found in Krippner (1998), who looks at the monetary policy framework has been materially different: the level of reserves of the banking sector with the central bank serving as an operational instrument for achieving the operational target was formulated in terms of an interest and exchange rate synthetic indicator called the monetary conditions index (MCI). Interested readers may refer to the discussion paper of the Reserve Bank of New Zealand (1997b), which notes that ‘... the Reserve Bank set out a proposal to alter key technical aspects of monetary policy implementation, involving a shift from targeting settlement cash balances to targeting the overnight interbank cash interest rate. After consulting with market participants, we were persuaded that the risks with the proposed change might not be sufficiently outweighed by the benefits.’ The initial proposal and detailed discussion are available in the aforementioned paper (Reserve Bank of New Zealand, 1997a).

A fully modern system of monetary policy implementation was not put in place until 1999, when the RBNZ explained, as cited in Archer et al. (1999), that it would be using the official cash rate to set the stance of the monetary policy, because ‘we increasingly recognised the unpredictability of the relationship between the quantity of settlement cash and overall monetary conditions’. According to the note, the ‘Official Cash Rate system is more transparent and easier to understand than the previous regime. [...] The setting of the official cash rate clearly and unambiguously expresses the desired stance of monetary policy, and between reviews the actual cash rate will stay close to the Official Cash Rate.’ Even following the implementations of this step, the initial announcement stated that the indicative conditional forward path for monetary policy, contained in the projections included with Monetary Policy Statements will continue to be expressed in MCI terms, which it did only for a brief period. This suggests that the more appropriate dating of the implementation of conventional forward guidance in New Zealand would be 1999 rather than 1997.
interest rate futures. While he finds that futures provide an unbiased estimate of the future money market rates, the root mean square forecast error is close to 1 pp for periods as short as 90-day forecast. Krippner (2002) proposes an approach to extracting policy rate expectations from the interbank lending yield curve. He finds a positive term premium and makes suggestions about the identification of the underlying expectations regarding official rates by the banks.

Drew and Karagedikli (2008) explore the communication policy of the RBNZ in order to find out whether interest rate forecast publications help financial markets better anticipate a central bank’s reaction to incoming data and not to overreact to macroeconomic data surprises. In particular, they use high frequency data to explore whether data and monetary policy surprises affect longer-term interest rates. They conclude that forward rates tend to increase in response to positive inflation or GDP surprises, anticipating the central bank’s reaction. They add that, while some of their contemporaries argued that transparent communication risks non-credibility, as policy-makers might not always follow through with prior commitments if conditions change sufficiently, their findings suggest that this concern is misplaced.

Fukač (2008) explores the monetary policy consequences of heterogeneous expectation formation. He concludes that, if the private sector’s beliefs are well-coordinated (if ‘knowledge is homogeneous’), then a central bank can decrease inflation volatility and speed up the learning of its reaction function, if it demonstrates inflation aversion. The power of the central bank to decrease inflation volatility if beliefs are heterogeneous is lower. Thus, they conclude that the central bank should coordinate interest rate expectations if it is able to do so.

Delbruck et al. (2008) provide an in-depth overview of the evolution of the RBNZ’s forecasting and modelling toolkit. They underline the role of the publication of the central bank’s endogenous interest rate path, which creates a natural focus for summarising changes in a projection relative to the previous quarter. Discussions structured around the review of the expected path of the policy rates help communicate how the medium-term outlook has changed and what has driven such a change. Notably, they provide the decomposition of contributions of various shocks to the reviewed interest rate trajectory.

Detmers and Nautz (2012) provide the most recent review of the RBNZ’s experience with forward guidance. Their review notes, that in order to achieve a highly efficient monetary policy implementation, the central bank needs to shape market expectations about the future path of short-term rates. This understanding lead the RNBZ to adopt a quantitative forward guidance strategy, which includes publication of long-term interest rate projections. The authors also show how the global financial crisis has changed the efficiency guidance, and conclude that the role of interest rate projections for futures rates has decreased, i.e. their estimates
show a sharp decline in the size and significance of all coefficients related to interest rate projections. They hypothesise that these findings might be related to: 1) data issues, i.e. unstable risk premiums, which make futures rates unsuitable proxies for market expectations; 2) deterioration of the central bank’s forecasting precision. Indeed, the RNBZ’s self-assessment (Reid, 2016) supports the latter view and reports that the RBNZ has ‘underpredicted’ the level of the trade weighted exchange rate, which led to the undershooting of the tradable and headline inflation rates.

The conclusion arising from the RBNZ’s experience is that conventional forward guidance helps coordinate the private sector’s expectations more efficiently, accelerates the learning of the central bank’s reaction function and decrease interest rate volatility. However, the ability of the central bank to manage interest rate expectations depends on the lack of persistent bias in its projections. If and when such biases arise, they will gradually diminish the power to coordinate expectations. We believe that this decline in the attention paid to the central bank’ communications would be present irrespectively of the type of communication chosen: biased inflation forecasts would be detrimental to the central bank’s credibility even if it limits itself to general comments.

### 3.2. Forward guidance research at Norges Bank

The evolution of the approach to monetary policy communication at Norges Bank (NB) offers another interesting glimpse into debates that are considered settled today.

In particular, prior to the adoption of conventional forward guidance, the monetary policy reports of the central bank not only omitted the bank’s interest rate projections but, also, all projections assumed a constant policy rate. This meant that the need for adjustment of the policy stance was judged on the basis of deviation of inflation from the target level, according to such projections. Gjedrem (2001) documents this debate in his address, noting that several academics had argued in favour of shifting from an unchanged to an optimal interest rate scenario as the basis for inflation projections, because the central bank itself does not regard an unchanged interest rate as the most probable scenario. Thus, other projections published by the central bank, including those relating to inflation, are indeed different from what the central bank in fact expects.

In an attempt to improve its communications, NB switched to conditioning its forecasts on the market-implied policy rate expectations, as Gjedrem (2005) notes, and started to comment on how the official expectations differ from those of the market. A typical commentary regarding this is provided by Norges Bank (2005a).

The next milestone was NB’s adoption of conventional forward guidance announced in a statement (Norges Bank, 2005c) that noted the published interest rate path provides a reasonable balance between the objectives of monetary policy.
Norges Bank (2005b) added two important comments. First, it highlighted that the motivation behind the adoption of conventional forward guidance was provided by NB’s evaluation of the accuracy of market expectations of money market rates. The estimated average deviation of the market implied rates and actual rates for horizons in excess of one year exceeded 2.5 pp for 1999 – 2005. Second, it provided the criteria that Norges Bank uses to evaluate the coherence of the projected interest rate path. The report presents six such criteria:

- The interest rate must be set so that inflation is stabilised near the target within a reasonable time horizon, normally 1 – 3 years.
- The inflation gap and the output gap should be in reasonable proportion to each other until they close, and normally not be positive or negative at the same time further ahead.
- Interest rate developments should result in an acceptable inflation and output, also under alternative assumptions regarding the model of the economy.
- The interest rate should be changed gradually so that we can assess the effects of interest rate changes and other new information about economic developments.
- Interest rate setting must take into account changes in property prices and credits.
- The interest rate path should be cross-checked with some simple monetary policy rules, and systematic and substantial deviations from these simple rules should be explained.

Qvigstad (2006) discusses these points in greater depth. Interestingly, he provides a useful definition of transparency in citing the European Central Bank’s president Duisenberg, who suggests that a central bank is transparent if its external communication reflects its internal deliberations.

Holmsen et al. (2008) review three years of experience with publication of the interest rate projections and find evidence of reduced volatility in market interest rates on the days with interest rate decisions. They show that the change in the 12-month money market interest rate on the day after the interest rate decision announcement nearly halved after the introduction of the endogenous monetary policy path in 2005. These results suggest that communicating policy intentions improve the market participants’ understanding of the central bank’s reaction pattern.

The paper also reviews two typical concerns regarding conventional forward guidance. The first one is voiced by Goodhart (2001) and Mishkin (2004), who believe that it would be more difficult for a monetary policy committee to agree on a path of future interest rates. Holmsen et al. (2008) see no evidence of the practical relevance of this concern. Additionally, theoretical literature explores
cases where there is a penalty for deviating from the forecasts in the central bank’s loss function. Typically, the finding is that this is suboptimal. Holmsen et al. (2008) underscore that their experience suggests that deviating from the announced interest rate forecast is not perceived as costly to the policy makers at NB, as long as reasons for it are provided.

Mirkov and Natvik (2016) bring this statement to the test and asks whether announced forecasts influence actual policy decisions. They examine the data provided by the RBNZ and NB and find some evidence of reluctance to deviate from previously announced interest rate paths. However, such dependence only extends as far as one quarter ahead, while forecasts older than one quarter have no effect on the current policy rate. These results, at least in the case of NB, are hardly surprising given that the criteria for an appropriate policy path listed by Norges Bank (2005b) explicitly state that the interest rate should normally be changed gradually.

The most recent evaluation of experience with conventional forward guidance are provided by Brubakk et al. (2017) for both Norway and Sweden. They find that Norway’s experience suggests that surprise revisions in the policy path significantly affect the yield curve, even as far as 10 years ahead. These effects are persistent. However, they note that central banks’ ability to shape the yield curve is not limited to conventional forward guidance, and information beyond what is revealed through the published interest rate path can have a significant impact. Additionally, the paper emphasises that the market understands that the guidance is issued conditionally based on the knowledge available and on the views of the committee, while the market’s own assumptions may differ from those of the central bank.

The bottom line is that NB adopted conventional forward guidance based on the understanding that it will provide a clearer communication strategy and appreciation of sizeable and persistent mismatches between the central bank’s interest rate and the market’s interest rate expectations. According to Holmsen et al. (2008), publication of the interest rate path allows market participants to gain information about the sign of the future interest rate decisions, but may have less information about the size. Evaluation of the accumulated experience suggests that the guidance issued by the central bank is able to shape most of the yield curve, but the private sector continues to form independent expectations regarding future economic developments, which preserves a diversity of outlooks.

3.3. Forward guidance research at Sverige Riksbank

An examination of the working papers produced at Sverige Riksbank provides another example of how thinking about modern inflation targeting evolved. In particular, just two decades ago Palmqvist (1998) explored the costs of disclosing
the level of the inflation target, noting that the signal is perceived as costly to the central bank because the central bank commits to bringing the inflation rate in line with the announcement.

The earliest working paper focused on interest rate expectations is that of Söderlind and Svensson (1997), which provides a methodological overview of the literature. The first original contribution, made by Faust and Svensson (2001), explores the implications of varying degrees of transparency of the monetary policy for the policy’s implementation and for social welfare. The study links transparency with the level of effort that is required of the public in order to deduce central bank intentions from observables. The authors conclude that high transparency induces the bank to follow a policy closer to the socially optimal one. Taken to the extreme, complete transparency is generally socially beneficial, but frequently not in the interest of the central bank. The authors are able to show that high transparency improves the public’s inferences about the central bank’s goals, and reduces errors in inflation expectations.

Ellingsen and Söderström (2001) explore the effects of unexpected interest rate changes on the yield curve. They find that the response of the market depends on the causes of the monetary policy shock: 1) if the reason is the better informed call by the central bank of the higher future inflation, then the yield curve shifts higher; 2) if the unexpected increase in the rate is due to the shift in the preferences of the central bank in favour of more stable inflation, then only the shorter-term rates increase, while the longer-term interest rates decline. Ellingsen and Söderström (2004) take this line of inquiry further by exploring the effects of monetary policy surprises on the long-term part of the yield curve. In a standard setting, longer-term yields would not respond to short-term surprises in the setting of key rates by the central bank, but in practice they are found to be sensitive to them. The authors show that this behaviour is consistent with the fact that the central bank possesses private information on the future path of inflation. This can help reproduce the excess longer-term yields’ volatility. We believe that these findings may imply that, unless the central bank communicates with the markets and shares its private information, then even in a situation where monetary policy is credible, longer-term interest rates would remain relatively volatile and, thus, the term premium would remain elevated.

Andersson et al. (2002) look at the impact of the central bank’s communications on the shape of the yield curve. Authors extend the scope of such communications to include speeches and press releases. They conclude that unexpected changes in the repo rate are the most important factors behind the movements of the short end of the yield curve. The authors go further and note that Riksbank’s monetary policy is rather unpredictable in the short run, which might suggest a problem with transparency. Their diagnosis of the cause of this disconnect is somewhat
original: they believe that much of the lack of predictability of the repo rate is due to the mismatch between the frequency of changes optimal policy rules imply.

Söderlind et al. (2003) ask an interesting question: if the central bank is following a rules-based policy and if the markets are good at forecasting fundamentals (GDP, inflation), then by implication they should also be good at forecasting the policy rate. They examine surveys of professional forecasters and financial market data and conclude that forecasts of fundamentals are relatively precise, but monetary policy decisions are less predictable, which was earlier established by Andersson et al. (2002), who found that both survey data and market expectations had a low predictive power. Researchers proceeded to test whether monetary policy is predictable based on theoretical statistical models and concluded that interest rate changes are not predictable according to the VAR model. One plausible reason for such an outcome may be that monetary policy responds to a wider set of factors than those included in the typical formulation of the monetary policy rule. This would mean that even the publication of the central banks model of the economy or its policy rule would be unlikely to align its policy decision with market expectations – thus, another way of communication is needed to fully convey its interest rate outlook.

Berg et al. (2004) revisit the theme of the policy rate predictability based on simple rules and the central bank’s own forecasts. They conclude that on average deviations are not very large, but also highlight that Riksbank occasionally deviates from the simple rules, driven by the factors absent from those rules. They suggest extending the scope of the published data beyond the forecasts of fundamentals and urge policy-makers to reveal the ‘secrets of the temple’ in order to increase the public’s understanding of the monetary policy strategy.

Evidence of the limited predictability of the monetary policy leads Riksbank’s governor (Ingves, 2006) to note that one possible future development could be to follow in the footsteps of central banks such as those of New Zealand and Norway, and actually publish Riksbank’s own views on future interest rates, instead of using market expectations as a basis. Later, the Duty Governor expanded on Sverige Riksbank’s position (Rosenberg, 2007) by stating: ‘I think it is unavoidable that a central bank sometimes surprises analysts when publishing its own forecast for the policy rate. The future is uncertain and different analysts may at times have vastly differing perceptions of economic developments. Publishing a forecast for the repo rate means that the central bank more clearly reports its view of future developments than with earlier more mechanical assumptions of the interest rate. At the same time, it means that the central bank risks being exposed to criticism to a greater extent. Even if this is the case, it is my firm opinion that it is better to go the whole way and publish a forecast for the repo rate. The advantages of an own interest rate path clearly outweigh the disadvantages’.
Riksbank’s monetary policy report for 2007 (Sverige Riksbank, 2007) marked the adoption of conventional forward guidance. The report highlighted the risks of confusing commitment with guidance, and attempted to prevent this confusion by including three alternative interest rate paths implied by different economic outlooks.

Svensson (2009b) explores various approaches to the evaluation of the consistency of monetary policy and finds that, although historical deviations of inflation from the target level could be a valid metric, it requires an exceedingly long time-series. On the other hand publication of the central bank’s policy rate forecasts can reveal monetary policy reaction function in real time. The author provides several case studies for Riksbank and concludes that the market both anticipated the repo-rate path reasonably well, and that expectations after the announcements were reasonably aligned with the new path.

Did the publication of the interest rate paths make verbal communication obsolete? Apel and Grimaldi (2012) explore whether information can be extracted from monetary policy meeting minutes, which can help forecast future decisions. Their results are supportive of this assumption and confirm the market’s awareness of the conditionality of interest rate forecasts. They show that markets are able to extract these conditions from the central bank’s verbal statements.

While central banks can more effectively steer interest rate paths using forward guidance, usually some discrepancy between the market-implied outlook remains. De Graeve and Iversen (2017) explore this discrepancy for the Swedish cash rate and report that monetary policy should not disregard deviations between announced and market forward rates, as they can have real effects. This disparity between the market outlook regarding rates and the central bank’s forecasts may be reasonable. As Alsterlind (2017) shows in his study of the four central banks reviewed in this article, all central banks have overestimated the actual policy rate on average since 2007. The reason behind the bias, as the authors suggest, has been a global decline in core rates.

The latest assessment of the efficiency of the central bank’s control over the interest rate expectations is provided in Ahl (2017). The author concludes that Riksbank has the ability to use the repo-rate path to affect market expectations up to between one and three years, noting that the effect is less than one-to-one and decreasing with the horizon. Another interesting finding is that the impact of a path surprise is larger if it narrows the current gap between the previously announced path and market expectations, while surprises that widen the gap are more readily dismissed by the markets.

The bottom line is that the adoption of forward guidance at Sverige Riksbank has been driven by careful study of the interest rate expectations and an appreciation of the uncertainty that surrounds monetary policy decision-making from the economy’s point of view. In particular, milestone findings by
Söderlind et al. (2003) show that even the private sector’s relatively tight handle on financial developments leaves the monetary policy unpredictable. Post-adoption evaluations show the ability of the guidance to steer longer maturity rates in the direction deemed appropriate from a monetary policy perspective.

Sveriges Riksbank’s Deputy Governor Svensson summarised his assessment of the central bank’s experience stating that not to publish the interest rate forecast would be to hide the most important information (Svensson, 2007).

3.4. Forward guidance research at the Czech National Bank

The earliest paper on market expectations regarding the CNB’s monetary policy interest rate was published by Hlušek (2002). The author uses the interbank interest rate and Forward Rate Agreement (FRA) data to identify probabilities of interest rate decisions. Hlušek then applies the model to five consecutive monetary policy meetings between November 2001 and March 2002. These case studies suggest that the market-favoured outcome coincided with the decision taken by the CNB only in one out of five studied cases. The note concludes with two observations regarding the post-decision revisions of the market-implied policy interest rate expectations: 1) the market does not adjust its monetary policy outlook if the central bank’s decision meets its expectation; 2) a surprising interest rate cut leads to a substantial decrease in the probability of additional cuts in the future. These effects may or may not match the intentions of the central bank.

These results demonstrated a need for a more transparent communication policy. As Kotlán and Navrátil (2005) note, since the launch of the new forecasting methodology in July 2002, the CNB limited itself to verbal comments on the forecast-consistent interest rate trajectory. Their paper reviews the results of Hlušek (2002) and reports that approximately 3/4 of the CNB’s decisions were in line with longer-term money market expectations, but over the sample available to researchers the repo rate expectations show an upward bias.

Fukač (2006) explores the consequences of the heterogeneity in expectations for the monetary policy. In particular, the paper devises a New Keynesian business cycle model in which monetary policy is driven by the central bank’s own forecasts, while the rest of the economy depends on the expectations of the private sector. He concludes that if knowledge and beliefs differ across various segments of the private sector, then to achieve a preferable equilibrium the central bank needs to focus on ‘knowledge homogenization’.

Bulir et al. (2007) explore how and whether verbal communication helps the public to understand the monetary policy outlook based on a panel of six countries over 2000 – 2005. They note that about one-third of all monetary policy decisions did not match expectations. This means that monetary policy decisions did not match prior communication and could not be explained away by inflation surprises.
Driven by this cumulative evidence of uncertainty regarding the monetary policy, the CNB decided to start publishing the interest rate forecasts, as stated in Filáček et al. (2007), noting that this should increase its accountability and credibility and improve the efficiency of its monetary policy. In 2008, the CNB announced its transition to conventional forward guidance (Czech National Bank, 2008).

The CNB’s approach to preventing confusion between guidance and commitment was to publish a chart showing its own historical policy rate forecasts as compared with actual outcomes. The chart was titled ‘The actual interest rate path often differs from the CNB forecasts’. The CNB’s reports also started to provide fan charts emphasizing that its forecast is not a single path, but distributions of outcomes.

Based on the 2008 working paper, Bohm et al. (2012) explore a tangent theme: what determines the ‘favourableness’ of press coverage of the monetary policy decisions between 2000 and 2007. One of their findings is relevant: they identify a high correlation between the indicator of surprising decisions and the indicator of the release of the new forecast. This means that before transition to conventional forward guidance, markets had a limited ability to foresee the revisions of the economic outlook by the central bank and monetary policy implications of such revisions.

Two years after the introduction of conventional forward guidance, Filáček and Saxa (2012), based on the 2010 working paper, explore changes in the diversity of the financial analysts’ forecasts depending on the distance from the publication of the central bank’s forecasts. They find that the standard deviation of the distribution of private forecasts is smallest immediately after the release of the CNB’s forecast, which suggests that such publications are an effective coordinating mechanism. They also explore the impact of interest rate publication on the efficiency of the central bank’s communication and find that publication of the interest rate forecast significantly increases the ability of the central bank to anchor inflation expectations.

Horvath et al. (2011) review another innovation: disclosure of voting records of the monetary policy committee. They find evidence of voting records to be informative about future monetary policy, concluding that this provides a case for publishing the records.

The price puzzle in conventional forward guidance setting is explored in Filáček and Matějů (2014). They report that it is possible that in the presence of information asymmetry, monetary policy surprises by the central bank can have paradoxical effects on inflation expectations: if the central bank raises the key rate by more than the markets expected, then the markets’ expectation of the inflation may increase too. This may happen if markets believe that the central bank is better positioned to understand future inflation. This result is in line with Filáček and Saxa (2012), who suggest that policy rate path publication improves the central bank’s ability to coordinate expectations.
The recent summary of the CNB’s experience with conventional forward guidance is summed up by Clinton et al. (2017). The authors note that the Czech experience shows that the published interest rate path has the potential to shape market expectations, supporting the transmission of the monetary policy. They also emphasise a caveat, stating that the efficiency of such communication might be undermined if the central bank provides alternative scenarios that point strongly in a direction different from a baseline.

The bottom line is that the CNB’s case provides more evidence of the ability of the central bank to promote monetary policy transmission efficiency through publication of the indicative paths of the policy rate. The key mechanisms are the coordination of the private sector’s expectations and the latter’s deeper understanding of the central bank’s decision-making patterns. CNB’s work also adds an important reminder, that monetary policy surprises may induce paradoxical responses, i.e. hawkish surprise might increase inflation expectations, which underscores the importance of expectation management.

4. Notes on concerns regarding conventional forward guidance

Yudaeva (2018) provides a lucid and concise summary of practitioners’ concerns regarding the publication of the policy rate path: “The publication of an interest rate path, even subject to reservations, could, in our view, be perceived exactly as a commitment, thus not only reducing policy flexibility but also shifting the focus of the discussion away from its end goal, inflation, to interim goals”.

We see here two distinct concerns:

1) Confusion: published interest rate paths would be perceived as commitment, not projections, based on the central bank’s assumptions regarding external environment, fiscal policy, etc.;

2) Lower effort: the publication of the interest rate forecasts will reduce the efforts made by the market and public to understand macroeconomic developments independently, as they move towards to an excessive reliance on the CBR’s forecast.

The reviewed literature provides three key conclusions regarding the problem of confusion.

---

4 Wider academic literature expresses concerns regarding other potential challenges related to publication of the interest rate path, i.e. (Gosselin et al., 2008) notes that opacity may be creative and raise welfare if the private sector is able to systematically predict central bank inflation forecast errors. Walsh (2007) notes that ‘optimal’ transparency depends on the relative size of cost and demand shocks, and Baeriswyl and Cornand (2011) finds a trade off between transparency and the weight on output stabilisation. We focus on the policy makers’ perspective in this note and thus emphasise the view in Yudaeva (2018), which provides a clear summary of such a perspective.
First, none of the cases provides evidence that the perceived commitment problem, as it is called by Clinton et al. (2017), has in fact materialised. Why? This leads us to the second conclusion.

Second, each of the central banks studied devised solutions to lower the confusion risk, which proved robust in all cases. They used the following strategies to convey the conditionality and uncertainty of the policy rate path:

1) Publish historical forecasts: in its first report containing conventional forward guidance, the CNB published a chart that superimposed its policy rate path onto historical forecasts, thus clearly showing that there can be discrepancies between the two;

2) Provide alternative scenarios: in its first report containing guidance, the NB provided three alternative economic outlooks with corresponding paths of the policy rate, explicitly showing the conditionality of the latter and providing additional data that would help the public to understand the reaction function of the central bank;

3) Provide confidence intervals: all central banks show confidence intervals around the policy rate forecast, showing that the indicative path is in fact a series of distributions and thus reflecting the uncertainty related to both various shocks and, quite simply, the mechanics of the economy.

Third, confusion between conditional guidance and unconditional commitment is not unique to explicit numerical guidance, but can also occur when the central bank decides to limit communication to general policy comments (see Table 2). In fact, the reviewed literature shows that the predictability of policy is lower for periods when central banks limit communication to verbal statements and increases when they decide to supplement it with explicit policy rate paths.

Table 2. Guidance: general comments vs. explicit path

<table>
<thead>
<tr>
<th>Communication</th>
<th>Can be perceived as commitment?</th>
<th>Susceptibility to misinterpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>General policy comments</td>
<td>Yes</td>
<td>High</td>
</tr>
<tr>
<td>Conventional forward guidance</td>
<td>Yes</td>
<td>Low</td>
</tr>
</tbody>
</table>

Source: authors.

Regarding the risk that the change in the focus of the monetary policy from its strategic goal to its instrument – the policy rate – the reviewed literature suggests that guidance can help coordinate inflation expectations, which is one of the key goals of monetary policy. Yet expectations remain diverse and market participants do hold outlooks regarding economic developments that are different from that of the central bank. Clinton et al. (2017) note that market reaction to macroeconomic news was not reduced by the introduction of published policy rate paths in 2007, controlling for the effect of the ZLB.
The power of the central bank to coordinate expectations is not constant but depends on its performance. For example, studies of the effects of guidance on the shape of the yield curve after the Great Recession found that it has declined, typically linking this observation to persistent undershoots in inflation and, by implication, emerging upward bias in rate guidance. Regarding forecast precision Holmsen et al. (2008) provide a useful comment: it is not possible to make forecasts that prove to be accurate in all respects, but by revealing errors, central banks provide a basis for improving the analysis.

The body of work review suggests three related concerns, which might arise with the publication of the policy rate forecast.

First, if policy makers are themselves overly attached to previously published forecasts, then guidance can result in excessive interest smoothing. The reviewed literature finds some evidence of interest rate smoothing, but such ‘path dependence’ does not exceed one quarter. Policy-makers themselves are demonstrably aware of the need to avoid excessive interest rate smoothing, i.e. Gjedrem (2001) notes that even if needed changes in the interest rates were to surprise the markets, that should not be an obstacle for adjusting the policy stance: ‘[...] a desire for predictability must not precede the demand for an interest rate setting that the central bank deems to be correct. The expectations of other economic agents must not control the setting of interest rates’.

Second, Goodhart (2001) and Mishkin (2004) suggest that it can be difficult for a monetary policy committee to agree on a whole path of future interest rates. This criticism might have been more salient in an era when central banks used to condition their published forecast on the assumption of the unchanged policy rate or market implied path, but is less applicable today, when the interest rate is endogenous and is agreed upon at each decision round, even if not published. Svensson (2009a) suggests an iterative procedure to coordinate the views of the members of the policy committee, but says that, in practice, the need for this has never arisen.

Lastly, recent papers (e.g. Filaček and Matějů, 2014) have argued that there might be risks associated with some unintended effects of signalling, i.e. if the market believes that the central bank is better informed about future inflation, then surprise tightening of the monetary policy may in fact increase inflation expectations, as the market corrects its inflation outlook. We believe that, while this mechanism is plausible, there is scarce evidence that the costs of the increase in inflation expectations offset the benefits of the tighter control of the yield curve.

5. Conclusion

The primary question that this review attempts to address is whether the central banks that opted to publish their policy rate forecasts experienced the drawbacks noted in Yudaeva (2018): confusion with a commitment to the
published path, and a reduction in the private sector’s effort to develop an economic outlook. We find no evidence supporting these concerns. Instead, the literature provides several robust approaches to diminishing these risks: publication of the historical forecasts, producing alternative scenarios, and wrapping the baseline in confidence bands.

Additionally, the reviewed body of works by the four central banks which have published policy rate forecasts for the last 10 to 20 years suggests that conventional forward guidance may increase the predictability of monetary policy and the efficiency of the transmission mechanism, accelerate the learning of the central bank’s reaction function by the private sector, and enhance the ability of the markets to understand policy implications of new data in real time, thus reducing interest rate volatility.

However, the literature suggests that the efficiency of the guidance may vary (i.e. decline if the central banks’ forecasts develop a bias), warning against excessive interest rate smoothing, and emphasising that policy rate path publication does not supplant verbal guidance, which remains relevant and closely watched.

6. References


The Bank of Russia and the Russian Journal of Money and Finance invite students and doctoral researchers of Russian universities to submit papers for our First Economic Research Competition.

Areas of interest for this competition include:

- Inflation
- Financial stability
- Banking
- Data analysis: methodology and results

Key requirements for the research:

→ Applied research.
  
  With either a direct result (forecast, macroeconomic analysis model) or an indirect result (testing of economic hypotheses that are important for central bank policy).

→ Meeting academic criteria.
  
  The research should include: clear objectives; a literature review and description of the study’s original contribution; a description of the mathematic or econometric model used and the decision quality criteria for the objective set by the author; a description of the statistical data used; a description of the results and a robustness check; and applied conclusions.

→ The study should have been carried out not earlier than 2016.

→ When carrying out the research, the author should have possessed a Bachelor’s or Master’s degree, or else been working towards a PhD.

How to apply:

Submit by email to dip1@cbr.ru:

1. The entire study in Russian or English, including in the email subject field the phrase: ‘Application for 2019 Bank of Russia Research Competition’.

Give your full name, course, and educational institution.

Papers are accepted in PDF format (file size not more than 5 MB).

2. A cover letter, containing the research objective, the contribution made by the authors (the originality of the results when compared against specific existing studies), results, and conclusions.

The cover letter should also be in PDF format and no longer than one page.

Competition dates:

The closing date is 1 March 2019.

The winners will be announced on 1 June 2019.

Competition jury:


Researchers from the Research and Forecasting Department, the Monetary Policy Department, and the Financial Stability Department of the Bank of Russia.

Publication of research papers:

The Russian Journal of Money and Finance will consider the winners’ papers for publication.

The competition winners will have the chance to participate in the Bank of Russia Macroeconomics Summer School, and the Bank of Russia International Conference in July, 2–3, 2019 in St. Petersburg (travel and accommodation costs covered).

For more detailed information, please see www.cbr.ru/eng/ec_research.
The Bank of Russia International Research Conference
Macroprudential Policy Effectiveness: Theory and Practice

July 3, 2019, St. Petersburg, Russia

The Bank of Russia invites researchers from academia, central banks, and policy institutions to attend our International Research Conference. The participants are welcome to present and discuss the results of their original research on the financial stability focusing on the effectiveness of macroprudential policy and on the interaction between monetary and macroprudential policy.

Submission topics

We invite submissions to the Bank of Russia conference on relevant topics, in particular (but not limited to) the following:

• Systemic risk, contagion, and financial stability.
• The global financial cycle, capital flows, and international spillovers of monetary policy with implications for macroprudential policy in small open economies.
• The exchange rate and its ambiguous role in macroeconomic stabilisation; invoicing currencies and dollarisation seen through the lens of financial and price stability.
• Global commodity cycles and their implications for financial stability.
• Macroprudential policy design: objectives, tools, horizons, communication, and implementation.
• Cross-sectional allocation of finance and its implications for financial stability; the housing sector, financial frictions, and macroprudential regulation; broad vs sectoral macroprudential measures.
• Unintended consequences of macroprudential policy; macroprudential policy spillovers.
• The interaction between monetary and macroprudential policy.
• Fiscal policy implications for financial stability.
• New data and tools in financial stability analysis: general equilibrium models, agent-based models, big data, and machine learning applications.

To submit a draft paper, please send it to board@cbr.ru, putting “Bank of Russia Conference 2019 Submission” in the subject line.

The closing date for submissions is March 1, 2019. Authors whose papers have been accepted will be notified by April 1, 2019.

The organisers will reimburse travel and accommodation costs for presenters and discussants taking part in the main and supplementary sessions.

Publication of Conference Proceedings

Conference participants will be invited to publish their papers in the Bank of Russia’s Russian Journal of Money and Finance, an open access peer-reviewed academic journal.

For more detailed information about the Conference, please see the Bank of Russia’s website: www.cbr.ru/eng/ec_research
20th APRIL INTERNATIONAL ACADEMIC CONFERENCE ON ECONOMIC AND SOCIAL DEVELOPMENT

On April 9-12, 2019 in Moscow, National Research University Higher School of Economics (HSE) with the support of the World Bank will be holding the 20th April International Academic Conference on Economic and Social Development.

The Conference’s Programme Committee is chaired by Professor Evgeny Yasin, HSE’s Academic Supervisor.

The Conference features a diverse agenda concerning social and economic development in Russia

The main thematic areas of the Conference:

- Arctic Research (special session)
- Demography and Labour Markets
- Economic Methodology
- Education
- Financial Institutions, Markets and Payment Systems
- Firms and Markets
- Healthcare Studies
- Instrumental Methods in Economic and Social Studies
- International Relations
- Macroeconomics and Growth
- Management
- Network Analysis
- Political Processes
- Regional and Urban Development
- Science and Innovation
- Social and Cultural Processes
- Social and Economic History
- Social Policy
- Sociology
- Sports Studies
- State and Local Government, and NPO sector
- Theoretical Economics
- World Economy

The Conference’s programme will also include special events such as roundtable discussions, PhD seminars and presentations of honorary papers.

The Conference’s working languages are Russian and English.

Online registration to attend the Conference will be open until March 25, 2019 at https://conf.hse.ru/en/2019/.

Should you have further questions feel free to contact the Organizing Committee at interconf@hse.ru.
CONTENTS

Inflation and Population Age Structure: The Case of Emerging Economies
Darya Antonova, Yulia Vymyatnina

Text Mining-based Economic Activity Estimation
Ksenia Yakovleva

Inflation Forecasting Using Machine Learning Methods
Ivan Baybuza

Review of the Bank of Russia – IMF Workshop

Recent Developments in Macroprudential Stress Testing
Elizaveta Danilova, Evgeny Rumyantsev, Ivan Shevchuk

Fear of Forward Guidance
Alex Isakov, Petr Grishin, Oleg Gorlinsky