# 1 Introduction

Accurate macroeconomic forecasts are extremely important for policy making. Central banks and government bodies monitor a large set of macroeconomic indicators to determine the policy (Beckner (1996), Bernanke and Boivin (2003)). Therefore, a model used for forecasting must be suitable for data-rich samples because large models might outperform low-dimensional ones by taking into account more potentially relevant information. This explains the recent resurgence in interest from academics, central bankers and private sector experts for macroeconomic forecasting in a data-rich environment.

In this paper, we forecast Russian macroeconomic indicators with Bayesian vector autoregressions (BVARs) of different sizes. Our goal is twofold. First, we compare the forecast accuracy of BVAR with that of unrestricted vector autoregressions (VARs) and random walk with drift models for 23 important macroeconomic indicators. Second, we question whether a high-dimensional model always outperforms a low-dimensional one in terms of forecasting accuracy.

For the last 30 years, VARs introduced by Sims (1980) have become a widely-used tool for forecasting. However, unrestricted VARs bear the risk of over-parametrization even for samples of moderate size. This risk stems from the fact that the number of parameters to be estimated increases nonlinearly with the number of equations. For this reason, in economic applications unrestricted VARs usually contain only up to eight variables, and this may potentially lead to the loss of some relevant information and undermine the forecast accuracy.

To deal with a data-rich environment researchers modify VARs and impose restrictions on the covariance structure. One strand of the literature focuses on dynamic factor models (DFM, Forni et al. (2000) and Stock and Watson (2002))and Panel VARs and Global VARs (PVARs, GVARs, Pesaran, Schuermann, and Weiner (2004) and Dees and Güntner (2014)). DFM are based on the idea that a relatively small set of indices extracted from a high-dimensional set of variables can summarize the information from this set. These factors are treated as variables in a VAR model either separately or in conjunction with several time series from the original information set in a factor-augmented VAR (FAVAR) model. For data sets with a panel structure a suitable choice is a PVAR or a GVAR with shrinkage done by exclusion, exogeneity or homogeneity restrictions.

Another method of shrinkage is the Bayesian one and we follow this approach. The shrinkage is done by imposing restrictions on the parameters in the form of prior distributions. While BVARs in a low-dimensional space were widely used for macroeconomic analysis, their use for data-rich environments was limited until recently. The reason was a general agreement that Bayesian shrinkage is insufficient to solve the over-parametrization problem in high cross-sectional dimension samples.

However, in their influential paper, De Mol, Giannone, and Reichlin (2008) show that Bayesian methods can be successfully applied to a data-rich environment if the degree of shrinkage is set relative to the cross-sectional dimension of the sample. Bańbura, Giannone, and Reichlin (2010) confirm and develop this assertion for BVARs applied to a large set of US time-series. Their main result is that high-dimensional models have better forecasting performance than small-dimensional models and even FAVARs. They also show that accurate forecasts can be already obtained using a medium-sized BVAR (20 variables in their case).

Several authors have recently shown that, in terms of forecasting accuracy, medium and large BVARs outperform their low-dimensional counterparts. For example, Beauchemin and Zaman (2011) present a medium BVAR with a good forecasting performance applied to the US data. Bloor and Matheson (2010) compare univariate autoregresions (ARs), unrestricted VARs and BVARs and show evidence that high-dimensional BVARs, in general demonstrate

better forecasting performance. Koop (2013) demonstrates that high-dimensional BVARs outperform factor models in terms of forecasting performance. Moreover, he argues that more complicated priors than those that are usually applied may not lead to more precise forecasts. Alessandri and Mumtaz (2014) underline the importance of financial factors for an accurate forecast of output and inflation, especially «for predicting «tail» macroeconomic outcomes». Carriero, Clark, and Marcellino (2015) study some characteristics of BVARs and find those providing the most accurate forecasts.

Our analysis delivers two important results. First, we show that most Russian macroeconomic indicators in our sample can be forecast by BVARs more accurately than by competing models. However, contrary to other studies (for example, Bloor and Matheson (2010), Bańbura, Giannone, and Reichlin (2010)) we do not confirm that relative forecast error monotonically decreases with the dimension of the sample. In almost half of those cases where a BVAR is the most accurate model, a small-dimensional BVAR outperforms its high-dimensional counterpart.

The paper is structured as follows. Section 2 presents our model and the prior distribution we apply. In Section 3 we describe our sample and the data transformations we use. Section 4 contains the results and their interpretation. Section 5 concludes.

# 2 Model

## 2.1 BVAR

Let $y_{it}$ be variables[1] stacked in a $m \times 1$ vector $y_t = (y_{1t}, y_{2t}, \ldots, y_{mt})'$. The reduced form VAR can be written as:

$$y_t = \Phi_{const} + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma) \tag{1}$$

where $\Phi_{const} = (c_1, \ldots, c_m)'$ is a $m \times 1$ vector of constants, $\Phi_l$ are autoregression $m \times m$ - dimensional matrices where $l = 1, \ldots, p$. Vector $\varepsilon_t$ is a $m$-dimensional vector of errors with covariance matrix $\mathbb{E}\,\varepsilon_t \varepsilon_t' = \Sigma$, and is uncorrelated with regressors. By grouping parameter matrices into one matrix $\Phi = [\Phi_1 \ldots \Phi_p\ \Phi_{const}]'$ and defining new vector $x_t = [y_{t-1}'\ \ldots\ y_{t-p}'\ 1]'$, the equation (1) can be written in a more compact form:

$$y_t = \Phi' x_t + \varepsilon_t \tag{2}$$

If the variables and shocks are grouped in the following way: $Y = [y_1, y_2, \ldots, y_T]'$, $X = [x_1, x_2, \ldots, x_T]'$, $E = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T]'$, the VAR can be written as:

$$Y = X\Phi + E \tag{3}$$

The Bayesian estimate combines a likelihood function $L(Y|\Phi, \Sigma)$ with a prior distribution $p(\Phi, \Sigma)$ and results in a posterior distribution of parameters $p(\Phi, \Sigma|Y)$:

$$p(\Phi, \Sigma|Y) \propto p(\Phi, \Sigma) L(Y|\Phi, \Sigma) \tag{4}$$

## 2.2 Conjugate normal — inverse Wishart prior

Our benchmark model for estimation and forecasting purposes is a BVAR with a conjugate normal — inverse Wishart prior. The prior can be written as:

[1]For the convenience of the reader, all the notations are also shown in Appendix 1.

$$\begin{cases} \Sigma \sim \mathcal{IW}(\underline{S}, \underline{\nu}) \\ \Phi|\Sigma \sim \mathcal{N}(\underline{\Phi}, \Sigma \otimes \underline{\Omega}) \end{cases} \tag{5}$$

The prior mean of the coefficient matrices is written with a $k \times m$ matrix $\underline{\Phi} = \mathbb{E}(\Phi)$, where $\underline{\Phi} = [\underline{\Phi}_1 \dots \underline{\Phi}_p \; \underline{\Phi}_{const}]'$.The matrices $\underline{\Phi}_l$ are defined as follows:

$$(\underline{\Phi}_l)_{ij} = \begin{cases} \delta_i \; i = j, l = 1; \\ 0, \quad \text{otherwise} \end{cases} \tag{6}$$

A matrix $\underline{\Omega}$ is assumed to be diagonal and it depends on several hyperparameters:

$$\underline{\Omega} = \text{diag}\{\underline{\Omega}_{lag=1}, \dots, \underline{\Omega}_{lag=p}, \underline{\Omega}_{const}\} \tag{7}$$

$$(\underline{\Omega}_{lag=l})_{jj} = \left( \frac{\lambda}{l^{\lambda_{lag}} \hat{\sigma}_j} \right)^2 \quad \underline{\Omega}_{const} = \lambda_{const}^2 \tag{8}$$

The hyperparameters have the following interpretation: $\lambda$ determines the overall tightness of the prior and it is responsible for the relative weight of the prior with respect to the information incorporated in the data, $\lambda_{lag}$ controls the velocity of the decrease of the prior variance with increasing the lag length, and $\lambda_{const}$ governs the relative tightness of the prior for the constant terms.

The scale matrix $\underline{S}$ is diagonal and its non-zero elements assure that the mean of $\Sigma$ is equal to the fixed covariance matrix of the standard Minnesota prior:

$$(\underline{S})_{ii} = (\underline{\nu} - m - 1)\hat{\sigma}_i^2 \tag{9}$$

The scale parameter $\sigma_i^2$ is usually set to be equal to the variance estimate of residuals in a univariate $AR$ model. The choice of degrees of freedom of inverse Wishart distribution $\underline{\nu}$ greater than or equal to than $\max\{m+2, m+2h-T\}$ guarantees the existence of the prior variance of the regression parameters and the posterior variances of the forecasts at horizon $h$ (Kadiyala and Karlsson (1997)).

It is possible to show that the posterior distribution formed by combining this prior distribution with a likelihood function is also normal — inverse Wishart (see, for example, Zellner (1996)):

$$\begin{cases} \Sigma|Y \sim \mathcal{IW}(\overline{S}, \overline{\nu}) \\ \Phi|\Sigma, Y \sim \mathcal{N}(\overline{\Phi}, \Sigma \otimes \overline{\Omega}) \end{cases} \tag{10}$$

with the following parameters:

$$\begin{aligned} \overline{\nu} &= \underline{\nu} + T \\ \overline{\Omega} &= (\underline{\Omega}^{-1} + X'X)^{-1} \\ \overline{\Phi} &= \overline{\Omega} \cdot (\underline{\Omega}^{-1}\underline{\Phi} + X'Y) \\ \overline{S} &= \underline{S} + \hat{E}'\hat{E} + \hat{\Phi}'X'X\hat{\Phi} \\ &\quad + \underline{\Phi}'\underline{\Omega}^{-1}\underline{\Phi} - \overline{\Phi}'\overline{\Omega}^{-1}\overline{\Phi} \\ \hat{\Phi} &= (X'X)^{-1}X'Y \\ \hat{E} &= Y - X\hat{\Phi} \end{aligned}$$

There is a popular alternative approach to calculate hyperparameters of the posterior distribution. We set $\underline{S}$ and $\underline{\Omega}^{-1}$ to be zero matrices and to compensate the difference we add supplementary observations into $X$ and $Y$ matrices according to:

$$Y^* = \begin{bmatrix} Y^{NIW} \\ Y \end{bmatrix} \quad X^* = \begin{bmatrix} X^{NIW} \\ X \end{bmatrix}, \tag{11}$$

where matrices $Y^{NIW}$ and $X^{NIW}$ are defined as follows[2]:

$$Y^{NIW} = \begin{bmatrix} \frac{\text{diag}(\delta_1\sigma_1,\ldots,\delta_m\sigma_m)}{\lambda} \\ 0_{m(p-1)\times m} \\ \text{diag}(\sigma_1,\ldots,\sigma_m) \\ 0_{1\times m} \end{bmatrix} \quad X^{NIW} = \begin{bmatrix} \frac{\text{diag}(1,2^{\lambda_{lag}},\ldots,p^{\lambda_{lag}})\otimes\text{diag}(\sigma_1,\ldots,\sigma_m)}{\lambda} & 0_{mp\times 1} \\ 0_{m\times mp} & 0_{m\times 1} \\ 0_{1\times mp} & \frac{1}{\lambda_{const}} \end{bmatrix} \tag{12}$$

This method permits the calculation $\overline{\Phi}$ as an OLS estimate of the regression of $Y^*$ on $X^*$: $\overline{\Phi} = (X^{*\prime}X^*)^{-1}X^{*\prime}Y^*$ and $\overline{S}$ as a sum of the squared residuals for this regression: $\overline{S} = (Y^* - \overline{\Phi}X^*)'(Y^* - \overline{\Phi}X^*)$.

## 2.3 Prior modifications

Doan, Litterman, and Sims (1984) and Sims (1993) propose complementing this prior distribution with additional information in form of two other priors. This modification reflects the belief that time series may have unit roots and cointegration relations. These elements in the prior allow avoiding an unreasonably large share of the variation in the data which is accounted for by deterministic components (Sims (1993)).

A sum-of-coefficients prior was introduced by Doan, Litterman, and Sims (1984). If all the time-series in a sample have a unit root, this information can be taken into account with a prior where a sum of all the lag parameters for each dependent variable is equal to one (Robertson and Tallman (1999), Blake and Mumtaz (2012)). In other words, when the mean of the lagged values of a variable is at a certain level, this level is a good forecast for future observations of this dependent variable. We implement this prior by combining the dataset given in 11 with artificial dummy-observations according to the following scheme:

$$Y^{SC} = \frac{1}{\lambda_{sc}}\left[\text{diag}(\delta_1\mu_1,\ldots,\delta_m\mu_m)\right] \tag{13}$$

$$X^{SC} = \frac{1}{\lambda_{sc}}\left[(1_{1\times p})\otimes\text{diag}(\delta_1\mu_1,\ldots,\delta_m\mu_m) \quad 0_{m\times 1}\right], \tag{14}$$

where $(1_{1\times p})$ is a unitary $[1\times p]$ vector, $\mu_i$ is $i$-th component of vector $\mu$, which contains the average values of initial observations of all variables in the sample[3]: $\mu = \frac{1}{p}\sum_{t=1}^{p} y_t$.

The dummy initial observation prior proposed by Sims (1993) expresses the belief that the variables have a common stochastic trend. Only one observation is added so that the values of all variables are equal to the average value of initial observations $\mu_i$ normalized to a scale coefficient $\lambda_{io}$. Therefore, this extra observation is defined as follows:

$$Y^{IO} = \frac{1}{\lambda_{io}}\left[\delta_1\mu_1,\ldots,\delta_m\mu_m\right] \tag{15}$$

$$X^{IO} = \frac{1}{\lambda_{io}}\left[(1_{1\times p})\otimes(\delta_1\mu_1,\ldots,\delta_m\mu_m) \quad 1\right], \tag{16}$$

[2]The similar formulae provided in Bańbura, Giannone, and Reichlin (2010), Berg and Henzel (2013) can be regarded as special cases of (12) for $\lambda_{lag} = 1$ и $\lambda_{const} \to \infty$.

[3]Some authors calculate $\mu$ using the average values of all observations in a sample, so that $\mu = \frac{1}{T}\sum_{t=1}^{T} y_t$ (Bańbura, Giannone, and Reichlin (2010) and Carriero, Clark, and Marcellino (2015)). However, following Sims and Zha (1998) we calculate $\mu$ using only initial $p$ observations.

This prior distribution reflects the belief that the average value for a variable is a linear combination of average values of all the other variables.

The hyperparameter $\lambda_{io}$ controls the tightness of this prior. When $\lambda_{io} \to 0$, the model implies that either all variables are stationary with the mean equal to sample mean of the initial observations or non-stationary without drift and cointegrated.

## 2.4 Choice of tightness hyperparameter: the algorithm of shrinkage

As shown by De Mol, Giannone, and Reichlin (2008) and confirmed in several other recent studies, a sample with a larger cross-sectional dimension requires a lower $\lambda$, so the prior must be tighter for a larger sample than for a smaller one. In this paper, we use the approach introduced by Bańbura, Giannone, and Reichlin (2010) to determine the optimal $\lambda$ for every model.

This algorithm is based on the idea that the shrinkage should be sufficiently tight to avoid over-parametrization. Moreover, it is assumed that a three-variable unrestricted VAR is parsimonious enough and does not require any additional shrinkage. This implies that the hyperparameter $\lambda$ can be chosen so that the model has the same in-sample fit as a three-variable VAR. In other words, a BVAR model of any dimension is shrunk to the size of a small unrestricted VAR. A detailed description of the procedure is laid out below. We denote the actual value of a variable $var$ at moment $T+h$ by $y_{var,T+h}$, and a forecast of the variable $var$ at moment $T$ for a horizon $h$ in a model with $m$ variables and an overall tightness parameter $\lambda$ by $y^{\lambda,m}_{var,T+h|T}$. The algorithm for choosing $\lambda$ has the following steps.

1. We make in-sample one-period forecasts with BVAR on a training sample and calculate the mean squared forecast error for the set of $M$ variables of central interest[4]:

$$MSFE^{\lambda,m}_{var,1} = \frac{1}{T_0 - p} \sum_{t=p}^{T_0-1} \left( y^{\lambda,m}_{var,t+1|t} - y_{var,t+1} \right)^2, \tag{17}$$

   where the BVAR coefficients are obtained using the training sample: $t = p+1, \ldots, T_0$ and $T_0$ is the last observation of the training sample: $T_0 = p + 120$.

2. In a similar way we calculate one-period forecasts according to the random walk model[5] for the same variables $\left(MSFE^0_{var,1}\right)$ and a new indicator $FIT^{\lambda,m}$:

$$FIT^{\lambda,m} = \frac{1}{M} \sum_{var \in \mathcal{M}} \frac{MSFE^{\lambda,m}_{var,1}}{MSFE^0_{var,1}} \tag{18}$$

3. We estimate VARs for the same set of $M$ variables of interest[6] and calculate MSFEs and an indicator $FIT^{\infty,M}$:

$$FIT^{\infty,M} = \frac{1}{M} \sum_{var \in \mathcal{M}} \frac{MSFE^{\infty,M}_{var,1}}{MSFE^0_{var,1}} \tag{19}$$

[4]Our benchmark set of variables of central interest ($\mathcal{M}$) includes the industrial production index, consumer price index and interbank interest rate so that $M = 3$. As a robustness check we excluded the interest rate from this set and there was almost no change in the vector of the optimal $\lambda$.

[5]We normalize MSFE for the BVAR and VAR models by MSFE obtained with the random walk model to take into account the different scales of the series. We use a superscript 0 for the random walk model as random walk may be considered as a special case of BVAR if $\lambda = 0$ and $\delta_i = 1, i = 1, \ldots, k$.

[6]We denote all results from VAR by a superscript $\infty$ as unrestricted VAR is a special case of BVAR with $\lambda \to \infty$. In this case the posterior coincides with the likelihood function.

4. The optimal lambda is the value minimizing the difference between $FIT^{\lambda,m}$ and $FIT^{\infty,M}$:

$$\lambda_m^* = \arg\min_{\lambda} |FIT^{\lambda,m} - FIT^{\infty,M}| \tag{20}$$

After the optimal $\lambda$ is chosen for every $m$, we keep it fixed and make out-of-sample forecasts on the evaluation sample.

### 2.5 Out-of-sample forecasting

We estimate BVARs with the optimal $\lambda$ on «rolling window» containing 120 observations, starting from observation $p+1$ and continuing until March 2015. The first $p$ observations are used as a pre-sample and the subsample $[p+1, p+120]$ is a training sample to determine the optimal $\lambda$ on a grid. We denote the last available observation as $T_1$, and the last observation of each evaluation subsample as $\tau$. The number of forecasts is equal to $T_1 - T_0 - h + 1$ where $h$ is the forecasting horizon ($h = 1, 3, 6, 9, 12$). Therefore, the number of one-period forecasts is greater than the number of three-period forecasts by two, etc.[7] For every model $m$ and forecasting horizon $h$ we calculate the out-of-sample MSFE for all $m$ variables included in the model:

$$OMSFE_{var,h}^{\lambda,m} = \frac{1}{T_1 - T_0 - h + 1} \sum_{\tau=T_0}^{T_1-h} \left( y_{var,\tau+h|\tau}^{\lambda,m} - y_{var,\tau+h} \right)^2, \tag{21}$$

Then we calculate the MSFE of out-of-sample forecasts obtained with random walk with drift ($OMSFE_{var,h}^{0}$) and unrestricted VAR ($OMSFE_{var,h}^{\infty,m}$):

$$OMSFE_{var,h}^{\infty,m} = \frac{1}{T_1 - T_0 - h + 1} \sum_{\tau=T_0}^{T_1-h} \left( y_{var,\tau+h|\tau}^{\infty,m} - y_{var,\tau+h} \right)^2 \tag{22}$$

$$OMSFE_{var,h}^{0} = \frac{1}{T_1 - T_0 - h + 1} \sum_{\tau=T_0}^{T_1-h} \left( y_{var,\tau+h|\tau}^{0} - y_{var,\tau+h} \right)^2, \tag{23}$$

To compare the forecast accuracy of different models we report the relative MSFE, that is, the ratio of the MSFE of the model in question by the MSFE of a reference model (random walk in our case):

$$RMSFE_{var} = \frac{OMSFE_{var,h}^{\lambda,m}}{OMSFE_{var,h}^{0}} \tag{24}$$

## 3 Data and Estimations

Our dataset consists of 23 time series running from January 1996 to April 2015. Our sample containing 232 observations is limited by data availability. The full list of the series and their sources is displayed in Appendix 2. We seasonally adjust data which demonstrate seasonal fluctuations with TRAMO/SEATS option in EViews and apply logarithms to the series, with the exception of those already expressed in rates.

We estimate models of different cross-sectional dimension. The industrial production index, CPI and interbank interest rate are forecast with three-variable, six-variable and 23-variable models. Monetary aggregate, the real effective exchange rate and the oil price index

[7] An alternative method is to calculate an equal number of forecasts for each horizon $h$, starting from $T_0 + 12$. However this implies the loss of some information about the forecasts and we do not proceed with this method here.

Table 1: List of models and variable sets

| | |
|---|---|
| VAR3/BVAR3 | $Y = \{IP, CPI, R\}$ |
| VAR4/BVAR4 | $Y = \{IP, CPI, R, Z\}$ |
| VAR6/BVAR6 | $Y = \{IP, CPI, R, M2, REER, OPI\}$ |
| VAR7/BVAR7 | $Y = \{IP, CPI, R, M2, REER, OPI, W\}$ |
| BVAR23 | $Y$ includes all 23 variables from the dataset |

where $IP$ is the industrial production index, $CPI$ is the consumer price index, $R$ is the nominal interbank rate, $M2$ is the monetary aggregate M2, $REER$ is the real effective exchange rate, $OPI$ is the Brent oil price index. $Z$ is any variable from the dataset besides $IP$, $CPI$ and $R$. $W$ is any variable from the dataset besides $IP$, $CPI$, $R$, $M2$,$REER$, and $OPI$.

are forecast with four-variable, six-variable and 23-variable models. All the other series are forecast with four-variable, seven-variable and 23-variable models. For all models with dimension less than eight we estimate both unrestricted VARs and BVARs. We estimate only a BVAR on the sample with 23 variables. The three-variable VAR is the simplest specification that can be justified by a textbook version of a New Keynesian model. A model with six variables is specified in line with many monetary models used previously for the structural analysis of different economies (Sims (1992), Kim and Roubini (2000), Bjørnland (2008), Scholl and Uhlig (2008)) and it contains the real effective exchange rate, monetary aggregate M2, and the oil price index in addition to three variables included in the smallest VAR. The oil price index is used as a variable in the model to reflect the belief that oil price index is an important explanatory factor for the other variables in the sample as Russia has a petroleum export-based economy. To forecast variables outside of these core sets we estimate four-variable and seven-variable VARs containing three or six-variable samples described above plus an additional variable of interest. We include all available time series in our 23-variable model. In totally, after the optimal $\lambda$ is chosen, we estimate 79 models. In a compact form the models used for forecasting are presented in Table 1. For VARs and BVARs we take all possible lags from 1 to 12.

# 4 Results

For every variable and every forecasting horizon we find a model with the lowest RMSFE. We compare 60 specifications for each variable and each forecasting horizon as we have 5 models (a VAR and a BVAR with 3 or 4 variables, a VAR and a BVAR with 6 or 7 variables, and a BVAR with 23 variables) and 12 lags for each of them. We visualize our results with color tables (Figures 1-2 ). The two tables in these Figures differ by the hyperparameter sets used for the BVAR priors. For models depicted in Figure 1 we take $\delta_i = 1$ for nonstationary series and $\delta_i = 0.5$ for stationary series while constructing the prior. We use the KPSS test to split the series into two groups. The parameters $\sigma_i$ are taken to be equal to the standard deviations of the residuals in the univariate $AR(p)$ model. This hyperparapeter set will be referenced as set A in what follows. Figure 2 is related to BVARs with the prior determined by the univariate $AR(1)$ model. We take $\delta_i$ as equal to the OLS estimates of the first lag parameter and $\sigma_i$ as equal to the standard deviations of residuals in AR(1) models. This set will be referenced as set B.

| | h=1 | h=3 | h=6 | h=9 | h=12 |
|---|---|---|---|---|---|
| ind product | 0.92 | | 0.96 | 0.82 | 0.7 |
| cpi | 0.44 | 0.38 | 0.46 | 0.38 | 0.33 |
| interb rate | | | 0.66 | 0.52 | 0.58 |
| agriculture | 0.93 | 0.82 | 0.7 | 0.67 | 0.57 |
| construction | 0.97 | | | | |
| employment | 0.67 | 0.42 | 0.43 | 0.6 | 0.72 |
| export | 0.59 | 0.61 | 0.76 | 0.81 | 0.89 |
| gas price | 0.73 | 0.43 | 0.22 | 0.29 | 0.5 |
| gov balance | 0.61 | 0.79 | 0.77 | 0.7 | 0.63 |
| import | 0.75 | 0.48 | 0.52 | 0.72 | 0.98 |
| labor request | 0.66 | 0.79 | 0.94 | 0.95 | 0.96 |
| lend rate | 0.94 | 0.84 | 0.77 | 0.77 | 0.8 |
| M2 | 0.53 | 0.51 | 0.71 | 0.95 | |
| nominal ER | | | | | |
| NFA of CB | 0.6 | 0.56 | 0.75 | 0.65 | 0.6 |
| oil price | 0.88 | 0.81 | 0.88 | 0.81 | 0.77 |
| ppi | 0.43 | 0.75 | 0.69 | 0.59 | 0.49 |
| real income | 0.87 | 0.84 | 0.83 | 0.71 | 0.73 |
| real invest | 0.78 | 0.61 | 0.73 | 0.88 | 0.91 |
| real ER | 0.72 | 0.68 | 0.6 | | |
| retail | 0.62 | 0.39 | 0.4 | 0.64 | 0.88 |
| unemp rate | 0.93 | 0.83 | 0.9 | 0.92 | 0.94 |
| wage | 0.74 | 0.51 | 0.46 | 0.42 | 0.41 |

Random walk
VAR3/4 - BVAR 3/4
VAR6/7
BVAR6/7
BVAR23

Figure 1: RMSFE of the best forecasting accuracy models, parameter set A: $\sigma_i$ are std of $AR(p)$ residuals, $\delta_i = 1$ for nonstationary series, $\delta_i = 0.5$ for stationary series

| | h=1 | h=3 | h=6 | h=9 | h=12 |
|---|---|---|---|---|---|
| ind product | 0.96 | | 0.96 | 0.82 | 0.7 |
| cpi | 0.38 | 0.37 | 0.46 | 0.36 | 0.27 |
| interb rate | | | 0.91 | 0.56 | 0.56 |
| agriculture | 0.93 | 0.82 | 0.7 | 0.67 | 0.56 |
| construction | 0.97 | | | | |
| employment | 0.7 | 0.54 | 0.59 | 0.7 | 0.81 |
| export | 0.57 | 0.62 | 0.71 | 0.8 | 0.86 |
| gas price | 0.7 | 0.42 | 0.22 | 0.31 | 0.51 |
| gov balance | 0.6 | 0.79 | 0.78 | 0.74 | 0.64 |
| import | 0.74 | 0.63 | 0.82 | 0.88 | 0.97 |
| labor request | 0.66 | 0.79 | 0.94 | 0.94 | 0.95 |
| lend rate | 0.95 | 0.89 | 0.79 | 0.71 | 0.66 |
| M2 | 0.55 | 0.6 | 0.8 | 0.97 | |
| nominal ER | | | | | |
| NFA of CB | 0.6 | 0.62 | 0.69 | 0.61 | 0.61 |
| oil price | 0.85 | 0.81 | 0.85 | 0.79 | 0.75 |
| ppi | 0.43 | 0.74 | 0.69 | 0.6 | 0.49 |
| real income | 0.91 | 0.93 | 0.84 | 0.73 | 0.75 |
| real invest | 0.81 | 0.63 | 0.76 | 0.88 | 0.92 |
| real ER | 0.72 | 0.69 | 0.8 | | |
| retail | 0.62 | 0.4 | 0.45 | 0.72 | 0.88 |
| unemp rate | 0.94 | 0.91 | 0.89 | 0.9 | 0.92 |
| wage | 0.75 | 0.53 | 0.47 | 0.42 | 0.41 |

Random walk
VAR3/4 - BVAR 3/4
VAR6/7
BVAR6/7
BVAR23

Figure 2: RMSFE of the best forecasting accuracy models, parameter set B: $\sigma_i$ are std of $AR(1)$ residuals, $\delta_i$ are first lag $AR(1)$ estimates

The color of a cell corresponds to the model that appears to outperform the others in terms of forecasting accuracy for a given variable and a given forecasting horizon. Most of cells are green (either light green or bright green) reflecting that a BVAR provides the most accurate forecast for the corresponding variables and forecasting horizons. An unrestricted VAR gives the most accurate forecast for variables and horizons indicated by yellow and orange cells. The procedure for choosing $\lambda$ is such that the BVAR and the unrestricted VAR necessarily coincide for the smallest sample (3 or 4 variables). This explains why orange represents both of these models. Blue means that neither BVAR nor VAR beat the random walk in terms of forecast accuracy.

The forecast accuracy is measured with RMSFE calculated according to (24) and is also shown in Figures 1-2. The numbers less than one indicate that the a VAR or a BVAR model provides a better forecast than the random walk and the smaller the number is, the more accurate the forecasts are relative to the random walk. We see that in most cases we have at least one model that provides a forecast much better than the reference model.

Despite different prior parameter sets, the two tables are very similar both in terms of the best forecasting models and the relative accuracy with respect to the random walk.

We interpret our results as follows. First, for many variables and forecasting horizons, BVARs outperforms the random walk and unrestricted VARs. Out of the 115 forecasting cases highlighted in the paper (23 variables times 5 forecasting horizons) BVARs appear to be best in terms of forecast accuracy in 71 cases for the prior hyperparameter set A and in 77 cases for the prior hyperparameter set B. There are variables in our sample that are forecast more accurately by BVARs for all forecasting horizons we try (such as employment, import and lending rate). For several variables BVARs are the best option for the shortest horizons (for example, monetary aggregate M2 and the real effective exchange rate). On the contrary, for the agricultural production index a BVAR model has the lowest forecast error only for a one-year horizon.

Second, among all cases where BVARs show their forecasting accuracy, a high-dimensional BVAR is the best option in about half of the cases (35 of 71 for set A and 39 of 77 for set B). In other cases it is beaten by a low-dimensional BVAR (6 or 7 variables).

Third, for some variables and some forecasting horizons neither unrestricted VARs nor BVARs outperform the random walk. For example, in all specifications we consider the nominal exchange rate cannot be forecast by either VARs or BVARs better than by the random walk, and it is a long-held consensus in economics remounting to Meese and Rogoff (1983). However, we question another wide-spread belief that the price of oil is a random walk process. We show that the oil price index can be forecasted by BVARs much better than by the random walk and the result is robust for different prior settings.

# 5 Conclusion

This paper evaluates the forecasting performance of BVARs on Russian data. We estimate BVARs of different sizes and compare the accuracy of their out-of-sample forecasts with those obtained with unrestricted VARs and random walk. Our sample consists of 23 variables and we forecast at 5 different horizons up to 12 months. We show that for the majority of the variables BVARs outperform the competing models in terms of forecasting accuracy. However, we cannot confirm the conclusion drawn in some other studies (for example, Bloor and Matheson (2010), Bańbura, Giannone, and Reichlin (2010)), where Bayesian methods were applied to data from developed countries, claiming that high-dimensional BVARs forecast better than low-dimensional models. Our results implies that a 23-variable BVAR performs most accurately in only about a half of cases where a BVAR is considered as a better forecasting tool with respect to its competitors. For the rest of those cases a

BVAR with a relatively small size (6 or 7 variables in our case) can outperform a 23-variable BVAR in terms of forecasting accuracy.

# References


Alessandri, Piergiorgio and Haroon Mumtaz (2014). *Financial conditions and density forecasts for US output and inflation*. Tech. rep. Working Paper, School of Economics and Finance, Queen Mary, University of London.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010). "Large Bayesian vector auto regressions". In: *Journal of Applied Econometrics* 25(1), pp. 71–92.

Beauchemin, Kenneth and Saeed Zaman (2011). *A medium scale forecasting model for monetary policy*. Working Paper 1128. Federal Reserve Bank of Cleveland.

Beckner, Steven K. (1996). *Back from the brink: The Greenspan years*. John Wiley & Sons Inc.

Berg, Tim Oliver and Steffen Henzel (2013). *Point and Density Forecasts for the Euro Area Using Many Predictors: Are Large BVARs Really Superior?* Tech. rep. ifo Working Paper.

Bernanke, Ben and Jean Boivin (2003). "Monetary policy in a data-rich environment". In: *Journal of Monetary Economics* 50(3), pp. 525–546.

Bjørnland, Hilde (2008). "Monetary Policy and Exchange Rate Interactions in a Small Open Economy". In: *Scandinavian Journal of Economics* 110(1), pp. 197–221.

Blake, Andrew and Haroon Mumtaz (2012). *Applied Bayesian econometrics for central bankers*. 1st ed. Centre for Central Banking Studies, Bank of England.

Bloor, Chris and Troy Matheson (2010). "Analysing shock transmission in a data-rich environment: a large BVAR for New Zealand". In: *Empirical Economics* 39(2), pp. 537–558.

Carriero, Andrea, Todd E Clark, and Massimiliano Marcellino (2015). "Bayesian VARs: specification choices and forecast accuracy". In: *Journal of Applied Econometrics* 30(1), pp. 46–73.

De Mol, Christine, Domenico Giannone, and Lucrezia Reichlin (2008). "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" In: *Journal of Econometrics* 146(2), pp. 318–328.

Dees, Stephane and Jochen Güntner (2014). *Analysing and forecasting price dynamics across euro area countries and sectors: a panel VAR approach*. Working Paper Series 1724. European Central Bank.

Doan, Thomas, Robert Litterman, and Christopher A. Sims (1984). "Forecasting and conditional projection using realistic prior distributions". In: *Econometric reviews* 3(1), pp. 1–100.

Forni, Mario et al. (2000). "The generalized dynamic-factor model: Identification and estimation". In: *Review of Economics and statistics* 82(4), pp. 540–554.

Kadiyala, K and Sune Karlsson (1997). "Numerical Methods for Estimation and Inference in Baesian VAR-Models". In: *Journal of Applied Econometrics* 12(2), pp. 99–132.

Kim, Soyoung and Nouriel Roubini (2000). "Exchange rate anomalies in the industrial countries: A solution with a structural VAR approach". In: *Journal of Monetary Economics* 45(3), pp. 561–586.

Koop, Gary (2013). "Forecasting with Medium and Large Bayesian VARS". In: *Journal of Applied Econometrics* 28(2), pp. 177–203.

Meese, Richard and Kenneth Rogoff (1983). "Empirical exchange rate models of the seventies: Do they fit out of sample?" In: *Journal of International Economics* 14(1-2), pp. 3–24.

Pesaran, M., Til Schuermann, and S.M. Weiner (2004). "Modeling Regional Interdependencies Using a Global Error-Correcting Macroeconometric Model". In: *Journal of Business & Economic Statistics* 22, pp. 129–162.

Robertson, John C. and Ellis W. Tallman (1999). "Vector autoregressions: forecasting and reality". In: *Economic Review-Federal Reserve Bank of Atlanta* 84(1), p. 4.

Scholl, Almuth and Harald Uhlig (2008). "New evidence on the puzzles: Results from agnostic identification on monetary policy and exchange rates". In: *Journal of International Economics* 76(1), pp. 1–13.

Sims, Christopher A. (1980). "Macroeconomics and Reality". In: *Econometrica* 48(1), pp. 1–48.

Sims, Christopher A. (1992). "Bayesian inference for multivariate time series with trend". In: *American statistical Association meetings*.

Sims, Christopher A. (1993). "A nine-variable probabilistic macroeconomic forecasting model". In: *Business Cycles, Indicators and Forecasting*. University of Chicago Press, pp. 179–212.

Sims, Christopher A. and Tao Zha (1998). "Bayesian methods for dynamic multivariate models". In: *International Economic Review*, pp. 949–968.

Stock, James and Mark Watson (2002). "Macroeconomic Forecasting Using Diffusion Indexes". In: *Journal of Business & Economic Statistics* 20(2), pp. 147–62.

Zellner, A. (1996). *An Introduction to Bayesian Inference in Econometrics*. Wiley Classics Library. Wiley. ISBN: 9780471169376.

# Appendices

## Appendix 1. Table of notations

| Notation | Dimension | Description | Formula |
|---|---|---|---|
| $p$ | scalar | number of lags | |
| $m$ | scalar | number of endogenous variables | |
| $d$ | scalar | number of exogenous variables | |
| $k$ | scalar | number of parameters in an equation | $k = mp + d$ |
| $h$ | scalar | forecast horizon | |
| $y_t$ | $m \times 1$ | vector of endogenous variables | $y_t = \Phi' x_t + \varepsilon_t$ |
| $x_t$ | $k \times 1$ | vector of all regressors | $x_t = [y'_{t-1} \dots y'_{t-p}\ z'_t]'$ |
| $\varepsilon_t$ | $m \times 1$ | vector of random errors | $y_t = \Phi' x_t + \varepsilon_t$ |
| $Y$ | $T \times m$ | all endogenous variables | $Y = [y_1, y_2, \dots, y_T]'$ |
| $X$ | $T \times k$ | matrix of regressors | $X = [x_1, x_2, \dots, x_T]'$ |
| $E$ | $T \times m$ | matrix of errors | $E = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_T]'$ |
| $\Phi_1, \dots$ | $m \times m$ | coefficients of VAR | $y_t = \Phi_1 y_{t-1} + \dots + \Phi_{const} + \varepsilon_t$ |
| $\Phi_{const}$ | $m \times d$ | matrix of constants | $y_t = \Phi_1 y_{t-1} + \dots + \Phi_{const} + \varepsilon_t$ |
| $\Phi$ | $k \times m$ | grouping of matrices $\Phi_1, \dots$ | $\Phi = [\Phi_1 \dots \Phi_p\ \Phi_{ex}]'$ |
| $\underline{\Phi}$ | $k \times m$ | prior mean $\Phi$ | $\underline{\Phi} = \mathbb{E}(\Phi)$ |
| $\overline{\Phi}$ | $k \times m$ | posterior mean $\Phi$ | $\overline{\Phi} = \overline{\Omega} \cdot (\underline{\Omega}^{-1}\underline{\Phi} + X'Y)$ |
| $\underline{\nu}$ | scalar | prior degrees of freedom | $\nu \geq \max\{m+2, m+2h-T\}$ |
| $\overline{\nu}$ | scalar | posterior degrees of freedom | $\overline{\nu} = T + \underline{\nu}$ |
| $\underline{S}$ | $m \times m$ | prior scale matrix | $\underline{S} = (\underline{\nu} - m - 1)\operatorname{diag}(\sigma_1^2, \dots, \sigma_m^2)$ |
| $\overline{S}$ | $m \times m$ | posterior scale matrix | $\overline{S} = \underline{S} + \hat{E}'\hat{E} + \hat{\Phi}' X' X \hat{\Phi} + + \underline{\Phi}'\underline{\Omega}^{-1}\underline{\Phi} - \overline{\Phi}'\overline{\Omega}^{-1}\overline{\Phi}$ |
| $\underline{\Omega}$ | $k \times k$ | matrix of prior scaling coefficients of covariance matrix $\Phi$ | $\underline{\Xi} = \Sigma \otimes \underline{\Omega}$ |
| $\overline{\Omega}$ | $k \times k$ | matrix of posterior scaling coefficients of covariance matrix $\Phi$ | $\overline{\Omega} = (\underline{\Omega}^{-1} + X'X)^{-1}$, $\overline{\Xi} = \Sigma \otimes \overline{\Omega}$ |
| $\Sigma$ | $m \times m$ | covariance matrix of errors | $\mathbb{E}\,\varepsilon_t \varepsilon'_t = \Sigma$ |

## Appendix 2. Data

| Name of serie | Type of series | Base period (if any) | Source |
|---|---|---|---|
| Industrial production index | base index | 2010 | IFS |
| Consumer price index | base index | 2010 | IFS |
| Employment in manufacturing index | base index | 2010 | IFS |
| Interbank interest rate | perc. per ann. | | IFS |
| Lending interest rate | perc. per ann. | | IFS |
| Real income index | base index | 01:1992 | FSSS |
| Unemployment rate | percent | | IFS |
| Crude oil (Brent) price index | base index | 2010 | IFS |
| Producer price index | chain index | | IFS |
| New houses commissioning | thous. of sq. met. | | FSSS |
| Real fixed investment index | base index | 01:1994 | UAESD |
| Real wage rates index | base index | 01:1993 | FSSS |
| Monetary aggregate M2 | bln. rub. | | CBR |
| Real effective exchange rate | base index | 2010 | IFS |
| Natural gas price | US$ for bln BTU | 2010 | IFS |
| International reserves excluding gold | Bln US$ | | IFS |
| Nominal exchange rate | rub. per US$. | | IFS |
| Declared need in workers | thous. of people | | UAESD |
| Real agricultural production index | base index | 01:1993 | UAESD |
| Real retail output index | base index | 01:1994 | UAESD |
| Total government budgetary balance | bln. rub. | | UAESD |
| Export of goods | mln US$ | | IFS |
| Import of goods | mln US$ | | IFS |

IFS - International Financial Statistics of IMF `http://www.imf.org/en/Data`
FSSS - Federal State Statistical Service`http://www.gks.ru/`
CBR - Central Bank of Russia `http://cbr.ru/`
UAESD - United Archive of Economic and Sociological Data `http://sophist.hse.ru/rstat/`