

УДК 330.43

## Comparing Forecasting Accuracy between BVAR and VAR Models for the Russian Economy

Yan Rudakouski

The commercial bank VTB,  
12, Presnenskaya naberezhnaya str., Moscow, 123317, Russian Federation.  
E-mail: yanrudakovski@mail.ru

This paper investigates variations in the accuracy of forecasting key macroeconomic indicators through the comparison of Frequentist and Bayesian vector autoregression (VAR) models. The primary aim of the study is to identify the most effective prior type in minimizing forecast errors for the key macroeconomic indicators in the context of the Russian economy. A significant aspect of this research involves elucidating the theoretical foundations of Bayesian methods and delineating the roles of different priors in the prediction of macroeconomic indicators. A pivotal consideration in the application of the Bayesian approach is the diversity of priors, such as Jeffreys and Minnesota, which may overlook economic considerations like inflation targeting and neutral money. Conversely, certain priors, such as steady-state or independent normal-Wishart priors, are grounded in economic policy. The study delves into the nuanced interplay between these priors and their implications for forecasting accuracy. The empirical findings reveal that all Bayesian VARs exhibit superior forecasting accuracy compared to the classical VARs. Furthermore, expanding the model's scope from a limited number of variables (4) to a more comprehensive set (18) enhances forecast precision, as evidenced by the escalating log-predictive scores, Model Confidence Sets, and The Diebold-Mariano test. Simultaneously, the BVAR with the steady-state prior has demonstrated the lowest forecast error over a two-year period, but the prediction with the Minnesota prior looks relatively stable in all horizons.

**Key words:** Bayesian approach; VAR; Minnesota prior; normal-Wishart prior; inflation; GDP; forecast accuracy.

**JEL Classification:** C11; C32; E27; E31.

---

**Yan Rudakouski** – senior analyst of Corporate Credit Risk Department in VTB.

The article was received: 04.09.2023/The article is accepted for publication: 09.11.2023.

**DOI:** 10.17323/1813-8691-2023-27-4-506-526

**For citation:** Rudakouski Y. Comparing Forecasting Accuracy between BVAR and VAR Models for the Russian Economy. *HSE Economic Journal*. 2023; 27(4): 506–526.

## 1. Introduction

Accurate economic forecasting serves as a compass for navigating the complex and interconnected economy. While forecasts are not always perfect due to the inherent uncertainties in economic systems, they provide valuable insights that guide decision-making, risk management, and strategic planning across various sectors of society. For example, accurate forecasts help governments anticipate economic trends, allowing them to implement timely and targeted interventions. For businesses, accurate economic predictions form the basis for strategic planning, production scheduling, and inventory management. By aligning their strategies with projected economic conditions, businesses can better position themselves in the market.

However, forecasting the economy is a complex task fraught with challenges. The quality and availability of data are key issues that researchers face. Economic data can be sparse, inconsistent, or subject to revisions, posing challenges when building models that rely on historical data for future predictions. Moreover, when working with economic data, researchers should keep in mind the interconnectedness between variables and their mutual influence. It creates endogeneity issues where the relationship between variables becomes difficult to disentangle. Feedback loops can amplify small initial changes into significant outcomes, making predictions uncertain.

Choosing the right economic model among various alternatives is also an issue. The choice involves balancing model performance on historical data, computational efficiency, and the model's ability to capture real-world economic dynamics. To solve these problems, economists have to strike a balance between model complexity and simplicity. Overly complex models might fit historical data well but fail to generalize to new economic conditions, while overly simplistic models might miss important nuances. A good example is taking into account the heterogeneity of economic agents' behavior. Households, firms, and policy-makers often exhibit heterogeneity, meaning they have diverse characteristics and behaviors. Ignoring this heterogeneity can lead to oversimplified and inaccurate conclusions, but accurately modeling their behavior significantly complicates the models.

Improving forecasting accuracy is one of the major challenges economists face. The specification of the choice of a given model is a vital part of forecast performance on which researchers spend considerable amounts of time. However, the evolution of economic and mathematical tools has significantly simplified this stage of the study.

The use of vector autoregression (VAR) is a standard approach for solving economic tasks, including forecasting. However, researchers have been faced with the issue of a choice of more important indicators due to over-parametrization. Bayesian inference can easily solve this problem, but this approach is computationally demanding. Regardless, the development of technology and data science has added a new impetus to the widespread application of Bayesian methods.

To solve problems by incorporating information from a large amount of data, researchers accordingly face the challenge of integrating economic theory into Bayesian models. This is why a variety of priors, such as Jeffreys, Minnesota, normal-Wishart, and so on, currently exist. However, this raises the issue of determining which of them is more suitable for specific economic

tasks. For example, if one were to make a prediction with many indicators, one should use the Minnesota prior, due to this being the simplest way to incorporate cross-variable shrinkage and shrinking parameters on further lags. By contrast, the independent normal-Wishart prior helps to incorporate neutral money into an estimated model.

In the paper, the research question focuses on a comparison of forecasting the results of applying the VAR and BVAR models. It is a classical research problem that we can find in numerous working papers. However, a lot of researchers using Russian economic indicators have concentrated on a narrow set of priors, notably the Minnesota prior. The motivation behind this study is to employ a diverse array of priors and subsequently compare their effectiveness in terms of forecast accuracy.

In so doing, BVAR models incorporate different types of priors beginning with those devoid of economic logic (Jeffreys prior) to including inflation targeting (steady-state prior). Apart from types of prior, we have compared accuracy based on model size, in particular a small model that included only four variables, whilst our medium and large models included 10 and 18, variables, respectively. Inflation and economic growth have been chosen as the main indicators to compare forecast accuracy. Therefore, when expanding the models, we used indicators that potentially affect prices and GDP.

The estimation of the models is based on quarterly data for the Russian economy from 2000 to 2021. There are four types of priors included in BVARs: Jeffreys, Minnesota, the independent normal-Wishart, and the steady-state. These priors are more commonly used in economics. The best model will be considered the one that minimizes the forecast error for the first quarter for the first and second years.

The research is organized as follows. In section 2, we present a literature review on applying Bayesian inference to address various macroeconomic tasks. In Section 3, we focus on the theoretical framework underpinning the Bayesian approach. This section illustrates the difference between types of prior and their interactions with each other. The data used, and the methodological aspects of the estimated models are described in section 4. Section 5 provides the empirical results of the evaluated models and, in particular, discusses which of them shows the best forecasting accuracy. Finally, the conclusions are presented in Section 6.

## **2. Literature review of the use of Bayesian models in macroeconomics**

BVAR models are popular tools to solve the forecasting tasks associated with key macroeconomic indicators. In general, these models have lower prediction errors due to the inclusion of large numbers of indicators. Despite having a high dimensional space, the overfitting issue is mitigated through the incorporation of priors.

The Bayesian approach has become highly popular among scientists, especially after the spread of machine learning. There are a lot of Bayesian applications in macroeconomics and finance described by [Koop, Korobilis, 2010; Carriero et al., 2019; Demeshev et al., 2016; Giannone et al., 2014] and others.

Solving the problem of over-parameterization in a VAR is one of the major benefits of using the Bayesian approach. Special emphasis needs to be placed on Large Bayesian VARs. This type of model is not widespread in macroeconomics due to the difficult interpretation of the associated economic results, but its popularity is increasing.

The study by [Bańbura, Giannone, Reichlin, 2008] is tangible evidence of the above. The authors ran a Large Bayesian VAR for monetary models with different sizes and priors, including 131 macro-indicators. Particular emphasis was placed on priors with regard to the sum of coefficients which allowed linear combinations of coefficients to be constructed. Adding this prior improved forecast performance in comparison with FAVAR models. A similar problem was solved by [Korobilis, 2013] who worked with a large number of correlated macroeconomic indicators (183 predictors) to predict the UK's GDP and inflation. However, Korobilis used the Bayesian variable selection algorithm, which posterior probabilities for importance predictors to be found. This algorithm was instrumental in reformulating the spike-and-slab prior for highly correlated data, providing more accurate results than typical spike-and-slab prior selection and principal component methods.

Large BVARs are generally based on the natural conjugate prior because they need the Kronecker product structure for the posterior covariance matrix. This significantly increases the speed of estimation of various parameters. However, the Kronecker structure requires a symmetric restriction on lags for own and other variables, a limitation that could distort the economic logic in some linkages.

One way to solve this problem was suggested by [Chan, Jacobi, Zhu, 2019], who used an asymmetric conjugate prior to the estimation of the BVARs. Firstly, the authors' prior simulation was faster than a natural conjugate prior. This finding was determined for BVARs with 100 variables and four lags on US data for the period 1985 to 2019. Secondly, the new prior was able to identify five structural shocks over the last 35 years. [Giannone, Lenza, Momferatoud, Onorante, 2014] constructed a Bayesian autoregressive model with different weights for the prior and sample. The key idea underlying the research was to predict the Euro area's inflation. The benchmark for the result was a random walk model with a normal-inverse Wishart prior.

In BVAR modelling, the autoregressive coefficients were determined using the Minnesota prior, while the covariance matrix of the residuals used an inverted Wishart. The variance of the prior decreased in the further lags, and the coefficients were correlated in BVAR equations, provided to the same variables and lags. The key hyperparameter  $\lambda$  controlled the scale of all the prior variances and covariances. If  $\lambda$  was set to zero, the posterior would be equal only to the prior. By contrast, when  $\lambda$  tends to infinity, the posterior would be effectively independent of the prior and equate to OLS estimates.

The BVAR models predict unconditionally and conditionally, where the latter means that some variables contain specific assumptions, such as oil price, policy rate, utility tariff, and so on. Evaluation results of the three models show that the mean errors for HIPCexp were larger than inflation without exception for the more volatile goods (energy) in all periods of the forecast. At the same time, both conditional and unconditional BVAR were found to be more accurate than random walks. Although the first model demonstrated better shorter-term predictions, in the longer run they yielded similar results.

An additional example highlighting the advantages of Bayesian inference is the work by [Dieppe, Legrand, 2016], who have gained estimates for a large number of models that use both Frequentist and Bayesian inferences. They presented the effects of increasing the policy rate of the US's monetary to GDP. According to their findings, all models consistently indicated that raising the policy rate results in a decline in GDP. The magnitude of the decline peaks after two quarters but continues to be felt for a duration ranging from 10 to 16 quarters.

The results of the Bayesian VAR models with the different priors were similar, with the exception of just a few moments. The BVAR with the dummy observation prior had a lesser impact on GDP than the others. This response was explained by the fact that prior information was passed to the model through the likelihood function, and not through the prior distribution. Also, dummy observations led to unit root processes, consequently leading to a weaker reaction of GDP in response to a shock in the impulse response.

Comparing the BVAR with the classical VAR and OLS shows very similar qualitative characteristics but a different numerical response to the monetary shock. Particularly, there was a shorter-lived reaction by the VAR than the BVAR, with fast bottoming and more moderate recovery.

In addition to the impulse response function, the authors compared the predictive power of the estimated models. The results obtained were more diverse. In a model with the normal-Wishart the unit root was permanently shifting prediction to the steady state. The panel VAR also had a larger mean error than the BVARs. This is associated with the additional information available in a panel structure (fixed and random effects).

In recent decades, various scientists have focused on the comparison between the forecasts obtained from the BVAR with QPM or DSGE models, which are used by the central banks as an important analytical tool. One of the most well-known pieces of research in this regard was written by [Domita, Monti, Sokol, 2019], who forecasted GDP growth and CPI via BVAR and the Bank of England's DSGE model (COMPASS). The implementation of BVAR was based on a simple prior which consisted of priors with dummy observations in a combination of conjugate priors. The combination contained the Minnesota and the sum-of-coefficients priors. Since most of the variables were stationary, according to the Minnesota prior each endogenous variable followed a random walk process with drift. Whilst the accuracy of GDP and its components in the BVAR was greater than COMPASS, CPI predictions were very similar in both models. According to the authors, this was achieved using a combination of conjugate priors, such as the Minnesota, the sum-of-coefficients, and the dummy-initial-observations. The last two of these priors were used in addition to Minnesota to decrease the significance of the deterministic component (trend and seasonality).

The same forecast exercise was performed by [Brázdik, Franta, 2017] according to the main Czech indicators. The authors compared predictions for seven quarters for BVAR and structural DSGE for the Czech economy. The main challenge to the study was the implementation of the structural changes, which were added to looser steady-state priors. Using this type of prior is associated with a size modification of inflation targeting. Particularly, the inflation target was changed from 3% to 2% by the end of 2008. Moreover, the estimation of the BVAR was in the mean-adjusted form, where the priors were superimposed on the steady-state variables.

The evaluation of the BVAR involves both level and growth rate variables. To achieve this, researchers have utilized the Minnesota prior, which generally supports the prior mean on the first own lag of variables for growth rates of zero and of one for variables in level. However, the traditional approach was not a good choice due to the implementation of the steady-state prior. According to the BEAR toolbox, [Dieppe, Legrand, 2016] used the same prior mean, set at 0.2 for the coefficients on the first own lag.

The key finding of the study was that the BVAR model showed a higher accuracy for inflation and interest rates in comparison to the DSGE model. This can be explained for the same reasons. Firstly, the DSGE included calibration by expert judgment (fiscal indicator or speed of

convergence to steady state), some of which may deviate from actual data. Secondly, the BVAR model was re-estimated at each prediction step while the DSGE model was evaluated only once.

The transition from monetary to inflation targeting since the 1990s has required the consideration of monetary authority goals. This has demanded a new type of prior that reflects a long-term expectation such as an unconditional mean (steady-state prior). The effectiveness of the forecasting accuracy due to adding steady-state belief was shown by [Villani, 2009], who ran models on the Swedish economy, having used a mean-adjusted form of the BVAR.

[Beechey, Österholm, 2010] showed the advantages of BVAR with the steady state prior for a wider range of developed countries. The researchers have shown a higher prediction accuracy for Bayesian AR, which included knowledge about the inflation target, than the traditional Bayesian AR and AR models for Australia, Canada, Sweden, New Zealand, and the United Kingdom.

The adoption of inflation targeting in Central and Eastern European countries stimulated researchers to use the long-run prior in a BVAR to improve forecast performance. One may note, for instance, the research of [Brázdik, Franta, 2017] about Czech inflation, [Stelmasiak, Szafranski, 2016] about Polish inflation. All these authors found that BVAR with the steady-state prior was very effective in terms of inflation prediction for about three to seven quarters, i.e., on the horizon of achieving goals by the monetary authorities.

Our research focuses on the Russian economy. The Bayesian methods are not a popular tool in Russian science. However, some economics apply these methods to estimate different shocks and to make predictions regarding macro-indicators. [Sheveleva, 2017] has shown that external factors, which included the Shanghai Stock Exchange Composite Index, CBOE Volatility Index, and the oil price, play a significant role in determining the dynamics of the Russian economy. Shevelev's BVAR was estimated via the Minnesota prior. Additionally, [Pestova, Mamonov, 2016] explored shocks to the Russian economy using the Independent normal-Wishart prior. A distinctive feature of their study was the more protracted nature of the impact of the shock than under the Minnesota prior.

Separately, it is worth mentioning the study by [Demeshev, Malakhovskaya, 2016]. They forecasted the industrial production, price index, and interest rates for the Russian economy. The BVAR model with the Minnesota prior gave consistently better results than a classical VAR when at least five variables were included. Moreover, the mean-square prediction errors decreased with increasing number of variables in the model. This was especially noticeable between models with 14 and five variables.

In recent years, the use of Bayesian models for forecasting economic indicators of the Russian economy has not lost its relevance. For example, [Sharafutdinov, 2023] forecasted the Russian GDP, inflation, key rate, and FX using the DSGE-VAR model. The scholar compared the predictive power between the DSGE model and the DSGE-VAR model in the form of the BVAR. A priori information of the BVAR based on the previous research of the Russian economy. The main conclusion of the study is that the use of a structural model as a priori information for the BVAR model has allowed for a better description of the data compared to an unconstrained VAR model. This was especially evident in the forecasting of key macro indicators for up to two years. Also, [Fokin, Polbin, 2019] estimated VAR-Lasso models for making scenario forecasts of the Russian GDP and its components to 2019–2024. The researchers have chosen as a basis for comparison a couple of models, such as ARIMA, VAR, and BVAR by Pestova and Mamonov, as well as the official predictions of the Ministry of Economic Development and IMF. The main as-

sumption of the estimated model was that GDP and its components had the same potential growth rate both before and after the structural break in 2008–2009. For some variables (GDP and consumption during 2 quarters) the VAR-LASSO, the model showed statistically significantly better predictions than the benchmark models. However, statistically significant improvement in prediction has not been universal among the analyzed variables. It could associate the low power of corresponding tests on a short-term period and the advantage in the predictive power of alternative models.

Fokin (2022) was also involved in forecasting the components of Russian GDP using a mixed-frequency Bayesian vector autoregression model (MFBVAR). The main advantage of this model is the ability to use both quarterly and monthly frequency data simultaneously. It allows the forecast properties to be improved with the arrival of new monthly information. In addition, such a model is resistant to the jagged edge problem, which is especially important when forecasting. The classical BVAR was like a benchmark. Both MFBVAR and BVAR were estimated with the Minnesota prior. According to the results of the Diebold-Mariano test, the MFBVAR model produced significantly higher forecast quality than the base naive forecast. It is especially noticeable when forecasting GDP, consumption, and foreign trade in 2020–2021. Thus, the works described above emphasize that Bayesian models have a clear advantage, especially in the short-term forecast.

### 3. Bayesian approach: theoretical framework

Before describing the Bayesian inference, we introduce the key concept of VAR models. A VAR is simply a set of equations, in which each variable depends on lags of own and other variables. Therefore, each equation in a VAR contains the same number of explanatory variables, estimated by the ordinary least squares [Dieppe, Legrand, 2016].

A general VAR model with  $n$  endogenous and  $m$  exogenous variables and  $p$  lags can be written like (1):

$$(1) \quad \begin{pmatrix} y_{1,t} \\ \vdots \\ y_{n,t} \end{pmatrix} = \begin{pmatrix} a_{11}^p & \dots & a_{1n}^p \\ \vdots & \ddots & \vdots \\ a_{n1}^p & \dots & a_{nn}^p \end{pmatrix} \begin{pmatrix} y_{1,t-p} \\ \vdots \\ y_{n,t-p} \end{pmatrix} + \begin{pmatrix} c_{11} & \dots & c_{1m} \\ \vdots & \ddots & \vdots \\ c_{n1} & \dots & c_{nm} \end{pmatrix} \begin{pmatrix} x_{1,t-1} \\ \vdots \\ x_{m,t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{n,t} \end{pmatrix}.$$

In compact form, the VAR is written like (2):

$$(2) \quad y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + C x_t + \varepsilon_t$$

where  $y_t$  is a  $n \times 1$  vector of endogenous variables,  $A$  is  $p$  matrices of dimension  $n \times n$ ,  $C$  is a  $n \times m$  matrix, and  $x_t$  is a  $m \times 1$  vector of exogenous variables which include constant, time or exogenous variables,  $\varepsilon_t$  is a vector of residuals with multivariate normal distribution with zero mean and variance-covariance matrix  $\Sigma$ .

It should be also noted that a variance-covariance matrix is symmetric positive definite with variance on the diagonal and covariance off-diagonal elements. Each equation of the VAR includes  $k = np + m$  coefficients. Therefore, the full VAR model consists of  $q = nk = n(np + m)$

coefficients for evaluation. An increase in the number of variables in the VAR leads to a non-linearly rise in a number of estimated parameters, but additional lags grow them linearly.

Collecting the regressors into a single matrix, one gets (3):

$$(3) \quad \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_T \end{pmatrix}}_{T \times n} = \underbrace{\begin{pmatrix} y_0 & y_{1-p} & x_1 \\ \vdots & \vdots & \vdots \\ y_{T-1} & y_{T-p} & x_T \end{pmatrix}}_{T \times k} \underbrace{\begin{pmatrix} A_1 \\ \vdots \\ A_p \\ C \end{pmatrix}}_{k \times n} + \underbrace{\begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_T \end{pmatrix}}_{T \times n},$$

where  $T$  is the sample size used for the regression.

Generally, the equation (3) is converted into a vector, stacked all VAR coefficients in one line. For example,  $y$  variables, which presents like a vector ( $T \times n$ ), are transformed to ( $nT \times 1$ ) vectors.

In compact vector form equation (3) looks like (4):

$$(4) \quad y = \bar{X}\beta + \varepsilon,$$

where  $y = \text{vec}(Y)$ ,  $\bar{X}$  is  $I_n \times X$ ,  $\varepsilon = \text{vec}(\varepsilon)$  with  $N(0, \bar{\Sigma})$ .

According to equation 5, estimation of parameters  $\beta$  can be computed like:

$$(5) \quad \hat{\beta} = (X'X)^{-1}X'y.$$

The concept of a Bayesian analysis is to combine assumptions about the parameters (distribution prior) with the observed data (likelihood function). Using of the Bayes rule, a combination of both components gives the posterior distribution:

$$(6) \quad p(\beta | y) = \frac{p(\beta | y)p(\beta)}{p(y)}.$$

Following (6), posterior distribution  $p(\theta|y)$  equals to the product of likelihood function  $p(y|\theta)$  with a prior distribution  $p(\theta)$ , divided by the density of data. The denominator  $p(y)$  is independent of  $\theta$ . It's called normalizing constant, which generally is ignored:

$$(7) \quad p(\beta | y) \sim p(y | \beta) p(\beta).$$

Probability density function summaries to prior beliefs about the VAR coefficients, which allow to rewrite Bayes' law like:

$$(8) \quad p(A, \Sigma | y_{T-p,t}) = \frac{p(y_{T-p,t} | A_p, \Sigma) p(A_p, \Sigma)}{p(y_{T-p,t})} \Rightarrow p(y_{T-p,t} | A_p, \Sigma) p(A_p, \Sigma).$$



The joint posterior distribution of the VAR coefficients  $p(A, \Sigma | y_{T-p:t})$  equals a product of the likelihood function  $p(y_{T-p:t} | A, \Sigma)$  and priors  $p(A, \Sigma)$ . In agreement with the autoregressive structure of a VAR and independent and identically distributed residuals, the likelihood function can be present like the conditional distribution of each observation:

$$(9) \quad p(y_{T-p,t} | A_p, \Sigma, y_{T-p,0}) = \prod_{t=1}^T p(y_t | A_p, \Sigma, y_{T-p,t-1}).$$

Given the assumption of Gaussian errors, the likelihood can be rewritten into:

$$(10) \quad p(y_{T-p,t} | A_p, \Sigma, y_{T-p,0}) = \prod_{t=1}^T \frac{1}{2\pi^{n/2}} |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(y - A'x_t)' \Sigma^{-1} (y - A'x_t)\right\}.$$

The connection between OLS and Bayesian methods lies in the fact that OLS seeks to find point estimates of the parameters by minimizing the sum of squared errors. In the Bayesian viewpoint, we formulate linear regression using probability distributions rather than point estimates. The response,  $y$ , is not estimated as a single value, but is assumed to be drawn from a probability distribution by incorporating prior information. Based on equation 4, rewritten equation 10 looks like:

$$(11) \quad p(y_{T-p,t} | A_p, \Sigma, y_{T-p,0}) = \prod_{t=1}^T \frac{1}{2\pi^{n/2}} |\bar{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(y - \bar{X}\beta)' \Sigma^{-1} (y - \bar{X}\beta)\right\}.$$

The Bayesian approach enables the incorporation of some knowledge about parameters by using priors. There are many types of priors, each of which solves specific tasks such as changing parameters over time, heteroscedasticity of residuals, shrinkage weak parameters, and so on.

A choice of a prior is a crucial step for the Bayesian inference. [Koop, Korobilis, 2010] have noted three issues for that. First of all, the probability of precise estimates will be low without a prior. Moreover, there is a posterior standard deviation to be large in the long run. Secondly, a prior leads to getting analytical results of models faster, otherwise an evaluation with MCMC is computationally demanding. In addition, there are kinds of prior distributions that ignore deviations in models. For example, an unconstrained VAR may use different explanatory variables in the equations, to allow time-varying coefficients or be a heteroscedastic structure.

The significance of different types of priors lies in their influence on the posterior distribution, which directly impacts parameter estimates, uncertainty quantification, and predictions. The choice of prior affects the trade-off between incorporating prior knowledge and letting the data speak for itself. It also influences the robustness of results and the level of trust stakeholders place in the analysis.

Jeffreys prior serves as a neutral starting point in Bayesian inference. Jeffreys prior is suitable when little prior information is available about the parameter being estimated. While Jeffreys prior avoids strong assumptions, it may not make full use of available economic theory or expert insights. It can be less effective in cases where some theoretical guidance exists. Using Jeffreys can provide flexibility and avoid bias from strong assumptions. However, it might not capture known economic relationships that can enhance forecast accuracy. In economic fore-

casting, this could be the case when the underlying relationships between variables are uncertain or not well understood.

Minnesota prior is the most popular used in the Bayesian inference. Its popularity is associated with the simplest form of implementation and calculation. The main idea of this prior is the behavior of macro indicators like a random walk with drift. That assumption ignores economic theory, but it describes movement in most time series.

According to the Minnesota prior, the residual variance-covariance matrix is known, so it is necessary to estimate only a vector of coefficients  $\beta$ .

For obtaining posterior distribution there are two elements: a likelihood function and a prior. The residual of the BVAR follows a multivariate normal distribution with zero mean and covariance matrix. Therefore,  $y$  also has the same distribution with mean and covariance matrix and its likelihood looks like equation 12:

$$(12) \quad p(\beta) \sim N(\beta_0, \Omega_0).$$

The prior also is a multivariate normal distribution with mean  $\beta_0$  and a covariance matrix  $\Omega_0$  conforming to the Minnesota prior, endogenous variables could include a unit root in their first own lags. The coefficients for more distant lags and cross-variable lag have to be equal to zero because they are less informative. Therefore,  $\beta_0$  is vector zeros, except the first own lag, which equals 1. However, that logic works for data in levels because, generally, these time series are not stationary. If the data are presented in growth rates, then it should set a value of the first lag less 1.

Also, the Minnesota prior does not exist covariance between variables, so the variance-covariance matrix  $\Omega_0$  is a diagonal. Moreover, the variance of further lags should be the smallest due to less information. In contrast, the variance of exogenous is large, because there is not enough information about them. For using the Minnesota prior, the resulting posterior distribution is normal, which means that you can easily get the value of any parameter function using the Monte Carlo methods.

Before evaluating the posterior distribution, it's necessary to determine the hyperparameters for the Minnesota prior [Dieppe, Legrand, 2016]:

1) the variance for own lags endogenous variables is

$$(13) \quad \sigma_{a_{ii}}^2 = \left( \frac{\lambda_1}{l^{\lambda_3}} \right)^2,$$

where  $\lambda_1$  is a general tightness of a parameter. If  $\lambda_1$  equals zero, a variable does not play any role for estimation, therefore posterior equals prior. By contrast, if  $\lambda_1$  tends to infinity, the prior is not value, and the posterior is a result of OLS estimation. Moreover,  $l$  is the lag considered by the coefficient,  $\lambda_3$  is coefficient controlling the speed for convergence more distance lag to zero.

2) the variance for cross-variable lag coefficients is

$$(14) \quad \sigma_{a_{ij}}^2 = \frac{\sigma_i^2}{\sigma_j^2} \left( \frac{\lambda_1 \lambda_2}{l^{\lambda_3}} \right)^2,$$

where  $\sigma_i^2$  and  $\sigma_j^2$  are the residual variance of the autoregressive models.  $\lambda_2$  is a cross-variable variance parameter.

3) the variance for exogenous variables is

$$(15) \quad \sigma_{x_i}^2 = \sigma_i^2 (\lambda_1 \lambda_4)^2,$$

where  $\lambda_4$  is a variance parameter of exogenous, which is supposed very large number ( $10^2$  or  $10^5$ ). It associates with an assumption about insufficient information of exogenous variables.

For example, if a VAR includes two endogenous with one lag and one exogenous (constant) variables, its variance-covariance matrix looks like:

$$(16) \quad \Omega_0 = \begin{pmatrix} \lambda_1 & 0 & 0 & 0 & 0 \\ 0 & \frac{\sigma_1^2}{\sigma_2^2} (\lambda_1 \lambda_2)^2 & 0 & 0 & 0 \\ 0 & 0 & \left( \frac{\lambda_1}{l^{\lambda_3}} \right)^2 & 0 & 0 \\ 0 & 0 & 0 & \frac{\sigma_1^2}{\sigma_2^2} \left( \frac{\lambda_1 \lambda_2}{l^{\lambda_3}} \right) & 0 \\ 0 & 0 & 0 & 0 & \sigma_1^2 (\lambda_1 \lambda_4)^2 \end{pmatrix}.$$

After determining the prerequisites for the prior's parameters, there can be computed the prior distribution for vector  $\beta$ :

$$(17) \quad p(\beta) \sim \exp \left[ -\frac{1}{2} (\beta - \beta_0)' \Omega_0 (\beta - \beta_0) \right].$$

The product of the likelihood function ( $f(y|\beta)$ , equation 11) with the prior ( $p(\beta)$ , equation 17) equals the posterior distribution:

$$(18) \quad p(\beta | y) \sim \exp \left[ -\frac{1}{2} (y - \bar{X}\beta)' \Sigma^{-1} (y - \bar{X}\beta) \right] \exp \left[ -\frac{1}{2} (\beta - \beta_0)' \Omega_0 (\beta - \beta_0) \right].$$

After a couple manipulation with equation 18 can be seen immediately multivariate normal distribution with mean  $\bar{\beta}$  and covariance matrix  $\bar{\Omega}$ :

$$(19) \quad p(\beta | y) \sim N(\bar{\beta}, \bar{\Omega}),$$

$$(20) \quad p(\beta | y) \sim \exp \left[ -\frac{1}{2} (\beta - \bar{\beta})' \bar{\Omega}^{-1} (\beta - \bar{\beta}) \right],$$

$$(21) \quad \bar{\beta} = \bar{\Omega} \left[ \bar{\Omega}_0^{-1} \beta_0 + (\Sigma^{-1} \otimes X') y \right],$$

$$(22) \quad \bar{\Omega} = \left[ \bar{\Omega}_0^{-1} + \Sigma^{-1} \otimes X'X \right]^{-1}.$$

The equations 21–22 demonstrate very well that posterior mean ( $\bar{\beta}$ ) and variance  $\bar{\Omega}$  are weight prior assumption and OLS estimation.

The Minnesota prior is a practical tool in economic forecasting when there is a reliable economic theory available to guide the model. It strikes a balance between data-driven inference and theoretical coherence, leading to forecasts that are both grounded in economic knowledge and informed by observed data.

A normal-Wishart prior (natural conjugate prior). The main drawback of the Minnesota prior is knowing the residual covariance. This premise may be uncomfortable for some economic tasks. One simple way to remove this flaw is to use a normal-Wishart prior where all parameters are unknown. The normal-Wishart prior supports a normal prior for the model's coefficients and an inverse-Wishart prior for the covariance matrix. Moreover, this type of prior is a conjugate prior for a VAR, which means the same family distribution for a prior, a likelihood function, and a posterior.

Because the data remain unchanged, the likelihood function is the same. So, the main focus will be on prior for the parameters (normal distribution for coefficients and an inverse-Wishart for a covariance matrix), which can be set as:

$$(23) \quad p(\beta | \Sigma) \sim N(\beta_0, \Sigma \otimes H_0),$$

$$(24) \quad p(\Sigma) \sim IW(S_0, \alpha_0),$$

where  $H_0$  – diagonal matrix;  $S_0$  – scalar matrix on which the diagonal elements have a first-order autoregressive process;  $\alpha_0$  – number of degrees of freedom.

Whereas the Minnesota prior has the full covariance matrix for the coefficients, the conjugate normal-inverse Wishart matrix  $H_0$  includes the variance for the parameters of only one equation. Therefore, it is necessary to use the Kronecker product to reflect the full prior covariance matrix.

Unlike the Minnesota prior, in the current distribution there is no difference between the parameter variances for dependent variable lags and other variable lags, particular  $\lambda_2$  equals 1. Therefore, the overall tightness ( $\lambda_1$ ) of the prior distribution should be lower than that of the Minnesota distribution.

Considering the above, the prior distribution for the coefficients and the covariance matrix is:

$$(25) \quad p(\beta) \sim \left| \Sigma^{-\frac{k}{2}} \right| \exp \left[ -\frac{1}{2} (\beta - \beta_0)' (\Sigma \otimes H_0)^{-1} (\beta - \beta_0) \right],$$

$$(26) \quad p(\Sigma) \sim \left| \Sigma^{-\frac{\alpha_0 + n + 1}{2}} \right| \exp \left[ -\frac{1}{2} \text{tr}(\Sigma^{-1} S_0) \right].$$

A combining the likelihood (equation 11) and the priors (equations 25–26) gives a posterior distribution:

$$(27) \quad p(\beta, \Sigma | y) = \left| \Sigma^{-\frac{k}{2}} \right| \exp \left[ -\frac{1}{2} (\beta - \bar{\beta})' (\Sigma \otimes \bar{H})^{-1} (\beta - \bar{\beta}) \right] \left| \Sigma^{-\frac{\bar{\alpha} + n + 1}{2}} \right| \exp \left[ -\frac{1}{2} \text{tr}(\Sigma^{-1} \bar{S}) \right],$$

$$(28) \quad \bar{H} = \left[ H_0^{-1} + X'X \right]^{-1},$$

$$(29) \quad \bar{\beta} = \bar{H} \left[ H_0^{-1} \beta_0 + X'y \right]^{-1},$$

$$(30) \quad \bar{\alpha} = T + \alpha_0,$$

$$(31) \quad \bar{S} = \left[ y'y + S_0 + \beta_0' H_0^{-1} \beta_0 - \bar{\beta}' \bar{H}^{-1} \bar{\beta} \right]^{-1}.$$

It should be noted that if  $S_0$ ,  $\alpha_0$ ,  $H_0$  tend to zero, then the posterior is very close with OLS estimation.

**An independent normal-Wishart prior.** The restrictions in the natural conjugate prior may be inconsistent with economic prerequisites because each equation has the same variables and a restriction of the prior covariance of the coefficients to be proportionally. The most popular example is money neutrality, which requires zero coefficients in front variables of real economy of in all equation with money growth.

Using the independent normal-Wishart prior gives opportunities for including this assumption in a BVAR. The key idea of this type of the prior is independence between coefficients and the error covariance from each other. This type of distribution does not have an analytical solution; therefore, numerical methods are used for this problem (for example, Gibbs sampling). The normal-Wishart prior is particularly useful for modeling multivariate economic relationships, incorporating existing knowledge about variability, and providing more accurate and robust forecasts. The choice of prior depends on the availability of prior information, the complexity of relationships among variables, and the balance between data-driven and knowledge-driven inference.

Each macroeconomic indicator in the long run has a steady state value [Villani, 2009]. There is potential growth, a neutral policy rate, the target of inflation, and others. This valuable knowledge has to be added to models for increasing accuracy. This is implemented through including information about the unconditional means, using steady-state prior.

To improve the accuracy of the forecasts, steady-state values are included in the prior distribution by adjusting the constant. At the same time, the parameters in such a model are independent, in particular, the coefficients have a Minnesota distribution, and the covariance matrix has the form of an independent normal-inverse Wishart distribution.

The types of prior distributions described above are usually not independent of each other. Many of them are special cases of the general distribution. For example, the Minnesota and normal Jeffreys priors are derived from the independent normal-inverse Wishart. The relationship between prior distributions is described in more detail in the research of [Demeshev, Malakhovskaya, 2015].

#### 4. Data description

Forecast accuracy testing was carried out on quarterly data of the Russian economy for 2000–2021 between VAR and BVAR models with two lags for inflation and real GDP growth. The VAR model includes one specification containing 4 variables (inflation, real GDP, nominal interest rates, exchange rate RUB/USD). The BVAR model is presented in four types of priors (Jeffreys, Minnesota, independent normal-inverse Wishart, steady state) and three sizes (4 variables, 10 and 18 variables (table 1)).

**Table 1.**

**Description of data for estimated models**

Variable	Type of variable	BVAR model			Source
		Small	Medium	Large	
Inflation	endogenous	*	*	*	FRED data
gdp growth	endogenous	*	*	*	FRED data
3-month interbank rates	endogenous	*	*	*	FRED data
exchange rate RUB/USD	endogenous	*	*	*	FRED data
gdp deflator	endogenous		*	*	FRED data
unemployment rate	endogenous		*	*	FRED data
broad money (M2)	endogenous		*	*	FRED data
oil price (Brent)	exogenous		*	*	FRED data
food index price	exogenous		*	*	FAO
real wage	endogenous		*	*	National agency
credit for economy	endogenous			*	Central bank
real retail	endogenous			*	FRED data
metal world index price	exogenous			*	World bank
producer price	endogenous			*	National agency
external debt	endogenous			*	Central bank
production index	endogenous			*	National agency
real effective exchange rate	endogenous			*	Central bank
EU gdp growth	exogenous			*	FRED data

Source: constructed by the author.

Forecast accuracy was assessed using root mean square error (RMSE), where the comparison base for VAR and BVAR models was an autoregressive model (AR). At the same time, a second-order autoregressive model was estimated for inflation – since price changes are more inertial than GDP. The next research was to add in the BVARs with different priors and more macro-indicators (plus six variables for a medium model and six more for a larger model).

The economic viability of the estimated models was tested using an impulse response function for inflation and economic growth to the shock of a one standard deviation rate hike. In general, after a monetary policy shock, inflation and GDP have to decline.

All variables in the models, with the exception of indicators reflecting interest rates, unemployment rate, and relative values (external debt), were seasonally adjusted and presented in annualized log differences. Due to time series being presented as growth rates, all of them were stationary in nature (KPSS and ADF tests).

## 5. Forecasting CPI and GDP with VARs and BVARs

In this section, we focused on the research question – particularly, making a prediction with different models to evaluate the best of them according to accuracy (in terms of minimizing the root mean squared errors).

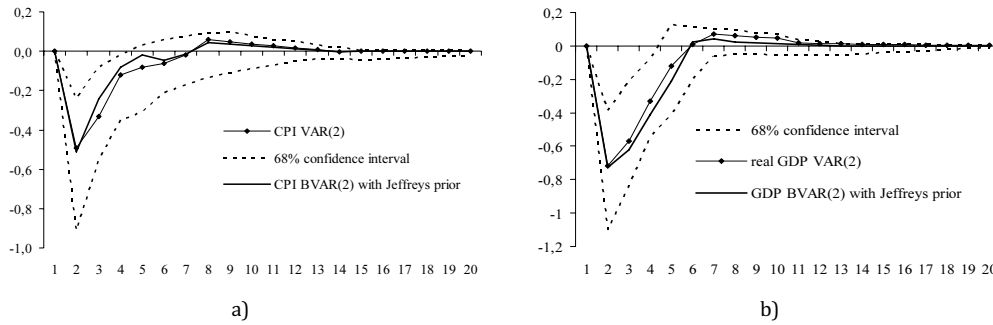
As a rule, a benchmark of the forecast accuracy is the random walk. In our research, we focused on inflation and economic growth; therefore, a selection of other explanatory variables in the medium and large BVARs was based on the potential impact on price and GDP. For this reason, the selection of the autoregression with two lags represented a more appropriate benchmark. Moreover, inflation is one of the most inert economic variables.

Before proceeding to forecasting, it is necessary to ensure the economic viability of the models. For vector autoregressive models, verification takes place through computing impulse response functions. The check is based on two small models with two lags and four variables (cpi, gdp growth, interest rate, and exchange rate), which included a classical VAR(2) and BVAR(2) with Jeffreys prior. Considering the fact that we used only four variables and a prior, where information did not depend on the set of parameters, the expected results of both models must be similar.

In our case, the hitting interest rate by 1 standard deviation has to decline in both inflation and GDP growth (Fig. 1). Simulation of the impulse-response functions of both macro-indicators illustrated the expected outcome. Raising the interest rate resulted in decreased CPI and output. At the same time, the GDP reaction was stronger, in particular –0.7 p.p. versus –0.44 p.p. for prices. Moreover, both models (the BVAR and the VAR) showed identical results.

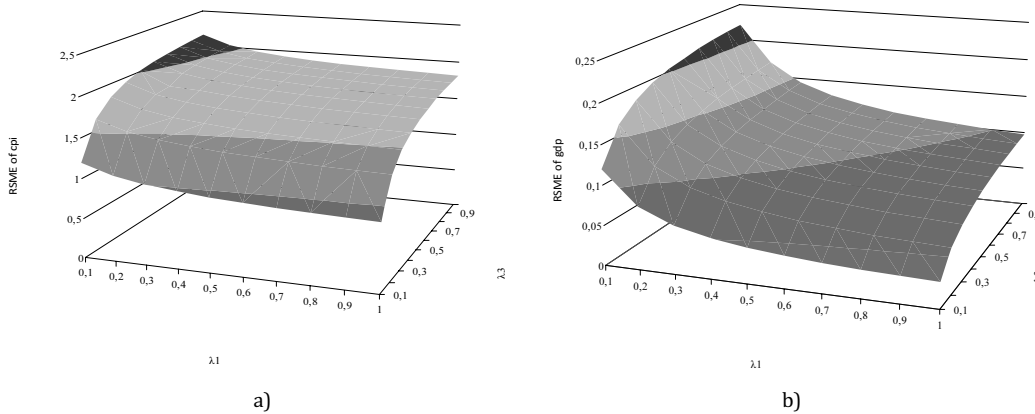
A key issue for the evaluation of BVARs is the selection of hyperparameters. Generally, most researchers set 0.1–0.2 for the overall tightness hyperparameter, a wider interval of 0.5–1.0 for the cross-variables hyperparameter, and larger for exogenous of about  $10^2$  or  $10^5$ . To answer the research question, the optimal hyperparameters were chosen according to the mean square errors of the small BVARs. The dependence of the forecast error on the parameters is shown in Fig. 2.

According to Fig. 2, as  $\lambda_3$  tends toward one and  $\lambda_1$  to zero, the RMSE decreases, indicating that prior should have only a limited influence on the optimal revised forecasting performance.



**Fig. 1.** Response of (a) inflation and (b) GDP to shock of interest rate using BVAR with Jeffreys prior and VAR(2)

Source: constructed by the author.



**Fig. 2.** Dependence Root Mean Square Error from  $\lambda_1$  and  $\lambda_3$  (a) CPI and (b) GDP

Source: constructed by the author.

Taking into account the fact that we have used hyperparameters that minimized the predicted errors, which equaled 0.2 for the overall tightness ( $\lambda_1$ ) and 0.9 for the lag decay ( $\lambda_3$ ). Also, in the Minnesota prior,  $\beta_0$  was 0.9 for the first own lags and zero for other variables; the diagonal element of  $\omega_0$  is the result of AR(2) for each variable. In the independent normal-Wishart,  $\beta_0$  equals zero, and  $S_0$  is an identical matrix by the number of endogenous variables.

The IRFs based on the Minnesota prior and the independent normal-Wishart with optimal hyperparameters also have illustrated the anticipated economic reactions of CPI and economic growth after interest rate shock. However, they slightly less affection by 0.1-0.3 p.p., but longer by 1-2 quarters.

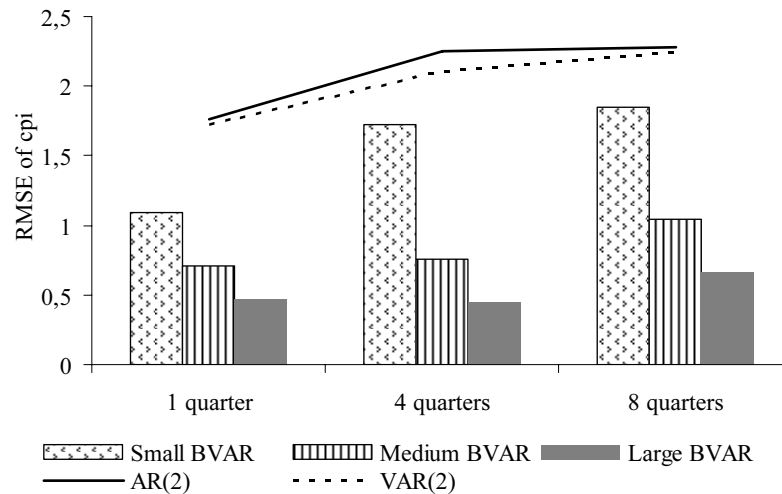
To make a prediction of inflation and GDP, there was a run MCMC algorithm for BVARs with the Jeffreys and Minnesota priors. Also, Gibbs sampling was applied for the independent



normal-Wishart and the steady-state priors due to complicated estimation for the associated analytical solutions.

To further examine our model, we use the posterior predictive distribution, which illustrates the difference between the observed and posterior predictive data. The posterior predictive distribution of the average of all models is computed through 20000 samples, 15000 of which are burned in.

Figure 3 shows the root mean squared errors of the BVARs forecast are substantially better than those of the autoregressive model and the classical VAR(2) for both economic indicators. Also, the further the forecast horizon, the lower the accuracy for all models. This confirms that uncertainty is increased for longer-term predictions. In terms of model size, additional variables help to reduce forecast errors. This finding is associated with the selected variables, but it is not a rule. There are many examples where small-dimensional BVARs are better than their high-dimensional counterparts [Demeshchev, Malakhovskaya, 2015; Koop, Korobilis, 2010].



**Fig. 3.** The average Root Mean Square Errors of CPI for models

Source: constructed by the author.

According to Model Confidence Sets, all estimated models have included in MCS p-value confidence intervals. Also, the medium and larger BVAR models have shown the better result for forecasting performance, because MCS p-value was equal nearly 1 (table 2).

The Diebold – Mariano test is an additional method used to assess the predictive power of a model. This test evaluates the null hypothesis that the accuracy of two forecasts is equal. The table 3 displays the results of the Diebold – Mariano test for various sizes of BVAR models in comparison to VAR(2).

At a 5% significance level, the null hypothesis is rejected for most models, except for the small BVAR models forecasting CPI at 4 and 8 quarters, as well as the medium BVAR model at 8 quarters, which reject the null hypothesis at the 10% significance level. In other words, the estimated BVAR models significantly enhance forecasting accuracy when compared to VAR(2).

**Table 2.**

Model	MCS p-values			
	GDP		CPI	
	RMSE	p <sub>MCS</sub>	RMSE	p <sub>MCS</sub>
AR(2)	1.39	0.691*	2.10	0.724*
VAR(2)	1.26	0.742*	2.02	0.747*
Small BVAR	0.40	0.945*	1.55	0.921*
Medium BVAR	0.55	0.971**	0.84	0.961**
Large BVAR	0.45	0.981**	0.53	0.981**

Note: Inclusion in the 95% and the 68% MCS is denoted by \* and \*\*.

Source: constructed by the author.

**Table 3.**

Horizons	P-value of Diebold – Mariano test					
	Small BVAR vs VAR(2)		Medium BVAR vs VAR(2)		Large BVAR vs VAR(2)	
	CPI	GDP	CPI	GDP	CPI	GDP
1 quarter	0.045	0.031	0.031	0.037	0.017	0.042
4 quarters	<b>0.076</b>	0.017	0.028	0.018	0.008	0.016
8 quarters	<b>0.088</b>	0.028	0.044	<b>0.057</b>	0.011	0.022

Note: \*\*\* – DM-test p-value < 0.01, \*\* – DM-test p-value < 0.05, \* – DM-test p-value < 0.1.

Source: constructed by the author.

The RMSEs determined for the BVAR with different priors are presented in table 4. The BVAR with the steady-state prior has the lowest forecast errors for forecasts one and four quarters ahead. However, this type of prior leads to the largest errors in the long term because the variables achieve their steady-state figures in the medium term. Most forecast errors fluctuate at about 0.3–0.7% from one period to another, but the prediction with the Minnesota prior looks relatively stable in all horizons.

Another measure showing a reasonable model accuracy is Log-predictive density (or log-likelihood), which should be proportional to the root mean squared error. According to our evaluation, the log-likelihood rises significantly from small to large models, reach by nearly 98 point, against 35 in the small models.

Should be noted, we have used the unconditional prediction for all models. However, if we support dynamic exogenous variables such as commodity prices or the GDP of main trade partners, then the forecast error will decrease.

Table 4.

## The result of root mean squared errors of the BVARs

Horizons	Variable	BVAR model											
		Small				Medium				Large			
		D	M	n-W	s-s	D	M	n-W	s-s	D	M	n-W	s-s
1 quarter	inflation	1.22	1.34	1.31	0.35	1.31	1.08	0.74	0.31	0.57	0.92	0.22	0.17
	gdp	0.31	0.25	0.61	0.55	0.67	0.37	0.41	0.54	0.42	0.22	0.17	0.04
4 quarter	inflation	1.46	1.48	0.78	0.77	0.66	0.72	0.53	0.64	0.64	0.58	0.34	0.27
	gdp	0.31	0.37	0.14	0.15	0.33	0.53	0.12	0.11	0.37	0.36	0.32	0.06
8 quarter	inflation	1.69	1.71	1.02	2.21	1.06	1.16	0.91	1.03	0.35	0.37	0.77	1.18
	gdp	0.53	0.58	0.31	0.66	1.55	0.59	0.24	1.14	1.31	0.49	0.86	0.82

Note: D – Diffuse (Jeffrey) prior; M – Minnesota prior; n-W – independent normal-Wishart prior; s-s – steady-state prior.

Source: constructed by the author.

## 6. Conclusions

In this paper, we have investigated the efficacy of Bayesian methods in addressing economic tasks, specifically focusing on forecasting accuracy in the context of BVARs and a classical VAR(2). Inflation and economic growth in Russia were chosen as macro-indicators for testing purposes. As a result, all BVAR forecasts were systematically more accurate than those for VAR(2) for both variables and horizons (one quarter, one, and two years ahead). This was confirmed by a significant decrease in the root square of the forecast errors. Moreover, extending the model size from a small (four variables) to a larger one (18 variables) consistently improved forecast accuracy, as evidenced by the increasing log-predictive scores. Among the BVAR models, the most accurate for the one-quarter and one-year horizons was the BVAR with the steady-state prior, but the independent normal-Wishart prior was more suitable for long-term planning. This suggests the importance of selecting a prior that aligns with the forecasting horizon and the specific characteristics of the economic variables under consideration.

Generally, this research has shown that Bayesian inference is a workable approach to avoiding overparameterization, using of priors about the model parameters. In order to obtain accurate forecasts, it is also necessary to consider the tightness of hyperparameters' own lags and cross-variable interaction, data type (in levels or growth rates), and a forecast horizon.

Bayesian Vector Autoregressive models have demonstrated lower forecast errors for Russian macroeconomic indicators compared to classical models. This outcome could prove valuable not only for the field of economic science but also for practical implications among policymakers and economists. Firstly, the BVARs allow for the incorporation of uncertainty in forecasting. Policymakers and economists can capture a range of potential outcomes by using different priors, reflecting the inherent uncertainty in economic predictions. Also, different priors allow for flexibility in capturing various economic scenarios. Policymakers can explore a wide array of possible outcomes based on different assumptions about the relationships between variables.

Secondly, the BVAR models with different priors can better handle data limitations, such as missing or incomplete data. Priors can help fill in gaps and provide reasonable estimates, enhancing the reliability of forecasts. It also includes incorporating expert knowledge or economic theory in the model. Therefore, BVAR models with expert-based priors can incorporate valuable insights, improving the model's accuracy.

Many research papers on the Russian economy using the Bayesian approach have focused on using a very limited number of prior distributions, particularly the Minnesota prior. Unlike previous studies, the current one showed the difference in prediction accuracy using the most commonly used prior distributions (Jeffrey prior, Minnesota prior, independent normal-Wishart prior, steady-state prior).

Future research directions may lie in comparing prediction accuracy between BVAR models and more complex models such as FAVAR, the time-varying VAR, Markov-Switching VAR, and MIDAS-VAR models. Which model will have the best results is an interesting question. For example, the Markov-Switching VAR and the time-varying VAR models tend to work better with macroeconomic data that has economic fluctuations, such as structural breaks. This is attributed to the inherent characteristics of these models, which are adept at identifying sudden shifts in economic data.

\* \*

\*

### References

- Banbura M., Giannone D., Reichlin L. (2008) *Larger Bayesian VARs*. Working paper ECB, 966.
- Beechey M., Österholm P. (2010) Forecasting Inflation in an Inflation-Targeting Regime: A Role for Informative Steady-State Priors. *International Journal of Forecasting*, 26, 2, pp. 248–264.
- Brazdik F., Franta M. (2017) *A BVAR Model for Forecasting of Czech Inflation*. Working Paper of the Czech National Bank, 7.
- Chan J., Jacobi L., Zhu D. (2019) *Efficient SElection of Hyperparameters in Large Bayesian VARs Using Automatic Differentiation*. Working Papers of Australian National University, 46.
- Carriero A., Galvao A., Kapetanios G. (2019) A Comprehensive Evaluation of Macroeconomic Forecasting Methods. *International Journal of Forecasting*, 35, 4, pp. 1226–239.
- Carriero A., Chan J., Clark T., Marcellino M. (2022) Large Bayesian Vector Autoregressions with Stochastic Volatility and Non-Conjugate Priors. *Journal of Econometrics*, 227, 2, pp. 506–512.
- Demeshev B., Malakhovskaya O. (2015) BVAR Mapping. *Applied Econometrics*, 43, 3, pp. 118–141.
- Demeshev B., Malakhovskaya O. (2016) Macroeconomic Forecasting with a Litterman's BVAR Model. *HSE Economic Journal*, 2016, 4, pp. 691–710.
- Dieppe A., Legrand R. (2016) *The BEAR Toolbox*. Working Paper ECB, 1934.
- Domita S., Monti F., Sokol A. (2019) Forecasting the UK Economy with a Medium-Scale Bayesian VAR. *International Journal of Forecasting*, 35, 4, pp. 1669–1678.
- Fokin N. (2023) Nowcasting and Forecasting Key Russian Macroeconomic Variables With the MFBVAR Model. *Ekonomicheskaya Politika*, 18, 3, pp. 110–135.
- Fokin N., Polbin A. (2019) Forecasting Russia's Key Macroeconomic Indicators with the VAR-LASSO Model. *Russian Journal of Money and Finance*, 78, 2, pp. 67–93.
- Giannone D., Lenza M., Momferatoud D., Onorantede L. (2014) Short-Term Inflation Projections: A Bayesian Vector Autoregressive. *International Journal of Forecasting*, 30, 3, pp. 635–644.

- Koop G., Korobilis D. (2010) Bayesian Multivariate Time Series Methods for Empirical Macroeconomics. *Foundations and Trends in Econometrics*, 3, pp. 267–358.
- Korobilis D. (2013) Bayesian Forecasting with Highly Correlated Predictors. *Economics Letters*, 118, 1, pp. 148–150.
- Pestova A., Mamonov M. (2016) A Survey of Methods for Macroeconomic Forecasting: Looking for Perspective Directions in Russia. *Voprosy Ekonomiki*, 6, pp. 45–76.
- Sharafutdinov A. (2023) Forecasting Russian GDP, Inflation, Interest Rate, and Exchange Rate Using DSGE-VAR Model. *Russian Journal of Money and Finance*, 82, 3, pp. 62–86.
- Sheveleva O. (2017) Bayesian Approach to the Analysis of Monetary Policy Impact on Russian Macroeconomics Indicators. *World of Economics and Management*, 17, 4, pp. 53–70.
- Stelmasiak D., Szafranski G. (2016) Forecasting the Polish Inflation Using Bayesian VAR Models with Seasonality. *Central European Journal of Economic Modelling and Econometrics*, 8, 1, pp. 21–42.
- Villani M. (2009) Steady-State Priors for Vector Autoregressions. *Journal of Applied Econometrics*, 24, pp. 630–650.