A BAYESIAN APPROACH TO VECTOR AUTOREGRESSIVE MODEL ESTIMATION AND FORECASTING WITH UNBALANCED DATA SETS

Davit Tutberidze
Institute of Economics and Business, Ilia State University, 0162, Georgia
davit.tutberidze@iliauni.edu.ge
Dimitri Japaridze
Institute of Economics and Business, Ilia State University, 0162, Georgia
japaridze@ iliauni.edu.ge

Abstract
One disadvantage of vector autoregressive (VAR) models is that they require time series to have equal lengths in the estimation process. This requirement induces a loss of potentially valuable information coming from time series that are longer than others. The issue is particularly evident in macroeconometric setups whenever variables have different starting points due to reasons grounded in various data recording and/or collection particularities. In many developing and emerging economies - especially those that were transitioned to market economies in the late 20th century - initial statistical observations on macro variables suffer from uneven availability and/or reliability. In this paper, we offer a remedy through a Bayesian approach: information in longer time series is aggregated into a prior which is then used in the estimation of parameters for the VAR process of clipped and equally-sized time series. Relative model performance is assessed by forecasting ability of resulting models gauged by mean absolute scaled errors (MASE). For illustration purposes, we employ time series from the Georgian economy and find that resulting (Bayesian) VAR models on average perform 7% better than standard alternatives with the same set of variables.

Key words: Bayesian econometrics; vector autoregressive models; data scarcity; Minnesota prior; empirical prior.

JEL Classification: *C11, C32, C52*

I. INTRODUCTION

Most econometric textbooks and guides discuss time-series estimation procedures for balanced datasets ignoring an issue of missing data (Baltagi 2006). Although particular workarounds have been offered in the literature, discussion on potential problems incomplete data may cause in the estimation of vector autoregressive (VAR) models is limited. Ordinary Least Squares (OLS) techniques for VAR models, by design, require a balanced set of data for coefficient estimators, which means that the procedure is implemented in software packages in such a way that lengths of time series are mechanically clipped down to the shortest one. This may potentially result in losses of valuable historical information that is simply discarded - eventually leading to estimates of lower significance and/or of lower predictive ability. For instance, in macro econometric setups, a researcher often faces situations where recording of country-level variables was launched at different points in time and/or they are available at lower frequencies in the beginning of the sample. This problem is particularly evident in macroeconomic data sets of a few developing and emerging economies. In Georgia, for example, reliable data on the gross domestic product (GDP) has been available since 2003 while the headline inflation and the money aggregates were started to be observed much earlier[1]. As another example, country-level inflows of foreign direct investment into Georgia are provided in annual amounts in the beginning of the sample while quarterly data become available later. Another notable observation is the time series for the monetary policy rate, which the National Bank of Georgia has been regularly publishing since 2008, i.e., around the time when the inflation targeting framework became operable in Georgia. In either case, a standard VAR setup would force a researcher to simply disregard whole chunks of the data with missing observations and to clip the dataset down to the earliest data point where all series are available at same frequencies. In the estimation part of the paper, we explicitly show this.

Reasons behind availability of data over different time spans can be various. Generally, one obvious reason was a wave of structural reforms of government institutions in countries that were transitioned to market economies in the late 20th century. This required them to completely overhaul frameworks for statistical data

[1] For reference, see “Statistics” section of the National Bank of Georgia, www.nbg.gov.ge/index.php?m=306&lng=eng

description, measurement and collection. Meanwhile, these institutions were encouraged to gradually increase transparency by sharing data with the public and, in some cases, were enabled to reach out to sources they had been unauthorized to tap before. Lastly, the reforms typically spawned completely new sets of data. For example, whenever modernization of monetary and financial systems in post-soviet countries commenced, making use of various data on interest rates, assets and liabilities became of the utmost importance for monetary policy and financial stability purposes.

We propose a Bayesian approach to the problem of uneven datasets by introducing a novel notion of 'empirical-iterative prior'. This type of prior is empirical in the sense that it is based on the data at hand unlike some common approaches of deriving priors from subjective views of a researcher. Further, its iterative nature is underpinned by step-by-step estimations of multiple VAR models with incrementing number of variables: at each step, the informative signal coming from the VAR model of longer-than-others series is accumulated into the Minnesota-type prior which is then used to estimate the same model appended with another variable of lower length. Thus, the estimation step of the final model utilizes all the information from the variables under consideration. We find that this prior remarkably improves out-of-sample properties of the VAR model compared to standard alternatives.

II. LITERATURE REVIEW

To our knowledge, virtually no direct solution has been offered in academic literature for the problem of estimation of VAR models lacking equisized time series. However, in general, modeling of scarce data has benefitted from a rigorous interest of researchers. For a detailed retrospective, readers are referred to Granger and Newbold (1986), Marcellino, Stock and Watson (2003), Banerjee, Marcellino and Masten (2005), Alba and Mendoza (2007), Stock and Watson (2017). Based on these contributions, Bayesian approach to overcome data issues such as noise, errors, and uncertainty has been widely recognized. One obvious reason is that Bayesian methods allow for subjective probabilistic judgments to be included in deriving inferences from data. This stands in stark contrast to traditional frequentist school of statistical inference which relies on conclusions largely drawn from pure data observations. Therefore, the latter typically performs apparently worse when data is scarce - yielding estimates of parameters of low significance and/or of low predictive ability.

Applications of Bayesian methods to VAR models dates back to seminal works by Litterman (1980, 1986). He argued that the structure and magnitude of true population parameters in the VAR model are unclear and implied that it is better not to give too much value (weight) to specific values of the model parameter (e.g., to those with outright zero constraints). Instead, he recommended describing this ambiguity of model parameters with some "prior" probabilistic distribution. As a result, the degree of initial uncertainty, represented by the prior, may be later improved by the information obtained from the data observations. In this case, the improvement is carried out through the informative "signal" from the data and not by the "noise", which ensures the reduction of the risk of overfitting (i.e. the situation when an estimated model excessively reflects random variation in the variables as compared to their underlying relationship). It is believed that for the reasons above, Bayesian vector-autoregressive models (BVARs) provide a much better prediction than reduced-form VAR classical alternatives or structural models (Canova 2007). The selection of a prior distribution is the most important step in starting a Bayesian evaluation. In general, preliminary information is essential even because two economists can quite rightly make two different statistical conclusions based on the same data (Leamer 1978). Due to the dependence of conclusions on the initial information, it is necessary to have a method that matches the sample of data to the prior, and the Bayesian approach offers exactly the desired rule.

A prior typically reflects a researcher's beliefs about relationships between the variables being modeled. These beliefs may stem from the economic theory, practical experience or simply, intuition. But a powerful alternative is to use the data at hand to directly estimate hyperparameters, i.e. parameters of the prior distribution. Although being derived from frequentist methods of estimation (in particular, maximum likelihood estimation), these estimators help determine the probabilistic nature of model parameters and instrumentalize the prior information (Giannone, Lenza and Primiceri, 2015). This type of approach yields so-called empirical priors in Bayesian setups, which we rely upon in this paper.

The scarcity of data (compared to the parameters to be estimated) naturally limits the desired number of variables to be included in the classical VAR model. In this regard, too, the Bayesian approach successfully tackles the problem. A widely acknowledged work published by the European Central Bank (Banbura, Giannone and Reichlin 2008) explicitly shows that BVAR is a full-fledged tool for large data panels under the conditions of proper Bayesian shrinkage of the probabilistic distribution of parameters.

III. DATA DESCRIPTION

The empirical part of the paper employs three variables from the Georgian economy to illustrate workings of the proposed approach. The annual inflation rate is based on the CPI measure published by GeoStat[2] in monthly frequency and is available from January 1996. The same institution releases quarterly data on GDP, the annual growth rate of which is available from Q1 2004. The last variable under consideration is the policy (refinancing) rate of the National Bank of Georgia with a starting point in January 2008. Monthly data is aggregated into quarterly through averaging. The last date for all transformed time series is Q1 2021. Later, it will be useful to have a visual understanding of uneven time dimensionality of these variables illustrated on Fig. 1 below.

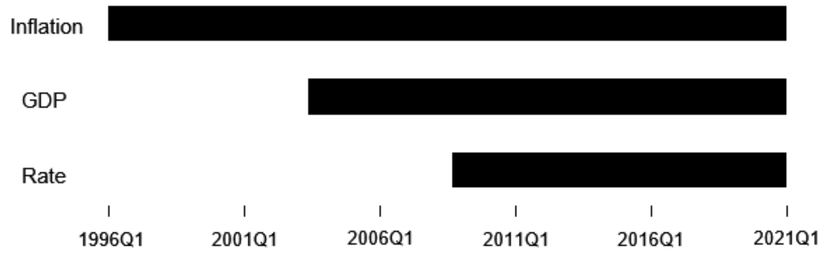

Figure 1 – Availability of data on the variables under consideration

IV. METHODOLOGY

In general terms, the idea behind Bayesian inference is to derive conclusions on unknown parameters, say, θ, of the model under consideration. However, unlike the frequentist approach to statistical inference, Bayesian methods view the parameters as random - described by probability distributions before and after observing the data. We denote prior ('before') knowledge by $f(\theta)$ probability density function and posterior ('after') knowledge by $f(\theta|y)$ where y denotes the data itself. Then, Bayes' theorem implies that

$$f(\theta|y) = \frac{f(\theta)f(y|\theta)}{f(y)}$$

where $f(y|\theta)$ is a probability density of the data given the parameters (frequentist terminology dubs it a likelihood function if viewed as a function of the parameters) while $f(y)$ is a probability density of the data over all possible values of the parameters. The latter plays a role of the normalizing constant to the posterior density.

For notational purposes, we first outline a concise form of a general VAR model to consider:

$$Y = ZB + U \tag{1}$$

where Y is used as a term for the data information on the dependent variable, Z summarizes the lagged values of Y, B collects all autoregressive and constant parameters of the system, and U is the error term. In particular, Y is a $T \times N$ matrix consisting of transposes of y_t – $N \times 1$ column-vector observations on N dependent variables at time $t = 1, .., T$. As an example, for the variables in our model specification ($N = 3$), this vector, at $t = 2$, would be

$$y_2 = \begin{bmatrix} inf_2 \\ gdp_2 \\ rte_2 \end{bmatrix}$$

where variable names are self-explanatory. Z is a $T \times (1 + pN)$ matrix (p is a lag-order) each column of which consists of either only ones, or zeros and lagged observations of the corresponding dependent variable. Again, continuing the example, the first column of Z would be a vector of three ones and the 2-nd column would look

[2] National Statistics Office of Georgia. Website: https://www.geostat.ge/en

like

$$\begin{bmatrix} 0 \\ inf_1 \\ inf_2 \end{bmatrix}$$

where we assumed $p = 2$. Further, B is a $(1 + pN) \times N$ matrix of intercepts and autoregressive parameters. Finally, U is a $T \times N$ matrix each column of which combines all error terms associated with one single dependent variable. Here we apply a common assumption of (conditional) multivariate normality to the error terms:

$$u_t \sim N(0, \Omega)$$

where 0 is a trivial expected value column-vector and Ω is a contemporaneous covariance matrix ($\Omega = E(u_t u_t')$). This assumption gives the following matric-variate distribution of U in the spirit of Karlsson (2012):
$U \sim MN(\mathbf{0}, \Sigma, I_T)$
where the first term of matric-variate normal distribution is a $T \times N$ matrix of zeros for expected values of errors, Σ is proportional to the contemporaneous covariance $N \times N$ matrix, Ω, between the error terms in the rows of U, and I_T is a $T \times T$ identity matrix related to the covariance between the error terms in the columns of U. The latter assumption implies that the errors are serially uncorrelated.

Let us now rewrite (1) in the form of

$$y = (I_N \otimes Z)\beta + u$$

where y, u and β are vertical stacks of all terms in y_t, u_t and B, respectively. It is obviously implied that $u \sim N(0, \Omega \otimes I_T)$. Then the likelihood function of β and Ω, given the y sample, allows for the following distributions of these parameters (Koop and Korobilis, 2010):

$$\beta | \Omega, y \sim N(\beta_{LS}, \Omega \otimes (Z'Z)^{-1}) \quad (2)$$

And

$$\Omega^{-1} | y \sim W\big((Y - ZB_{LS})'(Y - ZB_{LS})\big) \quad (3)$$

where LS subscript refers to the OLS estimators of the corresponding parameters and W denotes a Wishart distribution (with assumed respective degrees of freedom).

(2) and (3) fully summarize the knowledge of the researcher about β and Ω by purely observing the data. But the Bayesian estimation methods allow us to incorporate any prior beliefs on these parameters. Although there exists a number of alternatives, so-called Minnesota prior approach – originally proposed by Litterman (1980) - still enjoys a wide popularity due to its simplicity, tractability and ability to deliver accurate forecasts (Koop 2017). We, too, build our approach based on this framework under which the prior knowledge about β and Ω is given in a non-identical way. In particular, Ω is believed to be a priori a diagonal matrix with non-zero elements equal to the results of separate OLS estimations of VAR model equations. In our specification,

$$\Omega = \Omega_{OLS} = \begin{bmatrix} s_{inf}^2 & 0 & 0 \\ 0 & s_{gdp}^2 & 0 \\ 0 & 0 & s_{rte}^2 \end{bmatrix} \quad (4)$$

where s_{inf}^2 is a sample variance of residuals from a linear regression of the inflation equation. Other diagonal elements of the matrix are obtained from the GDP and policy rate equations. In turn, the prior for β is given in truly Bayesian fashion:

$$\beta \sim N(\beta_{PR}, B_{PR})$$

where it is clearly manifested that the parameter is a random quantity with a (normal) probability distribution and 'hyper-parameters' - prior mean of β_{PR} and prior variance of B_{PR}. Under Minnesota approach, elements of β_{PR}

are mostly set to zero to reflect a prior belief that time series (in case of growth rates) have low persistence, but any other theoretical value is also possible. As for B_{PR}, in subsequent derivations it is assumed to be diagonal and each non-zero element of it is determined based on whether it is related to a coefficient before the own lags of the dependent variable of the corresponding equation, to that of the lags of some other variable, or to that of an exogeneous variable and a constant (for details, readers are referred again to Koop and Korobilis (2010)).

Finally, Minnesota approach allows for an analytical posterior (i.e. after observing both the data and the prior) distribution for β:

$$\beta|y \sim N(\beta_{PS}, B_{PS})$$

where PS subscript refers to the posterior distribution parameters. Their estimators can be obtained in the following way:

$$B_{PS} = \left(B_{PR}^{-1} + \left(\Omega_{OLS}^{-1} \otimes (Z'Z)\right)\right)^{-1}$$

$$\beta_{PS} = B_{PS}(B_{PR}^{-1}\beta_{PR} + (\Omega_{OLS}^{-1} \otimes Z)'y)$$

Having set out the basic principles of the Minnesota prior, we now turn to a detailed description of our approach – empirical-iterative prior - which, in essence, is a rule for determination of numerical values of hyper-parameters β_{PR} and B_{PR}. The formation of empirical-iterative prior is carried out in several stages (or iterations). Without a loss of generality, we make use of the abovementioned model specification with three time series from the Georgian economy. The first iteration determines the one that starts from the earliest date. The data inspection reveals that the longest series is the CPI inflation. Let us denote it by $inf_t^{(\mathrm{I})}$, where the upper index indicates a sequence number of iterations. We write an autoregressive model for this variable with $p = 2$ lags and a constant with a sample size equal to the length of the inflation time series:

$$inf_t^{(\mathrm{I})} = \beta_0^{(\mathrm{I},inf)} + \beta_1^{(\mathrm{I},inf)} inf_{t-1}^{(\mathrm{I})} + \beta_2^{(\mathrm{I},inf)} inf_{t-2}^{(\mathrm{I})} + u_{t,inf}^{(\mathrm{I})} \quad (5)$$

Next, we apply a zero mean-prior, $\lambda^{(\mathrm{I},inf)} = (0,0,0)$, respectively to all $\beta^{(I,inf)}$ parameters in (5) and obtain posterior values: $\beta_{0,PS}^{(\mathrm{I},inf)}$, $\beta_{1,PS}^{(\mathrm{I},inf)}$ and $\beta_{2,PS}^{(\mathrm{I},inf)}$. Our particular interest lies in $\beta_{1,PS}^{(\mathrm{I},inf)}$ as it will be used in the second iteration.

The second iteration adds another – the longest - time series from the remaining two to the model specification – that is, the GDP growth rate. Now, with the sample size being clipped down to the length of the latter, we consider a 2-variable VAR:

$$inf_t^{(\mathrm{II})} = \beta_0^{(\mathrm{II},inf)} + \beta_1^{(\mathrm{II},inf)} inf_{t-1}^{(\mathrm{II})} + \beta_2^{(\mathrm{II},inf)} inf_{t-2}^{(\mathrm{II})} + \beta_3^{(\mathrm{II},inf)} gdp_{t-1}^{(\mathrm{II})} + \beta_4^{(\mathrm{II},inf)} gdp_{t-2}^{(\mathrm{II})} + u_{t,inf}^{(\mathrm{II})}$$

$$gdp_t^{(\mathrm{II})} = \beta_0^{(\mathrm{II},gdp)} + \beta_1^{(\mathrm{II},gdp)} inf_{t-1}^{(\mathrm{II})} + \beta_2^{(\mathrm{II},gdp)} inf_{t-2}^{(\mathrm{II})} + \beta_3^{(\mathrm{II},gdp)} gdp_{t-1}^{(\mathrm{II})} + \beta_4^{(\mathrm{II},gdp)} gdp_{t-2}^{(\mathrm{II})} + + u_{t,gdp}^{(\mathrm{II})} \quad (6)$$

Before posterior computations, we apply following mean-priors to the coefficients of each equation in (6):

$$\lambda^{(\mathrm{II},inf)} = (0,\ \beta_{1,PS}^{(\mathrm{I},inf)}, 0, 0, 0)$$

$$\lambda^{(\mathrm{II},gdp)} = (0, 0, 0, 0, 0)$$

The posterior quantities include, among others, four estimates of the first-lag parameters that we will use to construct mean-priors for the third and final iteration. The model specification at this stage is:

$$inf_t^{(\mathrm{III})} = \beta_0^{(\mathrm{III},inf)} + \beta_1^{(\mathrm{III},inf)} inf_{t-1}^{(\mathrm{III})} + \beta_2^{(\mathrm{III},inf)} inf_{t-2}^{(\mathrm{III})} + \beta_3^{(\mathrm{III},inf)} gdp_{t-1}^{(\mathrm{III})} + \beta_4^{(\mathrm{III},inf)} gdp_{t-2}^{(\mathrm{III})} + \\ + \beta_5^{(\mathrm{III},inf)} rte_{t-1}^{(\mathrm{III})} + \beta_6^{(\mathrm{III},inf)} rte_{t-2}^{(\mathrm{III})} + u_{t,inf}^{(\mathrm{III})}$$

$$gdp_t^{(\mathrm{III})} = \beta_0^{(\mathrm{III},gdp)} + \beta_1^{(\mathrm{III},gdp)} inf_{t-1}^{(\mathrm{III})} + \beta_2^{(\mathrm{III},gdp)} inf_{t-2}^{(\mathrm{III})} + \beta_3^{(\mathrm{III},gdp)} gdp_{t-1}^{(\mathrm{III})} + \beta_4^{(\mathrm{III},gdp)} gdp_{t-2}^{(\mathrm{III})} + +$$

$$+\beta_5^{(\mathrm{III},gdp)} rte_{t-1}^{(\mathrm{III})} + \beta_6^{(\mathrm{III},gdp)} rte_{t-2}^{(\mathrm{III})} + u_{t,gdp}^{(\mathrm{III})}$$

$$rte_t^{(\mathrm{III})} = \beta_0^{(\mathrm{III},rte)} + \beta_1^{(\mathrm{III},rte)} inf_{t-1}^{(\mathrm{III})} + \beta_2^{(\mathrm{III},rte)} inf_{t-2}^{(\mathrm{III})} + \beta_3^{(\mathrm{III},rte)} gdp_{t-1}^{(\mathrm{III})} + \beta_4^{(\mathrm{III},rte)} gdp_{t-2}^{(\mathrm{III})} + \beta_5^{(\mathrm{III},rte)} rte_{t-1}^{(\mathrm{III})} + \beta_6^{(\mathrm{III},rte)} rte_{t-2}^{(\mathrm{III})} + u_{t,rte}^{(\mathrm{III})} \quad (7)$$

with the following mean-priors:

$$\lambda^{(\mathrm{III},inf)} = \left(0, \beta_{1,PS}^{(\mathrm{II},inf)}, 0, \beta_{3,PS}^{(\mathrm{II},inf)}, 0, 0, 0\right)$$
$$\lambda^{(\mathrm{III},gdp)} = \left(0, \beta_{1,PS}^{(\mathrm{II},gdp)}, 0, \beta_{3,PS}^{(\mathrm{II},gdp)}, 0, 0, 0\right)$$
$$\lambda^{(\mathrm{III},rte)} = (0, 0, 0, 0, 0, 0)$$

Posterior computations of estimates for the parameters in (7) conclude the process. The described procedure ensures that the information obtained from one (smaller) iterative model in the form of the parameter estimator at lags in each equation is transferred to the new (larger) iterative model through λ prior, which plays the role of the initial assumption on the mean of the re-evaluated parameter. In this way, uneven time series contained in the data set are gradually "connected" to each other, and the final model is estimated in such a way as to account for the dynamics of all time series regardless of their unequal lengths. Also note again that a sample size at each iteration is clipped to the shortest one and, therefore, is changing from one stage to another. This fact is illustrated on Fig. 2 below.

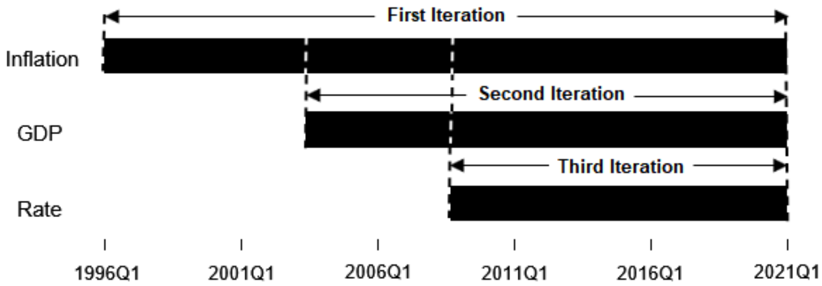

Figure 2. Sample Size at Each Iteration

However, the λ values determine only the first moments of the prior - the mean value of β_{PR}. As for the second moment – the variance (B_{PR}), which describes the degree of uncertainty around the mean-prior in Minnesota-type setups, it is controlled by three additional hyperparameters, ϕ, ρ and κ. In particular, the smaller they are, the smaller the statistical dispersion, and the estimation procedure tries to keep (or "shrink") the posterior estimators close to the mean value indicated by λ. In other words, the numerical magnitude of the posterior estimator is less affected by (new) data and is largely determined by the prior value. ϕ controls shrinkage around the prior on own lag parameters of the dependent variable in each equation, while ρ hyperparameter handles shrinkage around the prior on lag parameters of other variables in the same equation. As for κ, it controls shrinkage around the prior on the parameters before the exogeneous variables and the constant. In our illustrative example, using these hyperparameters, elements of B_{PR} corresponding to the inflation equation at the second iteration of our procedure are given in Table 3 of the Appendix where s values are standard deviations from (4), and ϕ and ρ hyperparameters are divided by the lag number. Hence, more distant lags lead to lower uncertainty in the prior, which is in line with the stylized fact that a first lag of a variable has better predictive power than other lags in an autoregressive process.

The question naturally arises: how to determine values for ϕ, ρ and κ? We apply a simple formula that defines ϕ and ρ as a ratio of a sample size at each iteration to the sample size at the first iteration divided by the number of β parameters to be estimated. Thus, we force the posterior to shrink more to the prior as less data becomes available. κ is taken to be equal to 100 as in Koop an Korobilis (2010).

In order to evaluate relative model performance, we compare final-iteration out-of-sample 1-step-ahead forecast of the model under consideration to that of two alternative models with the same set of variables. In particular, we choose reduced-form VAR and a simple Bayesian VAR as competing models. In the latter case, we apply commonly used priors as in Tutberidze and Japaridze (2017). The comparison is carried out based on a mean absolute scaled error (MASE) which is believed to yield superior results due to its scale invariance, symmetry, and interpretability (Franses, 2015). For a 1-step-ahead forecast, this measure is computed as

$$q_{T+1} = \frac{|d_{T+1} - \hat{d}_{T+1}|}{\frac{1}{T-1} * \sum_{t=2}^{T} |d_t - d_{t-1}|}$$

where d and $\hat{d}$ denote a data point and its forecast, respectively. In effect, q_{T+1} measures an absolute error of the model 1-step-ahead forecast relative to a sample average of absolute errors of naïve forecasts (i.e. forecasts of the "Random Walk"). The lower the MASE, the bigger the evidence in favor of better predictability (Hyndman and Koehler, 2006).

V. RESULTS

At the first iteration with a single equation in (5), our empirical-iterative procedure picks values for the hyperparameters as follows[3]:

$$\lambda^{(I,inf)} = (0, 0, 0)$$

$$\phi^{(I,inf)} = \rho^{(I,inf)} = \frac{1}{3} = 0.33$$

$$\kappa = 100$$

The prior variance matrix is[4]:

$$\begin{pmatrix} 1231.92 & 0 & 0 \\ 0 & 0.33 & 0 \\ 0 & 0 & 0.08 \end{pmatrix}$$

Resulting posterior estimate of the coefficient of interest is $\beta_{1,PS}^{(I,inf)} = 0.93$. Hence, at the second iteration, a mean-prior for the inflation equation parameters is constructed as

$$\lambda^{(II,inf)} = (0, 0.93, 0, 0, 0)$$

while applying a zero-mean prior to the GDP equation. ϕ and ρ are set to 0.64/(5*2)=0.06. κ, again, is equal to 100. Based on these values, diagonal elements of the prior variance matrix is presented in Table 4 in the Appendix where first five elements correspond to the parameters in the inflation equation, and the last five ones – to the GDP equation. The posterior estimates of the first-lag coefficients are:

$$\beta_{1,PS}^{(II,inf)} = 0.95, \beta_{3,PS}^{(II,inf)} = 0.17, \beta_{1,PS}^{(II,gdp)} = 0.20, \beta_{3,PS}^{(II,gdp)} = 0.64.$$

Therefore,

$$\lambda^{(III,inf)} = (0, 0.95, 0, 0.17, 0, 0)$$
$$\lambda^{(III,gdp)} = (0, 0.20, 0, 0.64, 0, 0)$$

while applying a zero-mean prior to the policy rate equation. ϕ and ρ are set to 0.49/(5*2)=0.05. κ, again, is equal to 100.

Finally, at the last iteration, based on the priors from the second stage, we obtain posterior estimates for β which are listed in the Table 1 below.

[3] Results hereafter are rounded to two decimal places.
[4] The diagonal elements of the prior matrix correspond to the constant, the first-lag, and the second-lag parameters of the equation.

Table 1. Posterior Estimates for Final-Iteration Parameters

	Inflation Equation			GDP Growth Equation			Policy Rate Equation		
Variable	**Constant**	**Lag 1**	**Lag 2**	**Constant**	**Lag 1**	**Lag 2**	**Constant**	**Lag 1**	**Lag 2**
Constant	2.27			6.45			1.66		
inf		0.92	-0.10		0.17	-0.01		0.08	-0.00
gdp		0.13	-0.03		0.59	0.00		0.02	0.01
rte		-0.11	-0.20		-0.95	0.08		0.73	-0.07

We use these estimates to calculate out-of-sample 1-step-ahead forecasts for the variables under consideration. Then we obtain the mean absolute scaled error (MASE) quantities of these forecasts and compare them to those from reduced-form VAR and simple BVAR models in Table 2.

Table 2. Mean Absolute Scaled Errors

Model	**Inflation**	**GDP Growth**	**Policy Rate**	**Average**
BVAR with Empirical-Iterative Priors (BVAR-EIP)	0.95	0.91	1.03	0.96
Reduced-Form VAR	1.09	0.98	1.21	1.09
Simple BVAR	1.04	0.96	0.97	0.99

As evidenced from Table 2, the Bayesian VAR model with empirical-iterative priors performs generally better than the alternatives. The error in BVAR-EIP forecasts, on average, is 7% lower than that in the competing models. In addition, the average MASE for BVAR-EIP is less than 1, which, according to the common practice, is a sign of a reasonable ability to predict through the underlying model.

VI. CONCLUSION

In this paper, we propose a novel forecasting method based on the Bayesian approach. In particular, information (a signal) coming from longer time series is progressively accumulated into a prior that is then used for a posterior estimation of a VAR model with original time series naturally clipped down to a single identical length. The advantage of the method is that it allows the model to be evaluated with any set of variables, regardless of unbalanced availability of historical data on them. For illustration purposes, the study selected three macroeconomic variables (CPI inflation, real GDP growth, monetary policy rate) from the Georgian economy with unequal observation periods and applied the novel method to forecast them. The results of cross-validation of the forecast revealed the appropriateness of the methodology used. In particular, the values of the mean absolute scaled error (MASE), according to the common practice, indicate a reasonable ability to predict under this approach.

The study contributes to the current intensive academic discourse on forecasting transition economies. Presumably, the algorithm developed within the study will successfully cope with the task of evaluating and forecasting the macroeconomic variables on examples of other transition countries. Notably, this type of economies are characterized by most of the features that portray the Georgian economy. In particular, macroeconomic time series typically suffer from fluctuating dynamics, low accuracy and, of course, unbalanced and scarce availability.

The approach utilized in this research allows for further insights: why not to take certain time series at lengths that a researcher deems feasible? This would let him avoid a negative impact of structural breaks or ‘unreliable’ historical data on the estimation and/or forecasting accuracy. Economic crises in the past, as well as COVID-19 pandemic, have affected dynamics of a number of macro and microeconomic variables severely. Adjusting a sample size of series and employing empirical-iterative prior in the Bayesian VAR environment may help successfully tackle the issue of poor estimates and large errors in forecasts. This is indeed an interesting avenue of research to follow in further studies.

VII. APPENDIX

Table 3. A General Form of the Matrix of Prior Variance of Parameters in the Inflation Equation at the Second Iteration

$$\begin{pmatrix} \kappa * s_{inf} & 0 & 0 & 0 & 0 \\ 0 & \frac{\phi}{1^2} & 0 & 0 & 0 \\ 0 & 0 & \frac{\phi}{2^2} & 0 & 0 \\ 0 & 0 & 0 & \frac{\rho}{1^2} * \frac{s_{inf}}{s_{gdp}} & 0 \\ 0 & 0 & 0 & 0 & \frac{\rho}{2^2} * \frac{s_{inf}}{s_{gdp}} \end{pmatrix}$$

Table 4. Estimated Diagonal Elements of the Prior Variance Matrix at the Second Iteration[5]

$$\begin{pmatrix} 483.59 & 0.07 & 0.02 & 0.03 & 0.01 & 1245.50 & 0.18 & 0.04 & 0.07 & 0.02 \end{pmatrix}$$

[5] The first element corresponds to the constant in the inflation equation, and the following four pertain to the coefficients of the lagged terms of the variables in the same equation.

VIII. REFERENCES

1. Alba, E., & Mendoza, M. (2007). Bayesian Forecasting Methods for Short Time Series. *Foresight: The International Journal of Applied Forecasting*, issue 8, 41-44.
2. Baltagi, B. H. (2006). *Panel Data Econometrics*. Van Haren Publishing, 's-Hertogenbosch.
3. Banerjee, A., Marcellino, M. & Masten, I. (2005). Forecasting macroeconomic variables for the new member states of the European Union. *Working Paper Series, No 482*, European Central Bank.
4. Canova, F. (2007). *Methods for Applied Macroeconomic Research*. Princeton University Press, Princeton.
5. Franses, Ph.H.B.F. (2015). *A note on the Mean Absolute Scaled Error.* International Journal of Forecasting, 32, pp. 20–22.
6. Giannone, D., Lenza, M., & Primiceri, G. E. (2015). Prior Selection for Vector Autoregressions. *Review of Economics and Statistics*, 97(2), 436–451.
7. Giannone, D., Reichlin, L. & Banbura, M. (2008), Large Bayesian VARs, *Working Paper Series, No 966*, European Central Bank
8. Granger, C. W. J., & Newbold, P. (1986). *Forecasting Economic Time Series* (Subsequent ed.). Academic Pr., Cambridge.
9. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679–688.
10. Karlsson, S. (2012), Forecasting with Bayesian Vector Autoregressions, *Working Papers, No 2012:12,* Örebro University, School of Business.
11. Koop, G. & Korobilis, D. (2010), Bayesian Multivariate Time Series Methods for Empirical Macroeconomics, *Foundations and Trends(R) in Econometrics*, 3, issue 4, p. 267-358.
12. Koop, G. (2017) Bayesian Methods for Empirical Macroeconomics with Big Data. *Review of Economic Analysis*, 9 (1), 33-56.
13. Leamer, E. E. (1978). *Specification Searches: Ad Hoc Inference with Nonexperimental Data* (1st ed.). Wiley, Hoboken.
14. Litterman, R. (1980). A Bayesian Procedure for Forecasting with Vector Autoregression. *Working Papers*. Massachusetts Institute of Technology, Department of Economics.
15. Litterman, R. (1986). Forecasting with Bayesian vector autoregressions — Five years of experience. *International Journal of Forecasting*, 2(4), 497–498.
16. Marcellino, M., Stock, J. H., & Watson, M. W. (2003). Macroeconomic forecasting in the Euro area: Country specific versus area-wide information. *European Economic Review*, 47(1), 1–18.
17. Stock, J. H., & Watson, M. W. (2017). Twenty Years of Time Series Econometrics in Ten Pictures. *Journal of Economic Perspectives*, 31(2), 59–86.

Tutberidze, D., & Japaridze, D. (2017). Macroeconomic Forecasting Using Bayesian Vector Autoregressive Approach. *Bulletin of Taras Shevchenko National University of Kyiv Economics*, 191, 42–49.