

NATIONAL RESEARCH UNIVERSITY HIGHER SCHOOL OF  
ECONOMICS

The Faculty of Economics Science

Department of Applied Economics

DIPLOMA

FACTOR-AUGMENTED VECTOR AUTOREGRESSION FOR  
FORECASTING RUSSIAN MACROECONOMIC TIME SERIES.

Petr Garmider, BEC 165

Advisor:

Boris Demeshev,  
Senior Lecturer,  
Department of Applied Economics

Moscow

28 February 2020

## Abstract

There is a number of models used for time series forecasts that focuses on incorporating a huge set of information. Conventional VAR fails to employ such an approach because of "degrees of freedom" problem. The focus of this paper is to test factor-augmented VAR's ability to forecast main Russian macroeconomic indicators. Results will be presented in a table of RMSEs for different horizons using time series cross validation procedure for each one. There are reasons to believe, that FAVAR will show better results than univariate models in long-term forecasts.

# 1 Introduction

## 1.1 Overview

Forecasting of macroeconomic time series is an important task for different economic institutions. Accurate forecasts for main macro indicators allows central banks timely react to possible dangers for the economy. Proper actions conducted by central bank may prevent a possible recession or the same way, wrong actions may result in initiating one. Due to the fact that some of its instruments have a delay time after its decision will have an impact on the economy, central banks especially interested in forecasts to be able influence a situation while it is possible.

Macro indicators forecasts are also useful for commercial companies as well. There are many scenarios in which corporations are particularly interested in understanding of possible future macro factor outcomes, such as: choosing the right price policy, signing long-term contracts with counterparties, making decision on a possibility of entering a new market and etc. One cannot deny that a huge amount of financial instruments heavily dependent on macroeconomic situation within the country. Investors in attempts to

value a particular financial asset always build a forecast of its determinants. By this way, almost everything to some extent shows comovement with one or another macro factor. This demonstrates that forecasting of macroeconomic time series is extremely crucial part in quite huge range of spheres.

There are a number of models used for time series prediction. Each model uses different approach of dealing with time series. Some approaches do not require any assumptions about data and aimed to minimise proper loss function in specific manner. The other includes restrictions on data, assuming, for example, presence of data-generating process with parameters that are to be estimated. State-of-the-art models, usually employ mixed approach such as forecast is based on results of two mentioned methods. Models of the first type show high accuracy of prediction, however fail to interpret its results and prone to outliers. Models of the second type, on the contrary, show moderate accuracy, however give an information about moving forces and produce quite robust results. The last approach, as one can understand tries to balance between advantages and disadvantages of such approaches. Of course, there are plenty of other models with different view on a data.

One can divide second type of models onto two subcategories: univariate and multivariate. To make a forecast for the considering time series more accurate, the last method tries to use additional data, while first assumes data generating process of a particular series to be a function of past realisations of itself and only. This paper mostly deals with multivariate models: vector auto regression (VAR) suggested by Sims (1980) and its variation factor-augmented VAR that aimed to eliminate drawbacks of classical version.

In this paper I will consider factor-augmented vector autoregression (FAVAR): method of estimation, its application for forecasting purposes and accuracy evaluation of produced forecast and its robustness to outliers. There

is no academic literature that uses FAVAR model for Russian macroeconomic time series prediction. The goal is to understand, whether factor model outperform univariate approach. One also may be interested in comparing classical VAR with its augmented version. It will be shown, that VAR is restricted model of FAVAR, therefore it is possible to test whether difference between the two is statistically significant.

## 1.2 Literature review

There are several studies that discuss FAVAR model and employ its potential for particular purposes. Bernanke et al. (2005) were first to introduce a model to incorporate a huge set of information to assess structural shocks effect on monetary policy. They discuss Sims (1992) and his interpretation of "price puzzle", when several papers showed that VAR models predict that contractionary monetary policy shock is followed by rise in price level, rather than decrease as traditional macroeconomic theory suggests. Authors believe that explanation proposed by the inventor of VAR yet had weak points, therefore suggest their own way to deal with "price puzzle". In the paper authors used 120 monthly time series pointing out that central banks while making decisions on interest rate act in "data rich environment", considering a larger set of time series than standard VAR approach allows to include. Therefore, econometrician, ignoring this fact, faces with a problem of biased estimates. They suggested two ways of how FAVAR can be estimated and one of them will be presented as pseudo-code in this work. The main goal of model inventors was to obtain correct impulse responses for variables of interest rather than producing forecasts for them.

Nevertheless, there are some works focused on applying considering model for prediction task as well, for example, Mönch (2008) in his work

uses FAVAR to forecast yield curve. Author mentions that state-of-the-art models in this sphere have a large number of parameters to estimate, which means that user cannot simply include as many explanatory variable as one wishes, so there is a risk to lose important information. Therefore, author used approach of latent factor extraction described in Bernanke et al. (2005) and mixed it with traditional models for yield curve prediction.

What for the time series prediction, most of the papers agree, that factor-augmented VAR shows better performance for long-term forecasts. In Figueiredo and Guillén (2013) FAVAR is used for time series forecasting of Brazilian consumer inflation (IPCA). Authors found that factor based models outperform those of traditional AR ones on long horizons. They also considered estimation methods pointing out that ML approach strictly outperforms two-step PCA method, that is the focus of study in this paper. Pang (2010) employed factor model for predicting GDP growth rate, unemployment and inflation rates of the Hong Kong economy. Author's results are in line with those obtained in Figueiredo and Guillén (2013) – FAVAR performs better than simple VAR and AR models especially for long horizons. Author used 76 mothly variables that are believed to reflect economic situation in Hong Kong. Transforming original data, author followed methodology suggested in Stock and Watson (2005).

Berggren and Lodenius (2016) in their paper computed Diebold-Mariano test of equal predictive accuracy, comparing FAVAR with benchmark models. Authors found, that forecast performance of FAVAR was not statistically better than traditional VAR or AR models. Nonetheless, they claim that further research should be done towards optimal number of lags for the factors, since they found that FAVAR with only one lag for factors performed statistically better than with twelve.

Overall, literature suggest that potential of FAVAR model for time series

prediction task still is not fully unlocked. There is a still space for improvements and experiments, for example, employing FAVAR for non-economic time series.

### 1.3 Methods

There are several approaches to measure quality of time series prediction models. Each method, however, captures one side of the model. Speaking about conventional time series cross validation, one may argue that it tests model's short-term prediction power, since it assesses one-step-ahead forecast. Traditional train test split is quite sensitive to split proportion and also to structural changes occurred after the end of training set. For this reason, it is not fully accurate to perform any of this method in their traditional form.

Nonetheless, it is possible to account for long-term prediction power performing time-series cross validation and treat a forecast for each horizon separately. Still results of such procedure may be argued, since a robust estimate of a model's accuracy requires a large number of observations that is problematic for Russian data available.

Forecasts are measured on RMSE (root of mean square error) metric as it is quite interpretable. Chosen measure will be evaluated for different horizons separately. The formula for h-th horizon as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=T+h-ts}^T [y_t - \widehat{y}_{t|h}]^2}$$

where  $ts$  – hold-out set size,  $\widehat{y}_{t|h}$  – forecast for period  $t$  having available information up to the moment  $t - h$ .

There are several ways to estimate FAVAR model. Bernanke et al. (2005) in their work consider two, that are two-step estimation using principal components analysis to recover latent factors and one-step Bayesian

likelihood approach estimation. Authors stated that they found no obvious evidence of one method being superior to another and weather computational cost of the first one worths it. Since cross-validation requires multiple reestimation, I simply choose approach that is less time consuming.

Particular properties of two-step approach estimation makes unclear distribution of estimated coefficients, therefore building right confident intervals becomes separate task. Bernanke et al. (2005) use bootstrap procedure for obtaining impulse response function intervals, the same is possible to apply for predictive confident intervals, however that is out of the scope of this paper. One need a rich dataset to test statistically correctness of build intervals, which is not the case in dataset available.

## 2 Model

### 2.1 VAR

Lets  $Y_t$  be  $M \times 1$  vector of variables under consideration. It is said that  $Y_t$  is VAR(d) process if it has following dynamic equation:

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \dots + A_d Y_{t-d} + \varepsilon_t, \quad (1)$$

where  $A_0$  –  $M \times 1$  vector;  $\forall i \neq 0$   $A_i$  –  $M \times M$  matrix;  $\varepsilon_t$  –  $M \times 1$  error vector with zero mean.  $\hat{A}_i$  can be obtained by equation-by-equation least squares. One can easily check that number of parameters to be estimated equals  $M(Md + 1) = O(M^2)$ ,  $M \rightarrow \infty$ . For example, econometrician should estimate 57 coefficients if one wants to use VAR(6) model with 3 variables for forecasting purposes. This means, that M is strictly limited to the number of periods. In practice, it is uncommon to see more then 6 variable included in VAR equation. This problem particularly relevant while dealing with short

time series, that is the case with Russian economic data. It is hard to believe that econometrician is able to select six variables without missing an important one. Consequently, researcher generally gets biased estimates which may result, for example, in popular VAR issue that is called "price puzzle". In Bernanke et al. (2005) authors propose explanation to this problem which completely rely on omitted equation assumption. Thus, ignoring one variable may result in incorrect economy dynamics modeling that may lead to weak forecasting performance. In order to address this issue, one can estimate FAVAR model, which allows incorporating a huge set of economic indicators without losing degrees of freedom<sup>1</sup>.

## 2.2 FAVAR

This section will be based on II.A part in Bernanke et al. (2005). Let  $Y_t$  be  $M \times 1$  vector of observable economic variables that are believed to reflect dynamics of the economy. Standard VAR approach implies that  $Y_t$  contains price indicator, observable variable reflecting economy's real output and policy instrument. Such specification of  $Y_t$  is mainly used for structural analysis. As discussed earlier, one may argue that barely information captured by  $Y_t$  reflects dynamics of the whole economy. Let  $F_t$  be  $K \times 1$  vector of unobservable variables, where in some way  $K$  is a small number, that are assumed to summarise all the information that is not captured by  $Y_t$ . These factors are said to reflect the situation of theoretical variables, such as: "economic activity", "investment environment", "price pressure" and etc.

Assume also that dynamics of vector  $(F_t', Y_t')$  follows equation:

$$\begin{bmatrix} F_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} F_{t-1} \\ Y_{t-1} \end{bmatrix} + v_t, \quad (2)$$

---

<sup>1</sup>Number of observations - number of model's parameters



where  $\Phi(L) = (\Phi_0 + \Phi_1 L + \Phi_2 L^2 + \dots + \Phi_d L^d)$ ;  $d$  – finite number;  $\Phi_i - (K + M) \times (K + M)$  matrix;  $v_t$  – error term vector, such that:  $\mathbb{E}(v_t) = 0$ ,  $\mathbb{E}(v_t v_t') = Q$ ;  $Q$  is symmetric positive semi-definite  $(K + M) \times (K + M)$  matrix;  $L$  – lag operator:  $LY_t = Y_{t-1}$ .

One may see, that if components of  $\Phi(L)$  that relate  $Y_t$  to  $F_{t-1}$  are zeros, then  $Y_t$  is simply VAR process. Otherwise, equation (2) is VAR in  $(F_t', Y_t')$ . The fact, that econometrician may obtain VAR in  $Y_t$  from (2) by putting restrictions on some coefficients makes model (2) unrestricted to VAR in  $Y_t$ . This makes it possible to perform statistical tests, for instance, likelihood ratio, that allow to check wheter unrestricted model is statistically different from restricted one. If so, (2) is commonly called factor-augmented vector autoregression, or simply FAVAR. In the paper Bernanke et al. (2005) authors correctly note that if the true model is FAVAR, but instead VAR in  $Y_t$  is estimated, then econometrician generally gets biased estimates, therefore observes poor forecasting power of estimated model.

Coefficients of  $\Phi(L)$  cannot be estimated directly since vector  $F_t$  is unobservable. If one was provided with real values of  $F_t$ , he could apply conventional techniques of VAR estimation, equation-by-equation least squares, for example, in order to estimate (2).

It is obvious, that in real world dynamics of  $Y_t$  is affected by a huge number of time series, such as exchange rate, price indices, investment climate situation, import/export shocks and many more. Let us introduce  $N \times 1$  vector  $X_t$  – set of "informational" variables, where  $N$  is in some way a "large" number, specifically such that  $K + M \ll N$ . Unlike  $F_t$ , vector  $X_t$  is assumed to be observable. It is also believed that  $X_t$  is a linear function of observable  $Y_t$  and unobservable  $F_t$ :

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t, \quad (3)$$

where  $\Lambda^f = N \times K$  matrix;  $\Lambda^y = N \times M$ ;  $e_t$  – error term vector, such that:  $\mathbb{E}(e_t) = 0$ ,  $\mathbb{E}(e_{ti}e_{tj}) \approx 0 \forall i \neq j$ .<sup>2</sup> Equation (3) reflects the idea that economy can be precisely described by just  $K + M$  main indicators. All the other variables that we may observe are generated from those key ones. The main point in (3) is that  $F_t$  is still vector of unobservable variables and the idea is to exploit the fact that  $(F_t', X_t')$  is observed to get estimates of  $F_t$  vector. Having in some sense accurate  $\hat{F}_t$ , econometrician is able to estimate (2) in  $(\hat{F}_t', Y_t')$ . It is important to note, that  $X_t$  depends only on  $F_t$  and  $Y_t$ , not from lags of these vectors. Though  $X_t$  itself may contain one step ahead series.

If  $N$  is a small number, it is possible to include  $X_t$  directly in (2) and estimate VAR in  $(X_t', Y_t')$ . However, barely it is the truth, that central bank, for example, monitor only three or four variables, setting the key rate. Though, if researcher want to include additional variable in a model, it is quite unlikely that he would have enough observations to get coefficients' estimates.

There are several possible FAVAR models may be constructed depending on what researcher initially assume. For example, it is possible to include in  $Y_t$  variables such as: CPI, GDP growth and central bank key rate, which means that researcher is uncertain about the rest structure of the economy and believe in the presence of finite number of latent factors that fill possible model misspecification. One may also argue, that in reality we don't observe truth GDP growth and CPI. Instead, we are provided with these series estimation by some federal institutions, that are still noisy measures for true variables. Such model specification is in line with the fact, that GDP growth and CPI series are subjects for several correction, even in a year. However, for sure, we observe truth central bank key rate.

---

<sup>2</sup>One can find formal requirements on  $e_t$  in Stock and Watson (2002)

## 2.3 Estimation

Authors in Bernanke et al. (2005) propose two ways of FAVAR estimation. The first is based on Gibbs sampling approach, that is not applicable if one desires to obtain cross-validation model's error, since such an approach is enormously time-costly. Estimation of one model may take even an hour. The second approach is referred in Bernanke et al. (2005) as two-step principal component method, that is much more computationally efficient in time and easy in implementation. Moreover, the last approach does not require special assumptions on error-terms distributions, as well allows some cross-correlation in it, which is more likely the truth dealing with real data.

Two-stage estimation approach includes:

1. Extraction of factors  $F_t$  from  $X_t$ , purifying it from the influence of  $Y_t$ .<sup>3</sup>
2. Estimation of (2), using obtained  $\hat{F}_t$  by standard VAR methods.

In order to perform stage 1 one should use the equation (3) and exploit the fact that  $(X'_t, Y'_t)$  is observable vector for each period  $t$ . Let us remind once again equation (3):

$$X_t = \Lambda^f F_t + \Lambda^y Y_t + e_t$$

We can rewrite (3) as:

$$X_t = \begin{bmatrix} \Lambda^f & \Lambda^y \end{bmatrix} \begin{bmatrix} F_t \\ Y_t \end{bmatrix} + e_t \quad (4)$$

$$X_t = \Lambda \alpha_t + e_t, \quad (5)$$

where  $\Lambda_{N \times (K+M)} \stackrel{\text{def}}{=} [\Lambda^f \Lambda^y]$  and  $\alpha_{t(K+M) \times 1} \stackrel{\text{def}}{=} \begin{bmatrix} F_t \\ Y_t \end{bmatrix}$ .

---

<sup>3</sup>To get only additional information about economy dynamics, that is not captured by  $Y_t$

The idea is if econometrician came up with such  $\hat{\Lambda}$  and  $\hat{\alpha}_t$  that  $\hat{e}_t$  is a small number for all  $t$ , then such estimates of (3) are "good". Assume  $X_t$  is normalized to have zero mean:  $\sum_{t=1}^T X_t = \begin{pmatrix} 0 \dots 0 \end{pmatrix}'_{N \times 1}$ . Then, let us then set the following minimization task:

$$\sum_{t=1}^T (X_t - \Lambda \alpha_t)' (X_t - \Lambda \alpha_t) \rightarrow \min_{\Lambda, \alpha_1, \alpha_2 \dots \alpha_T} \quad (6)$$

It is worth noting that (6) has in total  $(K + M)N + (K + M)T$  number of parameters for estimation. It is also important to note, that there are infinite number of solutions for such optimization task. If we came up with a solution to (6) in form  $(\tilde{\Lambda}, \tilde{\alpha})$  then for any  $Z$ :  $Z'Z = I \implies (\tilde{\Lambda}Z', Z\tilde{\alpha})$  is also solution of (6). Thus let us impose restriction on  $\Lambda$ :  $\Lambda'\Lambda = I$ , which basically means we are trying to obtain orthogonal and normalised unit-scale vectors. Problem (6) transforms to:

$$\begin{cases} \sum_{t=1}^T (X_t - \Lambda \alpha_t)' (X_t - \Lambda \alpha_t) \rightarrow \min_{\Lambda, \alpha_1, \alpha_2 \dots \alpha_T} \\ s.t. \Lambda'\Lambda = I \end{cases} \quad (7)$$

Assuming solution  $\hat{\Lambda}$  being known, optimal  $\hat{\alpha}_t = (\Lambda'\Lambda)^{-1}\Lambda'X_t$ . Since  $\alpha_t$  does not affect other periods error, then we can treat (6) as  $T$  independent minimization problems, such for fixed period  $t$ :

$$(X_t - \Lambda \alpha_t)' (X_t - \Lambda \alpha_t) \rightarrow \min_{\alpha_t} \quad (8)$$

With known  $\Lambda$ , problem (8) becomes nothing but simple least-squares task with known analytic solution. Thus, (7) transforms to:

$$\begin{cases} \sum_{t=1}^T (X_t - \Lambda(\Lambda'\Lambda)^{-1}\Lambda'X_t)' (X_t - \Lambda(\Lambda'\Lambda)^{-1}\Lambda'X_t) \rightarrow \min_{\Lambda} \\ s.t. \Lambda'\Lambda = I \end{cases} \quad (9)$$

$$\begin{cases} \sum_{t=1}^T (X_t - \Lambda\Lambda'X_t)' (X_t - \Lambda\Lambda'X_t) \rightarrow \min_{\Lambda} \\ s.t. \Lambda'\Lambda = I \end{cases} \quad (10)$$

Let  $L = \sum_{t=1}^T (X_t - \Lambda\Lambda'X_t)'(X_t - \Lambda\Lambda'X_t)$ , then:

$$\begin{aligned}
L &= \sum_{t=1}^T [(I - \Lambda\Lambda')X_t]' [(I - \Lambda\Lambda')X_t] = \\
&= \sum_{t=1}^T X_t'(I - \Lambda\Lambda')'(I - \Lambda\Lambda')X_t = \\
&= \sum_{t=1}^T X_t'(I - \Lambda\Lambda')X_t = \sum_{t=1}^T X_t'X_t - \sum_{t=1}^T X_t'\Lambda\Lambda'X_t
\end{aligned}$$

Thus, problem (10) is equivalent to:

$$\begin{cases} \sum_{t=1}^T X_t'\Lambda\Lambda'X_t \rightarrow \max_{\Lambda} \\ s.t. \Lambda'\Lambda = I \end{cases} \quad (11)$$

Let  $\tilde{L} = \sum_{t=1}^T X_t\Lambda\Lambda'X_t$ , then:

$$\begin{aligned}
\tilde{L} &= tr(\sum_{t=1}^T X_t'\Lambda\Lambda'X_t) = \sum_{t=1}^T tr(X_t'\Lambda\Lambda'X_t) = \\
&\stackrel{*}{=} \sum_{t=1}^T tr(\Lambda\Lambda'X_tX_t') = tr(\sum_{t=1}^T \Lambda\Lambda'X_tX_t') = tr(\Lambda\Lambda' \sum_{t=1}^T X_tX_t'),
\end{aligned}$$

where (\*) follows from the trace property:  $tr(AB) = tr(BA)$ , if  $A$  and  $B$  are matrices of suitable sizes.

Let us remind that if  $X_{T \times N} = (X_1X_2 \dots X_T)'$  is matrix that have zero mean by columns, then  $S_X = \frac{1}{T} \sum_{t=1}^T X_tX_t'$  - is empirical covariance matrix of  $X_t$  vector.

$$\tilde{L} = tr(\sum_{t=1}^T \Lambda\Lambda'X_tX_t') = tr(\Lambda\Lambda'TS_X) = Ttr(\Lambda'S_X\Lambda)$$

Thus (10) problem becomes equivalent to:

$$\begin{cases} tr(\Lambda'S_X\Lambda) \rightarrow \max_{\Lambda} \\ s.t. \Lambda'\Lambda = I \end{cases} \quad (12)$$

Still (12) has  $N(K+M)$  parameters along which econometrician should implement optimization procedure. Though, such an optimisation task has analytical solution. One should compute first  $K+M$  eigen vectors of  $S_X$  corresponding to largest eigen values. These vectors are columns of  $\hat{\Lambda}$  that solves maximization task (12) The fact that  $S_X$  is a symmetrical positive semi-definite matrix guarantees that it has real-valued eigen vector and eigen values. Columns of  $\hat{\Lambda}$  are nothing but first  $K+M$  principal components of  $X$ . Knowing solution to (12) one can easily find  $\hat{\alpha}_t$  from (8). Stock and Watson (2002) showed that under certain conditions on  $N$ , that should be "large", econometrician may use only first  $K$  principal and obtain consistent estimator of space spanned by  $Y_t$  and  $F_t$ . It is important to note, that at this moment we did not rely on the fact that  $Y_t$  is observable. For now we got  $K$  linear combinations of  $F_t$  and  $Y_t$ . In Bernanke et al. (2005) authors then perform purification procedure, that is removing  $Y_t$  from the space spanned by first  $K$  principal component of  $X_t$ .

In stage 2

## References

- Berggren, E. and Lodenius, E. (2016). Can favar improve swedish inflation forecasting?
- Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422.
- Figueiredo, F. M. R. and Guillén, O. (2013). Forecasting brazilian consumer inflation with favar models using target variables. Technical report, mimeo.
- Mönch, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics*, 146(1):26–43.
- Pang, I. A. J. (2010). Forecasting hong kong economy using factor augmented vector autoregression.
- Sims, C. A. (1980). Martingale-like behavior of prices.
- Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European economic review*, 36(5):975–1000.
- Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*, 20(2):147–162.
- Stock, J. H. and Watson, M. W. (2005). Implications of dynamic factor models for var analysis. Technical report, National Bureau of Economic Research.