

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

17 April 2020

## Abstract

There are several models used for time series forecasts that focus on incorporating a huge set of information. Conventional VAR fails to employ such an approach because of the "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 allow central banks to timely react to possible dangers for the economy. Proper actions conducted by the central bank may prevent a possible recession or the same way, wrong actions may result in initiating one. Since 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 to influence a situation while it is possible.

Macro indicators forecasts are useful for commercial companies as well. There are many scenarios in which corporations are particularly interested in understanding possible future macro factor outcomes, such as: choosing the right price policy, signing long-term contracts with counterparties, making decisions on the possibility of entering a new market, 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. In this way, almost everything to some extent shows comovement with one or another macro factor. This demonstrates that forecasting macroeconomic time series is an extremely crucial part of quite a huge range of spheres.

There are a number of models used for time series prediction. Each model uses a different approach to dealing with time series. Some approaches do not require any assumptions about data and aimed to minimize proper loss function in a specific manner. The other includes restrictions on data, assuming, for example, the presence of the data-generating process with parameters that are to be estimated. State-of-the-art models, usually employ mixed approaches such as forecast is based on results of the two mentioned methods. Models of the first type show high accuracy of prediction, however, they fail to interpret its results and prone to outliers. Models of the second type, on the contrary, show moderate accuracy, though, give 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 views on data.

One can divide the second type of model 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 the data generating process of a particular series to be a function of past realizations of itself and only. This paper mostly deals with multivariate models: vector autoregression (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. There is no academic literature that uses FAVAR model for Russian macroeconomic time series prediction. The goal is to understand, whether the 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 a restricted model of FAVAR, therefore it is possible to test whether the difference between the two is statistically significant.

## 1.2 Literature review

Several studies 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 the "price puzzle" when several papers showed that VAR models predict that contractionary monetary policy shock is followed by a rise in a price level, rather than decrease as traditional macroeconomic theory suggests. Authors believe that the explanation proposed by the inventor of VAR yet had weak points, therefore suggest their 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 a "data-rich environment", considering a larger set of time series than standard VAR approach allows to include. Therefore, econometrician, ignoring this fact, faces 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. The author mentions that state-of-the-art models in this sphere have a large number of parameters to estimate, which means that users cannot simply include as many explanatory variables as one wishes, so there is a risk to lose important information. Therefore, the author used the 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 a 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. The author used 76 monthly variables that are believed to reflect the economic situation in Hong Kong. Transforming original data, the author followed the 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. The authors found, that the forecast performance of FAVAR was not statistically better than traditional VAR or AR models. Nonetheless, they claim that further research should be done towards an 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, the literature suggests that the potential of FAVAR model for time series prediction task still is not fully unlocked. There is still space for improvements and experiments, for example, employing FAVAR for non-economic time series.

### 1.3 Methods

There are several approaches to measure the quality of time series prediction models. Each method, however, captures one side of the model. Traditional train test split is quite sensitive to split proportion and also to structural changes that occurred after the end of a training set. For this reason, it is not fully accurate to perform this method in their traditional form.

Nonetheless, it is possible to account for h-horizon 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. For this reason, I will measure models' performance using time series cross-validation with expanding window.

Forecasts are measured on RMSE (root of mean square error) metric as it is quite interpretable. The 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, which 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 whether the computational cost of the first one worths it. Since cross-validation requires multiple reestimation, I simply choose the approach that is less time-consuming.

Particular properties of the two-step approach estimation make unclear distribution of estimated coefficients, therefore building right confident intervals becomes a separate task. Bernanke et al. (2005) use a 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 needs a large dataset to test statistical correctness of build intervals, which is not the case in the 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 a VAR(d) process if it has the 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, an 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 than 6 variables included in a VAR equation. This problem particularly relevant while deal-

ing with short time series, which 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 an explanation to this problem which completely relies on omitted equation assumption. Thus, ignoring one variable may result in incorrect economic 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 the dynamics of the economy. Standard VAR approach implies that  $Y_t$  contains price indicator, observable variable reflecting the economy's real output and policy instrument. Such a specification of  $Y_t$  is mainly used for structural analysis. As discussed earlier, one may argue that barely information captured by  $Y_t$  reflects the 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 the 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, the likelihood ratio, that allows to check whether an unrestricted model is statistically different from a 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 an 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 real 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 a researcher wants to incorporate an 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 a 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 a 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 estimations 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 corrections, even in a year. However, for sure, we observe the 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, which is not applicable if one desires to obtain a 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 as 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 a 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:

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

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 optimization task has an exact analytical solution. One should compute the first  $K+M$  eigenvectors of  $S_X$  corresponding to the largest eigenvalues. 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 eigenvector and eigenvalues. Columns of  $\hat{\Lambda}$  are nothing but the 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  $K$  principle components 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, which is removing  $Y_t$  from the space spanned by the first  $K$  principal component of  $X_t$ .

For stage 2 authors divide variables into two groups: fast-moving and slow-moving. The first group assumed to contemporaneously react on unexpected shocks in  $Y_t$ , whereas the second one does not respond to unanticipated fluctuations in  $Y_t$  at period  $t$ . Such variables' separation is necessary for obtaining purified factors  $F_t$ , that contains all the information apart from that in  $Y_t$ . Since impulse response functions is the main focus in Bernanke et al. (2005), absence of immediate response in  $F_t$  to shocks in  $Y_t$  is one of the requirements in some SVAR identification schemes, such as recursive identification including Choleski decomposition of residuals covariance matrix. However, there is a place for experiments if one tries to employ FAVAR methodology for forecasting purposes. In literature, there are several ways to remove dependence of obtained in stage 1 factors from  $Y_t$ . The one used in Bernanke et al. (2005) will be described below.

Let  $\hat{C}(F_t, Y_t) \stackrel{\text{def}}{=} \hat{\Lambda} \hat{\alpha}_t \stackrel{(8)}{=} (\hat{\Lambda}' \hat{\Lambda})^{-1} \hat{\Lambda}' X_t \stackrel{(12)}{=} \hat{\Lambda}' X_t$ , then authors estimate following model:

$$\hat{C}(F_t, Y_t) = \beta_0 + \beta_{-0} \hat{C}^{slow}(F_t) + \gamma Y_t + v_t, \quad (13)$$

where  $\hat{C}(F_t, Y_t)$  – the first  $K$  principle components of  $X_t$ ;  $\hat{C}^{slow}(F_t)$  – the first  $K$  principle components of  $X_t^{slow}$ ;  $b_0$  –  $K \times 1$  vector;  $b_{-0}$  –  $K \times K$  matrix;  $\gamma$  –  $K \times M$  vector, though one may include subset of  $Y_t$  variables ;  $v_t$  –  $K \times 1$  error vector. One can obtain  $\hat{\beta}$  and  $\hat{\gamma}$  using simple OLS method by rows, then:

$$\hat{F}_t = \hat{\beta}_{-0} \hat{C}^{slow}(F_t) - \hat{\gamma} Y_t$$

Presence of  $F_t$  estimates allows one to substitute  $F_t$  by  $\hat{F}_t$  in (2), which result in:

$$\begin{bmatrix} \hat{F}_t \\ Y_t \end{bmatrix} = \Phi(L) \begin{bmatrix} \hat{F}_{t-1} \\ Y_{t-1} \end{bmatrix} + v_t \quad (14)$$

Note once again that (14) is nothing but standard VAR model in  $(\hat{F}_t, Y_t)$ , therefore  $\Phi(L)$  can be estimated using traditional VAR techniques. Summarising all, we finally get:

---

Algorithm 1 FAVAR estimation

---

$C(F_t, Y_t) \leftarrow$  the first  $K$  components of  $X_t$

$C^{slow}(F_t) \leftarrow$  the first  $K$  components of  $X_t^{slow}$

OLS estimation results  $\leftarrow \hat{C}(F_t, Y_t) = \hat{\beta}_0 + \hat{\beta}_{-0} \hat{C}^{slow}(F_t) + \hat{\gamma} Y_t$

$\hat{F}_t \leftarrow \hat{\beta}_{-0} \hat{C}^{slow}(F_t) - \hat{\gamma} Y_t$

$d \leftarrow \text{VAR}(\hat{F}_t, Y_t)$  lag selection according to AIC

FAVAR estimation results  $\leftarrow \text{VAR}(d)$  in  $(\hat{F}_t', Y_t')$

---

Explained procedure was used in Bernanke et al. (2005) for producing impulse response functions to policy shock. Purification is necessary for producing factors that has not respond contemporaneously to policy shock,

since authors used recursive identification with FFR (Federal Funds Rate) ordering last. For forecasting purposes, one may try to remove impact of another variables (from  $X_t$ ) or skip factor purification procedure at all. Using variables in  $Y_t$  at (13) will not change result of reduced form VAR (14) that is finally used for series prediction.

## 2.4 Model's specifications

Once econometrician has estimated  $\hat{\Phi}(L)$  in (14), h-horizon forecast for  $(\hat{F}_{t-1}, Y_{t-1})$  is nothing but:

$$\begin{bmatrix} \hat{F}_{t+h-1} \\ \hat{Y}_{t+h-1} \end{bmatrix} = \hat{\Phi}(L)^h \begin{bmatrix} \hat{F}_{t-1} \\ Y_{t-1} \end{bmatrix}$$

For a particular variable forecast, researcher takes corresponding component of  $\hat{Y}_{t+h-1}$  vector. One should also specify parameter  $K$  representing number of latent factors. I find, that  $K = 1$  is sufficient in most cases. However,  $K = 2$  sometimes helps to obtain more accurate results in short-term prediction, Interesting that  $K \geq 3$  generally increases model's error for almost all horizons. There is a space for experiments still, one may receive more accurate results at some horizons selecting proper variables for  $Y_t$ .

Another approach, is to treat variable of interest  $z_t$  as "informational" (included in  $X_t$ ), but I found no clear evidence that such method is more preferable than assuming  $z_t$  being observed. Such different FAVARs produce quite similar results, however there is a still place for experiments in choice of  $Y_t$  vector components.

## 3 Data

Data was collected from Thomson Reuters database. Dataset includes 199 monthly variables, that are assumed to represent economic situation in Rus-



sia, starting from 01/01/2000 and contains 240 observations maximum. Original data involved missings, such variables were dropped before the estimation procedure. To be in line with Bernanke et al. (2005) original time series were transformed (one or two times differencing) to induce stationarity<sup>4</sup>. In addition, for convenience original time series were standardised to have zero mean and unit standard deviation. Though, in order to obtain fair results, seasonal component in considered time series was not removed, since benchmark models are capable of handling seasonality in data. One can still decompose original time series on seasonal and trend components, though FAVAR methodology cannot be applied to produce seasonal component forecasts, so it becomes mix of FAVAR and other model for seasonal component. Probably such an approach will yield more accurate forecasts, but the goal of this paper to check pure FAVAR model prediction power.

Depending on variable being forecasted different number of observations was used: 238 in case of CPI and unemployment rate, 138 in case of GDP Y/Y % change forecasts.

## 4 Results

In order to understand FAVAR model's predictive power I decided to compare its results on cross-validation with popular univariate models, such as: ETS, ARIMA and simple random walk with drift; and restricted to FAVAR model – standard VAR ( $K = 0$ ). I treat variable of interest as being observed, so include it directly in  $Y_t$  vector. For ARIMA and ETS optimal model selection I used auto-ARIMA and auto-ETS functions in "forecast" R package, that select best model specification according to AIC. Optimal lag of resulting VAR models is also selected using AIC.

---

<sup>4</sup>According to KPSS test

|      | FAVAR | ARIMA | ETS  | VAR  | RWD  |
|------|-------|-------|------|------|------|
| h=1  | 0.70* | 0.86  | 0.76 | 0.83 | 0.93 |
| h=2  | 0.95* | 1.09  | 1.02 | 1.07 | 1.37 |
| h=3  | 1.01* | 1.07  | 1.06 | 1.12 | 1.54 |
| h=4  | 0.96* | 1.03  | 1.10 | 1.13 | 1.58 |
| h=5  | 0.91* | 1.03  | 1.05 | 1.13 | 1.49 |
| h=6  | 0.95* | 1.07  | 1.07 | 1.15 | 1.47 |
| h=7  | 1.00* | 1.05  | 1.19 | 1.17 | 1.65 |
| h=8  | 1.06  | 0.96* | 1.24 | 1.19 | 1.78 |
| h=9  | 1.05  | 0.99* | 1.26 | 1.18 | 1.75 |
| h=10 | 0.94* | 0.98  | 1.21 | 1.02 | 1.57 |
| h=11 | 0.90* | 0.99  | 1.23 | 0.97 | 1.35 |
| h=12 | 0.90* | 1.04  | 1.16 | 0.98 | 1.08 |

Table 1: RMSE  $\times 100$  for CPI forecasts

Table 1 shows cross-validation results for CPI forecasting. FAVAR specification includes two variables in  $Y_t = (aRUCPI_t, RUCBIR = ECI_t)'$  and one latent factor  $K = 1$ . Interesting that increasing number of factors deteriorates model's quality. One can see, that such FAVAR specification completely outperform other models almost at each horizon.

|     | FAVAR | ARIMA | ETS   | VAR   | RWD   |
|-----|-------|-------|-------|-------|-------|
| h=1 | 10.31 | 6.97* | 9.90  | 7.13  | 11.39 |
| h=2 | 10.25 | 6.95  | 9.75  | 6.74* | 10.97 |
| h=3 | 8.68  | 6.56  | 9.90  | 6.52* | 10.06 |
| h=4 | 8.04  | 6.60  | 9.86  | 6.36* | 12.06 |
| h=5 | 6.88  | 5.86  | 10.02 | 6.22* | 12.00 |

|      |       |       |      |      |       |
|------|-------|-------|------|------|-------|
| h=6  | 6.10  | 5.94* | 9.70 | 6.44 | 11.36 |
| h=7  | 5.74* | 6.00  | 9.27 | 6.56 | 12.64 |
| h=8  | 5.71* | 6.09  | 9.57 | 6.44 | 10.90 |
| h=9  | 6.15* | 6.15  | 9.66 | 6.17 | 10.40 |
| h=10 | 5.90* | 5.96  | 9.21 | 6.42 | 10.10 |
| h=11 | 5.86* | 5.89  | 9.67 | 6.52 | 9.32  |
| h=12 | 6.43  | 5.70* | 9.51 | 6.43 | 9.09  |

Table 2: RMSE  $\times$  100 for unemployment rate forecasts

Table 2 shows models' cross-validation errors in attempts to forecast unemployment rate in Russia. FAVAR specification is such that  $Y_t = (aRUCPI_t, RUUNR = ECI_t)'$  and  $K = 1$ . As with CPI, increasing of  $K$  leads to weak performance. As one can see, for such variable, FAVAR produces best forecasts only on long horizons. Yet one can experiment with proper  $Y_t$  selection to obtain lower error at short-term forecasts.

|      | FAVAR  | ARIMA | ETS    | VAR    | RWD   |
|------|--------|-------|--------|--------|-------|
| h=1  | 19.88* | 22.34 | 21.67  | 20.10  | 38.56 |
| h=2  | 20.44* | 23.22 | 22.07  | 20.79  | 25.65 |
| h=3  | 21.92* | 23.79 | 22.59  | 22.37  | 35.73 |
| h=4  | 22.40  | 23.13 | 22.35  | 22.31* | 31.02 |
| h=5  | 22.96  | 23.49 | 22.94  | 22.90* | 34.74 |
| h=6  | 22.75  | 22.98 | 22.82  | 22.68* | 30.00 |
| h=7  | 23.06* | 23.66 | 23.15  | 23.17  | 35.66 |
| h=8  | 23.46* | 24.45 | 23.53  | 23.51  | 34.69 |
| h=9  | 24.52  | 25.21 | 24.29  | 24.25* | 33.81 |
| h=10 | 22.74  | 25.06 | 22.72* | 22.75  | 33.99 |

|      |        |       |       |        |       |
|------|--------|-------|-------|--------|-------|
| h=11 | 22.98  | 24.59 | 22.74 | 22.70* | 33.43 |
| h=12 | 23.09* | 24.29 | 23.15 | 23.17  | 25.37 |

---

Table 3: RMSE  $\times$  100 for GDP forecasts

At table 3 one can find cross-validation error for corresponding models, that are used for GDP growth rate forecasts. It is important to mention again, that for this variable only 168 observations were used, because of missings for interested time series. That might one reason explaining model's performance. Surprisingly, FAVAR produces best forecasts mostly for short-term, rather than for long-term points as was expected. FAVAR in  $Y_t = (RUGDP = ECI_t, RUCBIR = ECI_t)'$  with  $K = 2$  shows best performance on 6 out of 12 horizons, though difference is significant only for  $h = \overline{1,3}$ .

## 5 Conclusion

Results showed that FAVAR is quite competitive model in a forecasting task, even though it was initially developed as the solution for "price puzzle", that is more about impulse response functions, rather than predictions. For some series FAVAR can completely outperform traditional univariate models as well as multivariate VAR. In addition, FAVAR allows incorporating as much variables as one wants and there are reasons to believe that this will improve model's performance. Nevertheless, it still remains a question whether cost of data collection outweigh potential forecast improvement for standard VAR.

# Appendixes

## A Data Description

Table below provides information about used dataset. An asterisk \*, after mnemonic of some series shows that such variables were assumed to be slow-moving. Each time series is provided with transformation applied to the original data:  $I$  – no transformations,  $\Delta$  – first difference and  $\Delta^2$  – second difference.

| Mnemonic        | Description   |          |
|-----------------|---|----------|
| 1. aRUPTTTT/C*  | Passenger Transport Turnover, Price Index   | $I$      |
| 2. aRUFTTCTT/C* | Freight Transport Turnover, Cargo transport, Volume Index   | $I$      |
| 3. aRUCEXBA*    | Exports of Goods, Balance of Payments Basis, Standardized, Current Prices                                       | $\Delta$ |
| 4. aRUCIMBA*    | Imports of Goods, Balance of Payments Basis, Standardized, Current Prices                                       | $\Delta$ |
| 5. aRUCBOPA*    | Visible Trade Balance, Balance of Payments Basis, Standardized, Current Prices                                  | $I$      |
| 6. aRUCEXBPA*   | Exports of Goods, Balance of Payments Basis, % month on month, Standardized, Chg P/P, Current Prices            | $I$      |
| 7. aRUCIMBPA*   | Imports of Goods, Balance of Payments Basis, % month on month, Standardized, Chg P/P, Current Prices            | $I$      |
| 8. aRUCBOPPA*   | Visible Trade Balance, Balance of Payments Basis, month on month, Standardized, Absolute change, Current Prices | $I$      |

|                  |   |            |
|------------------|---|------------|
| 9. aRUCEXBYA*    | Exports of Goods, Balance of Payments Basis, % year on year, Standardized, Chg Y/Y, Current Prices            | $\Delta$   |
| 10. aRUCIMBYA*   | Imports of Goods, Balance of Payments Basis, % year on year, Standardized, Chg Y/Y, Current Prices            | $\Delta$   |
| 11. aRUCBOPYA*   | Visible Trade Balance, Balance of Payments Basis, year on year, Standardized, Absolute change, Current Prices | $I$        |
| 12. RUCBIR=ECI   | Central bank key rate   | $\Delta$   |
| 13. aRUBCFBNKA   | Bank Lending: Loans to banks (in frgn. cur.), Current Prices  | $\Delta$   |
| 14. aRUDOMDET    | Domestic Debt, Current Prices   | $\Delta^2$ |
| 15. aRUBCDBNKA   | Bank Lending: Loans to banks (in rubles), Current Prices  | $\Delta$   |
| 16. RUCPIY=ECI*  | CPI, Chg Y/Y  | $\Delta$   |
| 17. aRUCPI*      | CPI, Price Index  | $\Delta$   |
| 18. aRUCCPIYF*   | CPI, % year on year, Standardized, Chg Y/Y, Price Index   | $\Delta$   |
| 19. aRUCCPIF/C*  | CPI, Standardized, Price Index  | $\Delta$   |
| 20. aRUCPFYYF*   | CPI, Food products, Chg Y/Y   | $\Delta$   |
| 21. aRUCPCORIF*  | CPI, Core inflation, Chg Y/Y  | $\Delta$   |
| 22. aRUCCPIYE/A* | CPI, % year on year, Standardized, Chg Y/Y, Price Index, SA   | $\Delta$   |
| 23. aRUCCPIPF*   | CPI, % month on month, Standardized, Chg P/P, Price Index   | $\Delta$   |

|                   |   |          |
|-------------------|---|----------|
| 24. aRUCCORYE/A*  | Core CPI, Standardized, Chg Y/Y, Price Index, SA              | Δ        |
| 25. aRUCCPIPE/A*  | CPI, % month on month, Standardized, Chg P/P, Price Index, SA | Δ        |
| 26. aRUCPNYYF*    | CPI, Non-food products, Chg Y/Y                               | <i>l</i> |
| 27. aRUCPIF*      | CPI, Food and beverages, Price Index                          | <i>l</i> |
| 28. aRUCPISERV*   | CPI, Services, Price Index                                    | Δ        |
| 29. aRUCPIXF*     | CPI, Non-food goods, Price Index                              | <i>l</i> |
| 30. aRUCPGOODF/C* | CPI, Goods, Price Index                                       | <i>l</i> |
| 31. aRUGOODSF*    | CPI, Goods, Chg Y/Y   | Δ        |
| 32. aRUCPFPFVF/C* | CPI, Food products without fruits and vegetables, Price Index | Δ        |
| 33. aRUCCPIE/CA*  | CPI, Standardized, Price Index, SA                            | Δ        |
| 34. aRUCCORF/C*   | Core CPI, Standardized, Price Index                           | Δ        |
| 35. aRUCCORE/CA*  | Core CPI, Standardized, Price Index, SA                       | Δ        |
| 36. aRUCCORPF*    | Core CPI, Standardized, Chg P/P, Price Index                  | Δ        |
| 37. aRUCCORPE/A*  | Core CPI, Standardized, Chg P/P, Price Index, SA              | Δ        |
| 38. aRUCCORYF*    | Core CPI, Standardized, Chg Y/Y, Price Index                  | Δ        |
| 39. RUCPIY=ECIX*  | CPI, Chg Y/Y  | Δ        |
| 40. aRUCPFWFVF*   | CPI, Food products without fruits and vegetables, Chg Y/Y     | Δ        |
| 41. aRUCPICORY/C* | CPI, Core CPI, Price Index                                    | Δ        |
| 42. aRUCPNPSVF*   | CPI, Paid services, Chg Y/Y                                   | Δ        |
| 43. aRUCPICOR/C*  | CPI, Core CPI, Price Index                                    | Δ        |
| 44. aRUCLEAD      | Composite leading indicators, Trend restored, SA              | Δ        |

|                |   |          |
|----------------|---|----------|
| 45. aRUEMPLMT* | Employment, Volume  | Δ        |
| 46. aRUBISRXBR | BIS, Real Broad Effective Exchange Rate Index   | Δ        |
| 47. aRUCXTWF/C | Trade Weighted nominal exchange rate, Standardized, Price Index                                     | Δ        |
| 48. aRUCXTWPF  | Trade Weighted nominal exchange rate, % month on month, Standardized, Chg P/P, Price Index          | <i>l</i> |
| 49. aRUCXTWYF  | Trade Weighted nominal exchange rate, % year on year, Standardized, Chg Y/Y, Price Index            | <i>l</i> |
| 50. aRUCXTRF/C | Trade Weighted real exchange rate, Standardized, Price Index  | Δ        |
| 51. aRUCXTRPF  | Trade Weighted real exchange rate, % month on month, Standardized, Chg P/P, Price Index             | <i>l</i> |
| 52. aRUCXTRYF  | Trade Weighted real exchange rate, % year on year, Standardized, Chg Y/Y, Price Index               | <i>l</i> |
| 53. aRUIRECE/C | Real effective exchange rate (REER) based on consumer price index, 2010=100, Not SA, Price Index    | Δ        |
| 54. aRUINECE/C | Nominal effective exchange rate (NEER) based on consumer price index, 2010=100, Not SA, Price Index | Δ        |
| 55. aRUBISNXBR | BIS, Nominal Broad Effective Exchange Rate Index  | Δ        |
| 56. aRUXRUSD   | Russian roubles to US \$  | Δ        |
| 57. RUGDP=ECI* | GDP, Chg Y/Y  | Δ        |
| 58. aRUTREV*   | Revenue, Federal budget, Current Prices   | Δ        |
| 59. aRUPFEXP*  | Expenditure, Current Prices   | Δ        |



|                 |   |          |
|-----------------|---|----------|
| 60. aRUCBUDIC*  | Revenue, Consolidated budget, incomes total, Current Prices                                       | $\Delta$ |
| 61. aRUEXPDT*   | Expenditure, Current Prices   | $\Delta$ |
| 62. aRUGDEF*    | Deficit/Surplus, Federal budget, Current Prices   | $l$      |
| 63. aRUCGOVA*   | Central government Deficit/Surplus, Standardized, Current Prices                                  | $l$      |
| 64. aRUCGOVPA*  | Central government Deficit/Surplus, month on month, Standardized, Absolute change, Current Prices | $l$      |
| 65. aRUCGOVYA*  | Central government Deficit/Surplus, year on year, Standardized, Absolute change, Current Prices   | $l$      |
| 66. aRSCGOVBLA* | Public Finances, Central Government, Budget, Balance, Deficit/Surplus, Current Prices             | $l$      |
| 67. aRUDSCBUD*  | Deficit/Surplus, Consolidated budget, Current Prices  | $l$      |
| 68. aRUDCTDA    | Dwellings Commenced, Total dwellings area, Volume   | $\Delta$ |
| 69. aRUDCTI     | Dwellings Commenced, Total dwellings area, Price Index  | $\Delta$ |
| 70. aRUPPICONs* | Producer Prices, Investment products, Chg P/P   | $\Delta$ |
| 71. RUTRD=ECI*  | Trade Balance, Total, Free On Board, Current Prices   | $l$      |
| 72. aRUEXP*     | Exports, FOB, Current Prices  | $\Delta$ |
| 73. aRUIMP*     | Imports, FOB, Current Prices  | $\Delta$ |
| 74. aRUTBAL*    | Trade Balance, FOB, Current Prices  | $l$      |
| 75. aRUEXPNGC*  | Exports, Natural gas, Current Prices  | $l$      |

|                  |  |          |
|------------------|--|----------|
| 76. aRUEXPCRD*   | Exports, Crude oil, Current Prices   | <i>l</i> |
| 77. aRUEXPDIESL* | Exports, Diesel fuel, Current Prices   | <i>l</i> |
| 78. aRUIMPPRO*   | Exports, Petroleum products, Current Prices  | <i>l</i> |
| 79. aRUIMP MED*  | Imports, Medicines, Current Prices   | <i>l</i> |
| 80. aRUIMPMNEQ*  | Imports, Machinery and equipment, Current Prices   | <i>l</i> |
| 81. aRUIMPCLOTH* | Imports, Clothing, Current Prices  | $\Delta$ |
| 82. aRUEXPMOM*   | Exports, Ferrous metals, Current Prices  | <i>l</i> |
| 83. aRUCEXPB/A*  | Merchandise Exports, Standardized, Current Prices, SA                                    | $\Delta$ |
| 84. aRUCIMPB/A*  | Merchandise Imports, Standardized, Current Prices, SA                                    | $\Delta$ |
| 85. aRUCEXPPB/A* | Merchandise Exports, % month on month, Standardized, Chg P/P, Current Prices, SA         | <i>l</i> |
| 86. aRUCEXPYB/A* | Merchandise Exports, % year on year, Standardized, Chg Y/Y, Current Prices, SA           | $\Delta$ |
| 87. aRUCIMPPB/A* | Merchandise Imports, % month on month, Standardized, Chg P/P, Current Prices, SA         | <i>l</i> |
| 88. aRUCIMPYB/A* | Merchandise Imports, % year on year, Standardized, Chg Y/Y, Current Prices, SA           | $\Delta$ |
| 89. aRUCVISB/A*  | Visible Trade Balance, Standardized, Current Prices, SA                                  | <i>l</i> |
| 90. aRUCVISPB/A* | Visible Trade Balance, month on month, Standardized, Absolute change, Current Prices, SA | <i>l</i> |
| 91. aRUCVISYB/A* | Visible Trade Balance, year on year, Standardized, Absolute change, Current Prices, SA   | <i>l</i> |
| 92. RUTRD=ECIX*  | Trade Balance, Total, Free On Board  | <i>l</i> |

|                    |  |          |
|--------------------|--|----------|
| 93. RUIP=ECI*      | Production, IP Total , Chg Y/Y   | <i>l</i> |
| 94. aRUIP*         | Production, Chg Y/Y  | <i>l</i> |
| 95. aRUIPOIL*      | Production, Oil output, including gas condensate, Volume                                 | $\Delta$ |
| 96. aRUCINDG/A*    | Industrial Production Index, Standardized, Volume Index, SA                              | $\Delta$ |
| 97. aRUIPNAG*      | Production, Natural gas output, Volume   | <i>l</i> |
| 98. aRUIPMANH/C*   | Production, Manufacturing, Volume Index  | <i>l</i> |
| 99. aRUIPMANG*     | Production, Manufacturing, Chg Y/Y   | <i>l</i> |
| 100. aRUINP/C*     | Production, Industry, Volume Index   | $\Delta$ |
| 101. aRUCINDPG/A*  | Industrial Production Index, Standardized, Chg P/P, Volume Index, SA                     | <i>l</i> |
| 102. aRUCINDYG/CA* | Industrial Production Index, % year on year, Standardized, Chg Y/Y, Volume Index, SA     | <i>l</i> |
| 103. RUIP=ECIX*    | Production, IP Total, Chg Y/Y  | <i>l</i> |
| 104. aRUPRATE      | Policy Rates, Minimum Rate on 7 Day Repo   | $\Delta$ |
| 105. aRUINTRES     | Reserves, Gross international, Current Prices  | $\Delta$ |
| 106. aRUFXRES      | Reserves, Foreign currency reserves, Current Prices                                      | $\Delta$ |
| 107. aRUFCRES      | Reserves, Foreign currency, Current Prices   | $\Delta$ |
| 108. aRUCRESA      | Official international reserves, Standardized, Current Prices                            | $\Delta$ |
| 109. aRURESGLD     | Reserves, Gold, Current Prices   | $\Delta$ |
| 110. aRUCRESPA     | Official international reserves, % month on month, Standardized, Chg P/P, Current Prices | $\Delta$ |
| 111. aRUCRESYA     | Official international reserves, % year on year, Standardized, Chg Y/Y, Current Prices   | $\Delta$ |

|                  |  |            |
|------------------|--|------------|
| 112. aRURTM1A1A  | Official reserve assets, Foreign currency, Current Prices                    | $\Delta^2$ |
| 113. aRURESSDR   | Reserves, Special Drawing Rights, Current Prices                             | $\Delta$   |
| 114. aRURTM1AA   | Official reserve assets, Overall, Current Prices                             | $\Delta$   |
| 115. aRURESIMF   | Reserves, Reserve position in the IMF, Current Prices                        | $\Delta$   |
| 116. aRUM2       | Money supply M2 by national definition, Current Prices                       | $\Delta$   |
| 117. aRUM0       | Money supply M0, Current Prices  | $\Delta$   |
| 118. aRUNM1      | Money Supply - M1, Current Prices  | $\Delta$   |
| 119. aRUMNBAS    | Monetary base (narrow definition), Current Prices                            | $\Delta$   |
| 120. aRUMBASMOC  | Monetary base (broad definition), Current Prices                             | $\Delta$   |
| 121. aRUMSMBBRA  | Broad money liabilities, Current Prices                                      | $\Delta$   |
| 122. aRUCMS2B/A  | Money Supply M2, Standardized, Current Prices, SA                            | $\Delta$   |
| 123. aRUCMS2PB/A | Money Supply M2, % month on month, Standardized, Chg P/P, Current Prices, SA | $l$        |
| 124. aRUCMS2YB/A | Money Supply M2, % year on year, Standardized, Chg Y/Y, Current Prices, SA   | $\Delta$   |
| 125. aRUCMS1B/A  | Money Supply M1, Standardized, Current Prices, SA                            | $\Delta$   |
| 126. aRUCMS0B/A  | Money Supply M0, Standardized, Current Prices, SA                            | $\Delta$   |

|                   |   |          |
|-------------------|---|----------|
| 127. aRUCMS1PB/A  | Money Supply M1, Standardized, Chg P/P, <i>l</i><br>Current Prices, SA              |          |
| 128. aRUCMS0PB/A  | Money Supply M0, Standardized, Chg P/P, <i>l</i><br>Current Prices, SA              |          |
| 129. aRUCMS1YB/A  | Money Supply M1, Standardized, Chg Y/Y, $\Delta$<br>Current Prices, SA              |          |
| 130. aRUCMS0YB/A  | Money Supply M0, Standardized, Chg Y/Y, $\Delta$<br>Current Prices, SA              |          |
| 131. RUPPI=ECI*   | Producer Prices, Chg P/P  | <i>l</i> |
| 132. RUPPIY=ECI*  | Producer Prices, Chg Y/Y  | <i>l</i> |
| 133. aRUPPI/C*    | Producer Prices, Price Index  | $\Delta$ |
| 134. aRUPPITCP*   | Producer Prices, Chg P/P  | <i>l</i> |
| 135. aRUPPIF*     | Producer Prices, Chg Y/Y  | <i>l</i> |
| 136. aRUCPPIE/CA* | Producer Prices, PPI, Standardized, Price In-<br>dex, SA                            | $\Delta$ |
| 137. aRUPPIAR*    | Producer Prices, Chg P/P  | <i>l</i> |
| 138. aRUCPPIF/C*  | Producer Prices, PPI, Standardized, Price Index                                     | $\Delta$ |
| 139. aRUCPPIPE/A* | Producer Prices, PPI, % month on month, Stan-<br>dardized, Chg P/P, Price Index, SA | <i>l</i> |
| 140. aRUCPPIPF*   | Producer Prices, PPI, % month on month, Stan-<br>dardized, Chg P/P, Price Index     | <i>l</i> |
| 141. aRUCPPIYE/A* | Producer Prices, PPI, % year on year, Standard-<br>ized, Chg Y/Y, Price Index, SA   | $\Delta$ |
| 142. aRUCPPIYF*   | Producer Prices, PPI, % year on year, Standard-<br>ized, Chg Y/Y, Price Index       | $\Delta$ |
| 143. RUPPI=ECIX*  | Producer Prices, Chg P/P  | <i>l</i> |
| 144. RUPPIY=ECIX* | Producer Prices, Chg Y/Y  | <i>l</i> |

|                   |   |          |
|-------------------|---|----------|
| 145. aRUPPIYAR*   | Producer Prices, Chg Y/Y                                    | <i>l</i> |
| 146. aRUPPITC*    | Producer Prices, Price Index                                | $\Delta$ |
| 147. RURSLY=ECI*  | Retail Sales YY, Price Index                                | $\Delta$ |
| 148. aRURSLS*     | Retail Trade Turnover, Current Prices                       | $\Delta$ |
| 149. aRURSLSTO*   | Retail Trade Turnover, Price Index                          | $\Delta$ |
| 150. aRUCRETF/C*  | Retail Sales, Standardized, Price Index                     | $\Delta$ |
| 151. aRUCRETE/CA* | Retail Sales, Standardized, Price Index, SA                 | $\Delta$ |
| 152. aRUCRETPF*   | Retail Sales, Standardized, Chg P/P, Price Index            | <i>l</i> |
| 153. aRUCRETPE/A* | Retail Sales, Standardized, Chg P/P, Price Index, SA        | $\Delta$ |
| 154. aRUCRETYF*   | Retail Sales, Standardized, Chg Y/Y, Price Index            | $\Delta$ |
| 155. aRUCRETYE/A* | Retail Sales, Standardized, Chg Y/Y, Price Index, SA        | $\Delta$ |
| 156. RURSLY=ECIX* | Retail Sales YY   | $\Delta$ |
| 157. aRURTTNFC*   | Retail Trade Turnover, Non-food commodities, Current Prices | $\Delta$ |
| 158. aRURSLS/C*   | Retail Trade Turnover, Volume Index                         | <i>l</i> |
| 159. aRURTTFC*    | Retail Trade Turnover, Food commodities, Current Prices     | $\Delta$ |
| 160. aRUGBOND     | Bid   | $\Delta$ |
| 161. aRUSHRPRCF   | MICEX, Composite Index, Price Index                         | $\Delta$ |
| 162. aRUWANMAVWG* | Wages, Nominal average monthly per worker, Current Prices   | $\Delta$ |
| 163. aRUWSAMTCPP* | Wages, Real average monthly, Volume Index                   | $\Delta$ |
| 164. aRUCWAGF/C*  | Wages, total, Standardized, Price Index                     | $\Delta$ |

|                   |   |          |
|-------------------|---|----------|
| 165. aRUCWAGPF*   | Wages, total, % month on month, Standardized, Chg P/P, Price Index      | <i>l</i> |
| 166. aRUCWAGYF*   | Wages, total, % year on year, Standardized, Chg Y/Y, Price Index        | $\Delta$ |
| 167. aRUAVGMANW*  | Average monthly accrued nominal wages, Current Prices                   | $\Delta$ |
| 168. RUUNR=ECI*   | Unemployment, Rate  | $\Delta$ |
| 169. RURWGE=ECI*  | Real Wages YY, Chg Y/Y  | $\Delta$ |
| 170. aRUUNR*      | Unemployment, Rate  | $\Delta$ |
| 171. aRUUNTOTR*   | Unemployment rate, ILO  | $\Delta$ |
| 172. aRUPOPEA*    | Economically active population - Unemployed, Volume                     | $\Delta$ |
| 173. aRUCUNPQ/A*  | Unemployment rate, Standardized, SA                                     | $\Delta$ |
| 174. aRUUEMPILO*  | Unemployment, Total persons (ILO), aged 15 and over, Volume             | $\Delta$ |
| 175. aRUUEMP*     | Unemployment, Officially registered, Volume                             | $\Delta$ |
| 176. aRUUNRAR*    | Unemployment, Rate  | $\Delta$ |
| 177. aRUCUNPPQ/A* | Unemployment rate, month on month, Standardized, SA                     | <i>l</i> |
| 178. aRUCUNPYQ/A* | Unemployment rate, year on year, Standardized, SA                       | <i>l</i> |
| 179. aRUCUNPO*    | Unemployment Level, Standardized, Volume                                | $\Delta$ |
| 180. aRUCUNPP*    | Unemployment Level, Standardized, Volume                                | $\Delta$ |
| 181. aRUCUNPPO/A* | Unemployment Level, % month on month, Standardized, Chg P/P, Volume, SA | <i>l</i> |
| 182. aRUCUNPPP*   | Unemployment Level, % month on month, Standardized, Chg P/P, Volume     | <i>l</i> |

|                   |  |          |
|-------------------|--|----------|
| 183. aRUCUNPYO*   | Unemployment Level, % year on year, Standard-<br>ized, Chg Y/Y, Volume | <i>l</i> |
| 184. aRUCUNPYP*   | Unemployment Level, % year on year, Standard-<br>ized, Chg Y/Y, Volume | <i>l</i> |
| 185. RURWGE=ECIX* | Real Wages YY, Chg Y/Y   | $\Delta$ |
| 186. RUUNR=ECIX*  | Unemployment, Rate   | $\Delta$ |
| 187. aRUUNRRILO*  | Unemployment, Rate, Officially registered<br>(ILO)                     | $\Delta$ |
| 188. .IRTS        | RTS Index  | $\Delta$ |
| 189. .IMOEX       | MOEX Russia Index  | $\Delta$ |
| 190. .RTSOG       | RTS Oil & Gas Index  | $\Delta$ |
| 191. .RTSTL       | RTS Telecom Index  | $\Delta$ |
| 192. BrentP       | Brent Price nominal  | $\Delta$ |
| 193. BrentPP      | Brent Price (%)  | <i>l</i> |
| 194. WTIP         | WTI Price nominal  | $\Delta$ |
| 195. WTIPP        | WTI Price (%)  | <i>l</i> |
| 196. RUEURP       | Russian roubles to EU €, Close   | $\Delta$ |
| 197. RUEURMAX     | Russian roubles to EsU €, Max  | $\Delta$ |
| 198. RUEURMIN     | Russian roubles to EU €, Min   | $\Delta$ |
| 199. RUEURPP      | Russian roubles to EU €, % change                                      | <i>l</i> |

---



## 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.