Forecasting with BVAR: an application on Russian data

Boris Demeshev[1] Oxana Malakhovskaya[2]

[1]Department of Applied Economics
National Research University Higher School of Economics

[2]Department of Theoretical Economics
National Research University Higher School of Economics

II International Conference: Modern Econometric Tools and Applications, September 2015

Motivation

- Accurate macroeconomic forecasts are important for policy making.
- Central banks monitor a large set of macroeconomic indicators to determine the policy
- Therefore, a model being used for forecasting purposes must be suitable for samples with large cross-sectional dimension to avoid the omitted-variable bias.

Motivation(2)

- Vector autoregressions have become a widely-used tool for forecasting. However, unrestricted VARs bear the risk of overparametrization even for samples of moderate size.
- Using of Bayesian estimation may alleviate this problem. The Bayesian shrinkage is done by imposing restrictions on parameters in the form of prior distributions.
- Recently many papers has claimed that, in terms of forecasting accuracy, medium and large BVAR outperform their small dimensional counterparts.
- Application of Bayesian econometrics on Russian data is scarce

Objective of the paper

The objective of the paper is:

- forecasting of macroeconomic indicators for Russian economy with BVARs of different size
- comparing their forecasting accuracy with one of competing models (RW and unrestricted VARs)

Our underlying hypotheses are:

- BVARs outperform the competing models in terms of forecasting accuracy
- high-dimensional BVARs forecast better than low-dimensional ones

Our baseline specification is a standard BVAR with a Normal-inverted Wishart conjugate prior.
VAR in reduced form:

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

where y_{it} is a $m \times 1$ vector of variables,
The model can also be written in a companion form:

$$Y = X\Phi + E, \tag{2}$$

where $Y = [y_1, y_2, \ldots, y_T]'$,$X = [x_1, x_2, \ldots, x_T]'$,
$x_t = [y'_{t-1} \ldots y'_{t-p}\ 1]'$, $\Phi = [\Phi_1 \ldots \Phi_p\ \Phi_{ex}]'$, $E = [\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T]'$

Bayesian estimate combines a likelyhood function $L(Y|\Phi,\Sigma)$, with a prior distribution $p(\Phi,\Sigma|Y)$ and results in a posterior distributions of parameters $p(\Phi,\Sigma|Y)$. To find the distribution, a corollary from the Bayes rule is employed:

$$p(\Phi,\Sigma|Y) \propto p(\Phi,\Sigma)L(Y|\Phi,\Sigma) \tag{3}$$

Conjugate Normal-inverted Wishart prior

A normal-inverted Wishart prior distribution can be written as:

$$\begin{cases} \Sigma \sim \mathcal{IW}(\underline{S}, \underline{\nu}) \\ \Phi|\Sigma \sim \mathcal{N}(\underline{\Phi}, \Sigma \otimes \underline{\Omega}) \end{cases}, \tag{4}$$

where $\underline{\Phi} = [\underline{\Phi}_1 \ldots \underline{\Phi}_p \ \underline{\Phi}_{ex}]'$ and

$$(\underline{\Phi}_l)_{ij} = \begin{cases} \delta_i \ i = j, l = 1; \\ 0, \quad \text{otherwise} \end{cases} \tag{5}$$

$$\underline{\Omega} = \begin{pmatrix} \underline{\Omega}_{lag=1} & 0_{m\times m} & \cdots & 0_{m\times m} & 0_{m\times 1} \\ 0_{m\times m} & \underline{\Omega}_{lag=2} & \cdots & 0_{m\times m} & 0_{m\times 1} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0_{m\times m} & 0_{m\times m} & \cdots & \underline{\Omega}_{lag=p} & 0_{m\times 1} \\ 0_{1\times m} & 0_{1\times m} & \cdots & 0_{1\times m} & \underline{\Omega}_{const} \end{pmatrix} \tag{6}$$

$\underline{\Omega}_{lag=l}$ is $m \times m$ matrix, and its diagonal elements are:

$$(\underline{\Omega}_{lag=l})_{jj} = \left(\frac{\lambda_{tight}}{l^{\lambda_{lag}}\sigma_j}\right)^2 \quad \underline{\Omega}_{const} = \lambda^2_{const} \tag{7}$$

So $\underline{\Phi} = \underline{\Phi}(\delta)$ and $\underline{\Omega} = \underline{\Omega}(\lambda_{tight}, \lambda_{lag}, \lambda_{const})$ We fix λ_{lag} at unity (as in an overwhelming majority of other studies) and $\lambda_{const} = Inf$ and we optimize on λ_{tight}

- While BVAR in low-dimensional space were widely used for macroeconomic analysis, their use for data-rich environment were limited until 2010. The reason was a general agreement that the Bayesian shrinkage is insufficient to solve the over-parametrization problem in high cross-sectional dimension samples.
- De Mol, Giannone, and Reichlin (2008) show that the Bayesian methods can be successfully applied in data-rich environment if the degree of shrinkage is set relative to cross-sectional dimension of the sample.
- Bańbura, Giannone, and Reichlin (2010) confirm and develop this claim for BVAR applied to a large set of US time-series.

Algorithms for choosing hyperparameters

Two ways of choosing optimal λ_{tight} in the literature are:

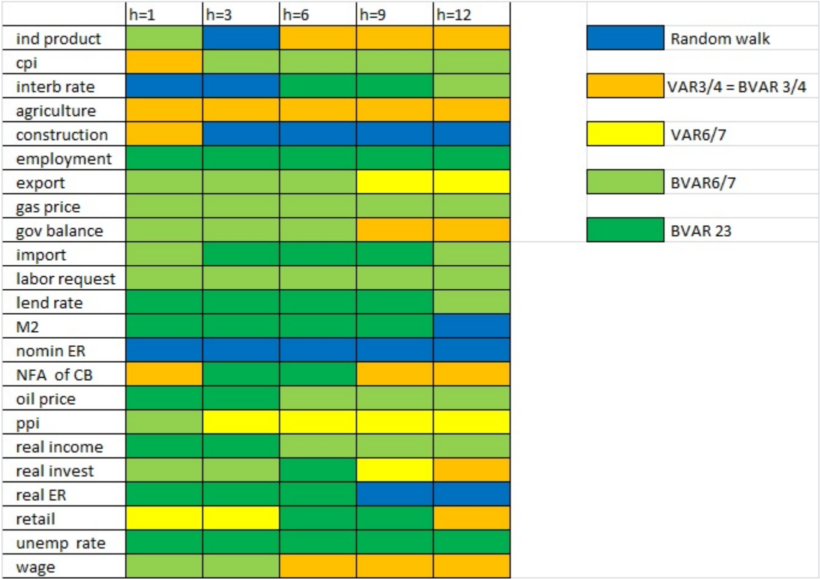
1. The shrinkage must be so tight to avoid an overparametrization. Three-variable unrestricted VAR is assumed to be parsimonious enough and it does not need any shrinkage. So λ_{tight} is chosen in such a way that average in-sample BVAR forecast on a training sample for several variables of interest is the same as 3-variable unrestricted VAR. This method was introduced by Bańbura, Giannone, and Reichlin (2010)
2. λ_{tight} is chosen so that to maximize the marginal data density. The algorithm was introduced by.... and used recently by Carriero, Clark, and Marcellino (2015)

In the paper we use both of these methods and compare the results

Our data

- 23 monthly time series running from January, 1996 to April, 2015
- series demonstrating seasonal behaviour are seasonally adjusted
- logarithms are applied to most of the series, with the exception of those already expressed in rates.

Results(1)


h=1
h=3
h=6
h=9
h=12
ind product
cpi
interb rate
agriculture
construction
employment
export
gas price
gov balance
import
labor request
lend rate
M2
nomin ER
NFA of CB
oil price
ppi
real income
real invest
real ER
retail
unemp rate
wage
Random walk
VAR3/4 = BVAR 3/4
VAR6/7
BVAR6/7
BVAR 23