Topic: **Bayesian VAR models**

Author: **Liana Isayan**

Bayesian Statistics, the 2nd Midterm

Introduction

The popularity of the Bayesian approach to econometric modelling continues to enjoy a growing number of followers and many macroeconomists argue that it has many important advantages over the classical (frequentist) approach dealing with identification issues, different data sources, misspecification, parameter uncertainty and a number of computational matters (Kotzé, 2017). Generally, VARs are frequently used in the study of macroeconomic data. Since VARs frequently require estimation of a large number of parameters, over-parameterization of VAR models is often a problem—with too few observations to estimate the parameters of the model. One approach for solving this problem is shrinkage, where we impose restrictions on parameters to reduce the parameter set. Bayesian VAR (BVAR) methods (Litterman, 1986; Doan, Litterman, and Sims, 1984; Sims and Zha, 1998) are one popular approach for achieving shrinkage, since Bayesian priors provide a logical and consistent method of imposing parameter restrictions.

The analysis on Bayesian VAR models here will be performed in the following order: firstly, the classical VAR models will be discussed, then the features of the Bayesian VAR models and their different types will be presented. At the end, with real data, a practical comparison of the classical and Bayesian models will be made.

Classical VAR models versus Bayesian VAR models

The vector autoregression (VAR) model extends the idea of univariate autoregression to k time series regressions, where the lagged values of k series appear as regressors. Put differently, in a VAR model we regress a vector of time series variables on lagged vectors of these variables. One may write the stationary, k-dimensional, VAR(p) process as:

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + C x_t + \epsilon_t ,$$

where

- $y_t = (y_{1t}, y_{2t}, \ldots, y_{Kt})'$ is a k x 1 vector of endogenous variables,
- $x_t = (x_{1t}, x_{2t}, \ldots, x_{dt})'$ is a d x 1 vector of exogenous variables,
- $A_1, \ldots, A_p$ are k x k matrices of lag coefficients to be estimated,
- C is a k x d matrix of exogenous variable coefficients to be estimated,
- $e_t = (e_{1t}, e_{2t}, \ldots, e_{Kt})'$ is a k x 1 white noise innovation process, with $E(e_t) = 0$, $E(e_t e_t') = \sum_e$, and $E(e_t e_t') = 0$ for $s \neq t$.

The last statement implies that the vector of innovations are contemporaneously correlated with full rank matrix $\sum_e$, but are uncorrelated with their leads and lags of the innovations and (assuming the usual orthogonality) uncorrelated with all of the right-hand side variables (IHS Markit, 2020).

VARs are now widely used in empirical macroeconomics (structural VARs, factor augmented VARs, time-varying parameter versions of these models, etc.). However, these models have a large number of parameters and, thus, over-parameterization problems may arise. This is particularly important because the number of parameters to be estimated grows qudratically to the number of variables modeled by the VAR. Bayesian methods have become increasingly popular as a shrinkage method to overcome these problems. Informative prior distributions can often mitigate the curse of dimensionality.

Bayesian VAR models combine a set of prior beliefs with the traditional VAR methods, the main target of which is to find the posterior moments of the parameter of interest. For instance, location and dispersion are the general estimates which are comparable to those obtained in classical estimation (namely the classical coefficient estimate and coefficient standard error). These point estimates can be easily derived from the posterior because the posterior distribution contains all the information available on the parameter.

Suppose that we have VAR(p) model given by:

$$y_t = a_0 + \sum_{j=1}^{p} A_j y_{t-j} + \epsilon_t ,$$

where y_t for t = 1,…,T is an vector containing observations on m different series and e_t is an vector of errors where we assume e_t is i.i.d. $N(0, \sum_e)$. For compactness we may rewrite the model as:

$$Y = XA + E,$$

or

$$y = (I_m \otimes X)\theta + e ,$$

where Y and E are T x m matrices and $X = (x_1, \ldots, x_t)'$ is a T x (mp + 1) matrix for $x_t = (1, y'_{t-1}, \ldots, y'_{t-q})$, I_m is the identity matrix of dimention m, θ = vec(A), and $e \sim N(0, \Sigma_\epsilon \otimes I_T)$. The likelihood function is:

$$l(\theta, \Sigma_\epsilon) \propto |\Sigma_\epsilon \otimes I_T|^{-1/2} \exp\left\{-\frac{1}{2}(y-(I_m \otimes X)\theta)'(\Sigma_\epsilon \otimes I_T)^{-1}(y-(I_m \otimes X)\theta)\right\}$$

To illustrate how to derive the posterior moments, let us assume Σ_ϵ is known and a multivariate normal prior for θ :

$$\Pi(\theta) \propto |V_0|^{-1/2} \exp\left\{-\frac{1}{2}(\theta-\theta_0)' V_0^{-1}(\theta-\theta_0)\right\} ,$$

where where θ_0 is the prior mean and V_0 is the prior covariance.

When we combine this prior with the likelihood function given above, the posterior density can be written as:

$$\Pi(\theta | y) = \exp\left\{-\frac{1}{2} \cdot ((V_0^{-1/2}(\theta - \theta_0))'(V_0^{-1/2}(\theta - \theta_0))\right.$$

$$\left. + \{(\Sigma_\epsilon^{-1/2} \otimes I_T)y - (\Sigma_\epsilon^{-1/2} \otimes X)\theta\}'\{(\Sigma_\epsilon^{-1/2} \otimes I_T)y - (\Sigma_\epsilon^{-1/2} \otimes X)\theta\})\right\}$$

which is a multivariate normal pdf. For simplicity, define

$$w \equiv \begin{bmatrix} V_0^{-1/2}\theta_0 \\ (\Sigma_\epsilon^{-1/2} \otimes I_T)y \end{bmatrix}$$

$$W \equiv \begin{bmatrix} V_0^{-1/2} \\ (\Sigma_\epsilon^{-1/2} \otimes X) \end{bmatrix}$$

Then the exponent in above equation can be written as

$$\Pi(\theta | y) \propto \exp\left\{-\frac{1}{2}(w - W\theta)'(w - W\theta)\right\} \propto$$

$$\exp\left\{-\frac{1}{2}(\theta - \bar{\theta})' W' W(\theta - \bar{\theta}) + (w - W\bar{\theta})'(w - W\bar{\theta})\right\}$$

where the posterior mean $\bar{\theta}$ is

$$\bar{\theta} = (W' W)^{-1} W' w = [V_0^{-1} + (\Sigma_\epsilon^{-1} \otimes X' X)]^{-1}[V_0^{-1}\theta_0 + (\Sigma_\epsilon^{-1} \otimes X)' y]$$

Since is known, the last part in exponent equation has no randomness about $\bar{\theta}$. The posterior therefore may be summarized as:

$$\pi(\theta | y) \propto \exp\left\{-\frac{1}{2}(\theta - \bar{\theta})' W' W(\theta - \bar{\theta})\right\}$$

$$= \exp\left\{-\frac{1}{2}(\theta - \bar{\theta})' \overline{V}^{-1}(\theta - \bar{\theta})\right\}$$

and the posterior covariance $\overline{V}$ is given as:

$$\overline{V} = [V_0^{-1} + (\Sigma_\epsilon^{-1} \otimes X'X)]^{-1}$$ (IHS Markit, 2020).

The Bayesian sampling approach to either impulse responses or forecasting from the VAR will differ depending upon the type of prior used during estimation. For priors that have a closed-form solution for the posterior: Litterman, normal-flat, normalWishart, Sims-Zha normal-flat and Sims-Zha normal-Wishart, although the posterior distribution of the VAR coefficients and covariances can be calculated without simulation, the posterior distribution of non-linear functions of the coefficients and covariance will still require calculation through Monte-Carlo integration simulations.

Popular prior distributions in the Bayesian VAR literature, with each incorporating differing beliefs on the underlying model are (IHS Markit, 2020):

1. The Litterman/Minnesota prior: A normal prior on β with fixed Σ.
2. The normal-flat prior: A normal prior on β that is independent of the distribution for Σ.
3. The normal-Wishart prior: A normal prior on β and a Wishart prior on Σ.
4. The independent normal-Wishart prior. A normal prior on β and a Wishart prior on Σ, where each endogenous equation's coefficients' distributions may be independent from each other.
5. The Sims-Zha normal-flat. A structural VAR equivalent of the normal-flat prior.
6. The Sims-Zha normal-Wishart prior. A structural VAR equivalent of the normal-Wishart prior.
7. The Giannone, Lenza and Primiceri prior. A prior that treats the hyper-parameters as parameters that can be selected through an optimization procedure,

etc.

Yet, a number of newly suggested priors with this or that advantages are continuously suggested by researchers (see for eg. Chan J., 2021).

For simplicity, in this analysis a widely used prior in the VAR literature - Minnesota prior (conjugate prior[1]) is used. This prior dates back to Litterman (1980) and Doan, Litterman, and Sims (1984). This priors are based on an approximation which leads to great simplifcations in prior elicitation and computation. This approximation involves replacing ∑ with an estimate, ∑^. The original Minnesota prior simplifes even further by assuming ∑ to be a diagonal matrix. In this case, each equation of the VAR can be estimated one at a time and we can set $\widehat{\sigma}_{ii} = s_i^2$ (where s_i^2 is the standard OLS estimate of the error variance in the i^{th} equation and $\widehat{\sigma}_{ii}$ is the ii^{th} element of ∑^). When ∑ is not assumed to be diagonal, a simple estimate such as ∑^ = S/T can be used. A disadvantage of this approach is that it involves replacing an unknown matrix of

[1] In Bayesian probability theory, if the posterior distribution p(θ | x) is in the same probability distribution family as the prior probability distribution p(θ), the prior and posterior are then called conjugate distributions, and the prior is called a conjugate prior for the likelihood function p(x | θ). Conjugate priors give rise to a range of useful analytical results, including a closed-form expression of the marginal likelihood (Chan J., 2021).

parameters by an estimate (and potentially a poor one) rather than integrating it out in a Bayesian fashion. The latter strategy will lead to predictive densities which more accurately react parameter uncertainty. However, replacing $\sum$ by an estimate simplifies computation since analytical posterior and predictive results are available. And it allows for a great range of flexibility in the choice of prior. If $\sum$ is not replaced by an estimate, the only fully Bayesian approach which leads to analytical results involves the use of a natural conjugate prior. Yet, the natural conjugate prior has some restrictive properties that may be unattractive in some cases (Koop et al., 2019).

The Minnesita/Litterman prior is given by:

$$\beta \sim N\left(\underline{\beta}_{Mn}, \underline{V}_{Mn}\right),$$

where $\underline{V}_{Mn}$ is assumed to be a diagonal matrix. The diagonal elements corresponding to endogenous variables, i,j at lag l are specified by:

$$\underline{V}^{l}_{Mn,i,j} = \begin{cases} \left(\frac{\lambda_1}{l^{\lambda_3}}\right)^2 & \text{for } i = j \\ \left(\frac{\lambda_1 \lambda_2 \sigma_i}{l^{\lambda_3} \sigma_j}\right)^2 & \text{for } i \neq j \end{cases},$$

where λ_1, λ_2 and λ_3 are hyper-parameters chosen by the researcher. The first two scalars are overall tightness and relative cross-variable weight, respectively, while λ_3 captures the lag decay that, as lag length increases, coefficients are increasingly shrunk toward zero. Note that changes in these hyper-parameter scalar values may lead to smaller (or larger) variances of coefficients, which is called tightening (or loosening) the prior. The exact choice of values for these three scalars depends on the empirical application, so that researchers can make trials with different values for themselves. The σi is the square root of the corresponding $(i,i)^{th}$ element of an initial estimate of Σ. Given this choice of prior, the posterior for θ takes the form:

$$\theta \sim N(\bar{\theta}, \overline{V}),$$

where

$$\overline{V} = [V_0^{-1} + (\hat{\Sigma}_\epsilon^{-1} \otimes X'X)]^{-1},$$

and

$$\bar{\theta} = \overline{V}[V_0^{-1}\theta_0 + (\hat{\Sigma}_\epsilon^{-1} \otimes X)'y]$$ (IHS Markit, 2020).

A big advantage of the Minnesota prior is that it leads to simple posterior inference involving only the Normal distribution. But, a disadvantage of the Minnesota prior is that it does not provide a full Bayesian treatment of ∑ as an unknown parameter. Instead, it simply plugs in ∑ = ∑^ , ignoring any uncertainty in this parameter.

There are a number of subtypes of Bayesian VAR models based on other model characteristics (rather than the prior selection) as well, including structural Bayesian VARs, structural Bayesian Global VAR (BGVAR), time varying BVARs, global VAR model with time-varying parameters and stochastic volatility (TVP-SV-GVAR model), Bayesian mixed-frequency VAR (MF-VAR), panel BVARs, etc. Each of these types has its specific usage and advantages. For example, MF-VAR allows some series to be observed at one bases (eg. monthly) and others at another bases (eg. quarterly) frequency. This is important, because in some cases, especially in macroeconomic forecasting, many series are available only in quarterly or yearly bases, while the advantage of using monthly information, on the other hand, is that the VAR is able to track the economy more closely in real time. The MF-VAR can be conveniently represented as a state-space model, in which the state-transition equations are given by a VAR at monthly frequency and the measurement equations relate the observed series to the underlying, potentially unobserved, monthly variables that are stacked in the state vector. Within-quarter monthly information leads to drastic improvements in the short-horizon forecasting performance (Schorfheide and Song, 2013). Another example is Bayesian Global VAR, the advantage of which is that it covers a wide array of cross section points (eg. counties). The model implies two distinct stages in the estimation process. In the first stage, N+1 country-specific multivariate time series models are specified, each of them including weakly exogenous regressors that aim to capture cross-country linkages. In the second stage, these models are combined using country weights to form a global model which then forms the basis for the impulse response analysis (Bock et al, 2021).

A numerical comparison: The impact of the economic shock in Russia on the economic indicators of Armenia

In this part of the work an practical implementation of BVAR and traditional SVAR models will be made. Thus, in order to assess the impact of the economic shock in Russia, first, a SVAR auto-regression model was developed with the following form:

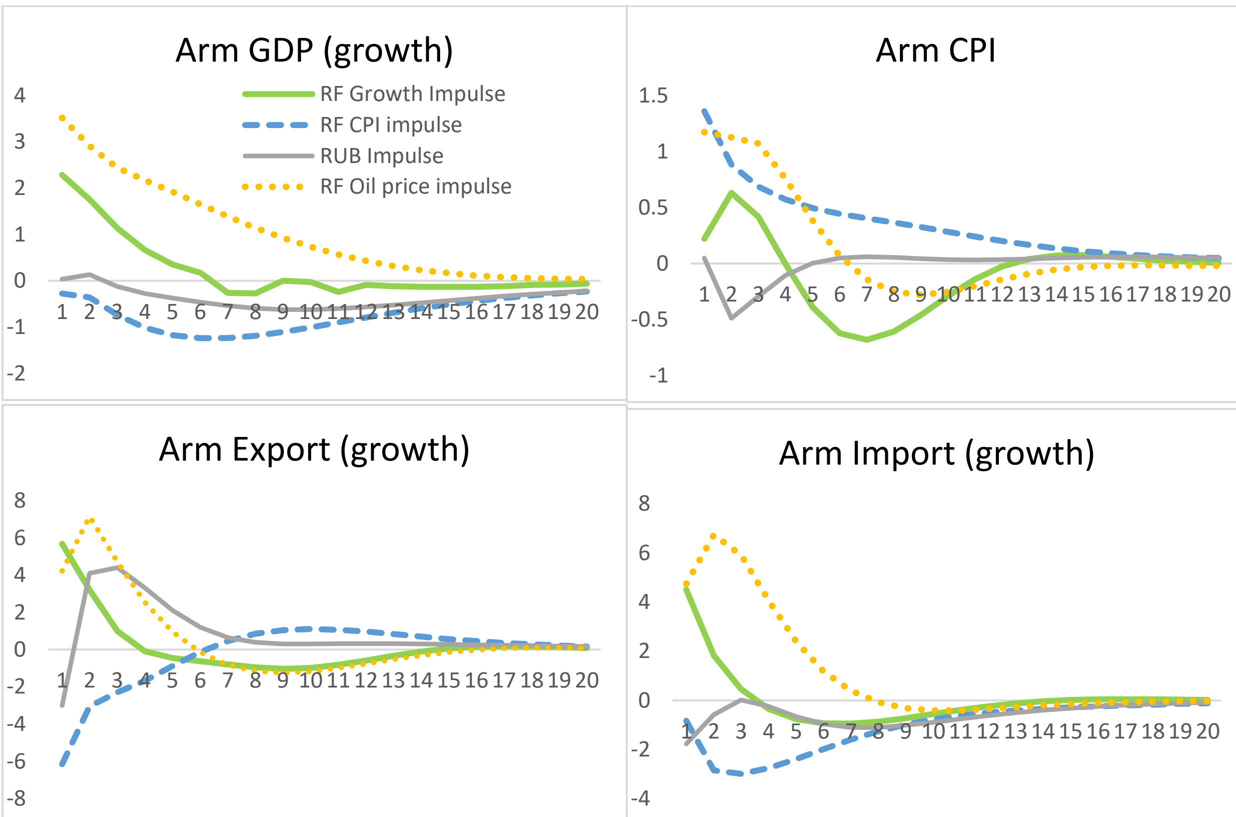
$$Ay_t = A_1^s y_{t-1} + \cdots + A_p^s y_{t-p} + C^s x_t + Bu_t ,$$

where A, all of the A_i^S, and C^S are the structural coefficients, and the u_t are the orthonormal unobserved structural innovations with $E(u_t\ u_t^{'}) = I_k$. Quarterly data of a number of indicators characterizing the RA and RF economy were used for the model from the first quarter of 2001 to the fourth quarter of 2021. Growths were calculated on an annual basis - each quarter relative to the corresponding quarter of the previous year. The stationarity of the series was checked using the Augmented Dickey-FullerTest. The vector of the variables used is as follows:

$$Y_t = VAR(OIL\_GR_t, USD\_RUB\_GR_t, RUS\_CPI\_INF_t, RUS\_GDP\_GR_t, USD\_AMD\_GR_t, REER\_GR_t, EXPORT\_GR_t, IMPORT\_GR_t, REM\_GR_t, CPI\_INF\_ARM_t, PCONS\_GR_t, ARM\_GDP\_GR_t),$$

where OIL_GR_t is the growth in oil prices in t period versus the same period of previous year, $USD\_RUB\_GR_t$ is the growth in RUB/USD exchange rate, $RUS\_CPI\_INF_t$ is the Russian CPI

(QoPYQ), RUS_GDP_GR$_t$ is the Russian GDP growth, USD_AMD_GR$_t$ Is the AMD/USD exchange rate growth, REER_GR$_t$, is the growth in RA REER, EXPORT_GR$_t$ is the RA exports growth, IMPORT_GR$_t$, is the RA imports growth, REM_GR$_t$, is the growth in net remittances, CPI_INF_ARM$_t$ is the RA CPI (QoPYQ), PCONS_GR$_t$, is the growth of private consumption in RA, ARM_GDP_GR$_t$ is the GDP growth in Armenia. Based on the test results of the AIC and SH criteria, the number of lags in the model is taken as 1. *Note that some restrictions in the model variables are imposed. Particularly, taking the impact of the changes in RA economy on RF economy is negligible, we restricted that opposite impacts in the model.*



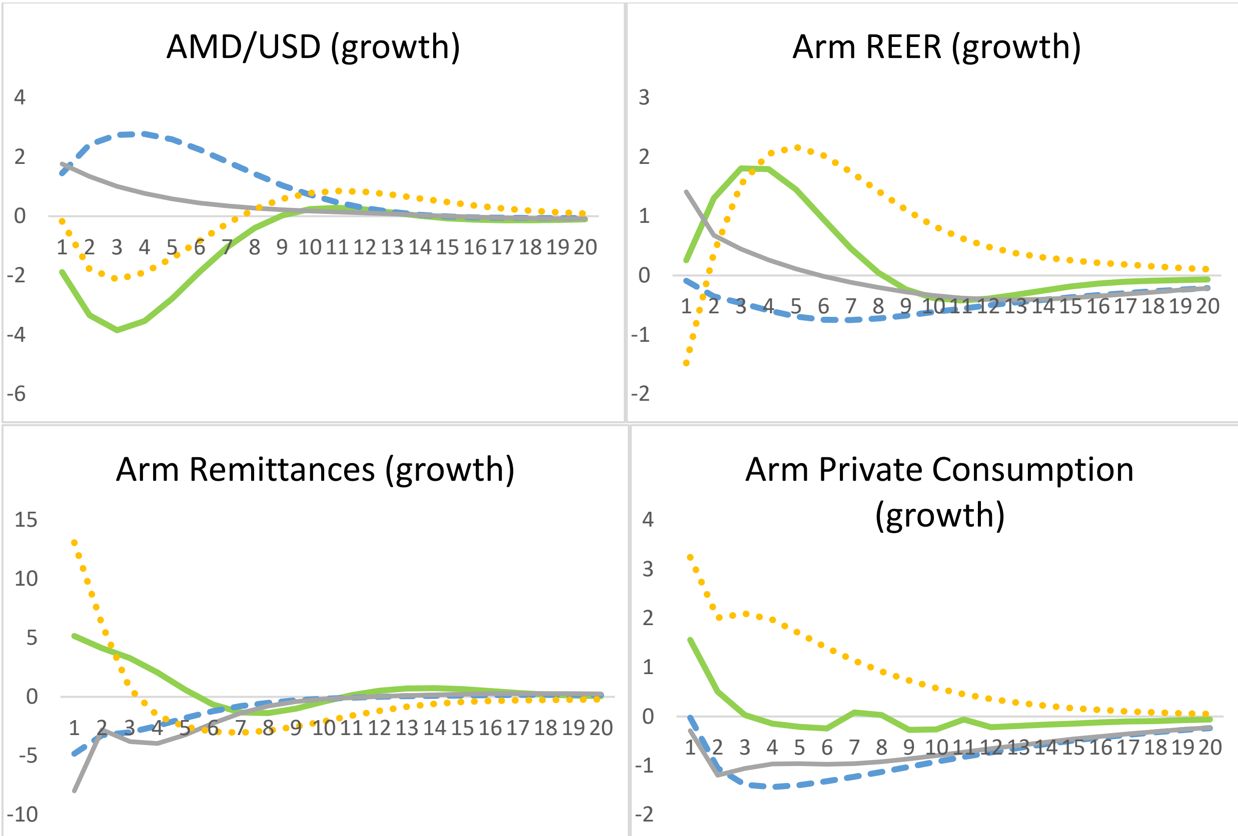
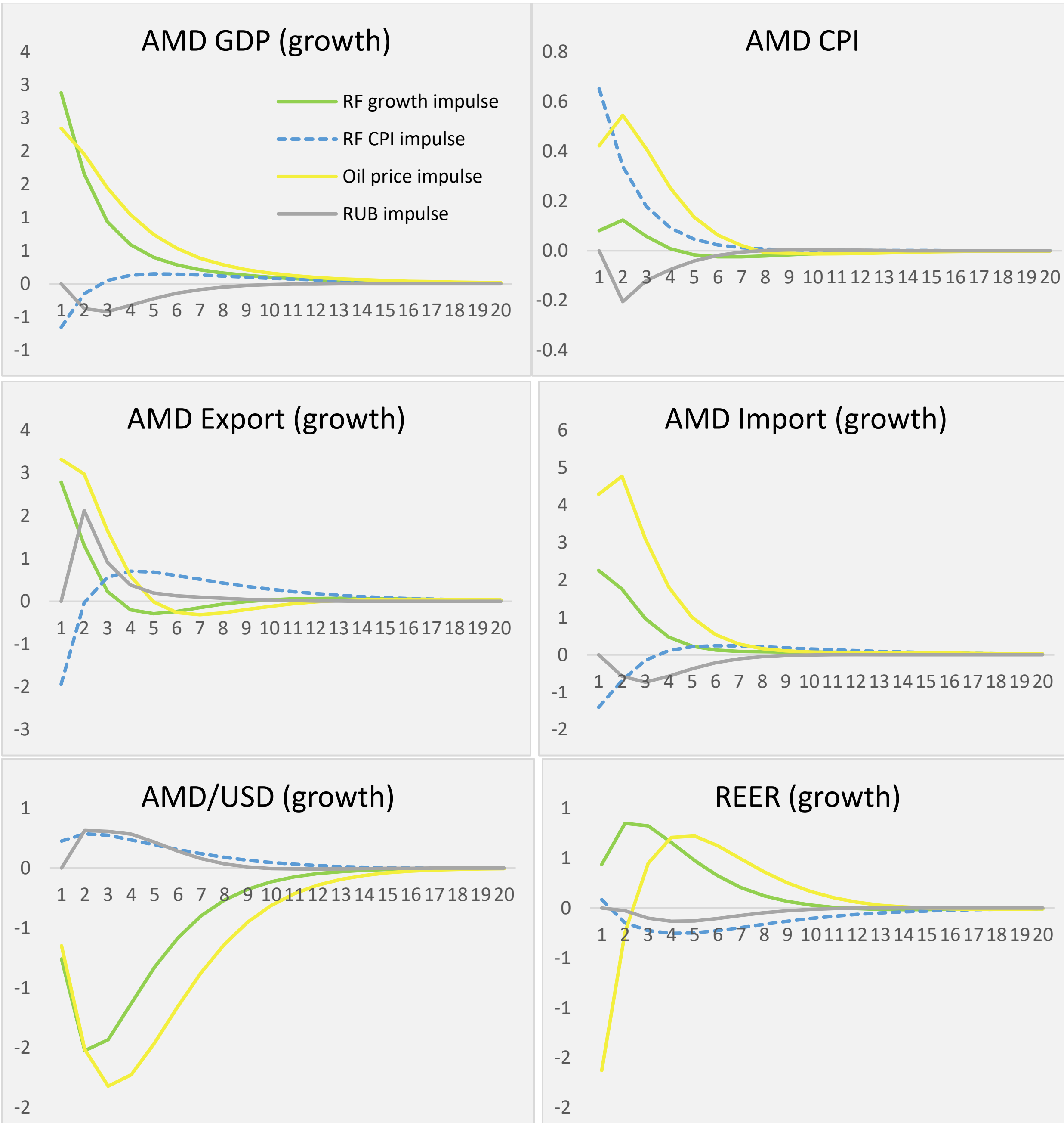


Figure 1. The response of the RA economic indicators to the shocks of the main economic indicators of the RF (Source: Author's calculations based on the SVAR model).

According to the results of the SVAR model, the one unit standard deviation change (i. e. 3.1pp change) of the economic growth of the Russian Federation in the short run leads to a change of the RA GDP growth by up to 2.28 percentage points. The impact of the unit sd change (25.8pp) in oil prices (3.5pp) on the GDP growth of the Republic of Armenia is also transmitted by influencing other variables. However, it fades quickly after 1-3 quarters. Meanwhile, the effects of the RUB/USD exchange rate and the GDP growth are more stable. Among the mechanisms of transmission of the economic shock in Russia, the impact of the growth rates of remittances, as well as imports and exports is significant. In the case of the former, 1 unit sd shock effect reaches up to 13.1 percentage points. The impact on the growth of private consumption in Armenia is similar to that of GDP growth.

In a short run (1 quarter) the one unit standard deviation increase (9.8pp) of the RUB/USD exchange rate shocks the CPI of the Republic of Armenia positively (-0.5 percentage points), while in a two year period it leads to a growth of up to 0.06 percentage points. The most significant impact on RA CPI has the inflation shock in Russia (up to 1.36pp), while the most significant impact on the AMD / USD exchange rate has the inflation shock in Russia (up to 2.8 percentage points). In a 1-1.5 year period the REER is also significantly affected by the oil price and RF GDP growth, while the impact of RUB / USD exchange rate shock is seen instantly.

Next, an one lag Bayesian VAR model was developed with Minessota/Litterman prior and with the following parameters: Mu1 (AR(1) coefficient for prior) = 0, $\lambda_1 = 0.1$, $\lambda_2 = 0.99$, and $\lambda_3 = 1$. Initial residual covariance option is selected to be Univariate AR estimate[2]. The impulse response functions for applied Bayesian VAR model are presented in Figure 2:



[2] Univariate AR estimates a univariate AR model (with number of lags matching those specified for the VAR) for each endogenous variable, then constructs the residual covariance matrix as a diagonal matrix with diagonal elements equal to the residual variance from the estimated univariate models.

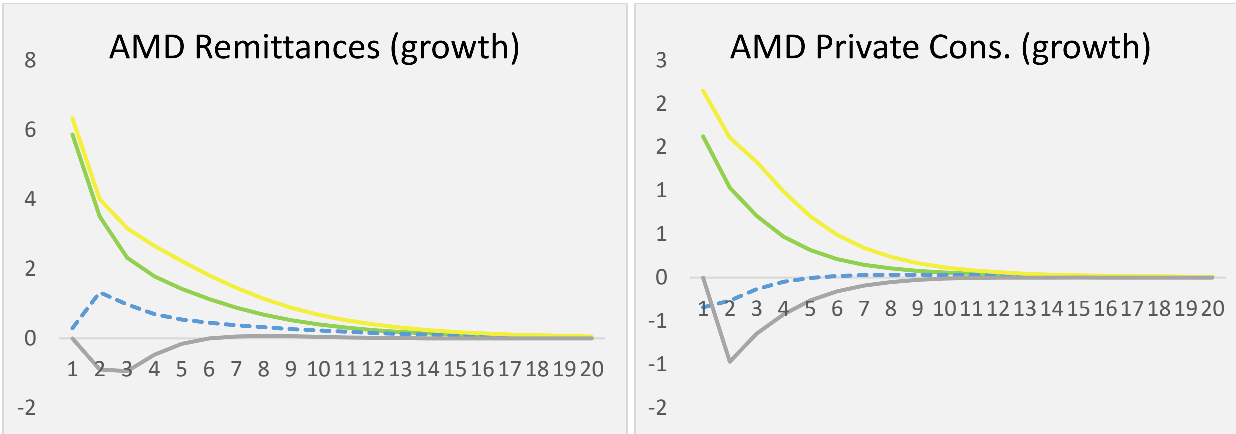
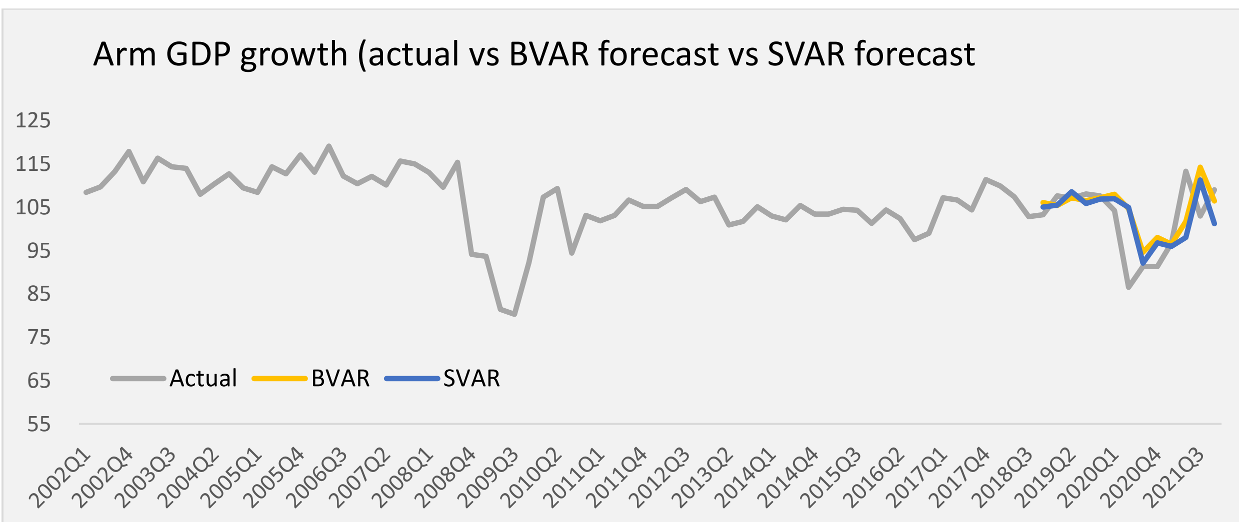


Figure 2. The response of the RA economic indicators to the shocks of the main economic indicators of the RF (Source: Author's calculations based on the BVAR model).

As we can see visually, the directions and amplitudes of the effects in the two models are almost the same. To compare the performance of the two models, a well-known method in time series analysis of out of sample forecasting is applied. Thus, we took 85% of the points for both models to fit, and used the other 15% (i.e. from 1Q2019 to 4Q2021) for the out of sample forecast (test data).



As a result, we see rather close predictions of the two models. Yet, the RMSE indicator for the BVAR model out of sample forecasts (7.26) is smaller than the RMSE indicator (7.64) for the SVAR model out of sample forecasts (note that total RMSE is 5.53 and 5.38 respectively).

In summary, one may note that the availability of flexible priors, reliable lag-selection criteria, and efficient sampling algorithms capable of producing precise Bayesian estimates makes Bayesian VAR inference a useful alternative to the traditional VAR analysis.

References

Bock M., Feldkircher M., Siklos P. (2021), International Effects of Euro Area Forward Guidance, *OXFORD BULLETIN OF ECONOMICS AND STATISTICS*, 0305-9049, doi: 10.1111/obes.12438.

Bock M., Feldkircher M., Huber F. (2020), BGVAR: Bayesian Global Vector Autoregressions with Shrinkage Priors in R. *Globalization Institute Working Paper No. 395*, Available at SSRN: https://ssrn.com/abstract=3682355 or http://dx.doi.org/10.24149/gwp395.

Chan J., (2021): "Minnesota-type adaptive hierarchical priors for large Bayesian VARs," *International Journal of Forecasting*, 37(3), 1212–1226.

Chan J. (2021), Asymmetric Conjugate Priors for Large Bayesian VARs, Draft, https://arxiv.org/pdf/2111.07170v1.pdf.

Chan J., Koop G., Yu X. (2021), Large Order-Invariant Bayesian VARs with Stochastic Volatility, Draft, https://arxiv.org/pdf/2111.07225v1.pdf.

Doan, T., R. Litterman, and C. Sims (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric reviews*, 3(1), 1–100.

Enders W. (2015), *Applied econometric time series*, John Wiley & Sons, Inc., USA (pp 285-305).

Hamilton J. (1994), *Time Series Analysis*, Princeton University Press, Princeton, New Jersey (pp 360-362).

IHS Markit (2020), *EViews 12 User's Guide II,* Seal Beach, CA, 1411 pages.

Kook J., Vaughn K., DeMaster D., Ewing-Cobbs L., Vannucci M. (2020), BVAR-Connect: A Variational Bayes Approach to Multi-Subject Vector Autoregressive Models for Inference on Brain Connectivity Networks, *Neuroinformatics*, https://arxiv.org/pdf/2006.04608v1.pdf.

Koop, G., Korobilis, D. Pettenuzzo, D. (2019). 'Bayesian compressed vector autoregressions', Journal of Econometrics, Vol. 210, pp. 135–154.

Kotzé K. (2017), Bayesian vector autoregression models, https://kevinkotze.github.io/ts-9-bvar/.

Litterman, R. B. (1986). Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. Journal of Business & Economic Statistics, 4(1), 25–38. https://doi.org/10.2307/1391384.

Nurfalah I., Aam Slamet Rusydiana A., Laila N., Cahyono E. (2018), Early Warning to Banking Crises in the Dual Financial System in Indonesia: The Markov Switching Approach, JKAU: Islamic Econ., Vol. 31 No. 2, pp. 133-156, DOI:10.4197/Islec. 31-2.10.

Schorfheide F., Song D. (2013), *Real-time forecasting with a mixed-frequency VAR*, Working Paper 19712, Cambridge, MA 02138, http://www.nber.org/papers/w19712.

Sims, C. A., and T. Zha (1998): "Bayesian methods for dynamic multivariate models," *International Economic Review*, 39(4), 949–968.

Appendix 1. The Bayesian VAR model: additional outputs.

Forecast Evaluation
Date: 05/22/22 Time: 20:45
Sample: 2002Q1 2021Q4
Included observations: 80

Variable	Inc. obs.	RMSE	MAE	MAPE	Theil
OIL_GR	80	26.68002	20.26156	20.24664	0.116186
USD_RUB_GR	80	10.81261	7.438419	6.737471	0.050401
RUS_CPI_INF	80	2.562801	1.949276	82.52901	0.389733
RUS_GDP_GR	80	3.791897	2.217284	401.3456	0.427244
USD_AMD_GR	80	5.833716	4.523339	160.7803	0.293975
REER_GR	80	4.697310	3.873658	3.848860	0.023063
EXPORT_GR	80	24.32010	16.67623	102.7956	0.750270
IMPORT_GR	80	11.54012	8.430824	8.470199	0.054583
REM_GR	80	26.91111	19.72932	108.9811	0.507154
CPI_INF_ARM	80	3.318230	2.792331	202.7482	0.500309
PCONS_GR	80	6.114204	4.295627	4.232168	0.029192
ARM_GDP_GR	80	5.530471	3.907050	3.839054	0.026038

RMSE: Root Mean Square Error
MAE: Mean Absolute Error
MAPE: Mean Absolute Percentage Error
Theil: Theil inequality coefficient