DOI: 10.20472/ES.2015.4.1.004

DO BAYESIAN VECTOR AUTOREGRESSIVE MODELS IMPROVE DENSITY FORECASTING ACCURACY? THE CASE OF THE CZECH REPUBLIC AND ROMANIA

VALERIU NALBAN

Abstract:

The supremacy of Bayesian VAR models over the classical ones in terms of forecasting accuracy is well documented and generally accepted in the literature on the grounds of overcoming the short sample and overfitting problems. However, the record is rather limited in case of emerging economies, ¬¬¬and more so for density (as opposed to point) forecasting. In this paper we compare the predictive accuracy of Bayesian and classical VAR models in case of density forecasting Czech and Romanian economic variables. The results show predictive densities are generally well calibrated, especially at shorter forecast horizons (less so for Romania). Log predictive density scores confirm the hypothesis of more accurate predictions of Bayesian VAR over classical VAR and naïve univariate models, the dominance of the former being more obvious in case of Romania. As such, the Bayesian approach to VAR yields a better approximation of the uncertainty surrounding the unknown future and minimizes the prediction errors.

Keywords:

Bayesian VAR, density forecasting, forecast evaluation, calibration, sharpness

JEL Classification: C11, C13, C32

Authors:

VALERIU NALBAN, Bucharest University of Economics Studies, Romania, Email: valeriunalban@gmail.com

Citation:

VALERIU NALBAN (2015). Do Bayesian Vector Autoregressive models improve density forecasting accuracy? The case of the Czech Republic and Romania. International Journal of Economic Sciences, Vol. IV(1), pp. 60-74., 10.20472/ES.2015.4.1.004

Acknowledgement: This work was cofinanced from the European Social Fund through Sectoral Operational Programme Human Resources Development 2007-2013, project number POSDRU/159/1.5/S/134197 “Performance and excellence in doctoral and postdoctoral research in Romanian economics science domain”.

1. Introduction

After the "incredible restrictions" critics brought by Sims (1980) to the large-scale macroeconomic models that were regularly used for policy analysis and forecasting in 1980s, Vector Autoregressive (VAR) models became the workhorse of applied economic research, being still extensively used by both scholars and policymakers. The great advantage of (purely statistical) VAR models is their simplicity and flexibility: the only input the researcher is required to supply are the endogenous variables and the number of lags. See Canova (2007) for a textbook description of VAR models.

Yet, two failures of VAR models are usually mentioned in the literature. The theoretical one is related to no structural interpretation or economic story telling in case of reduced-form models. To overcome this, some parameter restrictions are required, especially the ones related to the errors variance-covariance matrix, in order to disentangle the relationships between structural shocks and statistical residuals. However, this feature is not critical (and can be ignored) if one uses the model for forecasting purposes only, as stated in Felix and Nunes (2002). On the statistical ground, the overparameterization/overfitting issues arise if the number of parameters to be estimated is very large (which can be due to the number of endogenous variables included or due to the number of lags, or both) comparing to the sample information, which is rather limited in case of emerging economies (as Romania and the Czech Republic). As a result, the degrees of freedom become scarce, the model's solution might be explosive and the forecasts usually display erratic behavior and excessive variability, because the model does not properly capture the systematic relationships between the endogenous variables. See Ciccarelli and Rebucci (2003) for further discussion.

The Bayesian approach to VAR models is intended to overcome the overfitting problem by appending the sample information with the prior probabilities for the parameters. The functional principle of Bayesian VAR follows perfectly the Bayes rule, by forming a (subjective) prior probability for the object of interest and updating the prior with the (objective) information coming from the data-sample, thus forming the posterior distribution (see Geweke and Whiteman, 2006, and Karlsson, 2013 for textbook expositions of Bayesian methods). Formally, the priors act as restrictions on coefficients, shrinking and sharpening the estimates, and generating more accurate forecasts compared to Ordinary Least Squares (OLS) VAR, as in Litterman (1986) or Robertson and Tallman (1999). There are a number of ways for assembling a priori distributions, but Minnesota prior became the common approach in Bayesian VAR framework, following Doan et al. (1984) and Litterman (1986), as it can be quickly reproduced (thus minimizing the subjectivity of the prior). Moreover, the Bayesian VAR models include also OLS estimated VAR as a special case, if the priors are totally flat/uninformative (Sims, 2007).

In this paper we focus on the comparative density forecasting accuracy of Bayesian and classical (OLS) VAR models. Predictive densities emphasize the inevitable uncertainty associated to the unknown perspectives, allowing for a probabilistic assessment of the future and for a complete description of the risks surrounding the forecast (see Tay and Wallis, 2011). By their construction, point forecasts do not allow for such evaluations. Economic policymakers have recently moved to density forecasting (Bank of England and Sveriges Riksbank are the leading central banks in this field), publishing their predictive densities as “fan charts”. The fan charts are a useful communication tool to the economic policy authorities, according to Kascha and Ravazzolo (2010), because these can indicate the balance of risks by asymmetric widths of the bands and thus shaping the agents’ expectations. Bank of England (1998) realizes an ample description of applications of forecast densities and fan charts in their monetary policy assessment and forecast process.

However, as opposed to point predictions, the evaluation of out-of-sample forecasts in case of predictive densities is problematic by definition. This is because the researcher is bound to compare a continuous variable (the forecast density) with a discrete variable (the outturn) without knowing the true data generating process. The most common approach (see Diebold et al., 1998, and Mitchell and Wallis, 2011) is to assess if the ex-post realizations are uniformly distributed within the probability spaces associated to the predictive densities, i.e. to assess the degree of model’s calibration. Yet, Gneiting et al. (2007) obtain some “disconcerting results”, as calibration alone cannot perfectly rank the competing models. As such, they implement the concept of “maximizing the sharpness of the predictive distributions subject to calibration”. The definition of sharpness is related to the concentration of the forecast densities around some high mass point.

The concepts of “calibration” and “sharpness” are strongly connected with the ability of the model to correctly assess and forecast the uncertainty associated to the prediction. Thus, for economic policymakers a well calibrated model is essential for decision taking, especially for the central banks, whose actions should be anticipative given the monetary policy transmission lags.

We evaluate the comparative forecasting ability of the Bayesian and OLS VAR models in case of density forecasting Czech and Romanian economic data. The forecasting performance record of Bayesian VAR results is limited in case of emerging European countries, more so for density forecasts. This is in opposition with the United States and European developed economies, which serve as primary data sources for empirical applications, as in Litterman (1986), Kascha and Ravazzolo (2010), or Carriero et al. (2011). The main aim of the paper is to answer whatever shrinking the parameters via prior distributions improves the accuracy of the predictive densities in case of the Czech Republic and Romania. This purpose can be interpreting as testing the truth affirmation of Granger (1986) regarding the forecasting ability of competing approaches: “a good Bayesian will beat a non-Bayesian, who will do better than a bad Bayesian”.

Evaluating the models’ performance on two emerging countries which display both similarities (like floating exchange rate regime and inflation targeting monetary policy strategy) and specific features (with the Czech Republic being somehow more advanced, given earlier access to the European markets and technologies in late 1990s – early 2000s) is intended to provide more accurate and general conclusions. Given similar outcomes regarding the superiority of bayesian approach to VAR models for both the Czech Republic and Romania, the results can be extended and acknowledged as being generic for emerging economies as well, given previous literature records using advanced economies' data.

2. Methodology

A standard VAR model with *n* endogenous variables and *p* lags written in a compact form is represented as:

$$Y_t = X_t b + \varepsilon_t,\ t = 1,...,T \tag{1}$$

where Y_t is the $n \times 1$ vector of endogenous variables, $X_t = I_n \otimes (1, Y_{t-1},...,Y_{t-p})'$ of dimensions $n \times nk$ $(k = np + 1)$ holds the constant and lagged terms, $b = vec(B), B = (B_0, B_1,..., B_p)'$ is $nk \times 1$ vector of parameters, while ε_t is a $n \times 1$ vector of residuals which are independent and identically distributed $N(0,\Sigma)$. The parameters of interest are *b* (VAR coefficients) and Σ (residuals variance-covariance matrix).

Using the Bayes' rule, the posterior distribution of the parameters $p(b,\Sigma|Y)$ is equal to

$$p(b,\Sigma|Y) = \frac{p(Y|b,\Sigma)\,p(b,\Sigma)}{p(Y)} \propto p(Y|b,\Sigma)\,p(b,\Sigma) \tag{2}$$

where $p(Y|b,\Sigma)$ is the model likelihood, $p(b,\Sigma)$ is the joint prior distribution and $p(Y)$ is the marginal likelihood. The proportionality in (2) follows from the fact that the marginal likelihood does not depend on the parameters to be estimated, yielding the textbook expression for the posterior as being "proportional to the likelihood function times the prior" (Karlsson, 2013).

The most popular method for supplying the prior distribution $p(b,\Sigma)$ follows the works of Doan et al. (1984) and Litterman (1986), known as the Minnesota prior. It represents a very flexible way of forming the (subjective) a priori distribution, which can be straightforwardly reproduced. Only a reduced number of hyperparameters is required in order to construct the Minnesota prior for a model of any size. Moreover, the hyperparameters enclose some important statistical evidences specific for economic time series: unit roots, past values of a variable contain more information than past values of the other variables in the system, observations from the recent past are more important for current values than the ones in the more distant past, etc.

The prior mean of *b*, denoted with $\underline{b}$, in Minnesota setup is usually set to unity matrix, mapping a random walk process for each of the endogenous variable. The prior variance of *b*, denoted with $\underline{\Omega}$, is specified as a diagonal matrix with the following elements:

$$\mathrm{var}(b) = \underline{\Omega} = \begin{cases} \left(\dfrac{\lambda_1}{l^{\lambda_3}}\right)^2 & \text{for own lags} \\ \left(\dfrac{\lambda_1 \lambda_2}{l^{\lambda_3}}\right)^2 \dfrac{\sigma_m^2}{\sigma_q^2} & \text{for lags of variable } q \text{ in equation of variable } m \\ \lambda_4 \sigma_m^2 & \text{for the constant in equation of variable } m \end{cases} \quad (3)$$

In (3) $l = 1,..., p$ denotes the lag; $\sigma_i^2, i = 1,..., n$ are variances of residuals in univariate *p*th order autoregressions estimated for each endogenous variable, used to manage the different scales or units of measure of the variables; λ_1 controls the tightness of own lags; λ_2 governs the tightness of other variables lags; λ_3 controls the importance of own and foreign variables' lags (lag decay hyperparameter); and λ_4 manages the tightness of the constants. We follow Robertson and Talman (1999) and convert the specification of λ_3 hyperparameter from quarterly frequency model as in (3) to monthly frequency model used in this paper, replacing l^{λ_3} with $e^{\lambda_3 (0.13 l - 0.13)}$. The values of the hyperparameters are often taken from other empirical works, where these have been proven to increase forecasting performance. However, there exist data-driven procedures for searching the values that maximize the marginal likelihood in (2), as employed in Doan et al. (1984) or Carriero et al. (2011) among others.

For the prior of variance-covariance matrix Σ we use a inverse-Wishart distribution with $d > n+1$ degrees of freedom and $\underline{\Sigma} = (d - n - 1) \times diag(\sigma_1^2,..., \sigma_n^2)$ prior scale matrix. In order to have a conjugate posterior distribution (i.e. a distribution that comes from the same family as the prior and also has an analytical, known form, thus speeding-up computation as no cumbersome numerical simulation algorithms are required), we set $\lambda_2 = 1$ and obtain a convenient Kronecker structure (see technical details in Karlsson, 2013) on the prior:

$$\begin{aligned} p(b \mid \Sigma) &= N(\underline{b}, \Sigma \otimes \underline{\Omega}) \\ p(\Sigma) &= iW(\underline{\Sigma}, d) \end{aligned} \quad (4)$$

where $\underline{\Omega}$ is specified in (3). Note that in (4) the prior on *b* is conditional on Σ. Given the specification in (4), this type of prior distribution is known as the Normal-inverse-Wishart prior.

Given the prior distribution specified in (4), the posterior has the following closed-form solution (see mathematical derivations in Geweke and Whiteman, 2006, or Karlsson, 2013):

$$\begin{aligned} p(b \mid \Sigma) &= N(\bar{b}, \Sigma \otimes \bar{\Omega}) \\ p(\Sigma) &= iW(\bar{\Sigma}, T + d) \end{aligned} \quad (5)$$

where

$$\begin{aligned}&\bar{b} = vec(\bar{B}), \bar{B} = \bar{\Omega}(\underline{\Omega}^{-1}\underline{B} + X'XB_{OLS}), \bar{\Omega} = (\underline{\Omega}^{-1} + X'X)^{-1} \\ &\bar{\Sigma} = B'_{OLS}X'XB_{OLS} + \underline{B}'\underline{\Omega}^{-1}\underline{B} + \underline{\Sigma} + (Y - XB_{OLS})'(Y - XB_{OLS}) - \bar{B}'\bar{\Omega}^{-1}\bar{B}\end{aligned} \tag{6}$$

In (6) *OLS* subscript denotes Ordinary Least Square (OLS) estimates: $B_{OLS} = (X'X)^{-1}X'Y$. In practice, draws from the posterior distribution (5) are taken using the Direct Sampling algorithm, i.e. taking a draw from $p(\Sigma)$ first and then, conditioning on it, a draw from $p(b \mid \Sigma)$.

Bayesian approach to VAR includes also the classical OLS estimation as a particular case. It is easily proven (Karlsson, 2013 among many others) that a totally flat (uninformative) prior of the form $p(b,\Sigma) = \mid \Sigma \mid^{-(n+1)/2}$ results in the following posterior, whose mode is equal to OLS estimates:

$$\begin{aligned}&p(b \mid \Sigma) = N(b_{OLS}, \Sigma \otimes (X'X)^{-1}) \\ &p(\Sigma) = iW((Y - XB_{OLS})'(Y - XB_{OLS}), T - k)\end{aligned} \tag{7}$$

Again, Direct Sampling is used for drawing from the posteriors.

In the results section we describe the comparative forecasting performance of the OLS VAR and Bayesian VAR models estimated and evaluated on the Czech Republic and Romanian data, giving an explicit answer to the question raised in the title. In what follows, we denote the classical VAR with “OLS” and the Bayesian VAR with “NIW” (given the Normal-inverse-Wishart specification). For both types of models we use 4 lags ($p = 4$), a value that allows a richer dynamics than formal information criteria suggest. For the hyperparameters in Bayesian VAR model we use the most common values in relevant literature. λ_1 is set to 0.15, implying a moderate degree of tightness on own lags; Kadiyala and Karlsson (1997) set it to 0.1 for the Normal-inverse-Wishart prior model and 0.2 for other types of priors; Giannone et al. (2012) use a hierarchical model to estimate the mode of λ_1 between 0.15 and 0.2. As noted above, λ_2 is set to 1 in the conjugate prior we implement. Like in Kadiyala and Karlsson (1997) and Carriero et al. (2011), we set λ_3 = 1 (harmonic lag decay) and $d = n + 2$. The prior on constants is totally flat, letting the data speak: λ_4 = 10000.

Formally, the predictive density for future values between moments *t+1* and *t+H* computed at moment *t* is constructed using posterior distribution of the parameters given historical data between moments 1 and *t*, $p(b,\Sigma|Y_{1:t})$, and predicted values given these parameters, $p(Y_{t+1:t+H}|b,\Sigma,Y_{1:t})$, both integrated over *b* and Σ (see other technical elements in Geweke and Whiteman, 2006, or Karlsson, 2013):

$$p_t(Y_{t+1:t+H} \mid Y_{1:t}) = \int p(Y_{t+1:t+H} \mid Y_{1:t}, b, \Sigma)\, p(b, \Sigma \mid Y_{1:t})\, d(b,\Sigma) \tag{8}$$

In a typical VAR model one-step ahead forecast (*h*=1) is linear in parameters, but for *h*>1 the prediction becomes intractable in analytical form, requiring numerical algorithms. We use 1000 draws form the posterior distribution obtained via Direct

Sampling as mentioned above for approximating the out-of-sample predictive densities at any forecast horizon $h=1,...,H$. Moreover, we use a non-parametric kernel smoothing technique in order to allow for any non-normality arising in the predictive distributions.

The most common out-of-sample evaluation criteria for density forecasts calibration uses the probability integral transform (PIT) values (see Diebold et al., 1998), which represents the cumulative distribution function (CDF) of the predictive density $p_t(y_{t+h})$ evaluated at the corresponding outcome y_{t+h}:

$$PIT_{t,h} = \int_{-\infty}^{y_{t+h}} p_t(u)du \quad (9)$$

Diebold et al. (1998) explains that for a well calibrated density PIT should be uniform $U(0,1)$ or, equivalently, the inverse normal transformed PIT should follow a standard normal distribution $N(0,1)$. In addition, at one-step ahead horizon the PIT must not be autocorrelated. Besides visual inspection using histograms, some formal statistical test are usually employed in order to test the above mentioned hypothesizes, like one-sided Kolmogorov-Smirnov, Pearson chi-squared or Ljung-Box tests (see a more detailed explanation in Bache et al., 2011). As mentioned in the previous section, a model with well calibrated forecast densities is able to accurately evaluate the uncertainty surrounding the prediction.

While we use PIT for testing the calibration, we calculate the log predictive density scores (LPDS) in order to assess the degree of sharpness of the forecast distributions. Formally, LPDS is equal to the logarithm of the probability density function evaluated at the respective outcome:

$$LPDS_{t,h} = \log p_t(y_{t+h}) \quad (10)$$

The LPDS is equivalent to the forecast error in point forecasting literature, according to Mitchell and Wallis (2011).

Calculating and evaluating both (9) and (10) is in line with Gneiting et al. (2007) paradigm of “maximizing the sharpness subject to calibration”, allowing for an in-depth and comprehensive assessment of comparative forecasting performance of the models we estimate.

3. Data

For both economies we use the following four monthly frequency variables: industrial production index (seasonally adjusted), harmonized index of consumer prices (HICP), 3 months interbank market interest rates (PRIBOR 3M in the Czech Republic and ROBOR 3M in Romania), and nominal exchange rate of domestic currency (koruna and leu) vis-à-vis euro (expressed such that a rise indicates a depreciation of domestic currency). The source of the data is Eurostat, as available on 27 September 2014. The list of the variables is somehow standard for small open economy VAR, being included in Kim and Roubini (2000) among others. With the exception of interest rates, the series are transformed in 12-periods log-difference to induce stationarity. In addition, as the null hypothesis of a unit root in ROBOR 3M could not be rejected using the Augmented Dickey-Fuller test, we express it as 12-periods difference. Although stationarity is not a problem in bayesian framework, because posterior distribution of unit root models do not require a special treatment as in classical econometrics (see technical discussion in Canova, 2007), we follow Carriero et al. (2011) and use log-differenced data as it is less affected by eventual structural breaks in the series.

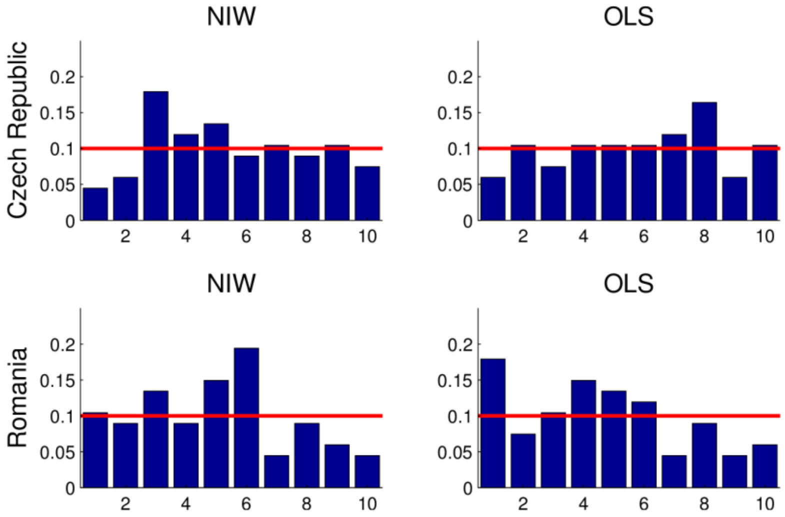
The sample covers January 2003 – July 2014 period, consisting of roughly three historical episodes: the boom period prior to late 2008, the recent 2000s financial crisis, and the subsequent recovery period starting 2010, characterized by a low growth environment. We prefer as initial observation the January 2003 one as prior to this date the inflation rate of Romanian HICP registered high values, inconsistent with the figures form 2003-2014 period. Also, since inflation targeting was officially implemented in Romania only in August 2005, the chosen sample is less affected by the monetary regime change than a longer one. Given the temporal dimension of the dataset is rather short (does not contain a single full business cycle), bayesian approach is expected to help overcoming small sample issues and sharpen the estimation. The expanding window recursive out-of-sample forecast procedure for up-to 12-months ahead is performed starting January 2009 (i.e. the first forecast is produced using data up-to-and-including December 2008 observations). We prefer the expanding window approach (i.e. adding one observation to the previous sample) instead of the rolling window one (i.e. adding one observation and deleting the first observation of the previous sample, thus keeping the length of the estimation sample constant) in order to benefit from increasing lengths of the samples and obtaining more relevant estimates. A note of caution should be made regarding the forecasting accuracy ranking of the competing models: as the forecasting sample is very short (January 2009 – July 2014), the results are specific to the period under analysis (i.e. the crisis and recovery episodes) and should not be generalized.

4. Results

First we assess the degree of calibration of the out-of-sample predictive densities obtained over January 2009 – July 2014 forecast horizon. As in Gneiting et al. (1997), the PIT sequences are calculated using formula (9) with $h = 1,...,12$. In order to check whatever these are uniformly distributed, we calculate the associated histograms using

10 bins. The small number of bins is determined by the short forecasting sample, due to limited temporal dimension of the time series considered.

Figure 1: PIT histograms for HICP inflation rate at one-step ahead forecast



An example of these histograms is presented in Figure 1, for the HICP inflation rate one-month ahead forecast. Given that there are four models with 4 endogenous variables each and we perform up-to 12-months ahead forecasts, there is a total of 192 histogram plots to be evaluated. As stated above, “NIW” indicates the Normal-inverse-Wishart prior Bayesian VAR and “OLS” denotes the Ordinary Least Squares VAR model. There are some clear signs of departures form uniformity (presented as the red line: given there are 10 bins, their expected frequency under uniformity is 0.1), but it is hard to decide whatever these are statistically significant.

In these circumstances, we perform the one-sided Kolmogorov-Smirnov test for uniformity of PIT sequences. Intuitively, the test evaluates if the largest difference between the outturn of an empirical distribution (PIT in our case) and the expected value corresponding to the theoretical distribution ($U(0,1)$ in our case) is statistically significant; see details in Diebold et al. (1998) for a comparative performance of the histogram-based tests. The p-values associated to the null hypothesis of uniformly distributed PIT are calculated and presented in Table 1 (p-values greater than 0.1, meaning the null hypothesis cannot be rejected, are shaded).

Table 1: Kolmogorov-Smirnov test p-values for uniformity of PIT sequences*

			h=1	*h*=2	*h*=3	*h*=4	*h*=5	*h*=6	*h*=7	*h*=8	*h*=9	*h*=10	*h*=11	*h*=12
The Czech Republic	OLS	IP	0.70	0.29	0.53	0.52	0.18	0.07	0.07	0.06	0.04	0.03	0.03	0.01
		HICP	0.99	0.69	0.68	0.82	0.81	0.65	0.49	0.48	0.47	0.22	0.21	0.08
		IR	0.41	0.20	0.28	0.52	0.51	0.37	0.35	0.34	0.23	0.22	0.14	0.20
		ER	0.70	0.69	0.53	0.52	0.66	0.65	0.25	0.10	0.10	0.22	0.21	0.08
		IP	0.55	0.83	0.39	0.13	0.27	0.17	0.11	0.10	0.10	0.09	0.09	0.08

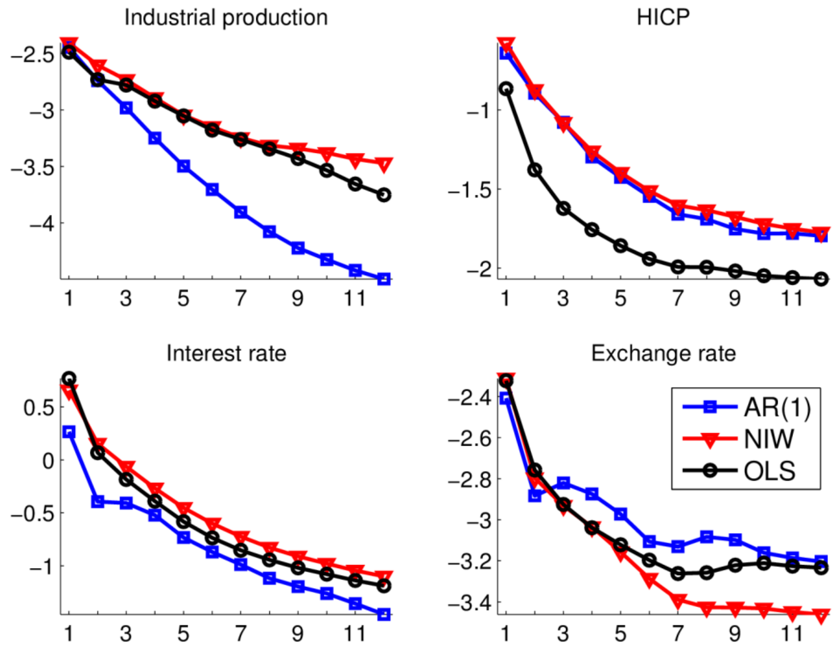
		HICP	0.84	0.83	0.82	0.82	0.81	0.80	0.92	0.91	0.91	0.90	0.76	0.58
		IR	0.06	0.29	0.13	0.19	0.08	0.07	0.07	0.10	0.10	0.09	0.05	0.05
		ER	0.70	0.40	0.93	0.38	0.18	0.26	0.07	0.04	0.02	0.06	0.05	0.02
Romania	OLS	IP	0.84	0.69	0.39	0.19	0.08	0.07	0.04	0.01	0.01	0.02	0.01	0.02
		HICP	0.21	0.14	0.08	0.05	0.08	0.11	0.04	0.04	0.01	0.00	0.00	0.00
		IR	0.06	0.06	0.13	0.19	0.08	0.07	0.07	0.04	0.06	0.15	0.05	0.08
		ER	0.70	0.83	0.93	0.99	0.93	0.80	0.64	0.63	0.62	0.77	0.44	0.43
	NIW	IP	0.41	0.29	0.05	0.13	0.03	0.03	0.02	0.01	0.01	0.00	0.00	0.00
		HICP	0.09	0.03	0.01	0.01	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
		IR	0.21	0.06	0.13	0.19	0.08	0.03	0.00	0.00	0.00	0.00	0.00	0.00
		ER	0.99	1.00	0.93	0.67	0.38	0.26	0.11	0.10	0.15	0.09	0.03	0.03

* p-values greater than 0.1 are shaded; the following notations are used: "OLS" for the Ordinary Least Square VAR model, "NIW" for the Normal-inverse-Whishart Bayesian VAR model, "IP" for industrial production index, "HICP" for harmonized index of consumer prices, "IR" for interbank three months interest rate, "ER" for exchange rate.

The shaded pattern in Table 1 indicates the predictive densities are well calibrated at short forecast horizons (less than 6 months) for the Czech Republic in case of both OLS and Bayesian VAR. At longer horizons, OLS VAR fails to accurately evaluate the uncertainty surrounding the industrial production index, while the Bayesian VAR does not perform well in case of interest rate also. Both models are less competitive in case of Romania. With the exception of HICP in Bayesian VAR, the p-values are larger than 0.05 for up-to 5-months ahead forecasts. Beyond this horizon, predictive distributions are generally not well calibrated, with minor exceptions (leu/euro exchange rate). A closer look reveals that in case of the interest rate variables the Bayesian VAR predictive densities are too wide when compared to actual data variability, i.e. the PIT histograms have the mass concentrated on central bins. The explanation of this outcome lies in the fact that the estimation sample contains periods of high volatility of the interest rates (which is further transmitted into higher forecast uncertainty and, correspondingly, large confidence bands), while in the evaluation sample these are rather stable. A similar result regarding better calibrated predictive densities at shorter-horizons is rather frequently obtained in relevant literature, as in Gerard and Nimark (2008) or Bache et al. (2011).

Next, we calculate the log-predictive density scores (LPDS) as specified in (10) and average them out over each forecast horizon. As stated before, these are formal indicators of the forecast densities sharpness, i.e. of the concentration of the forecasts around some central moments. Moreover, Mitchell and Wallis (2011) suggest interpreting LPDS as comparative forecast error. In addition to the OLS and NIW models, we add first order univariate autoregression models on each endogenous variable (denoted with "AR(1)"), usually used as a natural benchmark in practice. The corresponding LPDS are presented in Figure 2 and Figure 3 for the Czech Republic and Romanian respectively. Note that a higher LPDS suggest a smaller forecast error, indicating a more accurate model.

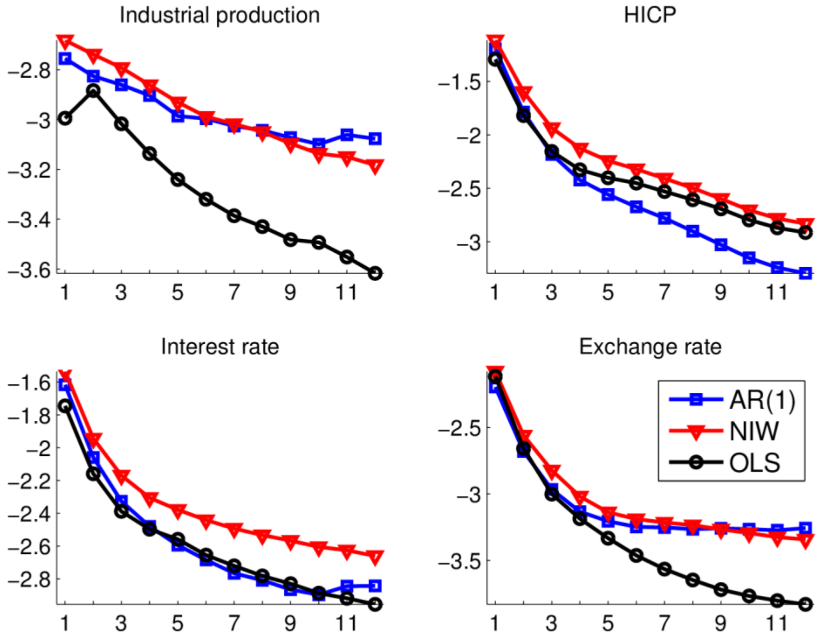
Figure 2: Average log predictive density scores for the Czech Republic



For the Czech Republic (Figure 2) Bayesian VAR model (NIW) outperforms its competitors in case of industrial production index, consumer prices inflation and interest rate at virtually all forecast horizons. OLS follows closely its Bayesian counterpart for interest rate and industrial production (up-to 9-months ahead forecasts), but is rather imprecise for HICP. AR(1) model performs well in case of HICP and exchange rate, in the latter case beating its rivals. Overall, Bayesian VAR seems to be the most skillful, while the other two models are hard to rank.

Turning to Romania (Figure 3), the comparative performance of Bayesian VAR model is more obvious, as it loses only to AR(1) in case of industrial production index and exchange rate at forecast horizons greater than 9 months (however the differences are minor and hardly significant). The OLS model is the least accurate one on average, being unable to beat the naïve univariate AR(1) model (with the exception of HICP inflation).

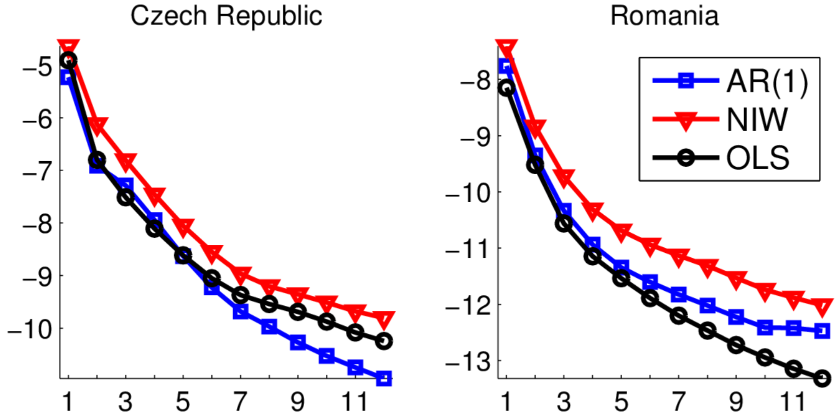
Figure 3: Average log predictive density scores for Romania



The industrial production index seems to display little connections with the other variables in the system in case of Romania, since the univariate model is rather robust, while the opposite is true for the Czech industrial output, where multivariate methods (both OLS and NIW VAR models) are clearly preferred. A somehow similar comparative description applies to the HICP inflation rate forecasts, but with Romanian data favoring multivariate models. The forecasting ability of Bayesian VAR in case of the interest rates is similarly superior for both economies. Meanwhile, the univariate models seem to be useful when forecasting the exchange rates: given high frequency of observing this variable (i.e. daily), it exhibits less interaction with the real economy variables that are observed monthly (industrial production and HICP).

In order to obtain a proper ranking of the three competing models, we construct a multivariate version of log predictive density scores by summing up individual variables LPDS and display them in Figure 4. As was somehow obvious from the univariate LPDS in Figure 2 and Figure 3, Bayesian VAR clearly outperforms the other two opponents for both countries, with Romanian NIW model showing a greater degree of supremacy. OLS and AR(1) perform similar up-to 6-months ahead forecasts, but beyond this horizon OLS is more accurate in case of the Czech Republic, while AR(1) is preferred over OLS in Romania. The results should, however, be treated as episode-specific for the forecasting sample considered (January 2009 – July 2014). Gerard and Nimark (2008) issue a warning regarding forecasting evaluation on short samples, mentioning the difficulty of properly distinguishing between predictive ability and luck. A replication of the procedure performed in this paper on another sample may give different results. Nevertheless, the conclusions drawn here are likely to be representative for the current low growth environment.

Figure 4: Sum of individual variables average log predictive density scores



The increased accuracy of Bayesian VAR models over the classical ones is well agreed in the literature, but most of the applications are related to developed countries: Litterman (1986) for the United States; Kadiyala and Karlsson (1997) for Sweden and the United States; Kascha and Ravazzolo (2010) for the United Kingdom, the United States, New Zeeland and Norway; Carriero et al. (2011) for the United States, Canada, France and the United Kingdom; etc. The results obtained in this paper show the results specific to more economically advanced countries apply also to the emerging economies.

5. Conclusions

In this paper we compare Bayesian VAR models with conjugate Normal-inverse-Wishart prior to the Ordinary Least Square (OLS) VAR models in terms of density forecasting accuracy. The models are estimated on monthly Czech and Romanian data. We use somehow standard values for the hyperparameters when constructing the prior distribution, granting a moderate degree of shrinkage in order to mitigate the small sample and overparameterization problems specific to OLS VAR models. The forecast evaluation period consists of the crisis and subsequent low growth environment of January 2009 – July 2014 sample. The out-of-sample predictive densities for up-to 12-months ahead are evaluated with respect to both calibration and sharpness, allowing for a complete assessment of the uncertainty the models forecast, with important insights for the policy makers.

Probability integral transform (PIT) sequences were converted to histograms and tested for uniformity with one-sided Kolmogorov-Smirnov test. The results are compatible with the empirical evidence in relevant literature, indicating that predictive densities are well calibrated at short forecast horizons for both Bayesian and classical VAR models, while for more distant ones there are signs of non-uniformity, especially for some Romanian variables.

Sharpness (i.e. the concentration of predictive densities) was assessed via average log predictive density scores (LPDS). In terms of relative forecast errors, for both the Czech Republic and Romania Bayesian VAR clearly outperforms OLS VAR and also naïve univariate first order autoregressive models at virtually any forecast horizon (the only significant exception is the Czech exchange rate forecast, where the univariate model is preferred). Multivariate forecasting accuracy evaluated with the sum of individual variables LPDS reinforce the conclusion regarding the increased accuracy of Bayesian VAR over both OLS VAR and univariate models, with the dominance of the Bayesian model being more obvious in case of Romania. As such, adding the Bayesian approach to VAR models yields reasonable approximation of the uncertainty surrounding the forecasts in case of the two emerging economies and minimizes the forecast errors, the results being similar to the ones obtained for more developed countries.

References

Bache, I. W., Jore, A. S., Mitchell, J., & Vahey, S., (2011). Combining VAR and DSGE forecast densities. *Journal of Economic Dynamics and Control*, 35(10), 1659-1670.

Bank of England, (1998). The Inflation Report projections: understanding the fan chart. *Bank of England Quarterly Bulletin: February 1998*, 30-37.

Canova, F., (2007). Methods for Applied Macroeconomic Research. Princeton University Press.

Carriero, A., Clark, T., & Marcellino, M., (2011). Bayesian VARs: Specification Choices and Forecast Accuracy. Federal Reserve Bank of Cleveland Working Paper 1112.

Ciccarelli, M., & Rebucci, A., (2003). Bayesian VARs: A Survey of the Recent Literature with an Application to the European Monetary System. IMF Working Papers 03/102.

Diebold, F., Gunther, T., & Tay, A., (1998). Evaluating Density Forecasts with Applications to Financial Risk Management. *International Economic Review*, 39(4), 863-83.

Doan, T., Litterman, R., & Sims, C., (1984). Forecasting and Conditional Projection Using Realistic Prior Distributions. *Econometric Reviews*, 3, 1-100.

Felix, M. F., & Nunes L. C., (2002). Bayesian Forecasting Models for the Euro Area. *Banco de Portugal Economic bulletin: March 2002*, 69-85.

Gerard, H., & Nimark, K., (2008). Combining multivariate density forecasts using predictive criteria. Economics Working Papers 1117, Department of Economics and Business, Universitat Pompeu Fabra.

Geweke, J., & Whiteman, C., (2006). Bayesian Forecasting. Handbook of Economic Forecasting, Elsevier.

Giannone D., Lenza, M., & Primiceri, G., (2012). Prior Selection for Vector Autoregressions. ECARES Working Paper 2012/002.

Gneiting, T., Balabdaoui, F., & Raftery, A., (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society*, 69(2), 243-268.

Granger, C. W. J., (1986). Comment (on McNees, 1986). *Journal of Business and Economic Statistics*, 4, 16–17.

Kadiyala, R., & Karlsson, S., (1997). Numerical Methods for Estimation and Inference in Bayesian VAR-Models. *Journal of Applied Econometrics*, 12(2), 99-132.

Karlsson, S., (2013). Forecasting with Bayesian Vector Autoregression. Handbook of Economic Forecasting, Volume 2, Elsevier.

Kascha, C., & Ravazzolo, F., (2010). Combining inflation density forecasts. *Journal of Forecasting*, 29(1-2), 231-250.

Kim, S., & Roubini, N., (2000). Exchange rate anomalies in the industrial countries: A solution with a structural VAR approach. *Journal of Monetary Economics*, Elsevier, 45(3), 561-586.

Litterman, R., (1986). Forecasting with Bayesian Vector Autoregressions: Five Years of Experience. *Journal of Business & Economic Statistics*, 4(1), 25-38.

Mitchell, J., & Wallis, K., (2011). Evaluating density forecasts: forecast combinations, model mixtures, calibration and sharpness. *Journal of Applied Econometrics*, 26(6), 1023-1040.

Robertson, J., & Tallman, E., (1999). Vector Autoregressions: Forecasting and Reality. *Federal Reserve Bank of Atlanta Economic Review First Quarter 1999*.

Sims, C., (1980). Macroeconomics and Reality. *Econometrica*, 48(1), 1-48.

Sims, C., (2007). Bayesian Methods in Applied Econometrics, or, Why Econometrics Should Always and Everywhere Be Bayesian. Princeton University Manuscript.

Tay, A., & Wallis, K., (2000). Density Forecasting: A Survey. *Journal of Forecasting*, 19, 235-254.

