Analyzing and Forecasting Output Gap and Inflation Using Bayesian Vector Auto Regression (BVAR) Method: A Case of Pakistan

Mian Abdullah Tahir[1]

[1] School of Economics, Jilin University, Changchun, China

Correspondence: Mian Abdullah Tahir, School of Economics, Jilin University, Changchun, China. Tel: 86-431-8516-6030. E-mail: mian.abdullah@gmail.com

Abstract

We attempt to forecast inflation and output gap of Pakistan using Bayesian VARs. We implement three different priors for this purpose. Analysis in this paper is conducted using Monetary Aggregates and Credit macro variables in order to forecast Output Gap and CPI Index based measure of Inflation. Output Gap used in our analysis is estimated in a State–Space framework using Kalman filter. Literature suggests that Bayesian shrinkage is an appropriate tool for forecasting using large number of Macro Economic variables. In addition, appropriate Prior selection is fundamental to robust forecasting in Bayesian VARs; in this backdrop, the 3 types of Priors implemented in our analysis are; 1: Minnesota Priors, 2: Independent Normal–Wishart Priors and 3: Independent Minnesota–Wishart Priors. Estimation and forecasting is conducted in conformity with Koop and Korobilis (2009). Diagnostics of Bayesian VAR models and robustness of forecast estimates show that Bayesian VARs provide robust forecasts and have suitable structural interpretation. This conclusion is especially relevant considering that Bayesian methods provide inherent solution to circumvent the problem of multicollinearity and over parameterization.

Keywords: Bayesian VAR, forecasting, monetary VAR, Kalman Filter

1. Introduction

Linear stochastic difference equations or commonly Vector Auto Regression are standard tools at the disposal of macro econometric researcher for forecasting and analysis. VARs provide both flexibility of model elicitation and the ability to capture intricate data linkages and relationships. This ease of model representation involves probable presence of model misspecification, multicollinearity, over parameterization, loss of degrees of freedom. In order to circumvent these issues VARs are generally implemented in a parsimonious variables setting.

Recently the use of Bayesian methods in time series has heralded in a new era of robust estimation, structural analysis and modeling complex inter-linkages. Litterman (1979) presented a new and revolutionary method of VAR specification. Litterman (1979) proposed biased estimation of VARs in order to avoid the problems of large sampling errors and the problem of multicollinearity (see also Doan, Litterman, & Sims, 1984; Litterman, 1986a). He introduced a method similar to Ridge regression (Hoerl & Kennard, 1970); Stein rule estimators (Stein, 1974), whereby prior information is incorporated in the model estimation. He argues that this may be interpreted as a form of Bayesian estimation, which entails incorporation of prior distribution for parameters for which the biased estimators (like Minnesota Priors) represent the posterior mean. Colloquially this VAR representation can be understood as one that uses prior information along with the data for robust estimation. See the next section for a detailed presentation.

Litterman's implementation of Bayesian methods to VAR representation has been interpreted in literature to be Quasi-Bayesian as he assumes the variance of the parameters to be known. This methodology greatly simplifies the task of estimation, as Bayesian estimation is notoriously hard and time intensive, consider for example a situation where analytical joint posteriors of the parameters are not available, in this case inference is carried out by first integrating out the conditional posteriors of all the parameters and then using a sampling technique, like importance sampler, particle sampler, Metropolis-Hastings Algorithm, Gibbs sampler, etc for drawing inference.

With the advent to fast computing hardware and software, Bayesian methods have gained popular use in all fields of study, particularly in Economics, see Hamilton, J. D. (2006). A vast range of priors can now be

implemented for estimation and structural analysis based on researcher's prior knowledge and research needs.

In out paper we implement 3 different priors for forecasting inflation and Output gap in Bayesian VAR methodology. Out of the three prior methods the latter two have complex posterior forms; i.e. analytical posteriors for Independent Normal–Wishart priors and Independent Minnesota–Wishart Priors are not available. We therefore utilize a Gibbs sampler consistent with Koop and Korobilis (2009) for obtaining inference.

The discussion on forecast results shows that our methodology and in-sample forecasts are robust. Finally, we discuss model diagnostics for comparison of the three models and conclude there from.

The paper is structured as follows; next section discusses the relevant literature on Bayesian methods in VARs, section 3 discusses the model as well as prior elicitation, along with the information about the use of data sets for the study while section 4 concludes by stating the model findings, results and interpretation.

2. Literature Review

Linear stochastic difference equations or common acronym Vector Auto Regression (VAR) specifications mainly suffer from the caveat of probable presence of multicollinearity and over parameterization (Litterman, 1979). This problem which impacts the model specification as well as forecast accuracy is exacerbated in the presence of data set which is characterized by low range. Earliest Bayesian VAR specifications are based on theory compliant with Ridge regressions (Hoerl & Kennard, 1970), Stein Rule estimators, see Stein (1974), and ridge-type estimators in uni-variate autoregressive context as Swamy and Rappaport (1975). These studies demonstrate that they generate biased estimators which possess statistically smaller Mean Squared Errors (MSEs) than OLS estimates.

Ridge estimators are argued to overcome large sampling error problem generally associated with multicollinearity by adding prior information to the existing data that; larger coefficients signify that they are more unreasonable, thereby implying that inherently ridge regression aims to penalize the unreasonably large estimation coefficients by "shrinking" then towards zero. Therefore this class of estimators is also called Shrinkage estimators and the respective priors are termed as "Shrinkage Priors".

Litterman (1979) and Litterman (1980) argue that each of the above stated estimators have Bayesian interpretation, whereby the researcher specifies the prior distribution for the parameters of interest and the posterior mean based on the priors can be interpreted as posterior biased with respect to priors. Essentially these procedures imply estimates based on data as well as researcher's prior beliefs on the subject matter.

Collectively Ridge estimators or Stein rule estimators are analogous in their treatment of multicollinearity and large sampling estimation errors by incorporating prior information in the regression equation; however they differ in the metrics through which the 'shrinkage' is applied. Below we present the various ridge regression methods along with the one implemented by Litterman(1979) in his novel study.

The normal linear model can be stated as;

$$Y = X\beta + \varepsilon, \ \varepsilon \sim N(0, \sigma^2) \tag{1}$$

Ridge estimators can be expressed as;

$$\beta^* = (X'X + KI)^{-1}X'Y, \text{Where } K > 0 \tag{2}$$

These estimators may be classified as posterior means corresponding to priors illustrated in the representation; $\beta = N(0, \lambda^2 I)$, with $K = \sigma^2/\lambda^2$; where σ^2 and λ^2 are assumed to be known.

Stein class of estimators is similar to ridge estimators and can be stated as;

$$\beta^* = (X'X + KX'X)^{-1}X'Y, \text{Where } K > 0 \tag{3}$$

Where the prior is assumed to be defined as; $\beta = N(0, \lambda^2(X'X))$

Initial application of Bayesian VARs or Bayesian VARs implementing the Minnesota / Litterman priors are in essence similar to a related and similar type of Stein estimators given by Leamer (1972). Where Leamer (1972) He considers a geometrically decaying response pattern in the model below;

$$Y = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 X_{t-2} + \cdots + \beta_n X_T \tag{4}$$

Where;

$$E(\alpha) = \alpha_0, \ E(\beta_i) = mr^i, \ Var(\alpha) = \lambda^2 a, \ Cov(\beta_i \beta_j) = \lambda^2 w^{(i-j)} r^{(i+j-2)}$$

Litterman (1979) simplified the task of computation and specification of priors by deriving the matrix of coefficients through Bayesian methods while he assumed constant variance of the coefficients which is also

assumed to be known, thereby drastically decreasing the task of computation. The above treatment of shrinkage priors proposed by Litterman (1979) in essence of Leamer (1972), implies that as lag length increases the coefficients of these lags are scaled to have tight marginal distributions around zero, in effect the farther the lag, the least its explanatory weight in the VAR regression equation. Leeper, Sims and Zha (1996), Sims and Zha (1998) and Robertson and Tallman (1999) implement Bayesian VARs similar to Litterman (1986a and 1986b), they conclude that such priors can be used effectively to circumvent typical problems with VARs, also Bayesian VARs exhibit robust forecast performance.

The assumptions of Litterman have since 1970a been developed and expanded in many directions. Kadiyala and

Marta B., Domenico G., & Lucrezia R. (2008), Koop and Korobilis (2009) extend the initial Litterman's Priors by relaxing the assumption of a known diagonal error variance-covariance term.

Kadiyala and Karlsson (1993, 1997) are of particular importance as they consider 4 different classes of distributions used to parameterize the prior beliefs. Particularly Kadiyala and Karlsson (1997) conclude that the priors based on Normal-Wishart priors and the Minnesota Priors exhibit robust forecasting performance.

Karlsson S. (2012) and Koop and Korobilis (2009) provide extensive and concise representation of a host of prior selection, computation, forecasting and model diagnostics.

Further developments on prior elicitation include examples like Giannone, Lenza and Primiceri (2012) and George, Sun and Ni (2008) these papers illustrate hierarchical priors, i.e. prior based on a prior; this is a prior structure which is influenced by the data, thereby providing "objective" selection procedure for prior selection.

As argued earlier, incase both the coefficients and their variances are assumed to be unknown, then the researcher has to specify probability distributions of both coefficients and their variances, the same holds for posterior distribution and for forecasting. With the introduction of fast computing hardware and software at researcher's disposal a lot of head way has been achieved in economic structural analysis and forecasting.

3. Methodology, Data and Estimation

Relevant literature for Bayesian VARs identify a wide set of priors methodologies available for implementation in various scenarios. Assumptions for probability distribution, hyper-parameters and posterior sapling methods differentiate one BVAR method from another. In this essence we perform a structural exercise on the impact of a Monetary Policy shock from a Monetary Aggregates channel on Output Gap and Inflation, in line with Marta, Domenico and Lucrezia (2008). In all we utilize three types of Priors for estimation and forecasting of Inflation and Output Gap.

Output Gap estimates are obtained using State-Space framework for quarterly total Output series and then using cubic-spline interpolation method for obtaining output gap for Pakistan economy on monthly basis (see Appendix 1).

Monetary and Credit variables used in the BVARs are ordered as follows; Discount Rate (Policy rate set by the Central Bank of Pakistan i.e. State Bank of Pakistan (SBP)), Reserve Money (M0), Broad Money (M2), Public Sector Borrowing (PSB), Private Sector Credit (PSC), Output Gap and CPI Inflation. All variables are used on monthly frequency and used at level. This paper implements the statement methodology for VARs from Koop and Korobilis (2009) and Lutkepohl (2007).

The main rationale for implementing this methodology using monetary aggregates is to access the dynamic relationship of discretionary Monetary policy in controlling inflationary tendencies in Pakistan. Discount rate is assumed to curb inflationary pressures though monetary aggregates, which are assumed to impact the credit appetite of the Government borrowing (PSB) leading to increase of decrease in credit off take by the private sector (PSC), which ultimately translates in changes in Aggregate Demand (Output Gap), leading to changes in CPI inflation.

1: Minnesota Priors;

2: Independent Normal - Wishart Priors;

3: Independent Minnesota–Wishart Priors.

We elaborate first the VAR notation methodology to simplify the BVAR elicitation. VAR with p number of lags can be denoted as;

$$y_t = a_o + \sum A_j y_{t-j} + \varepsilon_t \tag{5}$$

Where;

y_t: For $t = 1, \dots, T$ is a $M \times 1$ vector of observations about M number of variables.

a_o: $M \times 1$ vector of intercepts;

A_j: $M \times M$ matrix of coefficients;

ε_t: $M \times 1$ vector of errors assumed to be $iid \sim N(0, \Sigma)$;

We can write the VAR in matrix notation following the convention in Canova (2007), i.e. let Y be a $T \times M$ matrix which stacks the T observations on the M variables in columns next to each other, let

$$x_t = (1, y'_{t-1}, \dots, y'_{t-p}) \tag{6}$$

$$X = (x_1, x_2 x_3, \dots, x_T)' \tag{7}$$

Note that, if we let $K = 1 + Mp$ be the number of coefficients in each equation of the VAR; X is a $T \, x \, K$ matrix. Also let, $A = (\alpha_0 \, A_1 \dots A_p)'$ and $\alpha = vec(A)$ therefore is $KM \times 1$ matrix which stacks all the VAR coefficients and the intercept into a vector; therefore we can write the VAR as; $Y = XA + E$

Or,

$$y = (I \otimes X)\alpha + \varepsilon \tag{8}$$

Where ε is $iid \sim N(0, \Sigma)$, Likelihood function can be derived from the sampling density $P(y|\alpha, \Sigma)$ which can be solved as one distribution for α given Σ another where Σ^{-1} has a Wishart distribution.

$$\alpha | \Sigma, y \sim N(\hat{\alpha}, \Sigma \otimes (X'X)^{-1}) \tag{9}$$

And;

$$\Sigma^{-1} | y \sim W(S^{-1}, T - K - M - 1) \tag{10}$$

Where; $\hat{\alpha} = vec(\hat{A})$ and $\hat{A} = (X'X)^{-1}X'Y$ and $S = (Y - X\hat{A})'(Y - X\hat{A})$

3.1 Minnesota Priors

Economists at the University of Minnesota and the Federal Reserve Bank of Minneapolis in 1980s laid the foundations for great simplification in prior elicitation and evaluation (for reference in great detail see Doan, Litterman, & Sims (1984); Litterman, 1986b). The simplification sighted is that in the case of replacing Σ with an estimate $\hat{\Sigma}$ doing so greatly reduces the computational burden as in this case analytical posterior and predictive densities are available for further analysis and forecasting. In the context of Bayes' Theorem, It can be seen that the task of estimation is much simplified as there only remains prior elicitation for the case of α. For more information and examples of implementation of BVAR analysis in a Minnesota Prior setup refer to Banbura et al. (2010), Koop and Korobilis (2009) for excellent discussion. The original Minnesota Priors (see Doan, Litterman, Sims (1984) simplify the task of selection of an appropriate $\hat{\Sigma}$ by choosing the matrix Σ from an OLS estimate. Banbura et al. (2010) conclude that implementation of Minnesota Priors in BVARs leads to robust forecasts, even in the scenario of large number of variables in the VAR. In its essence Minnesota Priors (Litterman (1986)) assume all VAR equations to be "centered" i.e. following random walk with drift;

$$Y_t = c_0 + Y_{t-1} + \mu \tag{11}$$

Meaning that, in the VAR connotations, individual elements of coefficient matrix converge or "Shrink" towards 1 (Note 1). Technically utilizing a suitable $\hat{\Sigma}$ the diagonal elements of coefficient matrix A_j are shrunk towards one and the off diagonal elements are sought to be shrunk towards zero, doing so is advantageous in two major aspects;

1: Shrinking diagonal elements to 1 would lead to own lags to exhibit more explanatory power in VAR.

2: The impact of recent lags gets more weight in explanation of the dependent variable of the VAR.

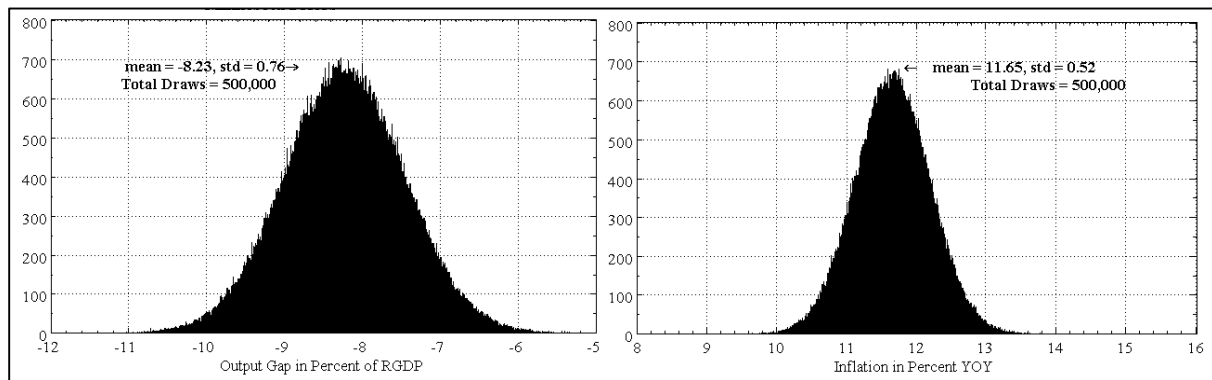The above objective is achieved by utilizing the following methodology;

Figure 1a. Predictive density for output gap
Nov-2013, Minnesota priors

Figure 1b. Predictive density for inflation
Nov-2013, Minnesota priors

$$E\big[(A_k)_{ij}\big] = \begin{cases} \sigma_i, & j = 1, K = 1 \\ 0, & otherwise \end{cases} \tag{12}$$

$$V\big[(A_k)_{ij}\big] = \begin{cases} \frac{\underline{a_1}}{r^2}, & for\ coefficients\ on\ own\ lag\ r, where\ r = 1,\dots,p \\ \frac{\underline{a_2}\sigma_{ii}}{r^2\sigma_{jj}}, & for\ coefficients\ on\ lag\ r\ of\ variable\ j \neq i, for\ r = 1,\dots,p \\ \underline{a_3}\sigma_{ii}, & for\ exogenous\ variables \end{cases} \tag{13}$$

Prior elicitation for the case of α is illustrated below, let;

$$\alpha \sim N(\alpha_{Mn}, V_{Mn}) \tag{14}$$

The Minnesota priors therefore simplify the task of specification of matrix V by choosing three scalars $\underline{a_1}$, $\underline{a_2}$, $\underline{a_3}$. The above specification allows for coefficient shrinkage towards zero and hence circumvents the over-specification or over-fitting problem of VARs with more than 3 or 4 variables. In this manuscript we use the following values in the BVAR methodology;

$$\underline{a_1} = 0.9, \ \underline{a_2} = 0.5, \ \underline{a_3} = 100 \tag{15}$$

Moving on the next step is the posterior formulation; One of the major advantages of Minnesota priors is that posterior inference involves only Normal Distribution, thus we can state;

$$\alpha|y \sim N(\overline{\alpha}_{Mn}, \overline{V}_{Mn}) \tag{16}$$

Where;

$$\overline{V}_{Mn} = \left[\underline{V}_{Mn}^{-1} + \left(\hat{\Sigma}^{-1} \otimes (X'X)\right)\right]^{-1} \tag{17}$$

$$\overline{\alpha}_{Mn} = \overline{V}_{Mn}\left[\underline{V}_{Mn}^{-1}.\underline{\alpha}_{Mn} + \left(\hat{\Sigma}^{-1} \otimes X\right)'y\right] \tag{18}$$

Using the monthly data at hand on the 7 variables described earlier we evaluate the BVAR with Minnesota priors, as with most VAR analyses the coefficients are not of primary concern, so we present here only the one period ahead forecast performance, see Table 1 and Figures 1a and 1b for details. For discussion on the results see Results section below.

3.2 Independent Normal-Wishart Prior and 3.3 Independent Minnesota-Wishart Priors

Development of Independent Normal-Wishart prior first appeared in the work of Zellner (1971) in the context of Seemingly Unrelated Regression Models, where he showed that marginal posterior for the matrix of coefficients can be expressed in a bi-modal stricture of Normal and Student's t distribution.

Let us express the advantages of implementing an independent prior setup as compared with a conjugate prior selection method. Natural Conjugate Priors impose an important restriction of dependence of probability distributions, doing so restricts the VARs to contain the exact same number of variables and lags, also, in Natural conjugate priors the prior covariance of the coefficients in any two equations is proportional to one another. However, Independent Normal-Wishart priors assume that VAR coefficients and the error covariance matrix to be "independent" of one another. This would lead to allow for the researcher to implement another variation of Priors i.e. Independent Minnesota-Wishart Priors. Theoretical underpinnings of the Independent Normal-Wishart

priors can be stated below. As we allow for VAR equations to contain different coefficients, we must change the notation pertaining to the VAR framework, following Koop and Korobilis (2009) and Karlsson (2012),

Here let β be the matrix of VAR coefficients such that;

$$y_{mt} = z'_{mt}\beta_m + \varepsilon_{mt} \tag{19}$$

Where;

y_{mt}: Is the tth observation on the mth variable.

z'_{mt}: Is the vector of dimensions $km \times 1$ containing the tth observation of explanatory variables for the mth variable.

β_m: Is the vector of dimensions kmx1 containing the coefficients conformable to z_{mt}.

Allowing for VARs with variable number of explanatory variables allows the researcher to impose automatic restrictions and also to allow for the desirable property of shrinkage of coefficients to zero. This is the major advantage of Independent Normal-Wishart Prior method over the Natural Conjugate priors. Moving on, we can thus write;
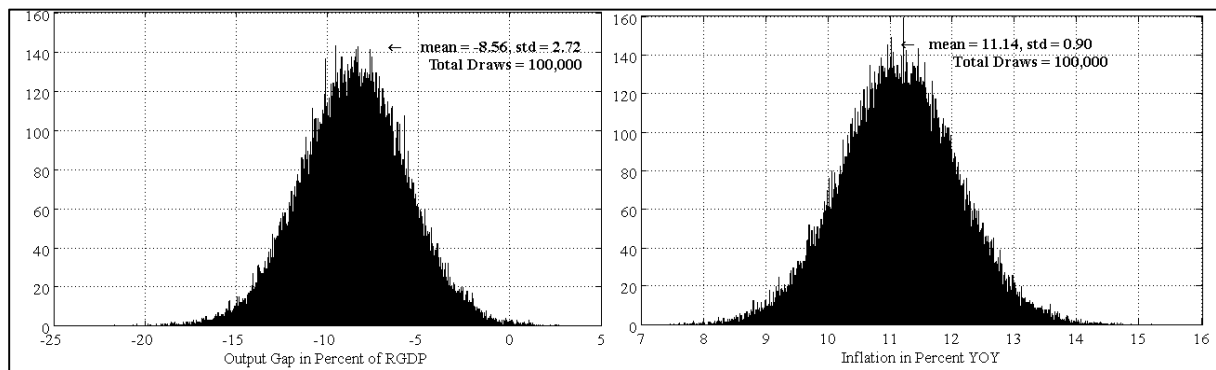

Figure 2a. Predictive density for output gap
Nov-2013, independent normal-Wishart priors

Figure 2b. Predictive density for inflation
Nov-2013, independent normal-Wishart priors

$$y_t = (y_{1t}, \dots y_{mt})' \tag{20}$$

$$\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Mt})' \tag{21}$$

$$\beta = (\beta_1, \dots, \beta_M)'_{kx1} \tag{22}$$

$$Z_t = \begin{bmatrix} z'_{1t} & 0 & \cdots & 0 \\ 0 & z'_{2t} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & z'_{Mt} \end{bmatrix}_{Mxk} \tag{23}$$

Where; $k = \sum_{j=1}^{M} k_j$, We can therefore write the VAR (which may possible be implemented in a restricted form) as;

$$y_t = Z_t\beta + \varepsilon_t \tag{24}$$

Where; $y_t = (y_1, \dots, y_T)'$, also, $\varepsilon = (\varepsilon_1, \dots, \varepsilon_T)'$, including $Z = (Z_1, \dots, Z_T)'$

And finally;

$$y = Z\beta + \varepsilon, \ \varepsilon \sim N(0, I \otimes \Sigma) \tag{25}$$

Utilizing the above notation that allows for VARs to be stated in the form of Restricted VARs we can now state the methodology for the Independent Normal-Wishart priors, Let;

$$p(\beta, \Sigma^{-1}) = p(\beta)p(\Sigma^{-1}) \tag{26}$$

Distributed as; $\beta \sim N\left(\underline{\beta}, \underline{V}_\beta\right)$ and $\Sigma^{-1} \sim W(\underline{S}^{-1}, \underline{v})$.

It is interesting to note here that we implement in our analysis two types of versions of the prior variance-covariance matrix $\underline{V}_\beta$;

Independent Normal–Wishart Priors we utilize the following convention; $\underline{V}_\beta = 10I_{KM}$.

Independent Minnesota–Wishart Priors we implement the Minnesota Prior for α (see equations 11–18).

Inference in Independent Normal-Wishart Priors and Independent Minnesota-Wishart Priors is computationally demanding as the joint posterior densities are not analytical and the inference must be carried out by first analytically integrating the conditional posterior distributions $p(\beta|y, \Sigma^{-1})$ and $p(\Sigma^{-1}|y, \beta)$ and thereafter using a type of Metropolis-Hasting algorithm, in this manuscript we implement a Gibbs sampler to draw inference of the Joint posterior densities. The Gibbs sampler used here is referenced in Koop, G. & Korobilis, D (2009). In order for Gibbs sampler to draw from the conditional distributions we specify them as;

A: Conditional distribution for β: $p(\beta|y, \Sigma^{-1})$;

$$\beta|y, \Sigma^{-1} \sim N(\overline{\beta}, \overline{V}_\beta) \tag{27}$$

$$\overline{V}_\beta = \left[\underline{V}_\beta^{-1} + \sum_{t=1}^{T} Z_t' \Sigma^{-1} Z_t\right]^{-1} \tag{)}$$

$$\overline{\beta} = \overline{V}_\beta \left[\underline{V}_\beta^{-1} \underline{\beta} + \sum_{t=1}^{T} Z_t' \Sigma^{-1} y_t\right] \tag{29}$$

B: Conditional Distribution for Σ^{-1}: $p(\Sigma^{-1}|y, \beta)$;

$$\Sigma^{-1}|y, \beta \sim W\left(\overline{S}^{-1}, \overline{v}\right) \tag{30}$$

Where;

$$\overline{v} = T + \underline{v} \tag{31}$$

$$\overline{S} = \underline{S} + \sum_{t=1}^{T}(y_t - Z_t\beta)(y_t - Z_t\beta)' \tag{32}$$

From here on a Gibbs sampler (Note 2) is used to draw sequentially from the Normal $p(\beta|y, \Sigma^{-1})$ and Wishart $p(\Sigma^{-1}|y, \beta)$. A Gibbs sampler is used to draw 10000 samples for the posterior distribution; the first 2000 are discarded as burn-in draws in conformation with relevant econometrics literature. In order to arrive at a point forecast; Predictive mean is used here.

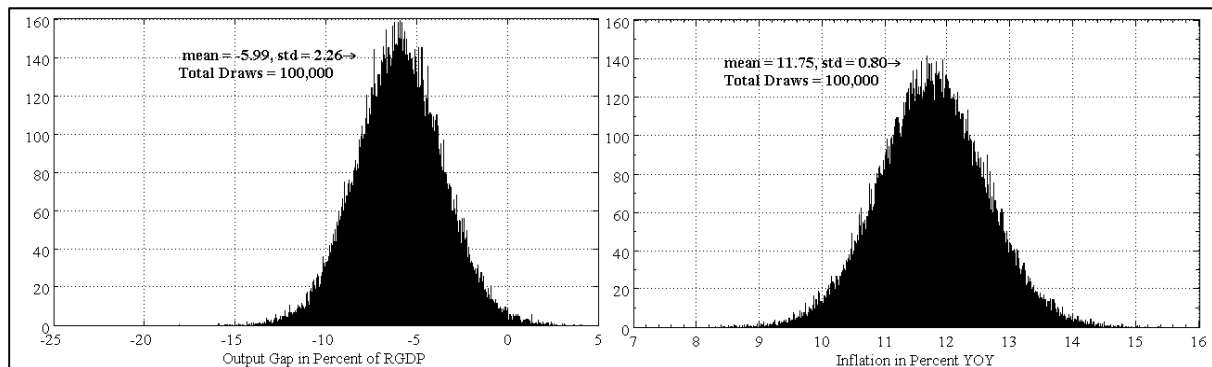$$Predictive\ Mean\ ; E(y_\tau|Z_\tau) = \sum_{r=1}^{R} Z_t \beta^{(r)}/R \tag{33}$$

Figure 3a. Predictive density for output gap
Nov-2013, independent Minnesota-Wishart priors

Figure 3b. Predictive density for inflation
Nov-2013, independent Minnesota-Wishart priors

For Predictive mean of one period ahead forecast of Potential Output and Inflation, and their standard deviation please refer to Figure 2a, 2b, 3a and 3b and Table 1 and 2.

4. Results and Discussion

The results described in the following Tables 1 and 2 are based on predictive draws estimated by the Gibbs sampler implemented for the three different priors. The forecasting exercise and the results thereof are conducted in sample, i.e. we use monthly sample of the range 1991M01–2013M11 for the seven variables in order to forecast the values of inflation and Output gap for the month of December 2014.

As discussed above, implementation of Minnesota prior simplifies the computational process whereby effectively the researcher only specifies the priors on the matrix of coefficients. Therefore the Minnesota prior representation of Bayesian VAR has analytical posterior and predictive densities. The posterior density in this prior setup effective comprises 500000 draws, whose mean represents the equivalent of point estimate for Bayesian analysis. See Table 1 and 2.

The last two Bayesian VARs based on Independent Normal-Wishart Priors and Independent Minnesota Wishart Priors are computationally demanding as described earlier. We therefore use a Gibbs sampler to obtain posterior densities for these two priors. The Gibbs sampler is programmed to draw 10000 draws from the predictive density, and in accordance with the literature first 2000 draws are discarded as burn-in draws.

Figures 1, 2 and 3 represent the predictive mean or Bayesian forecast for inflation and Output gap for the month of November 2013. All three methods yield reasonable and acceptable magnitude of forecast error (see Tables below). Where Bayesian VAR incorporating Independent Normal-Wishart Priors yields the least MSFE among the three methods proposed here.

Table 1. CPI inflation forecast (YOY) for Nov–2013 using data range Jan–1991 to Sep–2013

Priors	Lag Length	CPI Inflation Forecast for Nov–2013	Actual Value for Nov–2013	MSFE	MAE	Log Predictive Likelihood
Minnesota Priors	5	11.7	10.9	0.57	-0.76	-44.90
Independent Minnesota Wishart Priors	5	11.7	10.9	0.72	-0.85	-42.86
Independent Normal Wishart Priors	5	11.1	10.9	0.06	-0.24	-41.96

Output gap forecast from the three Bayesian VARs also yield robust predictive densities, signified by the relatively tight distribution of draws and the low magnitude of standard deviation (see Figures 1, 2, 3). Although all three VAR methods exhibit good overall performance, MSFE of Independent Minnesota-Wishart Priors appears to be least, which signifies this prior as a better candidate for forecasting in the horizon ahead.

Concluding we can observe from Table 1 and 2 that Bayesian VARS based on Independent Normal Wishart priors and Independent Minnesota-Wishart Priors perform better in comparison with Minnesota prior based VAR.

This is exemplified by the Log Predictive Likelihood, the value of the former two models is significantly better than the corresponding value of the Minnesota prior based Bayesian VAR.

Table 2. Output gap forecast for Nov–2013 using data range Jan–1991 to Sep–2013

Priors	Lag Length	CPI Inflation Forecast for Nov–2013	Actual Value for Nov–2013	MSFE	MAE	Log Predictive Likelihood
Minnesota Priors	5	-8.23	-5.60	6.94	-2.63	-44.90
Independent Minnesota Wishart Priors	5	-5.99	-5.60	0.15	-0.39	-42.86
Independent Normal Wishart Priors	5	-8.57	-5.60	8.80	-2.97	-41.96

Acknowledgements

References

Banbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian VARs. *Journal of Applied Econometrics, 25*, 71–92. http://dx.doi.org/10.1002/jae.1137

Doan, T., Litterman, R., & Sims, C. A. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews, 3*, 1–100. http://dx.doi.org/10.1080/07474938408800053

George, E., Sun, D., & Ni, S. (2008). Bayesian stochastic search for VAR model Restrictions. *Journal of Econometrics, 142*, 553–580. http://dx.doi.org/10.1016/j.jeconom.2007.08.017

Giannone, D., Lenza, M., & Primiceri, G. E. (2012). *Prior selection for vector Auto Regressions*. Universite Libre deBruxelles, Working Paper Series 2012–002.

Hamilton, J. D. (2006). *Computing Power and the Power of Econometrics*. Manuscript, University of California, San Diego.

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased regression for nonorthogonal problems. *Technometrics, 2*(1), 55–67. http://dx.doi.org/10.1080/00401706.1970.10488634

Kadiyala, K. R., & Karlsson, S. (1993). Forecasting with generalized Bayesian Vector Auto Regressions. *Journal of Forecasting, 12*, 365–378. http://dx.doi.org/10.1002/for.3980120314

Kadiyala, K., & Karlsson, S. (1997). Numerical Methods for estimation and Inference in Bayesian VAR models. *Journal of Applied Econometrics, 12*, 99–132. http://dx.doi.org/10.1002/(SICI)1099-1255(199703)12:2<99::AID-JAE429>3.0.CO;2-A

Karlsson, S. (2012). *Forecasting with Bayesian Vector Auto Regressions*. Orebro University, School of Business Working Paper Series No.12.

Koop, G., & Korobilis, K. (2009). Foundations and Trends. *Econometrics, 3*(4), 267–358.

Leamer, E. (1972). A class of informative priors and distributed lag analysis. *Econometrica, 40*, 1059–1081. http://dx.doi.org/10.2307/1913855

Leeper, E. M., Sims, C., & Zha, T. (1996). What does monetary policy do? *Brookings Papers on Economic Activity*, (2), 1–78. http://dx.doi.org/10.2307/2534619

Litterman, R. (1979). *Techniques of Forecasting using Vector Auto Regressions*. Federal Reserve Bank of Minneapolis, Working Paper Series No.115.

Litterman, R. (1986a). Forecasting With Bayesian Vector Autoregressions–Five Years of Experience. *Journal of Business and Economic Statistics, 4*, 25–38.

Litterman, R. (1986b). A statistical approach to economic forecasting. *Journal of Business and Economic Statistics, 4*(1), 1–4.

Litterman, R. B. (1980). *A Bayesian procedure for forecasting with Vector Auto Regressions*. Mimeo, Massachusetts Institute of Technology.

Lutkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis* (Vol. 2). New York, NY: Springer. http://dx.doi.org/10.1007/978-3-540-27752-1

Marta, B., Domenico, G., & Lucrezia, R. (2008). *Large Bayesian VARs*. European Central Bank, Working Paper Series No. 966.

Robertson, J. C., & Tallman, E. W. (1999). Vector auto regressions: Forecasting and reality. *Economic Review*, (Q1), 4–18.

Stein, C. M. (1974). *Multiple regressions*. Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling, Stanford University Press.

Swamy, P. A. V. B., & Rappoport, P. (1975). Relative Effeciencies of Some Simple Bayes Estimators of Coefficients in Dynamic Models. *Journal of Econometrics, 3*, 273–296. http://dx.doi.org/10.1016/0304-4076(75)90036-6

Notes

Note 1. For more explanation on coefficient "Shrinkage" see Litterman (1986a).

Note 2. For details on Gibbs sample utilized in the current analysis see appendix and for further reference Koop, G., & Korobilis D (2009).

Note 3. Abdullah Tahir (2014). *Estimating the Output Gap of Pakistan on Quarterly frequency using Structural Methods*. International Journal of Economics and Finance, Vol. 6, No. 5. 233–243.

Note 4. Quarterly estimates of real GDP are obtained using temporal disaggregation of annual real GDP using quarterly LSM index series as described in Enrique Q. (2005) and the accompanying Euro Trim software library. In all 5 different methods of temporal disaggregation of annual real GDP are implemented and simple average of the estimates from these 5 methods at each point in time is used for further analysis in the Bayesian VARs.

Note 5. YoY : Year on Year;

MSFE: Mean Squared Forecast Error;

MAE: Mean Absolute Error.

Appendix A

We implement a typical State Space representation utilizing real GDP (RGDP) and Inflation as observed variables and model RGDP as comprising a trend and cycle. Implementing standard assumptions on polynomials of lags in trend and cycle of real GDP series; trend is assumed to be modeled as $AR(1)$, while cyclical component is modeled as $AR(2)$. A typical Phillips curve equation is also included in the estimation. This representation is estimated by means of Kalman filter (Note 3).

$$Y_t = Y_t^{trend} - Y_t^{cycle} \tag{34}$$

$$Y_t^{trend} = Y_{t-1}^{trend} + \mu_{t-1} + \epsilon_t \tag{35}$$

$$Y_t^{cycle} = \tau_1 Y_{t-1}^{cycle} + \omega_t \tag{36}$$

$$\mu_t = \mu_{t-1} + \psi_t \tag{37}$$

$$\pi_t = \tau_2 \pi_{t-1} + \tau_2 \pi_t^* + \tau_3 Y_{t-1}^{cycle} + \varphi_t \tag{38}$$

Where $\epsilon_t, \psi_t, \omega_t, \varphi_t$ are Gaussian white noise terms with variances $\delta_\epsilon^2, \delta_\psi^2, \delta_\omega^2, \delta_\varphi^2$ respectively.

Y_t is real GDP, Y_t^{trend} is trend component of real GDP, Y_t^{cycle} is cyclical component of real GDP, π_t is CPI Inflation, π_t^* is the target level of Inflation.

The methodology adopted here is consistent with that in Harvey (1985), Watson (1986) Vineet (2004), Michal Anderle (2013) (Note 4).

Copyrights