Forecasting Monetary Policy Using Bayesian Vector Auto Regression

Matthew Koepke, Mehrdad Mohammadi, Michael Alexander Wieck-Sosa
STAT 429, Fall 2020

I. Introduction

Economic forecasting is often referred to as an art, perhaps because it involves not only data and groups of equations, or statistical models, but also the forecaster's personal beliefs about how the economy behaves and where it is heading at any moment. The Bayesian approach to statistics, a general method for combining beliefs with data, suggests an objective procedure for blending beliefs and data in economic forecasting models. This procedure provides a framework that forecasters can use to document and discuss their beliefs, which can help make economic forecasting more of a science and less of an art. One type of economic forecasting model, known as the Bayesian Vector Auto Regression (BVAR) model, has been developed explicitly along Bayesian lines and seems to be an improvement over other types of forecasting models. BVAR procedures give modelers more flexibility in expressing the true nature of their beliefs as well as an objective way to combine those beliefs with the historical record. A specific version of the procedures, developed by Minnesota researchers, has been used to build models whose unadjusted forecasts seem to be as accurate as the subjectively adjusted forecasts of other common models.

Moreover, relative to the widely used macroeconometric models, the BVAR approach has a distinct advantage in two respects. First and most important, it does not require judgmental adjustment. Thus it is a scientific method that can be evaluated on its own, without reference to the forecaster running the model. Second, it generates not only a forecast but a complete, multivariate probability distribution for future outcomes of the economy that appears to be more realistic than those generated by other competing approaches.

Our project will give a brief overview and justification of the Bayesian framework and discuss how a posterior distribution are constructed. Next, we will do an overview of the priors that are popular in the economic literature. Specifically, we will review the Minnesota prior (Litterman, 1986) and the Normal-inverted Wishart prior (Kadiyala, Karlsson, 1997) and discuss the modeling implications of using each prior. Next, we will use the Minnesota and Normal-inverted Wishart prior to forecast GDP, unemployment, and inflation (PCE). From a computation perspective, the Normal-inverted Wishart prior is ideal since it is a conjugate prior.

Finally, based on our forecasts, we will use the Taylor Rule and the Non-Accelerating Rate of Inflation (NAIRU) to come up with an appropriate path for monetary policy. The data used in this project in through Federal Reserve Bank of St. Louis's Fred Database .

A. Prior and Economic Forecasting

The problem of forecasting is to use past and current information to generate a probability distribution for future events. Generally speaking, this is one of the basic problems of statistical analysis, and many well-known statistical procedures have been developed and used successfully to forecast in a variety of contexts.

Some particular difficulties arise, however, in forecasting economic data. First, there is only a limited amount of data. Second, many complicated relationships that are only poorly understood and probably evolving over time interact to generate the data. Finally, it is generally impossible to perform randomized experiments to test hypotheses about those economic structures.

A realistic representation of the current state of economic theory requires a tremendous degree of uncertainty about the structure of the economy. Hence, a Bayesian procedure that can more accurately represent that uncertainty can produce a significant improvement over conventional techniques in the ability to generate a realistic probability distribution for future economic events.

Differently from frequentist statistics, Bayesian inference treats the VAR parameters as random variables, and provides a framework to update probability distributions about the unobserved parameters conditional on the observed data. By providing such a framework, the Bayesian approach allows to incorporate prior information about the model parameters into post-sample probability statements. The 'prior' distributions about the location of the model parameters summarise pre-sample information available from a variety of sources, such as other macro o micro datasets, theoretical models, other macroeconomic phenomena, or introspection.

In the absence of pre-sample information, Bayesian VAR inference can be thought of as adopting 'non-informative' (or 'diffuse' or 'flat') priors, that express complete ignorance about the model parameters, in the light of the sample evidence summarised by the likelihood function (i.e. the probability density function of the data as a function of the parameters). Often, in such a case, Bayesian probability statements about the unknown parameters (conditional on the data) are very similar to classical confidence statements about the probability of random intervals around the true parameters value. For example, for a VAR with Gaussian errors and a flat prior on the model coefficients, the posterior distribution is centred at the maximum likelihood estimator (MLE), with variance given by the variance-covariance matrix of the residuals.

The most commonly adopted macroeconomic priors for VARs are the the so-called 'Minnesota' priors (Litterman, 1980). They express the belief that an independent random-walk model for each variable in the system is a reasonable 'centre' for the beliefs about their time series behaviour. While not motivated by economic theory, they are computationally convenient priors, meant to capture commonly held beliefs about how economic time series behave. Minnesota priors can be cast in the form of a Normal-Inverse-Wishart (NIW) prior, which is the conjugate prior for the likelihood of a VAR with normally distributed disturbances.

B. The Problem of Dimensionality

The standard approach to specifying equations recognizes that given a limited number of observations, one must be very parsimonious about adding explanatory variables. Each additional coefficient must be estimated from the data; and even though doing this will always improve the fit in sample (though not always when adjustment is made for degrees of freedom), in the forecasts generated by the equation there will be a trade-off between decreased bias and increased variance. In a Bayesian specification framework, this trade-off disappears in that a mean squared error loss function is minimized by including all relevant variables along with prior information that accurately reflects what is known about the likely values of their coefficients.

One way to think about this problem is to view the forecasting equation as a filter that must pick out from the din of economic noise a weak signal that reveals the likely future course of the variable of interest. The standard approach takes the position that the best one can do is rely on economic theory to suggest at most a few places to look for useful information. The search for information becomes narrowly focused. The alternative BVAR approach is based on a view that useful information about the future is likely to be spread across a wide spectrum of economic data. If this is the case, a forecasting equation that captures and appropriately weights information from a wide range of sources is likely to work better than one with a narrow focus. The appropriate weights are the coefficient estimates, which combine information in the prior with evidence from the data.

VAR models are a generalization of univariate autoregressive (AR) models, based on the notion of interdependencies between lagged values of all variables in a given model. They are commonly resorted to as tools for investigating dynamic effects of shocks and perform very well in forecasting exercises. A VAR model of finite order p, referred to as VAR(p) model, can be expressed as:

$$y_t = a_0 + A_1 y_{t1} + \cdots + A_p y_{tp} + \epsilon_t \quad \text{with } \epsilon_t \sim N(0, \Sigma) \quad (1)$$

where y_t is an $M \times 1$ vector of endogenous variables, a_0 is an $M \times 1$ intercept vector, A_j $(j = 1, \ldots, p)$ are $M \times M$ coefficient matrices, and ϵ_t is an $M \times 1$ vector of exogenous Gaussian shocks with zero mean and variance-covariance (VCOV) matrix Σ. The number of coefficients to be estimated is $M + M^2 p$, rising quadratically with the number of included variables and linearly in the lag order. Such a dense parameterization often leads to inaccuracies with regard to out-of-sample forecasting and structural inference, especially for higher-dimensional models. This phenomenon is commonly referred to as the curse of dimensionality.

The Bayesian approach to estimating VAR models tackles this limitation by imposing additional structure on the model. Informative conjugate priors have been shown to be effective in mitigating the curse of dimensionality and allow for large models to be estimated (see Doan, Litterman, and Sims 1984). They push the model parameters towards a parsimonious benchmark, reducing estimation error and improving out-of-sample prediction accuracy. The flexibility of the Bayesian framework allows for the accommodation of a wide range of economic issues, naturally involves prior information, and can account for layers of uncertainty through hierarchical modeling.

II. Bayesian Vector Auto Regression

A. Model Setup

VAR model is specified as:

$$y_t = a_0 + A_1 y_{t1} + \cdots + A_p y_{tp} + \epsilon_t \quad \text{with } \epsilon_t \sim N(0, \Sigma) \quad (2)$$

where y_t is an $M \times 1$ vector of endogenous variables, a_0 is an $M \times 1$ intercept vector, A_j $(j = 1, \ldots, p)$ are $M \times M$ coefficient matrices, and ϵ_t is an $M \times 1$ vector of exogenous Gaussian shocks with zero mean and variance-covariance (VCOV) matrix Σ.

With flat priors and conditioning on the initial p observations, the posterior distribution of $\beta \equiv vec[a_0, A_1, \ldots, A_p]$ is centered at the Ordinary Least Square (OLS) estimate of the coefficients and it is easy to compute. However, working with flat priors leads to inadmissible estimators (Stein, 1956) and yields poor inference, particularly in large dimensional systems (see, for example, Sims, 1980; Litterman, 1986). One typical symptom of this problem is the fact that these models generate inaccurate out-of-sample predictions, due to the large estimation uncertainty of the parameters. In order to improve the forecasting performance of VAR models, Litterman (1980) and Doan, Litterman, and Sims (1984) have proposed to combine the likelihood function with some informative prior distributions. Using the frequentist terminology, these priors are successful because they effectively reduce the estimation error, while generating only relatively small biases in the estimates of the parameters. For a more formal illustration of this point from a Bayesian perspective, let's consider the following (conditional) prior distribution for the VAR coefficients

$$\beta | \Sigma \sim N(b, \Sigma \otimes \Omega \xi) \quad (3)$$

where the vector b and the matrix Σ are known, and ξ is a scalar parameter controlling the tightness of the prior information. The conditional posterior of β can be obtained by multiplying this prior by the likelihood function. Taking the initial p observations of the sample as given,the posterior takes the form;

$$\beta | \Sigma, y \sim N\Big(\hat{\beta}(\xi), \hat{V}(\xi)\Big) \quad (4)$$

$$\hat{\beta}(\xi) \equiv vec\Big(\hat{B}(\xi)\Big) \quad (5)$$

$$\hat{B}(\xi) \equiv \Big(x'x + (\Omega\xi)^{-1}\Big)^{-1}\Big(x'y + (\Omega\xi\nu)^{-1}\Big) \quad (6)$$

$$\hat{V}(\xi) \equiv \Sigma \otimes \Big(x'x + (\Omega\xi)^{-1}\Big)^{-1} \quad (7)$$

where $y \equiv [y_{p+1}, \ldots, y_T]'$, $x \equiv [x_{p+1}, \ldots, x_T]'$ and $x_t \equiv [1, y'_{t-1}, \ldots, y_{t-p}]'$ and ν is a matrix obtained by reshaping the vector b in such a way that each column corresponds to the prior mean of the coefficients of each equation (i.e. $b \equiv vec(\nu)$). If we choose a lower ξ, the prior becomes more informative, the posterior mean of β moves towards the prior mean, and the posterior variance falls.

B. The Minnesota Prior (Litterman, 1986)

Lütkepohl (2007) provides a derivation of the Minnesota prior is defined as:

$$p(\beta) = N(\beta_0, V_a) \tag{8}$$

Where β_0 is the prior mean for β. This value can be calculated by setting the first lag of each variable to one and all the other values in each equation to zero. V_a is variance-covariance matrix for the prior and is specified as

$$v_{a(i,j,l)} = \begin{cases} (\lambda/l)^2 \text{ if } i = j \\ (\lambda\theta\sigma_i/\sigma_j)^2 \text{ if } i \neq j \end{cases} \tag{9}$$

Where $v_{(l,i,j)}$ is the prior variance of $\beta_{l,i,j}$, λ is the prior standard deviation of the coefficients $\beta_{m,m,1}$ where $m = 1...M$, σ_i^2 is the i^{th} diagonal of Σ and $0 < \theta < 1$,. For this prior, λ controls how closely the coefficient for the first lag of the dependent variable is centered around zero. Additionally, the θ parameter is used to control the variance of variables other than the dependent variable. A θ value of 0.1 would shrink the variance variables not related to the dependent variable. The logic for this shrinkage is two-fold: first, variance declines with higher order lags and second, for most variables, much of the variation can be explained by the variable's own lags. Finally, with regards to Σ, it is important to note that Σ is usually unknown; therefore, the least-squares estimator for Σ is typically used in practice. Lütkepohl (2007). Given this prior, we can estimate the posterior distribution $p(\beta|Y) = p(\beta)L(Y|\beta, \Sigma_a)$. Where β and Σ_a are the unknown parameters. The likelihood function is given as:

$$p(\beta) \propto |V_a|^{-\frac{1}{2}} exp[-\frac{1}{2}(\beta - \beta_0)^t V_a^{-1}(\beta - \beta_0)]$$
$$L(Y|\beta, \Sigma) \propto |tr[\Sigma]|^{\frac{MT}{2}} exp[-\frac{1}{2}t(Y - X\beta)^t \Sigma^{-1}(Y - X\beta)]$$

$$\begin{aligned} p(\beta|y) \propto \exp[-\frac{1}{2}(V^{-\frac{1}{2}}(\beta - \beta_0))^t((V^{-\frac{1}{2}}(\beta - \beta_0)) \\ + (tr[\Sigma^{-\frac{1}{2}}](Y - X\beta))^t(tr[\Sigma^{-\frac{1}{2}}](Y - X\beta)] \end{aligned}$$

We can decompose the above marginal distribution into two vectors W and Z so the above equation can be written in the form of $-\frac{1}{2}(W - Z\beta)^t(W - Z\beta)$

$$W = \begin{bmatrix} V_a^{-\frac{1}{2}}\beta_0 \\ tr[\Sigma^{-\frac{1}{2}}]y \end{bmatrix} Z = \begin{bmatrix} V_a^{-\frac{1}{2}} \\ X^t \otimes \Sigma^{-\frac{1}{2}}\beta_0 \end{bmatrix} \tag{10}$$

The exponent for the posterior distribution, equation (3), can know be written as:

$$-\frac{1}{2}[(\beta - \hat{\beta})^t Z^t Z(\beta - \hat{\beta}) + (W - Z\hat{\beta})^t(W - Z\hat{\beta})] \tag{11}$$

$$\hat{\beta} = (Z^t Z)^{-1} Z^t W = [V_a^{-1} + (XX^t \otimes \Sigma)]^{-1}[V_a^{-1}\beta_0 + (X \otimes \Sigma)y] \tag{12}$$

$$\hat{\Sigma_a} = (Z^t Z)^{-1} = [V_a^{-1} + (XX^t \otimes \Sigma^{-1})y] \tag{13}$$

Note, in equation (5) the second term does not contain a β. Therefore, we can pull the second term out to get the resulting posterior marginal distribution for β:

$$p(\beta|Y) \propto exp[-\frac{1}{2}(\beta - \hat{\beta})^t \hat{\Sigma_a^{-1}}(\beta - \hat{\beta})] \tag{14}$$

Equation (15) has a multivariate normal distribution $p(\beta|Y) \sim N(\hat{\beta}, \hat{\Sigma})$

C. Important Considerations

According to Lütkepohl (2007), there are some practical considerations to take into account when using the Minnesota prior. Specifically, there are large computational costs when calculating the prior mean $\hat{\beta}$ since this requires calculating the inverse of $\Sigma_a^{-1} + (XX^t \otimes \Sigma^{-1})$, which is typically large. One solution is to estimate each equation separately; however, the implication of this approach, is that this won't take into consideration cointegration amongst the series.

D. The Normal-Inverted Wishart Prior (Kadiyala, Karlsson, 1997)

The Normal-Inverted-Wishart (NIW) prior was proposed as an alternative to the Minnesota prior. Unlike the Minnesota prior, the NIW prior is a conjugate prior for the posterior distribution; moreover, the Normal-Wishart prior relaxes the assumptions imposed on the residual variance-covariance matrix made by the Minnesota prior. The NIW prior is defined as:

$$\begin{gathered} p(\beta, \Sigma_a) = p(\Sigma_a)p(\beta|\Sigma_a) \\ p(\Sigma_a) \sim W^{-1}(V_0, n_0) \\ p(\beta) \sim N(\beta_0, \Sigma_a \otimes C^{-1}) \end{gathered} \tag{15}$$

Where V_0 is a m x m matrix, C is a (mp + 1) x (mp + 1) matrix, β_0 is a m x (mp + 1) matrix, and n_0 represents the degrees of freedom for the inverse Wishart distribution. Tsay (2014) provides a derivation of the NIW, which is used here. For our model defined in (1), the posterior distribution is defined as:

$$\begin{gathered} p(\beta, \Sigma_\beta|Y) \propto |\Sigma_\beta|^{-\frac{n_0+m+1}{2}} exp[-\frac{1}{2}tr(V_0\Sigma_\beta^{-1})] \\ x\ |\Sigma_a|^{-\frac{k}{2}} exp[-\frac{1}{2}tr[(\beta - \beta_0)^t C(\beta - \beta_0)\Sigma_\beta^{-1}] \\ x\ |\Sigma_\beta|^{-\frac{n}{2}}|exp[-\frac{1}{2}tr[(Y - X\beta)^t(Y - X\beta)\Sigma_\beta^{-1}]] \end{gathered} \tag{16}$$

In a similar fashion as equation (10), we can decompose the posterior distribution into two vectors W and Z by defining $C = U^t U$, where U is an upper triangular matrix.

$$W = \begin{bmatrix} X \\ U \end{bmatrix} Z = \begin{bmatrix} Y \\ U\beta_0 \end{bmatrix} \tag{17}$$

Based on this decomposition, we get the following relationships:

$$(\beta - \beta_0)^t C(\beta - \beta_0) + (Y - X\beta)^t(Y - X\beta) = (Z - W\beta)^t(Z - W\beta) \tag{18}$$

$$(Z - W\beta)^t(Z - W\beta) = \tilde{S} + (\beta - \tilde{\beta})^t W^t W(\beta - \tilde{\beta}) \tag{19}$$

$$\tilde{\beta} = (X^t X + C)^{-1}(X^t X\hat{\beta} + C\beta_0) \tag{20}$$

$$\tilde{S} = (Y - X\tilde{\beta})^t(Y - X\tilde{\beta}) + (\tilde{\beta} - \beta_0)^t C(\tilde{\beta} - \beta_0) \tag{21}$$

Using the relationship found in equation (20), we rewrite the posterior distribution as follows:

$$\begin{gathered} p(\beta, \Sigma_a|Y) \propto |\Sigma_a|^{-\frac{m}{2}} exp[-\frac{1}{2}tr[(\beta - \tilde{\beta})^t W^t W(\beta - \tilde{\beta}\Sigma_a^{-1}] \\ x\ |\Sigma_a|^{-\frac{(n_0+n+m+1}{2}} exp[-\frac{1}{2}tr[(V_0 + \tilde{S})\Sigma_a^{-1}]] \end{gathered} \tag{22}$$

The first term of equation (23) is a multivariate normal distribution, while the second term has an inverted Wishart distribution. The posterior distributions for β and Σ_a are:

$$\begin{gathered} \Sigma_a|Y \sim W^{-1}(V_0+\tilde{S}, n_0+n) \\ \beta|Y, \Sigma_a \sim N(\tilde{\beta}, \Sigma_a \otimes (X^t X + C)^{-1} \\ \text{where } n = T - p \end{gathered} \tag{23}$$

E. Important Considerations

One of the challenges with using this prior is the need to have prior information about correlations between the parameters. This is likely to be very difficult as the number of parameters increases. One potential solution is to select a vague prior by setting β_0 to zero and selection a large variance-covariance matrix for C. The implication of using a vague prior is that the estimates will closely resemble the least square estimates. Another potential solution is to set $C = \lambda I_{mp+1}$ and to perform a sensitivity analysis to test many priors (Tsay, 2014).

III. Economic Forecasting and Monetary Policy

The Federal Open Market Committee (FMOC) is the governing body responsible for determining U.S. monetary policy. The FMOC meets regularly eight-times a year to assess economic condition and to determine which policies it should adopt. Since the financial crisis in 2008, the policy tools at the FOMC's disposal have grown considerably. Prior to the financial crisis, the FOMOC's conventional policy tools included changing the Federal Funds rate, which is the rate banks lend to other bank on a short-term basis; changing the discount rate, which is the rate the Federal Reserve lends to banks on a short-term basis; and changing the reserve ratio, which the proportion of deposits makes must hold in reserve.

Prior to the crisis, the most commonly used policy tool was changing the Federal Funds rate. The Federal Reserve would toggle the Federal Funds rate using open market operations from which it would buy or sell bonds from primary dealers, which would either decrease (buy bonds) or increase (sell bonds) the Federal Funds rate. Taylor (1993) proposed a rule for which the FOMC could target the Federal Funds rate:

$$FF_{target} = r * + 1.5(\pi - \pi *) + .5(y - y*) \tag{24}$$

Where FF_{target} is the target Federal Funds rate; r^* is the long-run trend, usually 2.0 percent π is the inflation rate, measured in personal consumption expenditures; π^* is the target inflation rate, set at 2.0 percent; and y and y* are the logs of GDP and potential GDP.

IV. Results

In our experiment, we use Bayesian Vector Auto Regressions to forecast macroeconomic variables which are used to calculate the Federal funds rate using the Taylor Rule, which is a benchmark econometric formula used by economists at the Federal Reserve to set monetary policy.

A. Data

First, we gathered monthly data from the Federal Reserve Economic Data (FRED) website from January 1, 1959 to July 1, 2020 of the unemployment rate,[1] personal consumption expenditures excluding food and energy (PCEPILFE),[2] and personal consumption expenditures chain-type price index (PCEPI).[3] None of these monthly variables were transformed. Second, we collected quarterly data from FRED from January 1, 1959 to July 1, 2020 of the real gross domestic product (real GDP),[4] and used a spline interpolation method to get monthly real GDP data. Afterwards, we took a log transformation of the monthly real GDP data that was interpolated with the spline. Third, similar to the real GDP, we got quarterly data from the Congressional Budget Office (CBO) from August 1, 2020 to May 1, 2021 of the real potential gross domestic product (real potential GDPD),[5] and used a spline to get monthly real potential GDP data, then took a log transformation. Lastly, we conducted a Priestley-Subba Rao Test for stationarity with 5 tapers and 10 blocks. The null hypothesis of stationarity was rejected at the 0.05 significance level. The time series data are plotted in Figure 1 below.

B. Models

We built our BVAR models using the "BMR" library in R to forecast the unemployment rate, real GDP, PCEPI, and PCEPILFE variables. For the first BVAR model, we used a Minnesota prior using the bvarm() function. Following Canova (2007) and Koop et al. (2010), we set the coefficient prior to 0.9 for each variable, the var type to 1, the decay type to 1, hyperparameter 1 to 0.2, hyperparameter 2 to 0.5, hyperparamter 3 to 10,000, hyperparameter 4 to 1, the number of lags to 4, and we included the constant parameter in the equation. For the second BVAR model, we used a normal-inverse Wishart prior using the bvarinw() function. Following Canova (2007) and Koop et al. (2010), we set the coefficient prior to 0.9 for each variable, XiBeta to 4, XiSigma to 1, gamma to 4, the number of lags to 4, and we included the constant parameter in the equation. For both models, we took 10,000 draws for the Gibbs sampler. See the figures in the Appendix section below for the impulse response function plots.

C. Calculating the Federal Funds Rate

For both models, we forecasted 10 months ahead from July 1, 2020. See the figures below for the forecast plots. Using the forecast estimates for the PCEPILFE series, we calculated the year-by-year personal consumption expenditures (PCE) inflation rate as:

$$p = (PCE(t+12) - PCE(t))/PCE(t) \tag{25}$$

Using the forecast estimates for the real GDP and the original real potential GDP from the Congressional Budget Office (CBO), we calculated the percent deviation of real GDP from the real potential GDP as:

$$y = (RPGDP(t) - RGDP(t))/RGDP(t) \tag{26}$$

Lastly, we calculated the Federal funds rate using the Taylor Rule formula as

$$r = p + 0.5y + 0.5(p-2) + 2 \tag{27}$$

where r is the federal funds rate, p is the rate of inflation, and y is the percent deviation of the real GDP and the real potential GDP. See Table 1 and Table 2 in the Appendix section below for the Federal funds rate results for both the Minnesota and normal-inverse Wishart prior.

[1]https://fred.stlouisfed.org/series/UNRATE
[2]https://fred.stlouisfed.org/series/PCEPILFE
[3]https://fred.stlouisfed.org/series/PCEPI
[4]https://fred.stlouisfed.org/series/GDPC1
[5]https://www.cbo.gov/data/budget-economic-data6

V. Discussion

From the impulse response functions, variable forecasts, and Federal funds rate estimations, we observe that the BVAR model with the Minnesota prior gives much more reasonable values for a longer time period for the 10 month forecast than the BVAR model with the normal-inverse Wishart model. In fact, we see that the BVAR model with the normal-inverse Wishart prior explode to extreme values for the variable forecasts after the first 4 months. Nevertheless, for the first 4 months, both models give reasonable forecast values. Without further analysis, it is hard to say which of the models is to be preferred for the first 4 months in econometric practice. However, it is clear that for the 10 month forecast, the BVAR model with the Minnesota prior is superior.

VI. References

Beauchemin, K., & Zaman, S. (in press). A Medium Scale Forecasting Model for Monetary Policy. Federal Reserve Bank of Cleveland Working Paper.

Canova, Fabio, Methods for Applied Macroeconomic Research, Princeton, New Jersey: Princeton University Press, 2007.

Doan, T., Litterman, R., and Sims, C. (1984), "Fore- casting and Conditional Projection Using Realistic Prior Distributions," Econometric Reviews, 3,

Federal Reserve Board of Governors. (2018). Federal Reserve Board - Policy Rules and How Policymakers Use Them. Board of Governors of the Federal Reserve System. https://www.federalreserve.gov/monetarypolicy/policy-rules-and-how-policymakers-use-them.html

Litterman, Robert B. (1980), "A Bayesian Procedure for Forecasting With Vector Autoregression," Working Paper, Massachusetts Institute of Technology, Dept. of Economics.

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

Kadiyala, K. R. A. O., Karlsson, S. U. N. E. (1997). Numerical Methods for Estimation And Inference in Bayesian VAR-Models. Journal of Applied Econometrics, 12(2), 99–132

Koop, Gary and Dimitris Korobilis, "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics," Mimeo, 2010.

Stein, C. (1956): "Inadmissibility of the Usual Estimator for the Mean of aMultivariate Normal Distribution," Proc. Third Berkeley Symp. on Math. Statist. and Prob.,1, 197–206.

Lutkepohl, H. (2007): "New Introduction to Multiple Time Series Analysis", Springer

Tsay, R. (2014): "Multivariate Time Series Analysis", Wiley

VII. Appendix

DATE <date>	fedFundsRate <dbl>
2020-08-01	1.020861
2020-09-01	1.023069
2020-10-01	1.023718
2020-11-01	1.024525
2020-12-01	1.022028
2021-01-01	1.020016
2021-02-01	1.018366
2021-03-01	1.020801
2021-04-01	1.028950
2021-05-01	1.028014

Table 1:Minnessota Prior: Federal Funds Rate

DATE <date>	fedFundsRate <dbl>
2020-08-01	1.0186543
2020-09-01	1.0175375
2020-10-01	1.0159431
2020-11-01	1.0144945
2020-12-01	1.0089978
2021-01-01	1.0000223
2021-02-01	0.9860956
2021-03-01	0.9640597
2021-04-01	0.9366244
2021-05-01	0.8891390

Table 2: Wishart Prior: Federal Funds Interest Rate

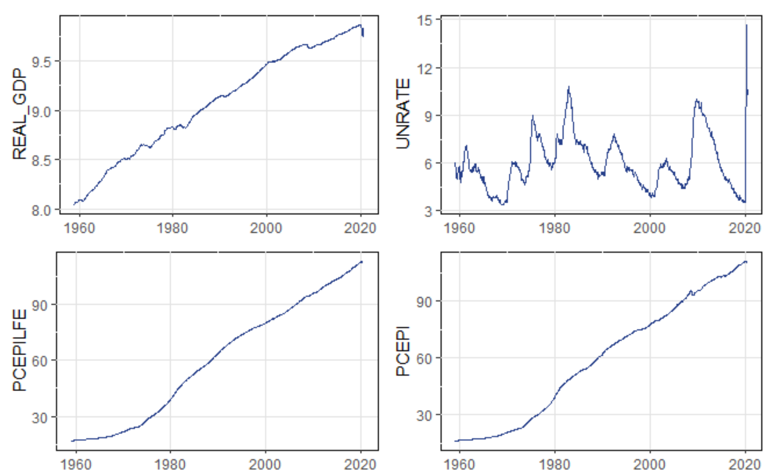

Figure 1. Time series plot of the variables

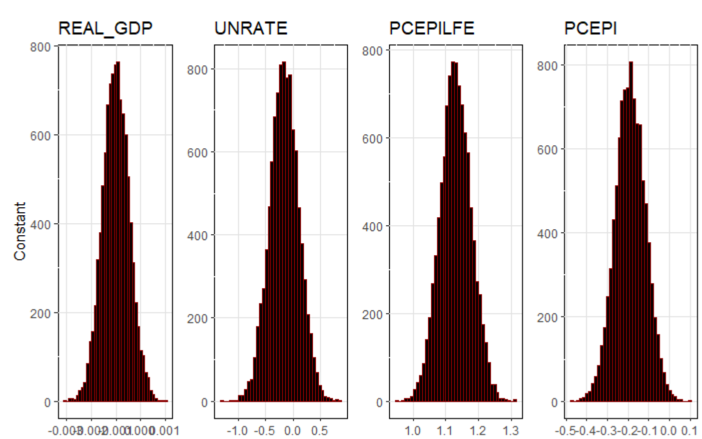

Figure 2. Minnesota Prior: Density functions of constant parameter estimations

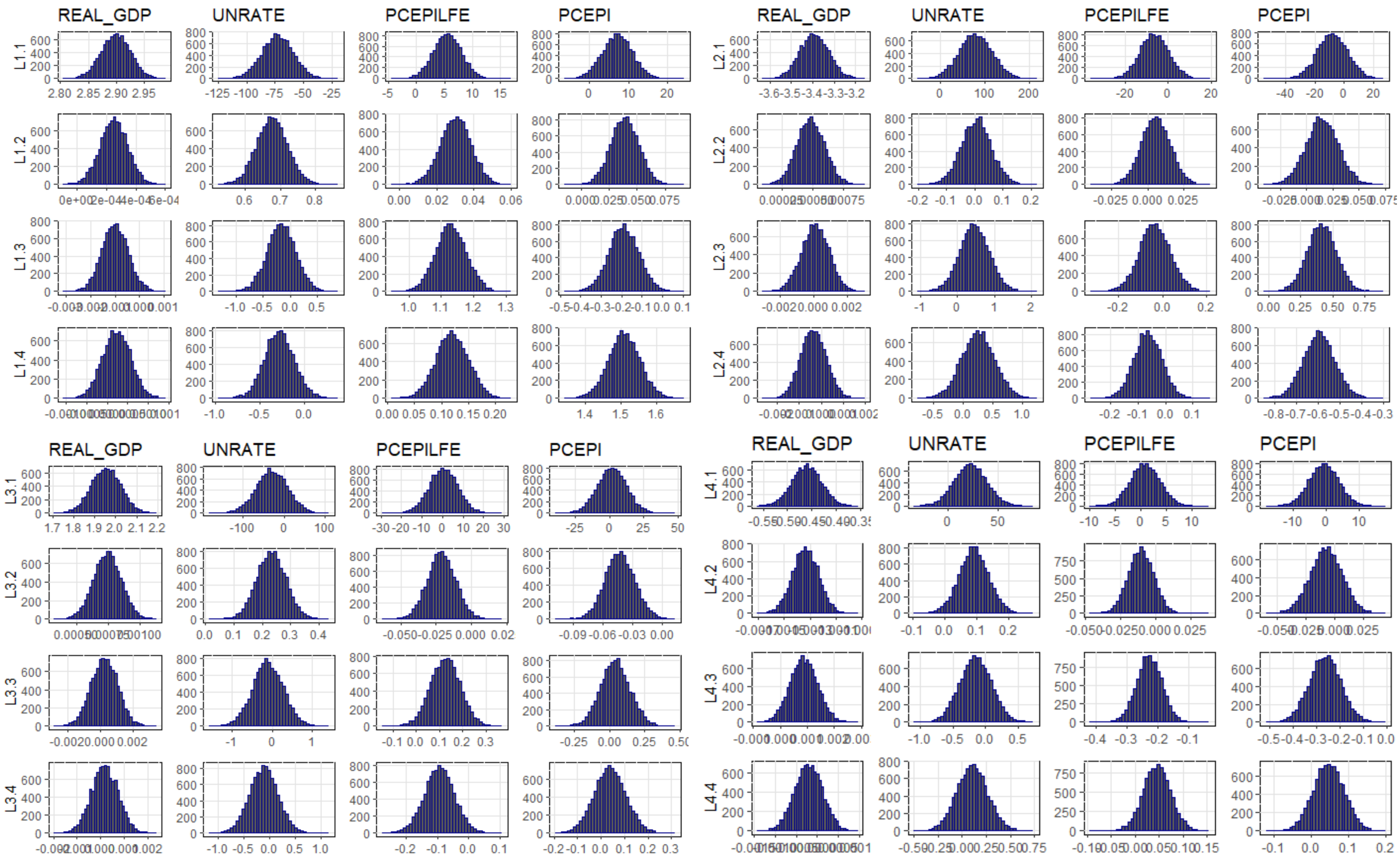

Figure 3. Minnesota Prior: Density functions of coefficient parameter estimations

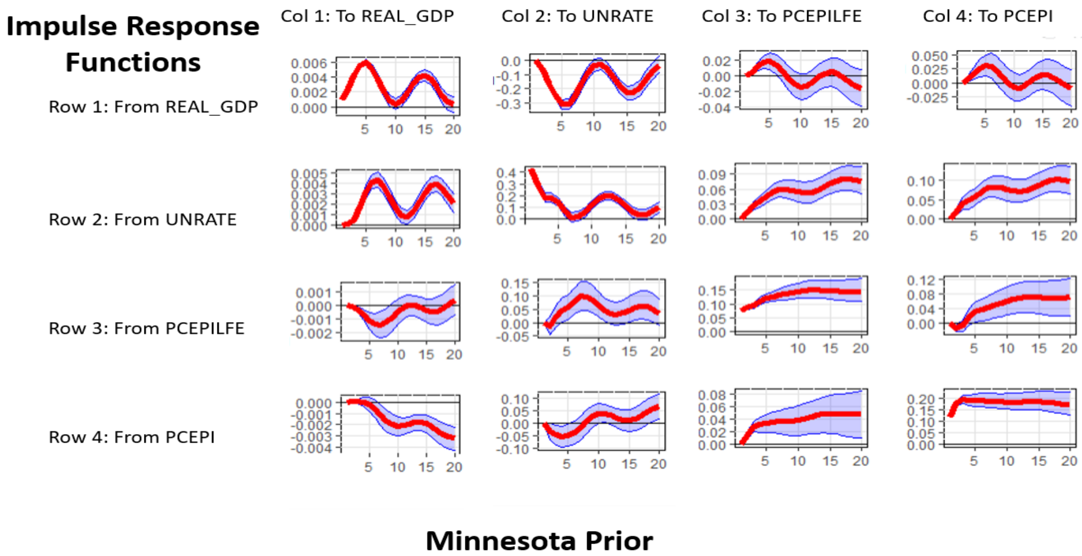

Figure 4. Minnesota Prior: Impulse response functions

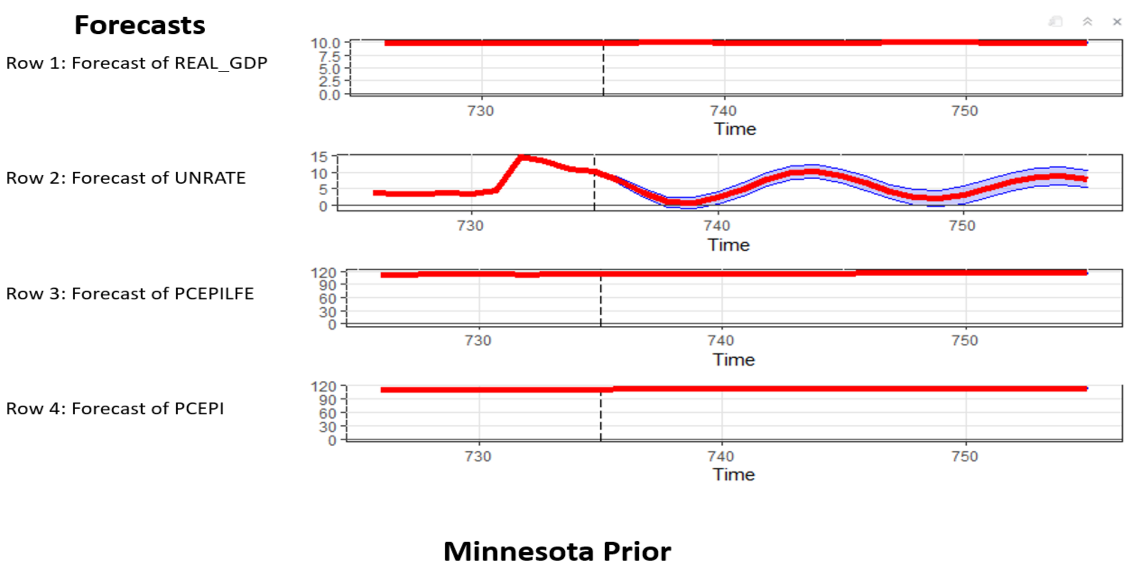

Figure 5. Minnesota Prior: Forecasts

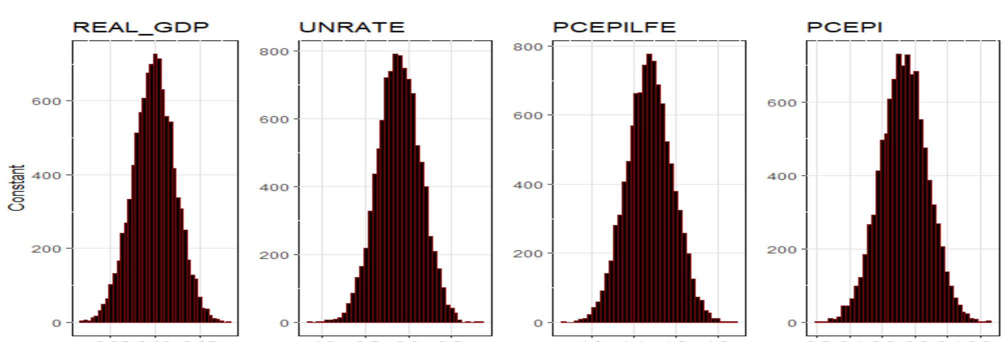

Figure 6. Wishart Prior: Density functions of constant parameter estimations

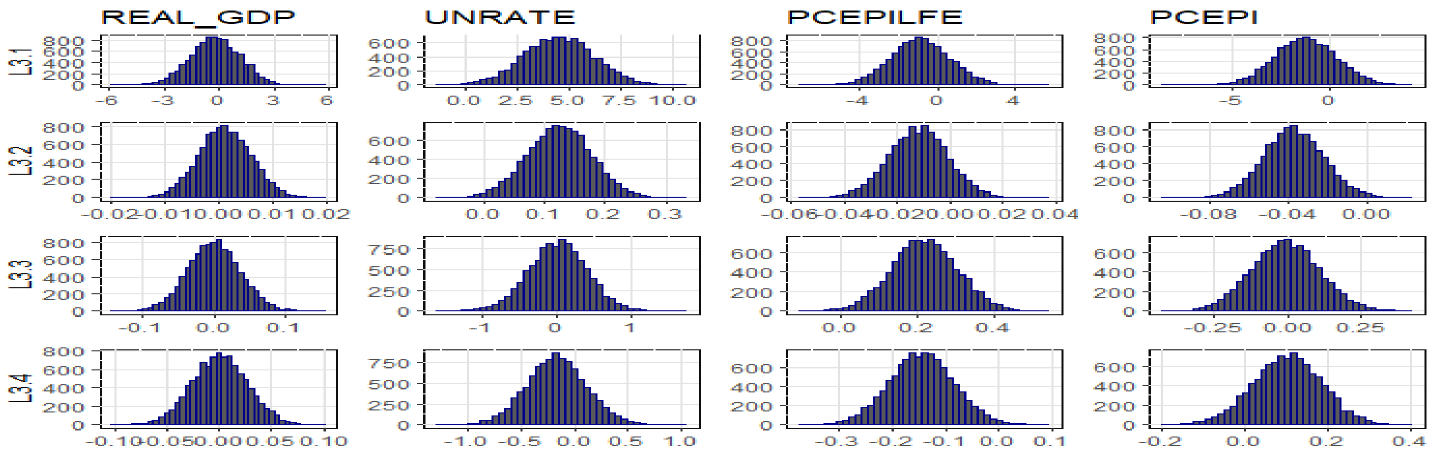

Figure 7. Wishart Prior: Density functions of Coefficient parameter estimations

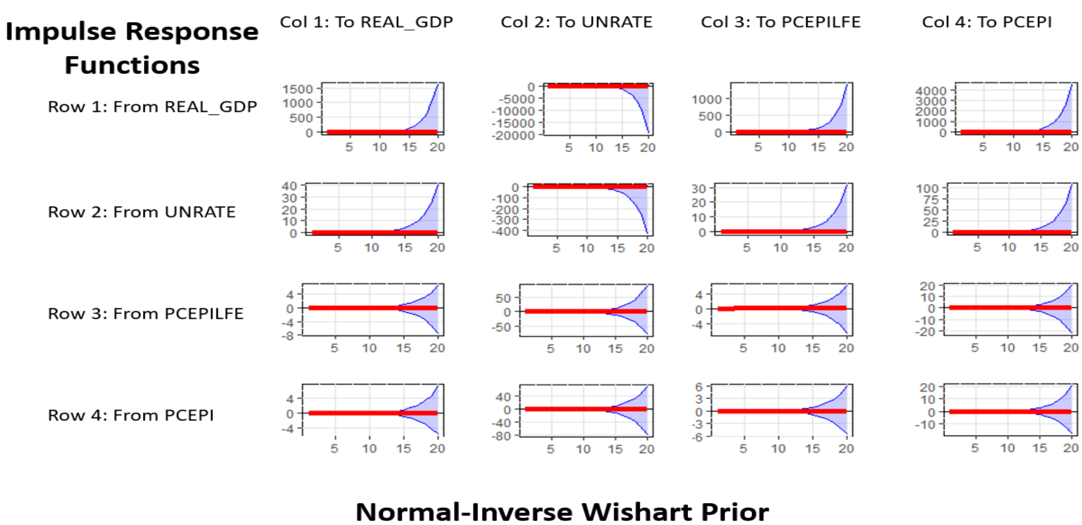

Figure 8. Wishart Prior: Impulse response functions

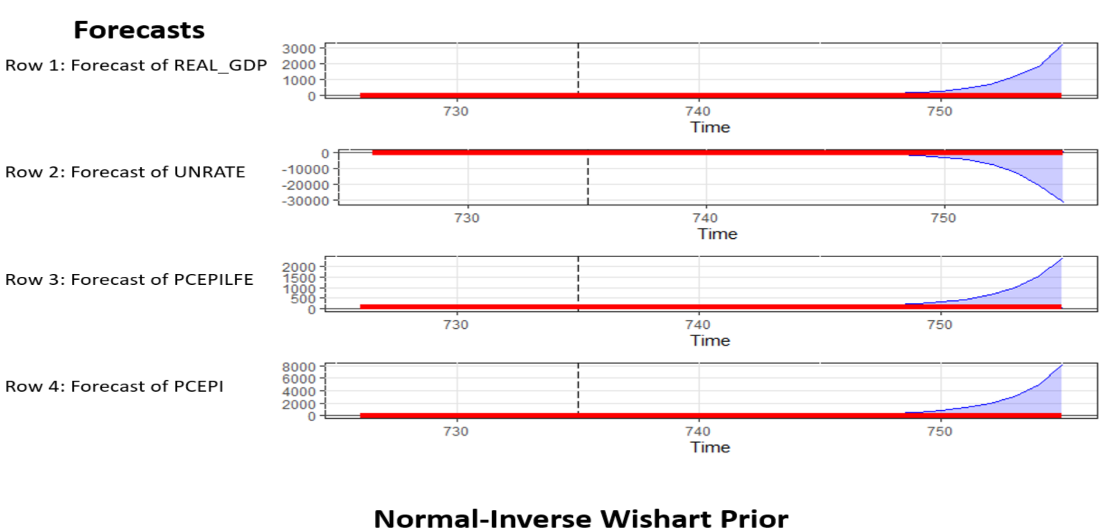

Figure 9. Wishart Prior: Forecasts