Rattanakosin Journal of Science and Technology: RJST
2020, Volume 2 Issue 3: 16-25
Research Article
Online ISSN:2673-0553

Bayesian Analysis of Unrestricted Vector Autoregressive Model with Non-Normality Innovations

Ojo O. Oluwadare[1*]

Abstract

Assumption of normality in innovations is often used in Vector Autoregressive (VAR) model, but this assumption is very sensitive to outliers. Outliers pose a major challenge to econometricians and practitioners. However, the unconditional distribution using t-distribution has fatter tails than the usual normality assumption. In this work, new efficient Bayesian approach when the innovations of VAR model follow a t-distribution was developed and discussed. The empirical example used a simulated data and data on USA economy by extending the usual normality to assume a t-distribution. The results show that the developed Bayesian approach is good for forecasting while small degree of freedom was capable to diminish the effects of outliers.

Keywords: Bayesian, Innovations, Normality, Outliers, VAR.

JEL classifications: C13, C15

1. Introduction

Since the development of Vector Autoregression (VAR) model by Sims [1] and followed by works of Lutkepohl [2] and Watson [3], VAR model had proved to be useful for describing the dynamic behaviour of both the economic and financial time series and forecasting. Apart from forecasting, it can also be used for structural inference and policy analysis.

The most prominent work on the use of Bayesian in VAR is Litterman [4]. Other works on Bayesian VAR model are the works of Sims and Zha [5], Sims and Ni [6]. Litterman [4] proposed the use of

[1] Department of Statistics, Federal University of Technology Akure, Nigeria.
*Corresponding author, Email: ojooo@futa.edu.ng

prior called Minnesota prior. This Minnesota is a kind of prior that entails set of data centric prior beliefs that shrinks the parameters towards a representation of macroeconomic variables.

Sims and Zha [5] also demonstrated on how to generate error bands on impulse responses in dynamic linear models using Bootstrap and Bayesian methods estimated from both reduced and identified forms of VAR model using flat priors. It was observed that their developed approach can be extended to over-identified case.

The properties of Bayes estimators of VAR coefficient and covariance matrix under two loss functions was investigated by Sun and Ni [6]. Two priors employed for the estimation were the reference and Jeffreys priors [7]. Their results show that the estimates with the use of constant prior for VAR dominate the constant Jeffreys prior.

A non-parametric VAR model that allows for non-linearity in heteroscedasticity in the conditional variance, conditional mean and non Gaussian innovations was proposed by Kalli and Griffin [8]. The proposed method was applied to both Euro-zone and USA macroeconomic time series and then compared to earlier proposed Bayesian VAR models. Results from the work shows that their proposed Bayesian Non-parametric VAR model is flexible.

Most of aforementioned Bayesian VAR models discussed earlier used the usual normality assumption, but the normality might not be able to cover all the degree of leptokurticity in the return series and it is also observed that this normality assumption is very sensitive to outliers. An Outlying observation is a major concern in any empirical analysis and it can have a great effect on mean and standard error significantly.

Herein, a new efficient Bayesian approach when the innovations of VAR model follow a non-normal assumption will be analyzed and discussed in this work.

The remainder of this paper is structured as follows. Section 2 reviews the VAR model and the distributional assumptions. Bayesian approach in VAR model for normality assumption and efficient Bayesian procedure with student-t innovations are described in section 3. In section 4, numerical studies will be conducted while the developed Bayesian method for t-distribution innovations will be compared with the Bayesian approach under normality using a real life data and simulated data. Section 5 discusses the results presented and section 6 renders the conclusion of the work.

2. Vector Autoregressive (VAR) model

Consider a VAR model given as:

$$Y_t = \alpha + \sum_{i=1}^{k} \beta_i \, Y_{t-i} + u_t \tag{1}$$

where $Y_t = (y_{1t}, y_{2t}, \ldots, y_{nt})'$ is the vector of time series variables

α and β_i are (n X n) coefficients matrices,

u_t is (n X 1) unobservable zero mean independent innovations which are serially uncorrelated with time invariant covariance matrix Σ.

It is assumed that $u_t \sim N(0, \Sigma)$ and $\beta_i = (\alpha, \beta_1, \beta_2, \ldots, \beta_i)'$.

Equation (1) can further be written as:

$$Y_t = \beta' X_t + u_t \tag{2}$$

where

$$X_t = (1, Y'_{t-1}, \ldots, Y'_{t-k})'$$

In matrix form, equation (2) becomes:

$$\mathrm{Y} = \mathrm{X}\,\beta + u \qquad (3)$$

where Y is the T X R matrices with t-th rows given by Y'_t

u is T X R matrices with t-th rows given by u'_t

X is T X K matrix with t-th row given by X'_t

u has a multivariate normal density function.

3. Materials and Methods

3.1 Bayesian method for with Normal distribution innovations

If u in equation (3) follows a matric-variate normal distribution i.e. $u \sim$ MN $(0, \Sigma \otimes I_t)$, the Bayesian method of estimation for VAR model can be derived in the following manner;

The likelihood function can be simply written as:

$$\mathrm{L}\,(\beta, \Sigma) \propto |\,\Sigma|^{-\frac{1\ t}{2}}\ \exp\{-\tfrac{1}{2}\,\mathrm{tr}\,[\Sigma^{-1}(\mathrm{Y} - \mathrm{X}\beta)'(\mathrm{Y} - \mathrm{X}\beta)]\} \qquad (4)$$

Discussion on the use of several priors in VAR model was given by Kadiyala and Karlsson [9]. However, we assume the use of the non-informative prior in this study and is given as:

$$\mathrm{P}\,(\beta, \Sigma) \propto |\,\Sigma|^{-\frac{M+1}{2}} \qquad (5)$$

Combining equations (4) and (5) yields the joint posterior distribution:

$$\mathrm{P}(\beta, \Sigma|\mathrm{Y}) \propto |\,\Sigma|^{-\frac{t+M+1}{2}}\ \exp\{-\tfrac{1}{2}\,\mathrm{tr}\,[\Sigma^{-1}\,(\mathrm{Y} - \mathrm{X}\beta)'(\mathrm{Y} - \mathrm{X}\beta)]\} \qquad (6)$$

Equation (6) has similar features to derivations of Normal linear regression model (see Koop [10]), the term in exponential can be written in terms of Ordinary Least Squares (OLS) quantities.

These OLS quantities are:

$$\hat{\beta} = (X'X)^{-1}\,X'Y$$

$$\mathrm{S} = Y'Y - Y'X\,(X'X)^{-1}\,X'Y$$

$$(\mathrm{Y} - \mathrm{X}\beta)'(\mathrm{Y} - \mathrm{X}\beta) = \mathrm{S} + (\beta - \hat{\beta})'\,X'X\,(\beta - \hat{\beta})$$

Hence, the posterior is given as:

$$\mathrm{P}\,(\beta, \Sigma|\,\mathrm{Y}) \propto |\,\Sigma|^{-\frac{t+M+1}{2}}\ \exp\{-\tfrac{1}{2}\,\mathrm{tr}\,[\Sigma^{-1}\,[\mathrm{S} + (\beta - \hat{\beta})'\,X'X\,(\beta - \hat{\beta})]]\} \qquad (7)$$

$$= |\,\Sigma|^{-\frac{t+M+1}{2}}\ \exp\{-\tfrac{1}{2}\,\mathrm{tr}\,[\Sigma^{-1}\,\mathrm{S}\,\}\ \exp\{-\tfrac{1}{2}\,\mathrm{tr}\,[\Sigma^{-1}\,(\beta - \hat{\beta})'\,X'X\,(\beta - \hat{\beta})]\} \qquad (8)$$

Ignoring the terms that do not depend on β, we have:

$$\beta|\mathrm{Y}, \Sigma \sim \mathrm{MN}\ (vec[\hat{\beta}]\ , \Sigma \otimes [X'X\ ]^{-1}\ ) \quad (9)$$

Equation (9) is a kernel of matric-variate normal distribution.

N.B: MN is a matric-variate normal distribution.

3.2 Bayesian method for Fat-tailed VAR model

The original VAR model assumes normality as seen under section 3.1. However, the use of student-t distribution can reduce the influence of outliers (William and Huang [11]). The use of student-t distribution will be assumed for innovation u.

The likelihood function of the model in (4) is given as:

$$\mathrm{L}\ (Y_t|\ \beta_t, \Sigma_t, v_t) = \frac{1}{\mathrm{C}_t}|\Sigma_t|^{-1/2}[v_t + (\ Y_t - X_t\,\beta_t)'\Sigma_t^{\ -1}(\ Y_t - X_t\,\beta_t)]^{-v_t+\mathrm{M}/2} \quad (10)$$

where

$$\mathrm{C}_t = \frac{\pi^{M/2}\Gamma\,(\frac{v_t}{2})}{\Gamma\left(\frac{v_t+\mathrm{M}}{2}\right)v_t^{\,v_t/2}}$$

Using a non-informative prior as introduced by Jeffrey (1946, 1961). As noted by Jeffrey that non-informative prior tend to be proper in most models and two rules have to be taken into consideration when applying such non-informative prior density function.

(a) If a parameter have any value in a finite range -∞ to ∞, the prior density function should be taken as uniformly distributed prior.

(b) If parameter by nature can take any value from the interval 0 to ∞; hence the prior probability of the logarithm should be taken as uniformly distributed prior.

It is assumed that β_t, Σ_t, and v_t are independently distributed, therefore the prior distribution can be simply be written as:

$$\mathrm{P}\ (\beta_t, \Sigma_t, v_t) = \mathrm{P}(\beta_t)\ \mathrm{P}\ (\Sigma_t)\ \mathrm{P}\ (v_t) \quad (11)$$

Using Jeffrey's invariant theory proposed by Zellner (1971), we can then write the prior as:

$$\mathrm{P}\ (\beta_t) = \text{constant} = 1 \qquad -\infty \text{ to } +\infty, \quad (12)$$

$$\mathrm{P}\ (\Sigma_t) \propto |\ \Sigma_t^{\ o}\ |^{-\frac{M+1}{2}}$$

$$\mathrm{P}\ (\beta_t, \Sigma_t, v_t) = |\ \Sigma_t^{\ o}\ |^{-\frac{M+1}{2}} \quad (13)$$

Equation (13) is proportional to square root of the determinant of fisher's information matrix.

Customarily, the posterior density is proportional to the likelihood function times prior and can be written as:

$$\mathrm{P}(\beta_t, \Sigma_t|\ Y_t) \propto \frac{v_t^{\,v_t/2}\ \Gamma\left(\frac{v_t+\mathrm{M}}{2}\right)}{\pi^{M/2}\Gamma\,(\frac{v_t}{2})}|\Sigma_t|^{-M+2/2}[v_t^{\ *} + (\ Y_t - X_t\ \beta_t^{\ *})'\Sigma_t^{\ -1}(\ Y_t - X_t\,\beta_t^{\ *})]^{-v_t^{\ *}+\mathrm{M}/2} \quad (14)$$

There is no way to simplify the expression in (14), we have to integrate out Σ_t and have the kernel of generalized t-distribution. Thus, the marginal posterior of β_t is given as:

$$P(\beta_t|Y_t) \propto t(\beta_t{}^*, D_t{}^* (x_t' x_t)^{-1}, v_t{}^*) \quad (15)$$

where

$$\beta_t{}^* = D_t{}^* (D_t{}^o B_t{}^o + x_t' x_t B_t{}^o)$$

$$v_t{}^* = v_t{}^o + N\text{-}M$$

$$D_t{}^* = ((D_t{}^o)^{-1} + x_t' x_t)^{-1}$$

Note: o and * denote the parameters of prior and posterior distributions.

4. Numerical results

In order to investigate the performance of the derived Bayesian VAR procedures, numerical results of both simulated data and real data application will be considered in this section.

4.1 Simulation study

Here, the properties of the derived Bayesian VAR procedures will be examined using a simulated data. The simulated data is generated for a VAR (2) process with student-t innovations with v_t degrees of freedom. The two-dimensional VAR (2) process is given as:

$$y_{1t} = 3 + 0.8\, y_{1t-1} + 1.5\, y_{2t-1} + 10\, y_{1t-2} + 50\, y_{2t-2} + u_{1t}$$

$$y_{2t} = 8 + 1.7\, y_{1t-1} + 3.5\, y_{2t-1} + 15\, y_{1t-2} + 25\, y_{2t-2} + u_{2t}$$

It is assumed that there is no correlation between the series. Hence, the correlation matrix structure by the series is given as:

$$H = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

The degree of freedom parameters are set as: v_t= 4 and 7 to enable the generation of time series data while the sample period is: 20 and 100. In the application, 10,000 posterior samples will be generated using the developed approach.

4.2 Real Data Application

The data on United States economy will be considered here. The data are annually for a period of 1987 to 2006. The data are extracted from the report of the president, 2007. The two variables that will be considered are Money supply (in billions of dollars) and interest rates (%). Thus the VAR model can be simple stated as:

4.3 Forecast evaluation

The forecast evaluation criteria for the VAR model are Root Mean Squares Forecast Error (RMSFE) and Mean Absolute Forecast Error (MAFE).

Root Mean Squares Forecast Error (RMSFE)

The RMSFE is simply defined as:

$$\text{RMSFE} = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t - \mathrm{y}_t^{\mathrm{f}})^2}$$

where $\mathrm{y}_t^{\mathrm{f}}$ is predicted in-sample of y and $y_t$is the time series data.

Mean Absolute Forecast Error (MAFE)

This MAFE can be computed as:

$$\text{MAFE} = \frac{1}{n}\sum_{t=1}^{n}|\mathrm{y}_t^{\mathrm{f}} - y_t|$$

5. Presentation and Discussion of Results

The results of the simulated and real data applications in this section. Both parameter and forecasts evaluation estimates are presented in Tables 1-5. Tables 1 and 2 report the means and standard error for the simulated data while Table 3 gives the forecasts evaluation estimates for Bayesian VAR with student-t and normal innovations across sample sizes 20 and 100. The values in parenthesis are the actual values of the parameters.

In Tables 4 and 5, the results of real data application for both the parameter estimates and forecast evaluation estimates of the proposed method and Bayesian VAR with normal innovations are also presented.

Table 1: Model parameter estimates for both the normal and student-t distribution for sample size 20.

	Normal		Student-t ($\boldsymbol{v_t}$ = 7)		Student-t ($\boldsymbol{v_t}$ = 4)	
variables	Mean	Standard error	Mean	standard	Mean	Standard error
	Equation 1: y_{1t}					
y_{1t-1} (0.8)	0.5903	0.0802	0.7124	0.2365	0.7391	0.2365
y_{2t-1} (1.5)	1.3492	0.1530	1.4282	0.0081	1.4458	0.0075
y_{1t-2} (10)	8.1238	0.0494	8.3182	0.1249	8.4129	0.1249
y_{2t-2} (50)	54.1234	0.0949	43.8129	0.0806	46.9123	0.0039
C (3.0)	19.4948	5.1055	5.8239	0.2645	5.2384	0.2814
Equation 2: y_{2t}						
y_{1t-1} (1.7)	2.5938	0.0495	1.4927	7.1034	1.5129	7.7080
y_{2t-1} (3.5)	3.4821	0.0495	2.3839	0.2434	2.4951	0.2440

	Normal		Student-t (v_t = 7)		Student-t (v_t = 4)	
variables	Mean	Standard error	Mean	standard	Mean	Standard error
y_{1t-2} (15)	3.9216	0.0944	14.2941	3.7516	14.2722	4.0723
y_{2t-2} (25)	24.1833	0.0299	24.1293	0.1283	24.7291	0.1269
C (8.0)	21.6735	3.1356	14.2819	8.8708	13.2911	9.1724

Table 2: Model parameter estimates for both the normal and student-t distribution for sample size 100.

	Normal		Student-t (v_t = 7)		Student-t (v_t = 4)	
variables	Mean	Standard error	Mean	standard	Mean	Standard error
	Equation 1: y_{1t}					
y_{1t-1} (0.8)	0.4382	0.1131	0.5298	0.1126	0.6219	0.1125
y_{2t-1} (1.5)	1.4191	0.1545	1.4328	0.1542	1.4282	0.1541
y_{1t-2} (10)	12.2718	0.0649	11.9184	0.0546	10.3827	0.0529
y_{2t-2} (50)	42.8491	0.0889	37.3815	0.0885	38.1847	0.0888
C (3.0)	1.0918	4.1019	1.2972	4.0187	1.3847	4.0170
Equation 2: y_{2t}						
y_{1t-1} (1.7)	2.1729	0.0829	2.1837	0.0810	2.0182	0.0818
y_{2t-1} (3.5)	3.4192	0.1132	3.4193	0.1122	3.4391	0.1118
y_{1t-2} (15)	16.3391	0.0475	17.2817	0.0446	17.1282	0.0438
y_{2t-2} (25)	20.2983	0.0652	23.0192	0.0622	23.1819	0.0426
C (8.0)	5.2938	3.0062	6.2938	3.0046	6.4918	3.0032

From Tables 1 and 2, it is obvious that the estimated results of the proposed method for parameter estimates are reasonable. The parameter estimates for the proposed method are closer to their real values than Bayesian VAR with normality innovations for the sample sizes considered. Results based on the standard error also indicate that the proposed method performed better than the Bayesian VAR with normality innovations having smaller standard error. There is similarity in results obtained for Bayesian with student-t innovations when the degree of freedom is 4 and 7. However, the proposed Bayesian VAR with student-t innovations when the degree of freedom is 3 gives a better result.

Table 3: Forecasts evaluation estimates for sample sizes of 20 and 100

ASSUMPTION DISTRIBUTION	SAMPLE	Equation 1: y_{1t}		Equation 2: y_{2t}	
		RMFSE	MAFE	RMFSE	MAFE
NORMAL	20	6.0193	5.9138	5.9817	4.4830
	100	8.1404	6.6861	0.1881	0.1520
Student-t ($\boldsymbol{v_t}$ = 7)	20	8.0429	6.1935	6.0008	4.9029
	100	8.0429	6.1935	6.0008	4.9029
Student-t ($\boldsymbol{v_t}$ = 4)	20	0.3940	0.3920	0.1882	0.1522
	100	7.8191	6.1839	5.9008	4.9179

From the results obtained in Table 3 using forecast evaluation criteria, the RMSFE and MAFE for the two methods for both sample sizes are high. It is clear to see that Bayesian VAR with student-t innovations model forecasts is better than Bayesian VAR with normality innovations model forecasts in terms of both RMFSE and MAFE having smaller estimates. The Bayesian VAR with student-t innovations model forecasts when the degree of freedom is 3 gave a better performance than degree of freedom of 1.

Table 4: Model parameter estimates for both the normal and student-t distribution for United States economy

	Normal		Student-t ($\boldsymbol{v_t}$ =4)	
Variables	Mean	Standard error	Mean	Standard error
	Equation 1: $M2_t$			
$M2_{t-1}$	-0.5336	0.3008	0.9049	0.0433
INT_{t-1}	37.3642	17.2198	0.2554	0.0844
$M2_{t-2}$	-0.5336	0.3008	0.1751	0.045
INT_{t-2}	-24.7069	20.7558	-1.2233	0.0477
c	-162.6923	241.8280	5.8239	0.2645
	Equation2: INT_t			
$M2_{t-1}$	-0.0017	0.0036	-0.0004	0.0005
INT_{t-1}	1.0707	0.2033	0.2554	0.0844
$M2_{t-2}$	-0.0017	0.0036	-0.0002	0.0005
INT_{t-2}	-0.8439	0.2451	-0.0232	0.0477
c	6.7291	2.8554	5.7623	1.2560

Table 5: Forecasts evaluation estimates for United States (US) economy

ASSUMPTION DISTRIBUTION	Equation 1: $M2_t$		Equation 2: INT_t	
	RMFSE	MAFE	RMFSE	MAFE
Normal	1.4910	1.2657	388.1253	337.3406
Student-t ($\boldsymbol{v_t}$ =4)	1.4842	1.2640	364.1908	316.7288

Based on the twenty annually observation, the proposed Bayesian VAR model with student-t innovations was applied. As shown in Table 4, the results of standard error for the proposed method are all smaller than the Bayesian VAR model with normal innovations while all the variables yield better prediction results. In Table 5, the RMFSE and MAFE for money supply equation are very low but RMFSE and MAFE vary significantly for the interest rate. It is also apparent that BVAR model with student-t innovations forecasts are better than the BVAR model with normal innovations forecast for both the money supply (M2) and interest rate (INT).

6. Conclusion

In Bayesian modelling framework, Vector Autoregressive (VAR) model with student-t innovations has been developed. In order to investigate the performance of this developed Bayesian VAR model, numerical results of both simulated data and real data application were considered. Both parameter and forecast evaluation estimates were used as criteria to examine the performance of the developed method.

It was found that a Bayesian Vector Autoregressive (BVAR) model with student-t innovations give better parameter estimates and forecasts when compared with that of Bayesian Vector Autoregressive (BVAR) model with usual normality innovations. It was also observed that the developed Bayesian approach with small degree of freedom was capable to diminish the effects of outliers.

This work, therefore recommend the use of Bayesian Vector Autoregressive (BVAR) model with student-t innovations in both empirical analysis and simulations.

7. References

[1] Sims, C.A. (1980). Macroeconomics and Reality, Econometrica, 48, 1-48.

[2] Lutkepohl, H. (1991). Introduction to Multiple Time Series Analysis. Springer-Verlag, Berlin.

[3] Watson, M. (1994). Vector Autoregressions and Cointegration, in Handbook of Econometrics, Volume IV. R.F. Engle and D. McFadden (eds.). Elsevier Science Ltd., Amsterdam.

[4] Litterman, R. B. (1979). Techniques of forecasting using vector autoregression. Minneapolis, MN: Federal Reserve Bank of Minneapolis, Research Department, Working paper paper 115.

[5] Sims, C. A. and Zha, T. (1995). Error Bands for impulse responses. Federal Reserve Bank of Atlanta, working paper 95-6.

[6] Sun, D. and Ni, S. (2004). Bayesian analysis of vector-autoregressive models with noninformative priors. Journal of statistical planning and inference, 121,2, 291-309.

[7] Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems, proceedings of the Royal statistical society of London, series A, 186, 453-461.

[8] Kalli, M. And Griffin, J. E. (2015). Bayesian non-parametric vector autoregressive models.

[9] Kadiyala, K. and Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models, Journal of Applied Econometrics, 12, 99–132.

[10] Koop, G. (2003). Bayesian Econometrics, Wiley, Chichester.

[11] William, R. B. and Huang, E. T. (2006). Using the t-distribution to Deal with outliers in small area estimation. Proceedings of Statistics Canada Symposium 2006: methodological issues in measuring population health.

[12] Bauwens, L., Lubrano, M. and Richard, J. F. (1999). Bayesian Inference in Dynamic Econometric Models, Oxford University Press, Oxford. Berger, J. (1985). Statistical Decision Theory and Bayesian.