Forecasting foreign tourist visits to Bali using Bayesian Vector Autoregression with Normal-inverse-Wishart Prior

I Wayan Sumarjaya[1] and Ni Ketut Tari Tastrawati[2]

[1]Department of Mathematics
Faculty of Mathematics and Natural Sciences, Udayana University
Bali, Indonesia
sumarjaya@unud.ac.id
[2]Department of Mathematics
Faculty of Mathematics and Natural Sciences, Udayana University
Bali, Indonesia
tastrawati@unud.ac.id

ABSTRACT
As a major tourist destination, Bali has become an icon for tourism in Indonesia. In general, the number of foreign tourist visits shows an increasing trend. However, there is considerable fluctuation in the number of visits which is affected by season. In another word, there is a stochastic trend in the number of tourist visits. Policy makers need a method to predict this tourist visits. A commonly used method to predict tourist visits is time series analysis. Time series analysis has been used in various fields such as finance, business, engineering, meteorology, geophysics, and tourism, to name but a few. Research on forecasting tourist visits usually uses univariate data. This research aims to forecast the number of foreign tourist visits from four major countries such as Asia Pacific, ASEAN, America, and Europe simultaneously using Bayesian vector autoregression with normal-inverse-Wishart prior. First data is plotted to see its characteristics. Then the data is modeled using Bayesian vector autoregression. In this stage normal-inverse-Wishart prior is used. Next, Markov chain Monte Carlo is conducted to make a prediction from the posterior distribution. The forecast suggests that the number of tourist visits in general increased, albeit some fluctuation in some months.

Keywords: autoregression, Bayesian vector autoregression, time series analysis, tourist visits forecasting

I. INTRODUCTION

As a major tourist destination, Bali has become an icon or role model for tourism in Indonesia. In general, the number of tourist visits shows an increasing trend. However, these visits vary from months to months or fluctuated due to seasonal effect. This means that there is a stochastic trend on the tourist visits data. Policymaker, for instance, the Bali tourism board, needs to predict these tourist visits so that a proper marketing strategy can be made. One way to forecast this visits is by using time series analysis.

Research on forecasting tourist visits to Bali using time series analysis has been carried out by [1], [2], and [3]. However, these authors mainly use univariate time series model. Univariate time series has few shortcomings. First, it cannot measure interdependence structure between the time series. Second, long-term relationship cannot be measured by using single series alone. These shortcomings can be overcome by using multivariate time series model.

This research aims to forecast the number of tourist visits of four major countries using Bayesian vector autoregression. Bayesian vector autoregression (BVAR) is different from classical vector autoregression in that the former uses prior information. This research uses foreign tourist visits data from four major countries: Asia Pacific, ASEAN, America (including Canada and Latin America), and Europe. The classification of these countries is based on grouping as mentioned in Bali Tourism Board.

This article is organized as follows. Section one contains motivation and background for forecasting foreign tourist visits to Bali. Section two discusses the brief concept of multivariate time series using Bayesian vector autoregression and prior selection. Forecasting and its analysis are discussed in section three. Section four concludes the article.

II. RESEARCH METHOD

This section discusses basic concept of vector autoregression, Bayesian vector autoregression, and normal-inverse Wishart prior. The following discussion is adapted from [4]. Let y_t for $t=1,\ldots,T$ be $M\times 1$ vector of observation on M time series variables, ε_t be $M\times 1$ vector of errors, β_0 be $M\times 1$ vector of intercepts, and B_j be $M\times M$. matrix of coefficients. Vector autoregression of order p is defined as

$$y_t=\beta_0+\sum_{j=1}^{p}B_j y_{t-j}+\varepsilon_t \tag{1}$$

where vector ε_t is assumed to be independent and identically distributed $N(0,\Sigma)$.

Equation (1) can be modified further. Define Y as $T\times M$ matrix which stack T observation of response variables, $x_t=(1,y'_{t-1},\ldots,y'_{t-p})$, and define

$$X=\begin{bmatrix}x_1\\ x_2\\ \vdots\\ x_T\end{bmatrix}. \tag{2}$$

Next, define $B=(\beta_0,B_1,\ldots,B_p)'$ and $\beta=\mathrm{vec}(B)$. Then (1) can be written in form of normal distribution variate matrix as

$$Y=XB+E \tag{3}$$

or

$$y_t=(I_M\otimes X)\beta+\varepsilon \tag{4}$$

where $\varepsilon\sim N(0,\Sigma\otimes I_T)$.

The likelihood function that corresponds to (4) is

$$\begin{aligned}&L(\beta,\Sigma\mid y,X)\\&=(2\pi)^{-TM/2}\left|\Sigma\otimes I_M\right|^{-1/2}\\&\times\exp\left\{-\frac{1}{2}(y-(I_M\otimes X)\beta)'(\Sigma^{-1}\otimes I_M)(y-(I_M\otimes X)\beta)\right\}\end{aligned} \tag{5}$$

and the log-likelihood has the form [5]

$$\begin{aligned}&\ln L\propto-\frac{1}{2}\ln\left|\Sigma\otimes I_M\right|\\&-\left\{-\frac{1}{2}(y-(I_M\otimes X)\beta)'(\Sigma^{-1}\otimes I_M)(y-(I_M\otimes X)\beta)\right\}.\end{aligned} \tag{6}$$

Define $\aleph:=I_M\otimes X$ and $\hat{\beta}=(\aleph'(\Sigma^{-1}\otimes I_M)\aleph)^{-1}\aleph'(\Sigma^{-1}\otimes I_M)y$ so that the log-likelihood (6) can be written as [5]

$$\begin{aligned}\ln L&=-\frac{T}{2}\ln|\Sigma|-\left\{\frac{1}{2}(\beta-\hat{\beta})\aleph'(\Sigma^{-1}\otimes I_M)\aleph(\beta-\hat{\beta})\right\}\\&-\left\{\frac{1}{2}\mathrm{tr}\left[(y-\aleph\hat{\beta})(y-\aleph\hat{\beta})'(\Sigma^{-1}\otimes I_{\mathrm{M}})\right]\right\}\\&\propto\ln(N(\beta\mid\hat{\beta},\Sigma,\aleph,y)IW(\Sigma\mid\hat{\beta},\aleph,y).\end{aligned} \tag{7}$$

Now the normal-inverse-Wishart prior has the form

$$p(\beta,\Sigma\mid\aleph,y)\propto p(y\mid\aleph,\beta,\Sigma)p(\beta,\Sigma). \tag{8}$$

Incorporating prior in (8) yields the following conditional posterior distribution:

$$\begin{aligned}&p(\beta\mid\Sigma,\aleph,y)=N(\widetilde{\beta},\widetilde{\Sigma}),\\&p(\Sigma\mid\beta,X,Y)=IW(\Xi_\Sigma+(Y-X\beta)'(Y-X\beta),T+\gamma).\end{aligned} \tag{9}$$

Markov chain Monte Carlo algorithm such as Gibbs sampling can be used to make inference based on (9). Detail derivation of (9) can be seen in [5] and [6].

III. RESULTS AND DISCUSSION

This research uses foreign tourist visits from four major countries: Asia Pacific, ASEAN, America (including Canada and Latin America), and Europe during the period from January 2010 to June 2016. The plot of the visits for the four countries can be seen in Figure 1.

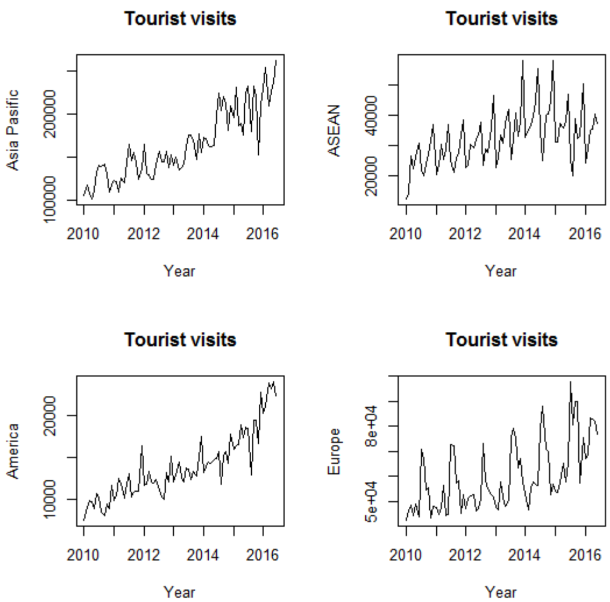

Fig. 1. Foreign tourist visits to Bali during January 2010—June 2016.

Figure 1 shows an increasing stochastic trend and pronounces seasonal effect on the four foreign tourist visits. To model and to forecast the four-time series simultaneously Bayesian vector autoregression with normal-inverse-Wishart prior is applied. The lag selection, in this case, 12, follows from an empirical research study carried out by [7].

The forecasts for foreign tourist visits can be seen in Table 1. The plot of the forecasts and the credible interval can be seen in Figure 2.

TABLE I

FORECASTS OF FOREIGN TOURIST VISITS FOR THE NEXT 12 MONTHS

Month	Asia Pacific	America	ASEAN	Europe
Jul. 2016	254,906	23,369	31,105	92,101
Aug. 2016	230,891	21,279	26,208	89,716
Sep. 2016	276,676	22,577	42,292	93,010
Oct. 2016	227,706	22,011	37,748	90,136
Nov. 2016	184,553	21,820	36,001	72,947
Des. 2016	250,688	27,660	57,285	90,310
Jan. 2017	252,709	23,998	30,620	87,931
Feb. 2017	252,574	23,509	33,713	77,574
Mar. 2017	228,883	27,894	47,204	87,810
Apr. 2017	264,296	27,415	42,981	100,601
May 2017	272,261	26,697	48,591	98,512
Jun. 2017	276,951	24,984	47,542	86,176

Figure 2 shows a plot of forecasts for the next 12 months. As can be seen from the figure there is an increasing (stochastic) trend for all the visits albeit some fluctuation in a certain month (see also Table 1).

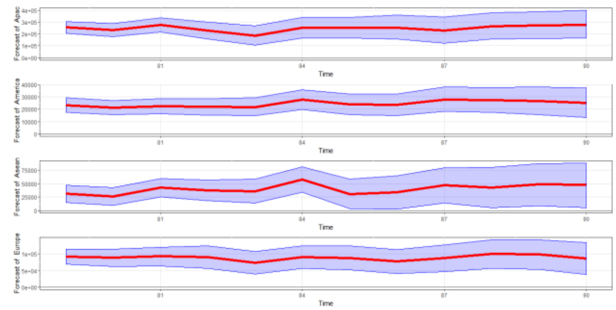

Fig. 1. Plot of forecasts for period June 2016 to July 2017.

IV. CONCLUSION

Forecasting foreign tourist visits to Bali using Bayesian vector autoregression shows an increasing stochastic trend albeit considerable fluctuation due to seasonal effect.

ACKNOWLEDGMENT

The authors would like to thank Udayana University for the financial support under the research scheme Hibah Unggulan Program Studi contract number 2731/UN14.1.28/LT/2016 dated 27 June 2016.

REFERENCES

[1] I. B. K. P. Arimbawa K., K. Jayanegara, and I. P. E. N. Kencana, “Komparasi metode Anfis dan fuzzy time series kasus peramalan jumlah kunjungan wisatawan Australia ke Bali,” *E-Jurnal Matematika*, vol. 2, no. 2, 2013, pp. 18—26.

[2] I. M. C. Satria, I. K. G. Sukarsa, and K. Jayanegara, “Peramalan jumlah wisatawan Australia yang berkunjung ke Bali menggunakan multivariate fuzzy time series,” *E-Jurnal Matematika*, vol. 4, no. 3, 2015, pp. 90—97.

[3] R. Rukini, P. S. Arini, and E. Nawangsih, “Peramalan jumlah kunjungan wisatawan mancanegara (wisman) ke Bali tahun 2019: metode ARIMA,” *Jurnal Ekonomi Kuantitatif Terapan*, vol. 8, no. 2, 2015, pp. 136—141.

[4] G. Koop and D. Korobilis, “Bayesian multivariate time series methods for empirical macroeconomics,” *Foundations and Trends in Econometrics*, vol. 3, no. 4, 2009, pp. 267—358.

[5] K. O’Hara (2016, November 1). Bayesian macroeconometrics in R [Online]. Available: http://bayes.squarespace.com/storage/bmr/BMR.pdf.

S. Karlsson, “Forecasting with Bayesian vector autoregression” in Handbook of Economic Forecasting, vol. 2B, E. Graham and T. Allan, Ed. New York: Elsevier, 20