Modelling tourists arrival using time varying parameter

1 Suciptawati P, 2 Sukarsa K G and 3 Eka N Kencana
Department of Mathematics, Udayana University, Kampus Bukit Jimbaran–Badung 80361, ID
E-mail: 1 putusuciptawati@yahoo.co.id, 2 sukarsakomang@yahoo.com, 3 i.putu.enk@unud.ac.id

Abstract. The importance of tourism and its related sectors to support economic development and poverty reduction in many countries increase researchers’ attentions to study and model tourists’ arrival. This work is aimed to demonstrate time varying parameter (TVP) technique to model the arrival of Korean’s tourists to Bali. The number of Korean tourists whom visiting Bali for period January 2010 to December 2015 were used to model the number of Korean’s tourists to Bali (KOR) as dependent variable. The predictors are the exchange rate of Won to IDR (WON), the inflation rate in Korea (INFKR), and the inflation rate in Indonesia (INFID). Observing tourists visit to Bali tend to fluctuate by their nationality, then the model was built by applying TVP and its parameters were approximated using Kalman Filter algorithm. The results showed all of predictor variables (WON, INFKR, INFID) significantly affect KOR. For in-sample and out-of-sample forecast with ARIMA’s forecasted values for the predictors, TVP model gave mean absolute percentage error (MAPE) as much as 11.24 percent and 12.86 percent, respectively.

1. Introduction
Tourism has been started attracts scholars and researchers to study this discipline and its related sectors in 1990s. At the beginning phase, tourism model regarding the visitors’ demand utilised multiple regression techniques (MRTs) to model and forecast the growth of visitors to a destination. Qiu and Zhang [1] used multiple regression analysis to model tourist arrivals and their expenditures in Canada. However, this technique is criticised noting tourist arrivals had non-stationary properties that is not suitable for MRTs. One consequence of applying MRTs for non-stationaries data is spurious regression function relates the dependent and its predictors. Since this works, alternative methods such as co-integration (CI) technique, error correction model (ECM), vector autoregressive (VAR) has been tried and tested to model tourist arrivals. These methods utilise ordinary least squares (OLS) to estimate the parameter(s) by assuming its are constant over time. This assumption is not completely fulfilled noting the tourist arrivals tend to fluctuate because of internal and/or external destination factors.

The insufficiency of constant parameters assumption for tourism demand modelling is demonstrated by Witt and Song [2] who proved price elasticity of tourism demand decreased when cross-price elasticity increased over time. To overcome non-constant behaviour of tourism demand, both authors introduced time varying parameter (TVP) to model number of visitors as targeted variables and several macroeconomic indicators as predictor variables. In addition, Guo [5] stated by applying Kalman Filter algorithm thorough its capability to find the estimators
adaptively, particularly by introducing an excitation condition, then the assumption regarding the stationarity or independence of predictors are not needed.

This work is aimed to model Korean tourists to Bali. According to Bali Provincial Statistics Office [3], the number of foreign tourists visiting Bali in year 2015 reached 4 million visitors, 6.65 percent growth compared to previous year. Classify according nationality, Australian tourists position in the first place while Korean visitors are in the fifth rank. To model Korean arrivals to Bali, TVP with Kalman Filter algorithm to 'purify' the state-spaces (time of visiting) is built. Refers to Welch and Bishop [4], a new state-space \( x_i \) is a linear combination of the previous state \( x_i \) and process noise \( u_i \).

2. Research Method

Secondary data from Bali Tourism Office were used to model Korean arrivals to Bali at time \( t \), and variables in the model are Monthly arrivals data on KOR (Secondary data from Bali Tourism Office were used to model Korean arrivals to Bali at time \( t \)). To model temporary disturbances \( H_t \) and inflation rate at Korea (INFKR\(_t\)) and Indonesia (INFID\(_t\)) for period January 2010–December 2015. Modelling steps are as follows:

(i) Defining state-space model that relates all of the variables under studied. Under normality assumption of errors in the model, refers to [6], TVP model in state-space form can be expressed by

\[
KOR_t = x_t\alpha_t + u_t; \quad u_t \sim NID(0, H_t).
\]

In equation (1) – is also known as measurement equation – \( x_t \) refers to a matrix of predictor variables, \( \alpha_t \) is a vector denotes an unobserved parameters or state vector, and \( u_t \) is a vector of temporary disturbances with mean 0 and variance matrix \( H_t \).

(ii) Defining state or transition equation that represents the transition from state \( t \) to state \( t + 1 \). This transition can be expressed by

\[
\alpha_{t+1} = \Phi_t\alpha_t + R_t\epsilon_t; \quad \epsilon_t \sim NID(0, Q_t).
\]

In equation (2) \( \epsilon_t \) represents temporary disturbances matrix with mean 0 and variance \( Q_t \) while \( \Phi_t \) and \( R_t \) are transition matrices initially assumed to be known. Regarding the measurement and transition equations, two assumptions have to be made, namely (a) the initial vector \( \alpha_0 \) has mean 0 and variance \( P_0 \); and (b) \( u_t \) and \( \epsilon_t \) are assumed serially independent. If the \( \Phi_t \) is an identity matrix, then \( \alpha_{jt} \) will follows random walk process \( \alpha_{jt} = \alpha_{jt-1} + \epsilon_t, j = 1, \ldots, \text{number of predictors} \).

(iii) Estimating the parameters. To overcome the limitation of OLR in estimating non-constant parameters, it is common Kalman Filter (KF) algorithm is applied. According to [2], this algorithm uses iteration procedure to find optimal solution from all information available at state-space at the given time \( t \). In brief, KF algorithm consists of three states:

(a) Initial state: at this state the initial values for \( a_0 \) and \( P_0 \) are determined. Whenever no information is available to init both values, then [2] suggested to use diffuse priors technique. This technique set \( P_0 \) as \( kI \) where \( k \) is an arbitrary large value range from \( 10^2 \) to \( 10^7 \) and I is an identity matrix. The initial values for \( a_0 \) is assign randomly;

(b) Prediction state: at this state the predicted values for \( y_{t+1} \), \( P_{t+1} \), and \( a_{t+1} \) are counted. The initial values of \( a_0 \) and \( P_0 \) as the values for state \( t \) is used to count

\[
a_{t+1|t} = \Phi_t a_t.
\]

\[
P_{t+1|t} = \Phi_t P_t \Phi_t' + R_t Q_t R_t'.
\]

Applying equation (3) and (4), we predict the value of \( \hat{y}_{t+1} \) by equation (5)

\[
\hat{y}_{t+1|t} = x_t a_{t+1|t}.
\]
The prediction error and mean squared error (MSE) for $y_{t+1}$ can be examined by using the equations $r_{t+1} = y_{t+1} - \hat{y}_{t+1|t}$ and $F_{t+1} = x_{t+1}P_{t+1|t}x'_{t+1} + H_{t+1}$, respectively;

(c) Update state: this state is used to refine the $a_t$’s until optimal vector is reached. The update state utilises Kalman gain $K_t$ that can be expressed as $K_{t+1} = P_{t+1|t}x'_{t+1}F^{-1}_{t+1}$ and $a_{t+1}$ as well as $P_{t+1}$ are updated using these equations:

$$a_{t+1} = a_{t+1|t} + P_{t+1|t}x'_{t+1}F^{-1}_{t+1}(y_{t+1} - \Phi_{t+1}a_{t+1|t}).$$  

(6)

and

$$P_{t+1} = P_{t+1|t} - P_{t+1|t}x'_{t+1}F^{-1}_{t+1}x_{t+1}P_{t+1|t}.$$  

(7)

(iv) Conducting goodness-of-fit test. To validate the model is sufficient, two tests are conducted. Kolmogorov-Smirnov’s test is used to check whether the measurement residuals follow normal distribution, and their correlation with residuals from transition matrix do not significant. The uncorrelated properties between measurement and transition residuals were checked by observing the correlation matrix.

3. Results and Discussion

3.1. Exploring the Data

Noting that TVP is suitable for time varying data – data that vary over time – it is necessary to explore before modelling is conducted. The plot for number of Korean tourists arrived at Bali is depicted in fig. 1. It is clear from fig. 1, the number of Korean tourists arrived at Bali vary over time. In addition, it looks like the arrivals do not follow a specific pattern. Based on this figure, we conclude that Korean arrivals has time-varying properties.
For the predictor variables, plot of data show similar result. The exchange rate of WON to IDR (EXCHANGE) and the inflation rate at both countries (INFKR and INFID) demonstrate time-varying properties as show in fig. 2 and 3.

![Figure 2. The EXCHANGE Rate of WON to IDR](image1)

**Figure 2.** The EXCHANGE Rate of WON to IDR

![Figure 3. The Inflation Rates in Korea and Indonesia](image2)

**Figure 3.** The Inflation Rates in Korea and Indonesia

By concluding the data matrix is appropriate to model using TVP, the state-space model were defined as follows:

\[ \text{KOR}_t = \alpha_{1t}\text{EXCHANGE}_t + \alpha_{2t}\text{INFKR}_t + \alpha_{3t}\text{INFID}_t + u_t. \]  \hspace{1cm} (8)

and

\[ \alpha_{it+1} = \alpha_{it} + e_{it+1}; \quad i = 1, \ldots, 3. \]  \hspace{1cm} (9)
3.2. Estimating the Parameters

To estimate the model parameters, we applied maximum likelihood estimation (MLE) to know the covariance of measurement and transition equations $H_t$ and $Q_t$. By using EViews software, the value for $H_t$ and $Q_t$ are:

\[
H_t = 2.05 \cdot 10^{-36}, \text{ and} \\
Q_t = \begin{pmatrix}
930.99 & 0 & 0 \\
0 & 917.68 & 0 \\
0 & 0 & 2.93 \cdot 10^{-22}
\end{pmatrix}
\]

The next step is to apply those values in estimating the variance of unobserved parameters $\alpha_t$ and its covariance state vector. KF algorithm is used with the initial value for $a_i$ and Kalman gain $k$ is set to 0 and $10^6$, respectively. The final state is listed on Table 1.

Table 1. Estimated the final state by using KF algorithm

| Log(Variance) | Std. Error | z-Statistic | Prob. |
|---------------|------------|-------------|-------|
| Constant      | 1.582      | 30.7 $10^{-5}$ | 5148.88 | 0.0000 |
| EXCHANGE      | -0.046     | 8.6 $10^{-5}$ | -535.44 | 0.0000 |
| INFKR         | 0.028      | 23.0 $10^{-5}$ | 121.25  | 0.0000 |
| INFID         | 1.233      | 6.4 $10^{-5}$ | 19151.66 | 0.0000 |

Refers to Table 1, as noted by Bossche [7], the output of EViews for each of predictors shows the logarithmic of its variance. Taking the exponential of these values, we found the inflation rate of Indonesia (INFID) has the biggest variance as much as $\exp(1.233) = 3.433$, and are followed by the inflation rate of Korea (INFKR) and the exchange rate between WON and IDR (EXCHANGE) as much as 1.028 and 0.955, respectively.

Our finding showed number of Korean tourists visiting Indonesia is more fluctuates on inflation of Indonesia and is less on the exchange rate between WON and IDR. Refers to Table 1, the measurement and state equation for Korean arrivals can be expressed as:

\[
\begin{align*}
KOR_t &= \begin{pmatrix}
EXCHANGE_t \\
INFKR_t \\
INFID_t
\end{pmatrix} \begin{pmatrix}
0.955 \\
1.028 \\
3.434
\end{pmatrix} \\
\end{align*}
\]

and final covariance matrix for $\alpha_t$ is given by equation 12.

\[
\begin{pmatrix}
1177.56 & -2911.28 & -8199.77 \\
-2911.28 & 1138650.00 & 30946.04 \\
-8199.77 & 30946.04 & 276626.30
\end{pmatrix}
\]

3.3. Validating the Model

To validate the resulted model, the residuals from measurement equation (10) is checked by using Kolmogorov-Smirnov normality test. The test showed p-value for test statistic is 0.056, slightly greater than type-I error ($\alpha$) 5 percent. We concluded, the residuals follow normal distribution. In addition, we also checked the correlation between the residuals and the predictors. Except for correlation between residuals and INFID as much as 0.266 (p-value = 0.024), the others were not significant with p-value as much as 0.864 and 0.862 for correlation
between residuals–EXCHANGE and residuals–INFKR, respectively. Observing the residuals follow normal distribution and two of correlation values were not significant, than the resulted model as expressed in equation (10) is qualified to represent the relationship between Korean arrivals and its predictors.

The other technique to validate our model is to count the Mean Absolute Percentage Error (MAPE) value. MAPE can be understood as how good the model represents the original data, and its value can be counted using this formula:

\[
\text{MAPE} = \frac{\sum_{t=1}^{m} \left( \frac{|e_t|}{y_t} \right)}{m} \tag{12}
\]

For in-sample forecasting, we got MAPE’s value is 11.24 percent. Although there is no consensus about the upper limit of MAPE to claim the model is good enough, we believe our model is sufficient to predict the number of Korean tourists come to Bali with the accuracy level as much as 88.76 percent. To validate, we make an out-of-sample forecast for six consecutive months in year 2016. In predicting Korean arrivals for January–June 2016, we forecast it on 2 different ways, i.e. (a) previously, we forecast all of the explanatory variables by using ARIMA models. Applying these forecasted values, then we make out-of-sample forecast of Korean arrivals for that period (Forecast\(^1\)); and (b) by utilising the real value for EXCHANGE, INFKOR, and INFID for that period (Forecast\(^2\)). The forecast’s result is listed on Table 2. The MAPE’s value for out-of-sample forecast using Forecast\(^1\) and Forecast\(^2\) as much as 12.86 percent and 12.28 percent, respectively.

**Table 2. Out-of-Sample Forecast for Korean Arrivals**

| Month  | EXCHANGE Rate | KOR Inflation | IND Inflation | Korean Arrivals |
|--------|---------------|---------------|---------------|-----------------|
|        | Forecast  | Real  | Forecast  | Real  | Forecast  | Real  | Forecast\(^1\) | Forecast\(^2\) | Real            |
| Jan. 2016 | 11.80    | 11.43  | 1.42    | 6.8   | 2.96    | 4.13  | 13623            | 12827           | 14257           |
| Feb. 2016 | 11.80    | 11.43  | 1.66    | 1.3   | 2.96    | 4.42  | 13775            | 13164           | 13247           |
| Mar. 2016 | 11.80    | 11.49  | 1.84    | 1.0   | 3.05    | 4.45  | 13906            | 12975           | 8327            |
| Apr. 2016 | 11.80    | 11.13  | 1.80    | 1.0   | 3.19    | 3.60  | 14880            | 12630           | 11900           |
| May 2016  | 11.80    | 11.10  | 1.61    | 0.8   | 3.20    | 3.33  | 13749            | 12461           | 14498           |
| June 2016 | 11.80    | 11.54  | 1.42    | 0.8   | 3.09    | 3.45  | 13625            | 12934           | 12419           |

Note:
Forecast\(^1\): The forecasted values for explanatory variables were used
Forecast\(^2\): The real values for explanatory variables were used

4. Conclusion
By demonstrating the use of time-varying parameter model to predict number of Korean tourists visit Bali, we found the inflation rate of Indonesia affects the Korean arrivals more fluctuate than the effects of inflation rate of Korea and the exchange rate between two countries. Furthermore, we found TVP model is qualified enough to model Korean demand to visit Bali with the in-sample MAPE’s is 11.24 percent. By using the ARIMA’s forecasted results for predictors’ values, we got the out-of-sample forecast as much as 12.86 percent, slightly greater when the real values for predictor were used to with MAPE’s as much as 12.28 percent.
Acknowledged
We would like to thank the Udayana University for funding this research through Grant No. 486.64/UN14.2/PNL.01.03.00/2016.

References
[1] Qiu H and Shang J 1995 Journal of Travel Research 34 43–49
[2] Witt S F and Song H 2000 Advances in Tourism Research: Tourism Demand Modelling and Forecasting. Modern Econometric Approaches (New York: Routledge)
[3] Badan Pusat Statistik Bali 2015 Statistik Wisatawan Mancanegara ke Bali Denpasar: BPS Provinsi Bali
[4] Welch G and Bishop G 2001 An Introduction to the Kalman Filter (University of North Carolina at Chapel Hill)
[5] Lei GUO 1990 IEEE Transaction on Automatic Control 35 (2) 141–147
[6] Gang Li, H Song and S F Witt 2005 Journal of Travel Research 44 82–89
[7] Filip A. M. Van den Bossche 2011 Journal of Statistical Software 41(8) 1–16