Forecasting China’s inbound tourist arrivals using a state space model

Zhiqi Xiong1, Jianxu Liu1,2,4, Songsak Sriboonchitta1,2 and Vicente Ramos3

1Faculty of Economics, Chiang Mai University, Chiang Mai 50200 Thailand
2Puey Ungphakorn Center of Excellence in Econometrics, Chiang Mai University, Chiang Mai, 50200, Thailand
3Department of Applied Economics, University of the Balearic Islands, Palma, 07122, Spain
4Email: liujuanxu1984@163.com

Abstract. The inbound tourism is one of the most important economic activities in China. Forecasting inbound demand is immensely helpful for policymakers and operators. The aim of this study is to evaluate the forecasting results of inbound tourist arrivals in China from six main tourist source markets: South Korea, Japan, the USA, Malaysia, Singapore and Canada, obtained from a state space model. The accuracy of forecasting results will be compared with the fixed linear regression, and ARIMA models, based on MAPE and RMSE. Empirical results suggest that all the variables have time varying character in six cases. And the income level has the most significant effect on tourist arrivals, followed by the substituted price in competitive countries. The price in China has the least impact on inbound tourist arrivals to China. And the accuracy of forecasting indicates that the state space approach performs better than the linear regression and ARIMA models for longer time frames.

1. Introduction
China’s inbound tourism has great economic contribution to China and the world. In view of the fluctuation in China’s inbound tourism market over the past several years, it will benefit governments and scholars to predict China’s inbound tourism in the future. Seasonality appears in most source countries, which affects investment and industry policies obviously. And the determinants - the income level, the tourism price in China and the price in substituted destinations are responsible for the quantity of tourism demand in Handbook of Tourism Economics [1]. Therefore, defining trend, seasonal components and explanatory variables in the model properly are essential to obtain accurate forecasting results. The state space approach is able to analyze tourist arrivals by unobserved components and define explanatory variables as time varying.

The aim of this study is to evaluate the forecasting results of inbound tourist arrivals in China from the six main source markets — which are South Korea, Japan, the USA, Malaysia, Singapore and Canada — by using the state space model. In order to accomplish this task, the objectives are specified as follows: a) To investigate the determinants of China’s inbound tourist arrivals from South Korea, Japan, the USA, Malaysia, Singapore and Canada. b) To assess the accuracy of forecasting tourist arrivals from six countries compared with the linear regression and ARIMA model. The quantitative analysis method is the state space model with time varying parameters. Variables include: explained variable is the number of inbound tourist arrivals (overnight), and explanatory variables are defined as
the income level of the country of tourists’ origin, the price in China and the substituted price. Quarterly data is used for this research. The great contributions of tourism to the Chinese economy make forecasting an important process; its results will contribute to the success of the economy in the future. The forecast of tourism is helpful for the government to plan, develop and implement policies and helpful for the business enterprises to plan their activities, and helpful for academics to develop and improve the science, theories and methods related to tourism in China.

The contribution of this paper can be listed as follows: first, it is the first time to apply the state space approach in China’s inbound tourism demand. Determinants with time varying parameters could help to get better forecast result than the OLS and ARIMA models about China’s inbound tourist arrivals. Second, it is proved that the income affects tourists’ decision to visit China most obviously, and the substituted price differs in different competing countries, and the price has the least effect on tourists to China.

2. Literature review
The studies on quantitative analysis of international tourist arrivals have become more diverse. But there are few quantitative studies in China’s inbound tourism market. Most studies of China’s inbound tourist arrivals applied models such as seasonal ARIMA [2] novel neural network technique, support vector regression [3], casual method combined with ADLM models [4], Grey-Markov model [5]. Also, the quantitative analysis methods for empirical forecast of international tourism were summarized by Courch [6][7][8][9]. Witt and Witt [10], Lim [11], Li et al. [12]. There is a comprehensive review in Song and Li [13] about tourist arrivals estimation and forecasting since 2000, with new research directions and methods emerging.

The state space approach modelling a dynamic system with unobserved variables has been employed in many other destinations. Song et al. [14] proposed the new method- TVP model on Hong Kong market. Song and Witt [15] verified that the TVP has a better performance than other models for short term forecasting on Denmark’s inbound tourism market. Li, Song and Witt [16] developed TVP and LAIDS (linear almost ideal demand system) models on Western European destinations, which appear better than the fixed-parameter counterparts. Li et al. [17] also forecast the demand by UK residents for Western European destinations by combining ECMs (error correction models) and TVP models, with the results outperforming many alternative models. Song et al. [18] concentrated on Hong Kong’s inbound tourism market by combining TVP and STSM to show that the TVP-STSM is the best time series model. Recently, Gunter and Onder [19] employed a number of models including TVP model for Paris’s inbound tourism market to prove TVP is a suitable model. In this paper, the state space approach with time varying parameter is to be conducted for forecasting China’s inbound tourist arrivals from six main originating countries.

3. Methodology

3.1. The Linear Gaussian state space model
The Linear Gaussian state space model (LGSSM) has three parts:

observation equation:
\[ y_t = Z_t \alpha_t + \epsilon_t, \epsilon_t \sim NID(0, H_t) \]

state equation:
\[ \alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \eta_t \sim NID(0, Q_t) \]

Initial state distribution:
\[ \alpha_1 \sim N(\alpha_0, P_0) \]

for \( t = 1, ..., n \), where the variables are explained as follows:

- \( y_t \) is a vector of \( n \) observations, \( \alpha_t \) is a state vector containing unobserved variables, \( Z_t \) parameter,
- \( \epsilon_t \) is a \((n \times 1)\) irregular vector containing the \( n \) observation disturbances, \( H_t \) is a variance matrix corresponding to \( n \) observation disturbances, \( T_t \) is a matrix called transition matrix, \( \eta_t \) is a vector containing the state disturbances with zero mean and unknown variance, \( Q_t \) is a \((r \times r)\) variance
matrix corresponding to $r$ state disturbances, $R_t$ is assumed as identity matrix $I_r$ in standard cases and selection matrix. In this study, we assume $r = 4$ and $T_t = I_4$ and $Q_t$ diagonal. And hyperparameters are set as $\epsilon_t$ and $\eta_t$. The Kalman Filter is used as the recursive filter algorithm for the state space. It enables the information to be updated in formulas, which could divide an observation into conditional mean and predictive residual constantly.

3.2. Model specification and variables
The following demand function is applied to model and forecast each tourists originating country in this research:

$$\ln D_t = f(\ln Y_t + \ln P_t + \ln SP_t + \text{level}),$$

where “ln” means the natural logarithm and the data is also taken by X12 seasonal adjustment; $D_t$ is tourist arrivals from source countries; $Y_t$ is, the gross domestic product (GDP) at constant prices, meaning the income level of source country $i$; and $P_t$ is China’s tourism price, exchange rate adjusted by CPI,

$$P_t = \frac{\text{CPI}_{\text{China}/\text{EX}_{\text{China}}}}{(\text{CPI}_i/\text{EX}_i)},$$

where, $\text{CPI}_{\text{China}}$ and $\text{CPI}_i$ are China’s and source countries’ Consumer Price indices (base year=2010); $\text{EX}_{\text{China}}$ and $\text{EX}_i$ are exchange rates (per dollar) of China’s and original countries’ currency; in fact, transportation, consumption and living cost are supposed to form the tourism cost, but the transportation cost to China is difficult to collect, and thus the exchange rate is chosen as tourism price in this paper; $SP_i$ is the substituted price of such selected substituted countries as: South Korea, Japan, Malaysia, Thailand, and the variable is calculated as:

$$SP_i = \sum_{j=1}^{5}(\text{CPI}_j/\text{EX}_j) \omega_j,$$

$$\omega_j = \frac{\text{TTS}A_j}{\sum_{j=1}^{5}\text{TTS}A_j},$$

where, $\text{CPI}_j$ and $\text{EX}_j$ are the Consumer Price index and exchange rate of substituted countries, and $\omega_j$ is the market share of substituted countries, $\text{TTS}A_j$ is the total number of arrivals to country $j$ from country $i$. In addition, the level component is used to help analyze tourist arrivals in time series.

In this paper, quarterly data has been collected during the period of time between 1999 Q1 and 2015 Q2, and between 1999 Q1 and 2011 Q2 — this data is used for estimating vectors. Data collected between 2011 Q3 and 2015 Q2 is used for forecasting and evaluation. The data on tourist arrivals has been collected from Chinese and other countries' tourism bureaus in monthly statistics aggregated to quarterly data. The GDP, CPI and Exchange Rate data are available from the International Monetary Fund.

4. Results and discussion

4.1. Model estimation
All the models in this study are based on the Linear Gaussian state space model, of which state equations are specified as random walk. Tourist arrival series are taken seasonal adjustment by using X12 [20] firstly, as well as component level and three variables are held in the final time varying model. R program ‘dlm’ package is applied for the state space approach to estimating and forecasting.

With regard to the hyperparameters in table 1, namely variance of the disturbance in observation and state equations — first, the level values of six countries are different from each other. Japan has the biggest and South Korea has the smallest, and therefore it can be proved that the arrivals from six source countries show different trends. Second, the number of all the variables are not equal to zero — it means that the time varying nature of three variables is successfully explained, which also can be seen in figure 1 since all the hyperparameters of the variables are different from each other.

Table 2 exhibits the statistical significance of variables in final state: only the effect of price on Korean arrivals is significant at the 5% significance level. From figure 1(a), the price elasticity dropped abruptly to -0.85 in 2008 when economic crisis occurred and then recovered in 2009. The price is inelastic as the absolute value keeps being less than one, but Korean people are the most
inbound tourists to China. Short distance and matching price level are two main reasons. It seems that Korean tourists prefer to travel on a tighter budget.

### Table 1. The estimators of hyperparameters (variance of disturbance).

|        | South Korea | Japan    | USA      | Malaysia | Singapore | Canada |
|--------|-------------|----------|----------|----------|-----------|--------|
| Y      | 0.00116     | 0.00119  | 0.0004   | 7.27216E-06 | 0.0004   | 0.0001 |
| P      | 0.00010     | 0.03607  | 0.1101   | 0.3188   | 0.11018   | 0.2673 |
| SP     | 0.20962     | 0.03536  | 0.0002   | 2.79263E-06 | 0.00025  | 0.0005 |
| Level  | 0.00019     | 0.30340  | 0.0139   | 0.0103   | 0.0139    | 0.00690|
| Irregular | 0.18545   | 0.01084  | 0.3115   | 0.4578   | 0.3115    | 0.01536|

### Table 2. The estimators of final states in 2011 Q2.

|        | South Korea | Japan    | USA      | Malaysia | Singapore | Canada |
|--------|-------------|----------|----------|----------|-----------|--------|
| Y      | -0.0172     | 1.0451***| 1.3202***| 1.0230***| 1.163     | 4.518***|
| S.E.   | 0.1886      | 0.0686   | 0.1531   | 0.0733   | 0.912     | 0.749  |
| P      | -0.4937**   | -0.1276  | -0.0607  | -0.2307  | -1.22     | -0.567 |
| S.E.   | 0.2328      | 0.3633   | 0.7264   | 0.9376   | 1.432     | 0.434  |
| SP     | 0.2023      | 0.2119*  | 0.2267** | -0.0323  | -2.270*** | -2.554***|
| S.E.   | 0.1372      | 0.1184   | 0.0956   | 0.1665   | 0.293     | 0.111  |

Notes: ***,*** denote significance at the 10%, 5% and 1% levels

### Figure 1. Time varying parameters at 2000Q3-2011Q2.

The estimates of income are at the 1% significance level in four of six countries: Japan, the USA, Malaysia and Canada, which are all sensitive to the income level. As shown in figure 1(b), the income elasticity becomes richer than obvious after 2008, as the economy of Japan has advanced unsteadily; and the estimates of income in Malaysia displayed in figure 1(e) are stable from 2004, which suggests
that Malaysia touring market matured after 2004. For these two countries, travelling to China are considered as a light luxury and affordable commodity by both Japanese and Malaysian people well. The time varying character is evident in both the USA and Canada as well. The USA away from China shows a different trend of income elasticity as the increasing speed of Chinese economy varies from time to time. Canada with its long distance from China exhibits quite high elasticity in income. By comparison, non-Asian countries are greatly affected by the income level because of the long distance and high transport fees. Asian countries are also quite sensitive to income as the demand for light luxury increases. Besides, at the end of 2008, all cases show that the economic crisis makes income elasticity to be larger than usual. The substituted price on Japanese tourists is significant at the 10% level and on American tourists at 5% level, and shows 1% on Singapore and Canada. The differences of the substituted price among four cases are based on the destinations that tourists decide to pursue.

4.2. Forecast accuracy evaluation
To evaluate accuracy of forecasting by the state space model estimated before, the two benchmark models are applied: the fixed parameter linear model (OLS), and the time series ARIMA model without any explanatory variables. The accuracy of forecasting is according to the mean absolute percentage error (MAPE) and root mean square error (RMSE). The forecasting results are compared in one quarter ahead, four quarters ahead, eight quarters ahead, and sixteen quarters ahead horizons within the period between 2011Q3 and 2015Q2.

Firstly, the MAPE and RMSE values of the state space of South Korea, the USA, Malaysia, Singapore and Canada are smallest. But the first- and fourth- quarter forecasting in Japan rank the second as shown in table 3. It indicates that the state space - time varying parameter model has a more accurate forecasting result than the other two models. But forecasting China’s inbound tourist arrivals in short term is not as good as long term by the state space approach.

Second, the superior performance of the state space compared to OLS suggests that defining coefficients of explanatory variables as stochastic is better than defining them as fixed in analyzing China's inbound arrivals. And the performance of the state space which is also superior to ARIMA indicates that influencing factors can explain tourist arrivals more than auto regression and moving average of time series. The wider error of long term forecasting of ARIMA also indicates that influencing factors have contribution to forecasting tourist arrivals.

| Table 3. Comparison of forecasting accuracy ranking over different horizons in Japan. |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
|                | MAPE            | ARIMA           | MAPE            | ARIMA           | MAPE            | ARIMA           | MAPE            | ARIMA           |
| Japan          |                 |                 |                 |                 |                 |                 |                 |                 |
| 1q             | 0.0046          | 0.0009(1)       | 0.00467         | 0.0636          | 0.0131(1)       | 0.06432         |                 |                 |
| 4q             | 0.0068          | 0.0008(1)       | 0.00447         | 0.0964          | 0.0152(1)       | 0.06182         |                 |                 |
| 8q             | 0.00973(1)      | 0.0113          | 0.01130         | 0.1562(1)       | 0.2286          | 0.19259         |                 |                 |
| 16q            | 0.01525(1)      | 0.0259          | 0.02234         | 0.2288(1)       | 0.4208          | 0.34945         |                 |                 |

Note: The figures (1) in parentheses denote rankings

5. Conclusions
In this study, the state space model with time varying parameters is proposed to estimate and forecast China’s inbound tourists from six main source countries, South Korea, Japan, the USA, Malaysia, Singapore and Canada, in the period 1999Q1-2015Q2. This approach defines the effect of determinants as stochastic, after arrivals get the seasonal adjustment. The accuracy of forecasting is compared with the fixed linear regression model and ARIMA time series model in quarter 1, quarter 4, quarter 8 and quarter 16 based on MAPE and RMSE methods. The empirical results demonstrate that:

Determinants of tourist arrivals have time varying effect on tourist arrivals, indeed. The income is elastic in four cases, long distance is the main reason for the USA and Canada countries, luxury tourism commodity is considered by people in Asian countries. The cross price differs in six cases assuming that the substituted destinations are the competing or complementary countries for China. In
addition, the price in China has less effect on both Asian and American countries except South Korea. Korean people are generally not too sensitive to prices in China, and China is always a popular travel destination for Korean people. Forecasting China’s inbound tourist arrivals by the state space with time varying parameter model offers superior performance compared to benchmark models, the fixed linear regression and ARIMA models, both in MAPE and RMSE. It’s obvious that long term forecasting is better than short term one. Forecasting with determinants has more superior performance than time series without any factors explaining tourist inbound arrivals. Besides, forecasting with time varying parameters is better than fixed parameter when introducing determinants into modelling China’s inbound tourist arrivals. The study is a reference for forecasting China’s inbound tourist arrivals by defining the influence of determinants as stochastic, which is helpful for Chinese governments, and operators to make policies and plans.

Acknowledgements
We are grateful to the editor and two anonymous reviewers for insightful comments that improved the paper significantly.

References
[1] Seetaram N 2014 Handbook of Tourism Economics: Analysis, New Applications and Case Studies Tourism Management 44 1
[2] Kulendran N and Shan J 2002 Forecasting China’s monthly inbound travel demand Journal of Travel & Tourism Marketing 13(1/2) 5-19
[3] Chen K Y and Wang C H 2007 Support vector regression with genetic algorithms in forecasting tourism demand Tourism Management 28(1) 215-226
[4] Ayeh and S Lin 2011 ‘Estimating tomorrow’s tourist arrivals’: forecasting the demand for China’s tourism using the general-to-specific approach Tourism and Hospitality Research 11(3) 197-206
[5] Sun X, Sun W S, Wang J Z, Zhang Y X and Gao Y N 2016 Using a Grey-Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China Tourism Management 52 369-379
[6] Crouch G I 1994 Demand elasticities for short-haul versus long-haul tourism Journal of Travel Research 33 2-7
[7] Crouch G I 1994 Price elasticities in international tourism Hospitality Research Journal 17 27-39
[8] Crouch G I 1994 The study of international tourism demand: A review of findings Journal of Travel Research 33 12-23
[9] Crouch G I 1994 The study of international tourism demand: A survey of practice Journal of Travel Research 33 41-54
[10] Witt S F and Witt C A 1995 Forecasting tourism demand: A review of empirical research International Journal of Forecasting 11 447-475
[11] Lim C 1999 A meta-analytic review of international tourism demand. Journal of Travel Research 37 273-284
[12] Li G, Song H and Witt S F 2005 Recent developments in econometric modeling and forecasting Journal of Travel Research 44 82-99
[13] Song H and Li G 2008 Tourism demand modelling and forecasting-A review of recent research Tourism Management 29 203-220
[14] Song H Y, et al. 2003 Tourism forecasting: accuracy of alternative econometric models International Journal of Forecasting 19(1) 123-141
[15] Song H Y, Li G, Witt S F and Fei B G 2010 Tourism demand modelling and forecasting: how should demand be measured? Tourism Economics 16(1) 63-81
[16] Li G, Song H and Witt S F 2006 Time varying parameter and fixed parameter linear AIDS: An
application to tourism demand forecasting *International Journal of Forecasting* **22** 57-71

[17] Li G, Wong K F, Song H and Witt S F 2006 Tourism demand forecasting: A time varying parameter error correction model *Journal of Travel Research* **45** 175-185

[18] Song H Y, Li G, Witt S F and Athanasopoulos G 2011 Forecasting tourist arrivals using time-varying parameter structural time series models *International Journal of Forecasting* **27**(3) 855-869

[19] Gunter U and Onder I 2015 Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data *Tourism Management* **46** 123-135

[20] Findley D F, Monsell B C, Bell W R, Otto M C and Chen B C 1998 New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program *Journal of Business and Economic Statistics* **16** 127-176