Forecasting Chinese outbound tourist to Singapore, Malaysia and Thailand destinations: a study of state space approach

Danhua Jiang1, Jianxu Liu2,3, Jirakom Sirisrisakulchai1,3,4 and Songsak Sriboonchitta1,3

1Faculty of Economics, Chiang Mai University, Chiang Mai 50200, Thailand
2Faculty of Economics, Shandong University of Finance and Economics, Jinan, China
3Puey Ungphakorn Center of Excellence in Economics, Chiang Mai University, Chiang Mai 50200, Thailand
4Email: sirisrisakulchai@hotmail.com

Abstract. This paper aims to forecast the tourism demand of mainland Chinese residents with regard to visiting Singapore, Malaysia and Thailand (SMT) from 2018 to 2022, and examine the factors that influence outbound tourism demand. The empirical results show that income level, tourism price, and substitute price all have an impact on tourism demand, of which income level and substitution price have a positive impact on tourism arrivals; in contrast, tourism price has a negative effect. Afterward, based on the forecasted performance, the TVP-STSM model in state-space form is better than other benchmark models (Linear Regression model, ARIMA model, BSM model, and CSM model). Finally, this study finds the best model to forecast the number of tourists to SMT in the next 5 years. The results show that the largest increase among the three countries goes to Thailand, followed by Singapore and Malaysia; recommendations are also made on the forecasting of trends.

1. Introduction
Chinese tourists have seen a rapid growth in South-East Asian outbound markets over the last few years, while Singapore, Malaysia and Thailand (SMT) have become the largest outbound destinations with absolute advantage. It has long been recognized that accurate and reliable international tourism demand forecasting is crucial for the planning and arrangement of the destination tourism service sector. So, the significance of this research is to analyze the tourism demands of Chinese outbound tourists to SMT and put forward relevant recommendations to optimize SMT tourism industries.

Firstly, we select different models to estimate, such as Naïve Forecasting Model (I/II), Linear Regression model, Autoregressive Integrated Moving Average (ARIMA), Basic Structural Model (BSM) and Causal Structural Model (CSM) with time varying parameter regression model. The empirical result displays that the Structural Time Series Model with time varying parameter (TVP-STSM) outperforms all six competitors under AIC criterion and error measures (RMSE, MAE and MAPE). Therefore, we use this model to measure the influencing factors and forecast the number of tourists.

This paper investigates the factors affecting tourism demand in SMT using time series data between 1999q1 and 2017q4. Using the impact of income levels, tourism prices, and alternative prices on tourism demand using elasticity. Then, this results show that the number of SMT tourists will maintain a general upward trend, the increase of Thailand are the largest and the fastest, they grow.
exponentially. The novel contribution of this paper is to find a more accurate tourism demand model to forecast the demand for SMT.

2. Literature

This article reviews a considerable amount of research on tourism demand, it also found a lot of analytical methods in econometrics to analyze the influencing factors.

Regarding tourism demand, Gunter and Onder [1] and Song et al. [2] used destination price, competing destination price, the relevant exchange rate, and tourist income as the variables to forecast tourism demand. Li et al. [3] and Lin et al. [4] indicated that the tourism demand and the elasticity can be obtained through the variables of income, tourism prices, and substitute price. It also includes a series of uncertain random factors such as policies and vacations form the article by Liu et al. [5].

About mathematical method, Song et al. [6] proposed that TVP model has superiority over its competitors as regards short-term forecasting. Therefore, this study incorporates TVP-STSM into the scope of research methods. Song et al. [7] proved TVP-STSM outperformed all seven competitors, including one-to-four-quarter-ahead ex-post forecasts and one-quarter-ahead ex-ante forecasts of the primary and causal STSM and TVP models, and BSM and CSM reflect the seasonal parameters in tourism demand better than the traditional models. Moreover, Gallant et al. [8] considered the Bayesian estimation of state-space models when the measurement density is not available and estimating equations for the parameters of the measurement density are available from moment conditions, but computing an accurate approximation to the measurement density is difficult. Hence, from the literature it is known that TVP-STSM is the advanced method employed in empirical studies of tourism demand and the state-space method uses a dynamic system with unobserved variables and is estimated by using a Kalman filter to provide a unified method for analyzing time series data. Since the TVP-STSM method in the form of state space is superior to other econometrics and it is expected to improve forecast accuracy when processing season data, hence it will be used for empirical research in this article.

To sum up that, BSM, CSM and TVP-STSM can better reflect the seasonal changes in tourism demand. Therefore, this paper selects them as well as the general predictive analysis method that linear regression and ARIMA with were compared.

3. Methodology

Improvements on the basis of the TVP-STSM, then we obtained the specification of BSM, CSM.

Equation (1) is TVP-STSM model, where \( y_t \) is a univariate time series and unobservable components. These include trend component \( \mu_t \), cycle component \( \psi_t \), seasonal component \( \gamma_t \), causal variables \( \chi_t \), corresponding vector of coefficients \( \Gamma_t \), and the last term irregular component \( \epsilon_t \).

BSM has no causal variables are incorporated, the TVP-STSM is reduced as: \( y_t = \mu_t + \psi_t + \gamma_t + \epsilon_t \). And CSM are estimated as time-invariant: \( y_t = \mu_t + \psi_t + \chi_t + \gamma_t + \epsilon_t \).

Besides, the most commonly used Time Varying Parameter (TVP) regression forms as follows:

\[
\begin{align*}
\beta_{0,t+1} &= \beta_{0,t} + \xi_t, \xi_t \sim N(0, \sigma_{\xi}^2). \\
\beta_{1,t+1} &= \beta_{1,t} + \zeta_t, \zeta_t \sim N(0, \sigma_{\zeta}^2).
\end{align*}
\]

Let \( \alpha_t = (\beta_{0,t}, \beta_{1,t})' \), \( x_t = (1, x_t)' \), \( H_t = \text{diag}(\sigma_{\psi}, \sigma_{\gamma}) \) and \( G_t = \sigma_{\epsilon} \). Thereby obtaining the state-space form of TVP:

\[
\begin{pmatrix}
\alpha_{t+1} \\
\gamma_t
\end{pmatrix} = \begin{pmatrix} I_2 \\ x_t \end{pmatrix} \alpha_t + \begin{pmatrix} H_t \eta_t \\ G_{\epsilon t} \end{pmatrix}.
\]
The parameters $\Phi_t = \left( \begin{array}{c} I \hskip -0.5cm \chi \end{array} \right)$, $\Omega = \begin{pmatrix} \sigma_1^2 & 0 & 0 \\ 0 & \sigma_2^2 & 0 \\ 0 & 0 & \sigma_3^2 \end{pmatrix}$ and the initial matrix is $\Sigma \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{pmatrix}$.

Then, it is possible to list the following causal relationship between tourism arrivals and subjective factors. Therefore, the specific demand function is proposed for modeling for the modeling of Chinese outbound tourism as follows:

$D_{1,t} = \beta_{1,t} Y_t + \varepsilon_{1,t}$  \hspace{1cm} (6)

$D_{2,t} = \beta_{2,t} P_t + \varepsilon_{2,t}$  \hspace{1cm} (7)

$D_{3,t} = \beta_{3,t} P_{si,t} + \varepsilon_{3,t}$  \hspace{1cm} (8)

$\beta_{1,t+1} = \beta_{1,t} + \nu_{1,t}$  \hspace{1cm} (9)

Therefore, the following specific function is used,

$D_{t} = \beta_{0,t} + \beta_{1,t} Y_t + \beta_{2,t} P_t + \beta_{3,t} P_{si,t} + \beta_{4,t} \text{Dummies} + \varepsilon_{1,t} + \varepsilon_{2,t} + \varepsilon_{3,t} + \varepsilon_{4,t}$  \hspace{1cm} (10)

where $D_{t}$ is the tourism demand, for which the variable of tourist arrivals to destination country $j$ is used as: $D_{t} = D_{1,t} + D_{2,t} + D_{3,t}$; $\beta_{0,t}$ is initial value of the number of tourists; $Y_t$ is the income level of the origin country $i$, measured by GDP at constant prices; $P_t$ is the tourism price in origin relative to that of the destination country $j$, adjusted by the relevant exchange rates as follows, $\varepsilon_{1,t}, \varepsilon_{2,t}, \varepsilon_{3,t}$, all are error terms; and $\beta_{1,t}$ is the time series parameter.

$P_t = \frac{\text{CPI}_i}{\text{CPI}_j}$  \hspace{1cm} (11)

$P_{si,t} = \frac{\sum_{i=1}^{3} \text{CPI}_i}{\sum_{i=1}^{3} \text{CPI}_j} \omega_j$  \hspace{1cm} (12)

where the substitute price variable $P_{si,t}$ is defined as the weighted average index of the selected countries’ tourism prices; $\text{CPI}_i, \text{CPI}_j$ are the exchange rate of the origin country (China) and substitute country (SMT); CPI$_i$, CPI$_j$ is the consumer price index of the origin country and the substitute country (base year is 2010); $\omega_j$ is the market share of the substituted destination $j$, which is calculated from

$\omega_j = \frac{\text{TTA}_j}{\sum_{j=1}^{3} \text{TTA}_j}$  \hspace{1cm} (13)

where TTA$_j$ is the total number of tourists arrival country $j$. And the variable often is proved to have a significant impact on tourism demand. For the unknown parameters in above matrices, this paper estimates them using maximum likelihood estimation.

4. Empirical results

4.1. Model estimation

All the data are secondary data. Quarterly tourist arrivals, GDP ($2010 = 100$), CPI and exchange rate ($2010 = 100$) are collected from the DataStream database. Almost all data exhibit strong negative skewness after X-12 processing, except Thailand tourists arrivals and tourism price. Also, most of the Kurtosis values are less than 3, this means light-tailed and Jarque-Bera test results show that the distribution of most sequences rejects the normality assumption.

In accordance with the practice that the commonly used models for evaluating the fitting effect, the goodness of fit, are AIC, BIC, and HQC, the values obtained using these models were compared with the values obtained from other tests. The statistics finally determine the optimal model as the TVP-STSM.

Additionally, Figure 1 also can show that TVP-STSM has absolute accuracy at the five models. In addition, all the trend items are significant at the 1% level in the BSM and the CSM models. This is why this article puts the “trend” in the TVP-STSM model. Although not significant in the final model, it still has some impact on tourism demand.
Obviously, consumption of tourists will increase naturally due to higher incomes. This is reflected in both the linear Regression model and the CSM model with income factors.

### Table 1. Estimation and diagnostics for SMT.

| Estimation Model | Singapore | Malaysia | Thailand |
|------------------|-----------|----------|----------|
| **Linear Regression** |           |          |          |
| Income           | 0.868*** (0.108) | 1.340*** (0.262) | 1.377*** (0.270) |
| $P_i$            | 0.268 (1.485) | -1.189 (1.526) | -0.184 (0.752) |
| $P_s$            | -0.262 (0.176) | -1.247** (0.512) | -2.178*** (0.575) |
| C                | 4.962** (1.967) | 4.837* (2.053) | 7.635*** (1.849) |
| $R^2$            | 0.711 | 0.697 | 0.432 |
| $\overline{R^2}$ | 0.694 | 0.679 | 0.455 |
| S.E.             | 0.290 | 0.367 | 0.356 |
| AIC              | 0.429 | 0.903 | 0.842 |
| BIC              | 0.575 | 1.049 | 0.988 |
| HQC              | 0.485 | 0.959 | 0.899 |
| **ARIMA**        |           |          |          |
| C                | 0.028*** (0.004) | 0.0329*** (0.0120) | 0.020 (0.023) |
| AR(1)            | 0.157 (0.260) | 0.436** (0.190) | 0.087 (0.245) |
| MA(1)            | -1.0 (1255.567) | -1.0 (1704.238) | -0.678*** (0.180) |
| SIGMASQ          | 0.075 (2.073) | 0.120 (4.747) | 0.127*** (0.015) |
| $R^2$            | 0.427 | 0.290 | 0.260 |
| $\overline{R^2}$ | 0.393 | 0.247 | 0.215 |
| S.E.             | 0.284 | 0.360 | 0.371 |
| AIC              | 0.459 | 0.925 | 0.933 |
| BIC              | 0.606 | 1.073 | 1.080 |
| HQC              | 0.516 | 0.982 | 0.989 |
| **BSM**          |           |          |          |
| Trend            | 13.377*** (0.199) | -54.931 (605.329) | 13.813*** (0.267) |
| Seasonal         | 0.121 (0.276) | 67.796 (605.3270) | -0.291 (0.315) |
| Irregular        | 5.88E-13 (9.13E-05) | -1.66E-10 (0.001) | 2.81E-20 (4.52E-08) |
| $\beta_1$       | 0.007 (0.008) | -0.001 (0.004) | 0.005 (0.005) |
| AIC              | 5.357 | 3.222 | 5.847 |
| BIC              | 6.152 | 4.018 | 6.642 |
| HQC              | 5.665 | 3.531 | 6.155 |
| **CSM**          |           |          |          |
| Trend            | 15.244*** (0.137) | 12.567*** (0.145) | 18.074*** (0.199) |
| Seasonal         | -0.180** (0.085) | -0.063 (0.271) | -0.083 (0.302) |
| Irregular        | -0.020 (0.233) | -1.39E-06 (0.002) | -2.96E-06 (0.010) |
| Income           | -0.342 | 0.236 | 0.621 |
| $P_i$            | 1.373 | -1.362 | 1.505 |
| $P_s$            | -0.373 | -0.260 | -2.278 |
| $\beta_1$       | 15.206*** (0.126) | 12.517*** (0.133) | 17.976*** (0.173) |
| AIC              | 5.273 | 5.005 | 5.383 |
| BIC              | 6.185 | 5.918 | 6.296 |
| HQC              | 5.626 | 5.358 | 5.736 |
| **TVP-STSM**     |           |          |          |
| Income           | 0.373*** (0.085) | 0.857*** (0.300) | 1.451*** (0.186) |
| $P_i$            | -3.023*** (0.405) | -0.506*** (0.036) | -3.157*** (0.249) |
| $P_s$            | -0.794*** (0.132) | 1.780* (0.916) | 3.4920*** (0.423) |
| Trend            | -0.110 (0.203) | 0.336 (0.120) | -0.043 (0.107) |
| Dummy            | -0.174 (0.702) | -0.333 (0.169) | -0.134 (0.132) |
| C                | 13.387 (2.214) | 10.161 (2.023) | 17.258*** (1.982) |
| AIC              | 0.394 | 0.864 | 0.833 |
| BIC              | 0.689 | 1.158 | 1.127 |
| HQC              | 0.508 | 0.977 | 0.946 |

Source: By Author
Notes: *, ** and *** denote significance at the 10%, 5% and 1% levels, respectively (one-tailed tests for the income and own-price variables); the above models are estimated for the initial estimation sample 2013q1–2017q4; the values in brackets are standard errors.
For Singapore, the Income and Psi are within unit flexibility. When they increase by 1%, the percentage change in demand of Chinese tourists to Singapore will be less than 1%. Different from the full flexibility of Pi, visitors are more sensitive to tourism prices, this is also the case for Malaysia. However, for Thailand, the elasticity of its three factors is greater than one and all significant, indicating that Thai tourism is a luxury for Chinese tourists. Especially for the price (tourist price and substitute price), its elasticity far exceeds “one”. However, when the price of tourism products rises, it is not excluded that tourists with higher loyalty for Thailand. Because of luxury goods, the income and Psi of the source country (China) also have a great impact on tourism demand.

In a sense, Pi can indirectly indicate the income level of tourists. Besides, Table 1 demonstrates that the impact of tourism prices will be tremendous on tourists, and this kind of influence is even worse for tourists visiting Thailand and Malaysia because the impact is almost the same. In fact, tourists who want to go to Thailand or Malaysia may abandon Singapore because of different levels and services for the same price.

Moreover, as shown in Table 2, it is evident that both the Naïve I and Naïve II models have large errors, and the prediction lags so behind the actual model. Hence, Table 2 does not compare the AIC and the BIC values of the Naïve model with those of the other five models.

### 4.2. Model forecasting

The results of values are reported in Table 2, the RMSE, MAE, and MAPE of the TVP-STSM in state space are minimum with tourism income and tourism prices of Singapore. And it almost ranked first among the five models. This shows that the TVP-STSM has more accurate forecast results than the other models. In order to more intuitively express the degree of forecast in the Figure 1, this paper will select the top three optimal performance prediction methods for forecast.

**Table 2. Comparison of forecasting accuracy ranking over different horizons.**

| Index   | Singapore | Malaysia | Thailand |
|---------|-----------|----------|----------|
| Linear  | RMSE      | 0.263    | 0.963    | 0.710    |
|         | MAE       | 0.231    | 0.853    | 0.656    |
|         | MAPE      | 1.752    | 6.548    | 4.556    |
| Regression  | RMSE    | 1.820    | 0.350    | 1.170    |
|         | MAE       | 1.264    | 0.241    | 0.757    |
|         | MAPE      | 0.097    | 0.021    | 0.055    |
| BSM     | RMSE      | 0.073    | 0.146    | 0.519    |
|         | MAE       | 0.140    | 0.106    | 0.410    |
|         | MAPE      | 0.011    | 0.008    | 0.028    |
| CSM     | RMSE      | 0.189    | 0.183    | 0.420    |
|         | MAE       | 0.161    | 0.166    | 0.381    |
|         | MAPE      | 1.212    | 1.268    | 2.655    |
| ARIMA   | RMSE      | 0.211    | 0.007    | 0.427    |
|         | MAE       | 0.176    | 0.005    | 0.339    |
|         | MAPE      | 0.013    | 0.001    | 0.024    |

In summary, according to Figure 1, it can be seen that red lines (State-Space approach) are the best match to blue lines (Real tourists), the other two models are different optimal models for each country. so the state-space approach is the best method to solve this issue in both the long and short term. It is more consistent with the actual arrival numbers and trends, and it is the model closest to the actual state, although sometime it did not coincide with the actual trend at the beginning, slowly improved the accuracy in the later stage.
5. Conclusions

5.1. Forecast results
Before forecasting the number of tourists, it is first necessary to forecast the influencing factors (GDP, Pi and Psi) using polynomial. Moreover, this study used the TVP-STSM model to set the trend and explanatory variables to time variations. It has been concluded above that this model is the best predictive model among the five econometric models with time series. It finally gets the curve of prediction of the Figure 1.

5.2. Policy implications
For Singapore. Figure 2 reveals that the number of people arriving in Singapore is on the rise, and it can be seen that this visitor arrival shows an "inverted U" shape. The peak in 2020q4. After that, there
is a slight downward trend. This may be due to Chinese GDP growth has begun to slow down. Now, growth in Chinese tourism to Singapore may be less robust than in previous excitation due to the income factor. Therefore, Singapore’s primary task is to maintain a relatively pretty growth trend from the beginning and avoid subsequent recessions. It is known that Singapore is a normal consumer product according to the income elasticity, but the it cannot control the income level of China. That’s why it is necessary to pay attention to the impact of own country income. Besides, Singapore's substitute price is negative, which means that although the tourist prices in Malaysia and Thailand are rising, the number of tourists in Singapore will decrease. This indicates that Singapore's tourism prices are quite higher than others. So, lowering tourism service departments’ price is a good policy. For example, reducing the cost of scenic spots, adjusting the transportation and immigration services.

For Malaysia. The growth rate is not obvious. The reason for this trend is related to the environment of security in Malaysia. It also is a regular product in terms of income elasticity. But according to tourism prices and alternative prices, it is the opposite of Singapore. Malaysia's tourism price elasticity is less than one, which means that Chinese tourists are not very sensitive to its price level. This explains that when the tourism price of the others changes, it will lead to a major change in Chinese tourists. This reflects that Chinese tourists are disloyal to Malaysia from the side and Malaysia needs more popular attractions to keep and improve the loyalty of tourists. In a sense, simplifying visa procedures is a good way to attract tourists. And then, it also needs to promote the image to expand awareness and break the information market asymmetry.

For Thailand. The most surprising forecast is Thailand. The number of Chinese tourists will reach a new peak around 10,825,909 by the end of 2022 under the forecast results. Thailand differs from Singapore and Malaysia in that all three coefficients are greater than one. That’s means it is luxury for Chinese tourists, more flexible and sensitive to prices fluctuation. That’s mean that Thai government needs to hold the pretty image through marketing, media and trade agency to tourism. In addition, hardware devices also need to be improved, such as increasing the number of appropriate airlines and hotel rooms. It is worth noting that since the development of the “Belt and Road” Initiative, a lot of opportunities can be available for Thailand. After all, so many tourists into Thailand is not necessarily a perfect thing, perhaps it should be adjusted according to the local ecological environment and the level of reception.

References
[1] Gunter U and Onder I 2015 Forecasting international city tourism demand for Paris: Accuracy of uni- and multivariate models employing monthly data Tourism Management. vol. 46 pp. 123-135 Feb.
[2] Song H and Li G 2008 Tourism demand modeling and forecasting-A review of recent research Tourism Management vol. 29 pp. 203-220 Apr.
[3] Li G, Song H and Witt S F 2005 Recent developments in econometric modeling and forecasting Journal of Travel Research Volume 44 82-99
[4] Lin V S, Liu A and Song H 2015 Modeling and Forecasting Chinese Outbound Tourism: An Econometric Approach Journal of Travel & Tourism Marketing vol. 32 pp. 34–49 Apr.
[5] Liu J, Sripoonchitta S, Nguyen H T and Kreinovich V 2017 Chinese outbound tourism demand to Singapore, Malaysia and Thailand destinations: A study of political events and holiday impacts. SCI vol. 692
[6] Song H and Kevin K and Wong F 2003 Tourism Demand Modeling: A Time-Varying Parameter Approach. Available: http://journals.sagepub.com/doi/abs/10.1177/0047287503253908?journalCode=jtrb
[7] Song H, Li G, Witt S F and Athanasopoulos G 2011 Forecasting tourist arrivals using time-varying parameter structural time series models International Journal of Forecasting vol. 27 pp. 855-869 Sep.
[8] Gallanta A R, Giacomini R and Ragusac G 2017 Bayesian estimation of state space models using moment conditions Journal of Econometrics vol. 201 pp. 198-211 Dec.