Chin Yi Fang[1], Chen-Lun Hsu[2]

EXPLORING FORECASTING MODELS FOR TOURIST ARRIVALS IN INTERNATIONAL TOURIST HOTELS[*]

Accurate tourism forecasting is particularly crucial not only for governments and practitioners but also for investor resource allocation and decision-making. The main objective of this study was to obtain more accurate forecasts of tourist arrivals for a specific international tourist hotel in Taiwan by comparing the autoregressive integrated moving average (ARIMA) model and the unrestricted vector autoregressive (VAR) model, which have rarely been employed in the hotel industry. Monthly data covering the period 2000 M1 to 2010 M12 were collected from the Monthly Report on Tourism, which is published by the Tourism Bureau of Taiwan. The studied variables included the gross domestic product (GDP), the consumer price index (CPI), the exchange rate, and the hotel operating characteristics. Forecasting performance is assessed in terms of mean absolute percentage error (MAPE). The superior performance of the VAR model implies that the inclusion of endogenous variables is required in forecasting international tourist hotel demand. An impulse response analysis was performed to assess the impact level of tourist arrivals in response to shocks in the economic variables in the VAR model.

Keywords: *ARIMA; VAR; impulse response analysis; tourist arrival; forecasting; international tourist hotel.*

Чінь Ї Фанг, Чень-Лунь Сюй

ВИВЧЕННЯ МОДЕЛЕЙ ПРОГНОЗУВАННЯ ПРИБУТТЯ ТУРИСТІВ У МІЖНАРОДНІ ТУРИСТИЧНІ ГОТЕЛІ

У статті доведено, що точне прогнозування у сфері туризму має особливо важливе значення не лише для уряду і практиків, але й для розподілу інвестиційних ресурсів і ухвалення рішень. Тому розроблено точніші прогнози прибуття туристів на прикладі окремого міжнародного туристичного готелю на Тайвані методом порівняння інтегрованої моделі авторегресії – ковзаючого середнього (ARIMA) і необмеженої моделі векторної авторегресії (VAR), які рідко використовуються в готельній індустрії. Щомісячні дані, що охоплюють період з початку 2000 до кінця 2010 року, було отримано зі щомісячного звіту з туризму, який публікується Бюро туризму Тайваню. Змінні, що вивчаються, включають валовий внутрішній продукт (ВВП), індекс споживчих цін (ІСЦ), обмінний курс і характеристики діяльності готелю. Прогнозування продуктивності оцінено з точки зору середньої абсолютної помилки у відсотках (MAPE). Відмінні результати VAR-моделі показують, що для прогнозування світового попиту на туристичний готель потрібне включення ендогенних змінних. Аналіз імпульсного відгуку проведено з метою оцінювання рівня прибуттів у відповідь на потрясіння в економічних змінних у VAR-моделі.

Ключові слова: *ARIMA; VAR; аналіз імпульсної перехідної функції; прибуття туристів; прогнозування; міжнародний туристичний готель.*

Чинь Йи Фанг, Чень-Лунь Сюй

ИЗУЧЕНИЕ МОДЕЛЕЙ ПРОГНОЗИРОВАНИЯ ПРИБЫТИЯ ТУРИСТОВ В МЕЖДУНАРОДНЫЕ ТУРИСТИЧЕСКИЕ ОТЕЛИ

В статье обосновано, что точное прогнозирование в сфере туризма имеет особенно важное значение не только для правительства и практиков, но и для распределения инвестиционных ресурсов и принятия решений. Поэтому были разработаны более точные

[1] PhD, Assistant Professor, Graduate Institute of Hospitality Management, National Taiwan Normal University, Taiwan.
[2] Graduate Institute of Hospitality Management, National Taiwan Normal University, Taiwan.
[*] Financial support from Taiwan's National Science Council (NSC99-2410-H-003-002) is gratefully acknowledged.

прогнозы прибытий туристов на примере отдельного международного туристического отеля на Тайване методом сравнения интегрированной модели авторегрессии – скользящего среднего (ARIMA) и неограниченной модели векторной авторегрессии (VAR), которые редко используются в гостиничной индустрии. Ежемесячные данные, охватывающие период с начала 2000 до конца 2010 года, были получены из ежемесячного отчета по туризму, который публикуется Бюро туризма Тайваня. Изучаемые переменные включают валовой внутренний продукт (ВВП), индекс потребительских цен (ИПЦ), обменный курс и характеристики деятельности отеля. Прогнозирование производительности оценено с точки зрения средней абсолютной ошибки в процентах (MAPE). Отличные результаты VAR-модели предполагают, что для прогнозирования мирового спроса на туристический отель требуется включение эндогенных переменных. Анализ импульсного отклика проведено с целью оценки уровня прибытий в ответ на потрясения в экономических переменных в VAR-модели.

Ключевые слова: *ARIMA; VAR; анализ импульсной переходной функции; прибытия туристов; прогнозирование; международный туристический отель.*

Introduction. Products in the tourism industry are notably perishable, e.g., hotel rooms, restaurant seats, meals, and airline seats (Witt and Witt, 1995). Therefore, efficient and effective forecasts of tourist demand are important requirements for tourism planning, cost management, employee scheduling, and risk hedging. Asia has recently gained an increasing interest in tourism demand modeling and forecasting studies (UNWTO, 2010). Tourism forecasting is now considered crucial, due to the large amounts of assets and human capital involved (Song and Witt, 2006).

The tourism industry is becoming increasingly important in Taiwan. Also known as Formosa, Taiwan is an island located off the southeast coast of mainland China, being separated from mainland China by the Taiwan Strait. Taiwan covers an area of 36,000 square meters with the population of 23 mln. and the population density of approximately 640 persons/km^2. Travellers can experience the wonders of traditional Chinese culture, enjoy the authentic cuisine, outdoor recreation and great shopping. In 2008, the revenue generated by the tourism industry was 5,936 mln., and its growth rate was 13% (Annual Statistical Report on Tourism, 2008). According to the official data of the Tourism Bureau, international tourist arrivals have seen a dramatic increase in recent decades, from 2,624,037 in 2000 to 5,567,277 in 2010.

Since 2008, two benefits of deregulation have been impressive; not only have the flows of tourists to Taiwan increased, but revenue and employment in the tourism industry have also been growing steadily. In Taiwan, there are 41 tourist hotels scheduled for construction between now and 2018, among which 29 are international tourist hotels and 12 are standard tourist hotels (Taiwan Tourism Bureau, 2011). Therefore, the expected outcomes of appropriate models for tourism demand forecasting are useful and even crucial not only for academic researchers but also for tourism practitioners.

Previous research established forecasting models for inbound tourist volumes or tourist expenditures based on destination or national perspectives (e.g., Vanegas and Croes, 2000; De Mello and Nell, 2005; De Mello and Fortuna, 2005; Min, 2005; Song and Witt, 2006; Wong, Song, Witt and Wu, 2007; Song, Li, Witt, and Fei, 2010). Carnage (2003) was the first to establish a model for forecasting gross restaurant sales. However, little attention has been given to forecasting hotel tourist demand.

Therefore, the objectives of this work are fourfold: (1) to establish tourism demand forecasting models for the observed international tourist hotels, (2) to validate their forecasting accuracy based on historical data, (3) to forecast tourist arrivals for the observed hotels, and (4) to compare the planed room capacity with the forecasted tourist arrivals and suggest specific actions for the observed hotels. The forecasting methods selected for this study are the autoregressive integrated moving average (ARIMA) model and the unrestricted vector autoregressive (VAR) model. In this study, forecasts based on these models for the period January, 2000 to March, 2010 are compared with the actual volume of arrivals in the selected hotels from April, 2010 to December, 2010 obtained from the official tourist hotel publications.

The rest of this paper is organised as follows. The next section reviews recent publications in the area of tourism demand forecast modeling. The theoretical models are described in the methodology section, which briefly introduces the ARIMA and VAR models. The results section presents the results of the forecasting performance in comparison with the empirical results. The paper ends with concluding remarks and suggestions for future research.

Literature Review. Due to the perishable quality of products in the tourism industry, an accurate demand forecast is important for both private practitioners and government bodies as a source of useful information for strategic and operational planning and decision-making (Witt, Song and Wanhill, 2004; Wong, Song, Witt and Wu, 2007). Tourism forecasts may be broadly generated by either quantitative or qualitative approaches (Song and Li, 2008). The quantitative approaches can be divided into two subcategories: non-causal time-series models and causal econometric approaches. The main difference in these approaches lies in whether the demand model identifies any causal relationship between the demand and its influencing factors. This study focuses on quantitative methods, specifically, econometric methods.

Witt and Witt (1995) discussed the existing empirical literature on tourism demand forecasting. Additionally, the authors evaluated the accuracy of tourism forecasts generated by various models. Song and Li (2008) provided a comprehensive review of 121 tourism demand-forecasting studies published since 2000. These authors suggested that the latest developments in quantitative forecasting techniques are time-series models, econometric methods and other emerging approaches.

Many researchers have noted that no single forecasting method can be viewed as perfect across different situations (Song and Li, 2008; Witt and Witt, 1991; Witt and Witt, 1995). Carnage (2003) tested time-series models for forecasting gross restaurant sales. The empirical results offered line managers a scientific source of information for strategic and operational planning and decision-making. Min (2005) utilised a seasonal autoregressive integrated moving average (SARIMA) model to establish a demand model for forecasting inbound Taiwan tourist volume during and following the outbreak of severe acute respiratory syndrome (SARS).

De Mello and Nell (2005) used a VAR model to construct long-term relationships among tourism shares, tourism prices and tourism budgets. Song and Witt (2006) used the VAR model to forecast the Macau inbound tourist demand in the period 1993 Q1 to 2003 Q2. The results showed that China is expected to show the strongest increase in tourism demand. Thus, business practitioners should devote more effort to satisfying the needs of tourists from mainland China.

Other researchers have demonstrated that econometric models tend to generate more accurate forecasts than simple univariate time series models do (Witt and Witt, 1991; Witt, Song, and Louvieris, 2004). Recently, most of the forecasting researches have focused on international inbound arrivals (Vanegas and Croes, 2000; Song, Li, Witt, and Fei, 2010; Song and Witt, 2006; Wong et al., 2007). Furthermore, several studies have concentrated on tourist expenditures as forecasting variables (De Mello and Nell, 2005; De Mello and Fortuna, 2005; Song et al., 2010). To the best of our knowledge, no published research on tourism demand modeling has focused specifically on hotel demand forecasting.

Methodology.

Time-series Box-Jenkins models. Time-series models for predicting future patterns are composed of historic trends and random disturbances. Because time-series models require only historical observations of a variable, their advantage lies in less costly data collection and model building.

Univariate Box-Jenkins (1976) autoregressive integrated moving average (ARIMA) analysis has been widely used for modeling and forecasting. The autoregressive and moving average models can be combined to form the ARMA (p,q), indicating that

$$\mu_t = \rho_1\mu_{t-1} + \rho_2\mu_{t-2} + ... + \rho_p\mu_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q}, \quad (1)$$

where the parameters ρ and θ are the serial correlation coefficients, ε_t is the disturbance term, and t is the innovation of disturbance.

Econometric models. The VAR model is a system estimation approach that was first developed by Sim (1980). Compared with time-series models, econometric methods have the ability to analyse the causal relationships between the tourism demand and its influencing factors. Among many forecasting approaches, the VAR model can generate relatively accurate medium- and long-term forecasts for tourism demand (Song and Witt, 2006; Witt et al., 2003, 2004). The VAR model treats all variables as endogenous variables, and each variable is specified as a linear combination of the others. The general VAR model is illustrated as follows:

$$Y_t = \alpha_1 Y_{t-1} + ... + \alpha_p Y_{t-p} + \beta X_t + \varepsilon_t, \quad (2)$$

where p is the lag length of the VAR model, Y_t is the k vector of endogenous variables, X_t is the vector of d exogenous variables, $\alpha_1 ... \alpha_p$ and β are the matrices of coefficients to be predicted, and ε_t is the vector of innovations that may be contemporaneously correlated but are uncorrelated with their own lagged values and uncorrelated with all the right-hand side variables.

Analytical models. 3 endogenous variables are discussed and collected from the official Accounting and Statistics, Taiwan. The cost of living for tourists in Taiwan is measured by the consumer price index (CPI) for the basis year 2006. The income level is measured by the gross domestic product (GDP), the growth rate and the exchange rate. We use the following vector to represent the variables defined in the system:

$$Y_i = f(P_i, X_i R_i, \varepsilon_i), \quad (3)$$

where Y_i is the quantity of the tourism demand in Taiwan for a particular international tourist hotel i, P_i is the cost of living for tourists in Taiwan, X_i is the income level in Taiwan, R_i is the exchange rate with the US dollar.

The lag lengths of the VAR model are determined by the Akaike information criterion (AIC) and Schwarz information criterion (SIC). The most popular accuracy measurement of the forecasting models is the mean absolute percentage error (MAPE). Witt and Witt (1992) indicated that MAPE is the most appropriate error measure for evaluating the forecasting performance of tourism models. The MAPE is calculated as follows:

$$\text{MAPE} = (\Sigma_{i=1} | Y_i - \hat{Y}_i | / Y_i)/n \times 100\%, \quad (4)$$

where Y_i and $\hat{Y}_i$ are the true and predicted values, respectively. The MAPE can be interpreted as the degree to which the average forecast differs from the average actual arrivals figures.

Empirical Results. Tourism demand is measured by the number of arrivals at the observed international tourist hotels in Taiwan. The TTB has already established a star-rating system in Taiwan to unify global hotel rating standards. The initial stage of the TTB star-rating system identified 8 international tourist hotels as the top-ranked (five-star) hotels. This study selected 6 international tourist hotels as our observed samples because there were two recently launched hotels for which no historical data was recorded during 2000-2010. The research is based on continuous data over the 10-year period (2000 M01 to 2010 M12) supplied by the Taiwan Tourism Bureau. First, we plotted the data, as shown in Figure 1.

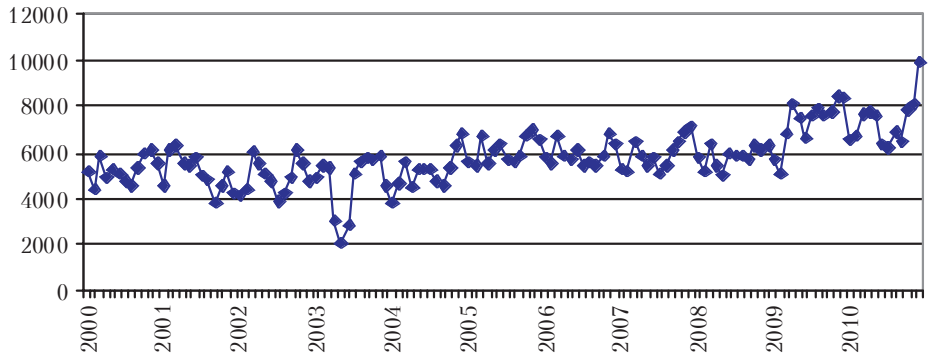

Figure 1. **Tourist arrivals from period 2000 M01 – 2010 M12 for a specific hotel**

The monthly data from January, 2000 to March, 2010 (123 points) are used as the training data set. The out-of-sample forecasting period from April, 2010 to December, 2010 (9 points) is used as the test data set for both models. First, because the time-series data is clearly nonstationary, a differencing is required. To confirm this property, the autocorrelation function (ACF) was positive and decayed fairly slowly.

To select the best suitable forecasting lag length, two criteria are discussed, listed in Table 1. Lower values for the Akaike information criterion (AIC) and the

Schwarz information criterion (SIC) indicate better fitness of the forecasting models. Therefore, the ARIMA(2,1,2) model is the best fitting for the time-series model in our example hotel.

Table 1. **ARIMA Model Fitness**

Model	AR(1)	AR(2)	AR(3)
MA(1)	AIC-0.9507	AIC-1.0374	AIC-0.8797
	SIC-0.8813	SIC-0.9677	SIC-0.8096
MA(2)	AIC-1.1534	AIC-1.1608*	AIC-1.2159
	SIC-1.0841	SIC-1.0911*	SIC-1.1458
MA(3)	AIC-0.8740	AIC-1.0901	AIC-0.8906
	SIC-0.8047	SIC-1.0204	SIC-0.8205

Note: AIC = Akaike info criterion; SIC = Schwarz info criterion.

The forecast values of the ARIMA and VAR models for the out-of-sample period from April, 2010 to December, 2010 were calculated and compared with the actual values. Better forecasting values are provided by the VAR model with a forecasting performance accuracy of 0.069529 according to the MAPE criterion. In addition, the results show that the guest-forecasting model fits the data well according to MAPE, as presented in Table 2.

Table 2. **Actual and predicted values of guests in ARIMA and VAR models**

Date	Actual	ARIMA		VAR	
		(predicted)	diff	(predicted)	diff
Apr-10	7778	7381	-397	7685	93
May-10	7639	7441	-198	7324	315
Jun-10	6400	7456	1056	6543	-143
Jul-10	6200	7479	1279	5501	699
Aug-10	6916	7501	585	5811	1105
Sep-10	6477	7523	1046	6358	119
Oct-10	7828	7545	-283	7130	698
Nov-10	8086	7567	-519	8305	-219
Dec-10	9921	7589	-2332	8500	1421
MAPE (%)		11.31154		0.069529	

Figure 2 illustrates annual forecasts of tourist arrivals generated by the ARIMA model. The trend in Figure 2 indicates that the number of tourist arrivals will increase to 8,000 per month for our sample hotel.

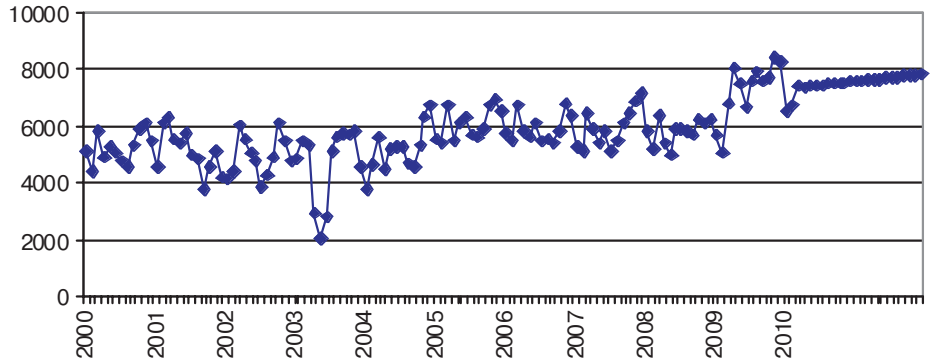

Figure 2. **Actual and Forecasted Number of Guests using ARIMA(2,1,2)**

According to the parameters generated by the VAR model, the predicted values of demand in the hotel were calculated as displayed in Figure 3. In estimating the VAR model, a number of exogenous variables are included to capture the influences of specific characteristics of the hotel. These exogenous variables include the average room price, the occupancy rate and the average monthly sales revenue of the hotel.

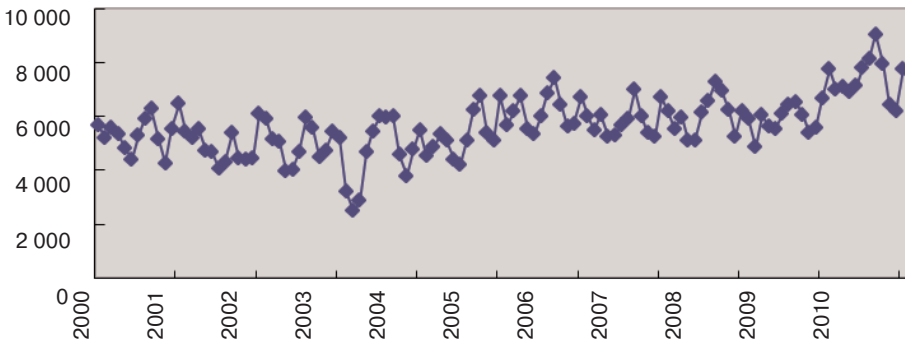

Figure 3. **Forecasted Number of Guests using VAR(2)**

The responses of tourist arrivals to the shocks in each of the explanatory variables are plotted via EVIEW 7.0. These demonstrate the impulse-response relationships between the dependent variables and the shocks in the other variables. The shocks in this study are measured by the Cholesky one-standard-deviation innovations (Hamilton, 1994). The number of tourist arrivals responds negatively to a shock in GDP in the first 3 months and then becomes positive; the momentum of this impact lasts for more than 30 months. This result indicates that GDP is an essential factor influencing tourist arrivals to hotels. The number of tourist arrivals responds positively to shocks in CPI index, and this impact gradually becomes minor after 25 months. The implication of the impulse response analysis is that Taiwanese government must realise that the influence of shocks in the economic variables in the VAR model tend to continue for more than 30 months.

Conclusions. This article has investigated the forecasting demand for a five-star international tourist hotel in Taiwan. Forecasting accuracy is notably important for resource allocation and decision-making for both academic and private-sector interests. Two popular and robust forecasting approaches were employed, and both methods efficiently provided consistent results.

The superior performance of the VAR model implies that the inclusion of endogenous variables is required in forecasting international tourist hotel demand. The empirical results indicate that the time-series ARIMA model and the econometric VAR can generate accurate demand forecasting. In comparison, the VAR model yielded more accurate performance, which is consistent with the results of Witt et al. (2004).

This study demonstrated that these forecasting methods can be successfully applied to predict the number of tourist arrivals for specific hotels. An impulse response analysis was performed to assess the impacts on the level of tourist arrivals in response to the shocks in economic variables within the VAR model. A better-per-

forming forecasting model provides benefits with respect to resource scheduling and capacity planning in international tourist hotel management.

The study is limited to qualitative forecasting. This study does not take the quality of service into account at the example hotel, which can also influence forecasting accuracy. It is hoped that future researchers will study the emotional factor, which has a significant influence on total tourist arrivals. Finally, another possibility for future research would be to consider the impact of one-off events, which can lead to long-term impacts on habitual consumer behaviour; therefore, these factors should be evaluated appropriately.

References:

Box, G. E. P., & Jenkins, G. M. (1976). Time series analysis, Forecasting and control (revised edition). San Francisco: Holden-Day.

Cranage, D. (2003). Practical time series forecasting for the hospitality manager. International Journal of Contemporary Hospitality Management, 15, p. 86-93.

De Mello, M. M., & Fortuna, N. (2005). Testing alternative dynamic systems for modeling tourism demand, Tourism Economics, 11, p. 517-537.

De Mello, M. M., & Nell, S. K. (2005). The forecasting ability of a cointegrated VAR system of the UK tourism demand for France, Spain and Portugal. Empirical Economics, 30, p.277-308.

Hamilton, J.D. (1994). Time series analysis. Princeton: Princeton University Press.

Min, C. H. (2005). The effect of the SARS illness on tourism in Taiwan: an empirical study, International Journal of Management, 22, p. 497-507.

Sim, C. (1980). Macroeconomics and reality, Econometrica, 48, p. 1-48.

Song, H., & Witt, S.F. (2006). Forecasting international tourist flows to Macau, Tourism Management, 27, p. 214-224.

Song, H., & Li, G. (2008). Tourism demand modeling and forecasting - a review of recent research, Tourism Management, 29, p. 203-220.

Song, H., Li, G., Witt, S.F., & Fei, B. (2010). Tourism demand modeling and forecasting: how should demand be measured? Tourism Economics, 16, p. 63-81.

Witt, S. F., & Witt, C. A. (1992). Modeling and forecasting demand in tourism. London: Academic Press.

Witt, S. F., & Witt, A. C. (1995). Forecasting tourism demand: A review of empirical research, International Journal of Forecasting, 11, p. 447-475.

Witt, S. F., Song, H., & Louvieris, P. (2003). Statistical testing in forecasting model selection, Journal of Travel Research, 42, p. 151-158.

Witt, S. F., Song, H., & Wanhill, S. (2004). Forecasting tourism-generated employment: the case of Denmark, Tourism Economics, 10, p. 167-176.

Wong, K. K. F., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine?, Tourism Management, 28, p. 168-178.

Vanegas, Sr. M., & Croes, R. R. (2000). Evaluation of demand for US tourists to Aruba, Annuals of Tourism Research, 27, p. 946-963.

Стаття надійшла до редакції 05.03.2012.