Impacts of Volatility on Forecasting Inbound Tourists into China

Ying-Chih Chen[1], Hsin-Hong Kang[2] and Tzer-Chyun Yang[3]

The study aims at examining if the errors of the estimation of the inbound tourist numbers into China are heteroscedastic. Three volatility models, GARCH, EGARCH and GJR GARCH, are used to estimate the conditional volatility of monthly arrivals of inbound tourists into China. Heteroscedasticity of errors, asymmetric effects and leverage effects are found in this study. It indicates that volatility should be taken into account in future studies so as to improve accuracy of international tourism demand forecasting.

Field of Research: Hospitability Industry Management,
Tourism Economics

1. Introduction

Tourism demand forecasting is very important because tourism products cannot be stored and thus are perishable. Once an airline or bus seat, a hotel room or a theme park ticket is unsold on a particular day, it perishes. Accurate forecasting will help avoid loss of unsold products as well as unfilled demand. Many studies forecast tourism demand by using time series analysis models and regression models. However, most of these studies assume variance of errors to be a fixed constant. In fact, time series of economic data have often exhibited the volatility clustering phenomenon. That is, large changes tend to

[1] Graduate School of Business Administration, National Cheng Kung University E-mail: cycjim@mail2000.com.tw
[2] Graduate School of Business Administration, National Cheng Kung University E-mail: hhkang@mail.ncku.edu.tw
[3] Graduate School of Business Administration, National Cheng Kung University E-mail: yong@mail.ncku.edu.tw

be followed by large changes, of either sign (i.e., positive or negative), and small changes tend to be followed by small changes. This means that volatility changes over time; as a result, a linear regression model will fail to make accurate forecast in most cases. One of the purposes of this study is to examine if inbound tourism demand volatility in China has ARCH (Autoregressive Conditional Hetetroskedasticity) effects.

Tourism demand is highly sensitive to the sudden events, especially to those concerning travellers' health and safety. Crises such as fatal infectious diseases, earthquakes, typhoons, floods, terrorist attacks, political instability and wars will stop leisure travellers from visiting a destination suffering one of those. On the contrary, special fairs, cultural festivals or sporting events, like the Chinese New Year Celebration in China, the cultural festivals in Japan, the Olympic Games, the Asia Games, FIFA World Cup, etc., will attract huge numbers of tourists. However, most of the studies on the influence of special events on tourism demand are based on intervention analysis, developed by Box & Tiao (1975), i.e., solving a structural change problem by using dummy variables. For example, Qu and Lam (1997) studied the impact of the Tiananmen Square tragedy, the Gulf War, and the change in visa issuance on the tourist arrivals from mainland China to Hong Kong; Goh and Law (2002) investigated the impact of the Asia financial crisis and the outbreak of bird flu on tourist arrivals to Hong Kong; Lim and McAleer (2002) explored how incidents such as the 1979 Oil Crisis affected Australia tourism; Chen & Kang & Yang (2005) found that the SARS threats were credited with decreasing the number of the inbound tourists into China.

The advantage of intervention analysis that is used in forecasting tourism demand is that it presents the level of influence of unexpected incidents on tourist arrivals, as well as helping to reduce model errors and thus increase the accuracy of forecast. On the other hand, its weakness is that the respective influence of good and bad news on volatility is not taken into account. In fact, good and bad news differently affects volatility, which refers to asymmetric effects. What's more, the researchers in economics and finances have found leverage effects, which means that bad news to economic and financial time series tends to cause volatility to rise by more than the good news of the same magnitude. The difficulty in forecasting, consequently, increases. Therefore, another purpose of this study is to find if the asymmetrical effect and the leverage effect exist in the impacts on tourism demand from both the good and the bad news.

The subject of this research is the number of tourist arrivals to China. World

Tourism Organization predicts that China will become the top tourist destination in the world and the fourth largest source market in 2020. However, not many studies have been done on the forecasting of the tourism demand in China. Therefore, this study aims to fill the gap in current studies by introducing better models of tourism demand forecast in China, and then the models can also be applied to forecast the demand in other countries.

2. Literature review

Many past studies have used the time series model as their tourism demand forecasting model (like Martin & Witt, 1989; Witt, Witt & Wilson,1994 ; Turner, Kulendran & Pergat, 1995; Kulendran & King, 1997; Chu, 1998, 2004; Kim, 1999; Lim & McAleer, 2001 ; Lim & McAleer, 2002; Cho, 2003). The forecasting results differ according to the variance in selection of models, countries of study, methodology of data collection (quarterly data or monthly data), and study periods.

Seasonal fluctuation is also a key point in tourism demand forecasting. Hyllebert (1992) suggests three major seasonal factors: the climate factor, the calendar (i.e., holiday) factor, and the festival factor. In practice, using seasonal adjustment, including seasonal dummy variables in the model, and building seasonal ARIMA models are the three methods used to process seasonal factors. Goh and Law (2002) argue that there are certain defects in the first two methods. Although seasonal adjustment frees the estimation from the interference of seasonal factors, it may cause a loss of a considerable amount of useful information. Using seasonal dummy variables, on the contrary, can rectify this shortcoming and obtain precise seasonal fluctuation data. However, its weaknesses are that the degree of freedom of estimation is decreased if the monthly data are used, and that, according to Abeysinghe (1994), the use of seasonal dummy variables will result in spurious regression if the time series are not stationary. In view of the weaknesses of the first two methods, the seasonal ARIMA model is built in this study to handle seasonal fluctuations.

Those studies are mainly based on the mean-variance analysis, but they are unaware that variance of errors changes over time. Engle (1982) believes that in an econometric model, the present conditional variance is influenced by those of the past periods. He thus proposes the autoregressive conditional heterokedasticity (ARCH) model, which successfully explains volatility clustering by having the conditional variance change over time. Compared to the traditional econometric model, it reflects more characteristics of the time series data. Bollerslev (1986) expands the ARCH model and introduces the generalized ARCH (GARCH) model, which contains fewer lag orders. Having a simpler structure, a GARCH model is appropriate for most of the economic time series data. Later scholars, like Akgiray (1989), Baillie & DeGennaro (1990), Schwert & Seguin (1990), also prove that the GARCH model is capable of capturing the changes of long-term volatility with only a few parameters.

Both ARCH and GARCH models assume symmetry in the conditional variance structure so they are unable to investigate the asymmetric effects of volatility. Then, using the GARCH model to forecast the volatility might lead to a wrong result (Antoniou, Holmes and Priestley, 1998). Therefore, some asymmetrical GARCH models have been proposed, such as EGARCH, GJR GARCH, TGARCH, etc.

All the above-mentioned literature on tourism demand forecast shows that although there are various models used in forecasting, they ignore the impacts of volatility on the forecast. Therefore, this study aims to build up GARCH, EGARCH, and GJR GARCH models to make more accurate forecast.

3. Methodology

In this study, different volatility models are introduced to forecast the tourism demand in China and to validate the existence of asymmetric effects and leverage effects.

3.1 The ARCH Model

Engle (1982) in his model assumes that the conditional variance is a positive function of the value of the previous error terms instead of a fixed constant. In other words, the error variance depends on p lags of squared errors, and the ARCH(p) model is specified as follows:

$$\sigma_t^2 = \alpha_0 + \sum \alpha_i \varepsilon_{t-i}^2$$

where α_0 >0, α_i ≧0(i = 1,…p)

Here, σ_t^2 is the conditional variance at time t, α_0 is a constant parameter, α_i are coefficients, and ε_{t-i}^2 are the ARCH terms. Since σ_t^2 is a conditional variance, its value must always be positive. In order to ensure that the equation is meaningful, all the coefficients in the RHS of the equation, namely, α_i are required to be non-negative. Since the conditional variance, σ_t^2, is affected by the past error terms, ε_{t-i}^2 and α_i are always non-negative, the present volatility is positively correlated with the past error terms, which is known as volatility clustering.

3.2 The GARCH Model

Bollerslev (1986), incorporating the concept of the ARMA (autoregressive moving average) model in the ARCH model by adding the conditional variance of the previous lags into the ARCH model, proposes the GARCH model. The GARCH (p, q) model specification for the conditional variance equation is

$$\sigma_t^2 = \alpha_0 + \sum \alpha_i \varepsilon_{t-i}^2 + \sum \beta_j \sigma_{t-j}^2$$

where α_0>0, $\alpha_i \geq 0$, $\beta_j \geq 0$ (i = 1,…,p; j = 1,…,q)

Here, σ_t^2 is the conditional variance at time t, α_0 is a constant parameter, α_i are coefficients, ε_{t-i}^2 are the ARCH terms, and σ_{t-j}^2, the last period's forecast conditional variances, are the GARCH terms. Just like the ARCH model, all the coefficients in the RHS of the equation, namely α_i and β_j, are required to be non-negative so that the equation would not be meaningless. In the GARCH model,$\Sigma\alpha_i + \Sigma\beta_j$ must be less than 1 to satisfy the stationary condition. If $\Sigma\alpha_i + \Sigma\beta_j$ is close to 1, it means that the impact of news on volatility will last for a long time.

3.3 The EGARCH (Exponential GARCH) Model

One of the weaknesses of The ARCH and GARCH models is that they enforce a symmetric response of volatility to good and bad news. Nelson (1991), thus, introduces the Exponential GARCH (EGARCH) model to solve the problem and also fulfill the requirement that a coefficient shall not be negative.
The EGARCH model is written

$$Log(\sigma_t^2) = \alpha_0 + \alpha\left[\frac{|\varepsilon_{t-1}|}{|\sigma_{t-1}|} - \sqrt{\frac{2}{\pi}}\right] + \beta\log(\sigma_{t-1}^2) + \gamma\frac{\varepsilon_{t-1}}{\sigma_{t-1}}$$

where α_0 is a constant parameter,and α, β ,γ are coefficients. In this model, γ reflects the asymmetric effect from the good and the bad news on the conditional variance $\log(\sigma_t^2)$. When $\gamma \neq 0$, there is an asymmetrical effect; when γ is negative, a leverage effect occurs.

3.4 The GJR GARCH Model

The GJR GARCH model, proposed by Glosten, Jagannathan & Rukle (1993), is capable of capturing the asymmetric effect in regard to the conditional volatility, which is what the GARCH model cannot explain.
The GJR GARCH model is specified as follows:

$$\sigma_t^2 = \alpha_0 + \sum \alpha_i \varepsilon_{t-i}^2 + \sum \beta_j \sigma_{t-j}^2 + \gamma \varepsilon_{t-i}^2 I_{t-i}$$

where

$I_{t-i} = 1$ if $\varepsilon_{t-i} < 0$

$I_{t-i} = 0$ if $\varepsilon_{t-i} \geq 0$

$\alpha_0 > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$ (i = 1,…,p; j = 1,…,q)

I_{t-1} is a dummy variable. If $I_{t-1} = 1$, it means there is an impact from negative news on tourism environment. On the other hand, If $I_{t-1} = 0$, the tourism environment is facing a positive news impact. This model captures the asymmetrical effect with the dummy variable, I_{t-1}. That is, if $\gamma \neq 0$, it means an asymmetric effect occurs, and if $\gamma > 0$, there is a leverage effect.

4. Discussion of findings

This study focuses on the period from January 1985 to November 2008, and the data set consists of 287 monthly data of the inbound tourists into China. The sources of data are the *Yearbook of China Tourism Statistics* and the official website of National Tourism Administration of the People's Republic of China.
In building a model, most of the economic time series data are processed with the use of the logarithmic transformation. Lim (1997) believes that the logarithmic transformation is the most appropriate method to deal with economic time series data. One advantage is that, with the logarithmic transformation, if the series are non-stationary, first-order difference equations

would be used and the result is the monthly growth rate of visitor arrivals, as follows:

$$\log(Arr_t) - \log(Arr_{t-1}) = \log(\frac{Arr_t}{Arr_{t-1}}) \approx \frac{Arr_t - Arr_{t-1}}{Arr_{t-1}}$$

where Arr_t is the number of tourist arrivals. Therefore, this study will apply the logarithmic transformation to the number of inbound tourists into China.

4.1 Unit Root Test

When using time series for forecasting, the first step is to examine if this time series is stationary. This study uses augmented Dicker-Fuller (ADF) unit root test to do the examination. Given the ADF statistic is -3.30 and the critical value at 5% is -3.43, it fails to reject the null hypothesis of nonstationarity. Thus, the number of inbound tourists into China is a non-stationary time series.

A non-stationary time series requires differencing to induce stationarity. After the first difference, the result shows that the ADF statistics is -4.56, and the critical value at 5% is -3.45, so it rejects the null hypothesis. It means that the time series becomes stationary after the first difference.

4.2 The specification of conditional mean equation

Before building the volatility models of the number of the inbound tourists into China, an appropriate conditional mean equation must be specified to investigate both serial correlation and seasonality. This study uses the seasonal ARIMA model, proposed by Box and Jenkins (1976). Seasonal autoregressive (SAR) and seasonal moving average (SMA) terms are used for monthly or quarterly data with systematic seasonal movements. The model is expressed by the notation ARIMA (p, d, q) $(P, D, Q)_s$, where the term (p,d,q) represents the order of an ARIMA model, (P,D,Q) represents the order of a seasonal ARIMA model and s is the length of the seasonal cycle. In this study, the monthly data are used, so s is 12. The order of the model is selected by two criteria—the Akaike information criterion (AIC) and the Schwarz information criterion (SIC). After the test, ARIMA$(0,1,1)(1,0,1)_{12}$ is shown to be the best conditional mean equation. All the coefficients are significant, and the t values are very high (See Table 1 for the other statistics).

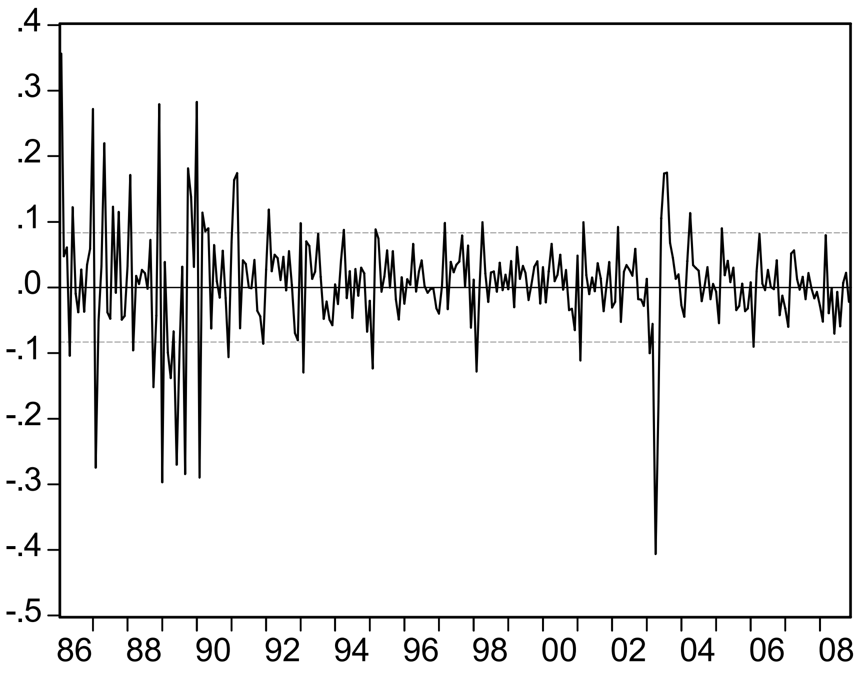
Table 1. : Estimation of ARIMA(0,1,1) $(1,0,1)_{12}$

Variable	Coefficient	t-value	AIC/SIC
Constant Term	-0.001	-0.12	AIC = -2.12
MA(1)	-0.4998****	-9.73	SIC = -2.07
SAR(12)	0.9513****	59.98	
SMA(12)	-0.8728****	-35.35	

* p<0.1; **p<0.05; ***p<0.01; ****p<0.001

After the model specification, the errors need to be checked for evidence of serial correlation. In Figure 1, the error plot exhibits a volatility clustering, for example, 1988 to 1990, and 2003 to 2005. Moreover, the p value of Ljung-Box Q(10) statistic is 0.267, which is not significant, i.e., none of the error is serial correlated. Yet, the p value of Ljung-Box Q(10) statistic of the error square is less than 0.001, which is significant, i.e., the error square is serial correlated. The LM test (Lagrange multiplier) is applied to examine if the residual is higher order series correlated. The result shows that the F statistic is 2.61 whereas p value is 0.025. The result of the statistics is consistent with what the error plot exhibits so the ARCH effects are likely to exist. After the ARCH LM test, the result shows that the F statistic is 9.34, and p value is less than 0.001, which rejects the hypothesis that there is no ARCH effect in the errors. Therefore, conditional volatility models have to be built to investigate the ARCH effects.

Figure 1 time series plot of ε_t



4.3 Estimation of the GARCH model

GARCH(1,1), proposed by Bollerslev, Chou & Kroner (1992), is preferred to explain the GARCH model in most of the economic time series. Parsimonious as it is, this model is sufficient to capture the volatility clustering in the data. Therefore, this study uses GARCH(1,1) to estimate the volatility.

The result of the estimation shows that ARCH parameter α and GARCH parameter β are both significant at the 0.05 level of significance, which indicates the existence of the ARCH effects. Moreover, $\alpha + \beta = 0.96$. Although it is less than 1, which satisfies a stationary condition, it is still close to 1, i.e., the impact of the news has a long-term effect on the number of the inbound tourists into China.

Table 2.: Estimation of GARCH(1,1)

Variable	**Coefficient**	**z-Statistic**	**AIC/SIC**
Constant Term	0.009	1.02	AIC = -2.69
MA(1)	-0.5179****	-9.15	SIC = -2.58
SAR(12)	0.9639****	95.67	
SMA(12)	-0.7954****	-32.46	
α_0	0.006**	2.12	
α	0.474***	3.30	
β	0.486****	4.55	

* p<0.1; **p<0.05; ***p<0.01; ****p<0.001

4.4 Estimation of the EGARCH model

The result shows that all the parameters in the variance equation are significant at the 0.05 level of significance. Most importantly, parameter $\gamma \neq 0$, and its sign is negative. It means that the volatility of monthly tourist arrivals into China has asymmetric effects and leverage effects.

Table 3.: Estimation of EGARCH(1,1)

Variable	**Coefficient**	**z-Statistic**	**AIC/SIC**
Constant Term	0.002	0.27	AIC = -2.69
MA(1)	-0.522****	-9.19	SIC = -2.57
SAR(12)	0.956****	98.70	
SMA(12)	-0.802****	-32.72	
α_0	-1.272***	-3.76	
α	0.710****	5.18	
β	-0.158**	-2.07	
γ	0.865****	15.30	

* p<0.1; **p<0.05; ***p<0.01; ****p<0.001

4.5 Estimation of the GJR GARCH model

The result shows that all parameters in the variance equation, except for γ, are significant at the 0.05 level of significance, but γ is significant at the 0.1 level of significance. The sign of γ, being positive, shows the asymmetric and leverage effects might exist.

Table 4.: Estimation of GJR GARCH(1,1)

Variable	**Coefficient**	**z-Statistic**	**AIC/SIC**
Constant Term	0.004	0.59	AIC = -2.69
MA(1)	-0.5302****	-9.7	SIC = -2.58
SAR(12)	0.9592****	94.42	
SMA(12)	-0.7997****	-32.31	
α_0	0.0006***	2.17	
α	0.296**	2.09	
β	0.493****	4.63	
γ	0.359*	1.73	

* p<0.1; **p<0.05; ***p<0.01; ****p<0.001

In diagnostic checking of the three volatility models, both Ljung-Box Q(10) and Ljung-Box $Q^2(10)$ statistics are not significant. The result of the ARCH LM test shows that the F statistic is not significant. It shows that none of the error term is serial correlated and there is no ARCH effect in the three volatility models.

Table 5.: Diagnostic checking of three volatility models

	GARCH(1,1)	EGARCH(1,1)	GJR GARCH(1,1)
Ljung-Box Q(10)	5.55	6.34	6.20
Ljung-Box $Q^2(10)$	3.59	3.44	4.06
ARCH LM	F = 0.37	F = 0.29	F = 0.44

5. Conclusion

This study validates that inbound tourism demand volatility in China has ARCH effects. That is, in order to increase the accuracy of tourism demand forecast, volatility cannot be assumed as a constant, which does not change over time, because volatility clustering occurs in tourist arrivals. In this study, the GARCH, EGARCH, and GJR GARCH models, which can estimate volatility, are introduced to forecast tourism demand. The AIC and SIC of the three models are much lower than those of the seasonal ARIMA model, meaning that the three models do have better accuracy in forecasting tourism demand.

The result of this study also indicates the existence of asymmetrical effect and leverage effect in the number of inbound tourists into China. If the EGARCH or GJR GARCH models are not used, volatility will be underestimated when the tourism demand is influenced by bad news, whereas overestimated when influenced by good news. The impact from the bad news will significantly increase volatility as well as the difficulty in forecasting the number of the tourists into China, which would cause many problems in the investment and marketing of the tourism-related businesses in the future. Therefore, Chinese government and the tourism industries must face this problem and try to minimize the impact from bad news. Although a natural disaster is inevitable, a good disaster rescue and prevention system will strengthen the confidence of the tourists. Man-made disasters, such as infectious diseases or political instability, must be solved by the authorities of China. For instance, without taking proper measurements, the outbreak of H5N1 bird flu in the future may deter tourists from visiting China, just like the outbreak of SARS in 2002.

To increase the accuracy of a forecasting model is a very important, but also a difficult, task. By introducing the GARCH, EGARCH, and GJR GARCH models, which take volatility into account, this study hopes to contribute to the relevant studies. The future research should use the three volatility models to forecast the number of inbound tourists into China or other countries more accurately so that the government and tourism industries can make appropriate policies and decisions.

Reference

Abeysinghe, T. 1994. "Deterministic seasonal models and supurious regression", *Journal of Econometrics*, vol. 61, pp. 259-72.

Akgiray, V. 1989. "Conditional heteroscedasticity in time series of stock returns : evidence and forecasts", *Journal of Business*, vol. 62, pp. 55-80.

Antoniou, A., Holmes, P. & Priestley, R. 1998. "The effects of stock index futures trading on stock index volatility: an analysis of the asymmetric response of volatility to news", *Journal of Futures Markets*, vol. 18, no. 2, pp. 151-66.

Baillie, R.T. & DeGennaro, R. P. 1990. "Stock returns and volatility", *Journal of financial and Quantitative Analysis*, vol. 25, pp. 203-14.

Bollerslev, T. 1986. "Generalized autoregressive conditional heteroscedasticity", *Journal of Econometrics*, vol. 31, pp. 1307-27.

Bollerslev, T., Chou, R. Y. & Kroner, K. F. 1992. "ARCH modeling in finance: a review of the theory and empirical evidence", *Journal of Econometrics*, vol. 51, pp. 5-59.

Box, G. E. P. & Tiao, G. C. 1975. " Intervention analysis with application to economic and environmental problems", *Journal of the American Statistical Association*, vol. 70, pp. 70–9.

Box, George E. & Jenkins, G. M. 1976, 'Time series analysis: forecasting and control', Holden-Day, San Francisco.

Chen, Kang & Yang 2007. "A study on the Impact of SARS on the forecast of visitor arrivals to China", *Journal of Asia-Pacific Business*, vol. 8, no. 1, pp. 31-50.

Cho, V. 2003. "A comparison of different approaches to tourist arrival forecasting" *Tourism Management*, vol. 24, pp. 323-30.

Chu, F. L. 1998. "Forecasting tourism demand in Asian-Pacific countries", *Annals of Tourism Research*, vol. 5, pp. 597–615.

Chu, F. L. 2004. "Forecasting tourism demand: A cubic polynomial approach", *Tourism Management*, vol. 25, pp. 209–18.

Engle, R.F. 1982. "Autoregressive conditional Heteroskedasticity with estimates of the variance of UK inflation" *Econometrica*, vol. 50, pp. 987-1008.

Glosten, L., R. Jagannathan & Runkle, D. 1993. "On the relation between the expected value and the volatility on the nominal excess returns on stocks" *Journal of Finance*, vol. 19, pp. 3-29.

Goh, C. & Law, R. 2002. "Modeling and forecasting tourism demand for arrivals with stochastic nonstationary seasonality and intervention", *Tourism Management*, vol. 23, pp. 499–510.

Hyllebert, S. 1992, *Modeling Seasonality*, Oxford University Press, Oxford, UK.

Kim, J. H. 1999. "Forecasting monthly tourist departure from Australia" *Tourism Economics*, vol. 5, pp. 277–91.

Kulendran, N. & King, M. L. 1997. "Forecasting international quarterly tourist flow using error-correction and time-series models" *International Journal of Forecasting*, vol. 13, pp. 319-27.

Lim, C. 1997. "Review of international tourism demand models", *Annals of Tourism Research*, vol. 24, pp. 835–49.

Lim, C. & McAleer, M. 2001. "Monthly seasonal variations: Asian tourism to Australia", *Annals of Tourism Research*, vol. 28, pp. 68–82.

Lim, C. & McAleer, M. 2002. " Time series forecasts of international travel demand for Australia" *Tourism Management*, vol. 23, pp. 389-96.

Martin, C. A. & Witt, S. F. 1989. "Forecasting tourism demand: a comparison of the accuracy of several quantitative methods" *International Journal of Forecasting*, vol. 5, pp. 7-19.

Nelson, D.B. 1991. "Conditional heteroskedasticity in asset returns : a new approach", *Econometrica*, vol. 59, pp. 347-70.

Qu, H. & Lam, S. 1997. "A travel demand model for Mainland Chinese tourist to Hong Kong", *Tourism Management*, vol. 18, pp. 593–7.

Schwert, G.W. & Seguin, P.J. 1990. "Heteroscedasticity in stock return" *Journal of Finance*, vol. 45, pp. 1129-51.

Turner, L. W., Kulendran, N. & Pergat, V. 1995. " Forecasting New Zealand tourism demand with disaggregated data" *Tourism Economics*, vol. 1, pp. 51-69.

Witt, C. A., Witt, S. F. & Wilson, N. 1994. "Forecasting international tourist flows" *Annals of Tourism Research*, vol. 21, pp. 612-28.