This is the Pre-Published Version.

An Empirical Study of Forecast Combination in Tourism

Haiyan Song [1]
Stephen F. Witt
Kevin K. F. Wong
Doris C. Wu

School of Hotel and Tourism Management
The Hong Kong Polytechnic University
Hung Hom, Kowloon
Hong Kong SAR

An Empirical Study of Forecast Combination in Tourism

Abstract

The performance of forecast combination techniques is explored at different time horizons in the context of tourism demand forecasting. Statistical comparisons between the combination and single model forecasts show that the combined forecasts are significantly more accurate than the average single model forecasts across all forecasting horizons and for all combination methods. This provides a strong recommendation for forecast combination in tourism. In addition, the empirical results indicate that forecast accuracy does not improve as the number of models included in the combination forecasts increases. It also appears that combining forecasts may be more beneficial for longer-term forecasting.

Key Words: Forecast combination; forecasting accuracy; tourism demand

1. Introduction

With the fast growth of the tourism industry in many developed and developing economies, tourism forecasting has attracted much attention from marketers as well as researchers. Accurate forecasts of tourism demand are of great importance not only for the private sector, such as hotels, airlines and tour operators in terms of their business planning and investment, but also for destination governments in terms of tourism policy formulation and implementation. A number of quantitative techniques have been applied to tourism demand forecasting over the last three decades, and these methods include time series analysis, econometric models, and nonlinear modeling approaches (see Li et al. 2005).

This study aims to examine whether forecasting performance could be improved by combining tourism forecasts generated by individual forecasting models. Witt and Song (2001) noted that the forecasting accuracy of individual forecasting methods varies across origin-destination pairs and over different forecasting horizons. Since tourism planners and business decision-makers attach high importance to the accuracy of forecasting, it is crucial for researchers to explore what are the best techniques for tourism demand forecasting. This study is a further attempt to look at forecasting accuracy by examining whether forecast combination could improve the overall forecasting accuracy of tourism demand models. The demand for Hong Kong tourism by ten major origin countries/regions is used as the base of the study. Combination forecasting techniques focus on combining the individual forecasts generated by different models through appropriate weighting schemes, which have been developed in the general forecasting literature. Published general forecasting studies show that the combination of individual forecasts can improve forecasting accuracy (Winkler & Makridakis, 1983), but this conclusion is not supported by some of the more recent studies, such as Hibon and Evgeniou (2005) and Koning et al. (2005).

Although forecast combination has attracted wide attention in the general forecasting literature, very little research on this topic has appeared in the tourism forecasting

literature. This study aims to make a major contribution to the tourism forecasting literature by providing a more comprehensive empirical investigation of combination forecasting in tourism; than has previously been undertaken.

The empirical analysis of this paper follows five steps. Firstly, four modern forecasting methods are employed to generate single model forecasts at different time horizons: one, two, four and eight quarters(s) ahead. Secondly, three combination methods are utilized to obtain combination forecasts at different time horizons. Thirdly, the differences in accuracy between combination forecasts and single forecasts are tested for statistical significance. Fourthly, forecasting accuracy differences are examined in terms of both the combination method used and the length of the forecasting horizon. Lastly, the expected improvement in forecasting accuracy resulting from including extra models in the forecast combination is investigated for different time horizons. This is the first time that the issues considered in steps three, four and five have been addressed in the tourism forecasting literature.

The rest of the paper is organized as follows. Section 2 reviews the recent developments in the tourism forecasting and combination techniques literature. Section 3 introduces the data, the forecasting models and the combination techniques. Section 4 presents the empirical results and the conclusions are given in Section 5.

2. Literature Review

During the past two decades, econometric techniques have advanced significantly. These new developments have also played an important role in the understanding of tourists' behavior and their demand for tourism products/services. Li et al. (2005) reviewed eighty-four studies on tourism demand analysis published since the 1990s and found that a majority of these studies used econometric methods. For example, the general-to-specific approach was used by Song and Witt (2003) to build an ADL (Autoregressive Distributed Lag) model to forecast inbound tourism to South Korea from four major tourism origin countries. A vector autoregressive (VAR) model was applied

by Song and Witt (2006) to forecast *ex ante* tourist flows to Macau from eight major origin countries/regions. Law and Au (1999) used the neural network model to forecast Japanese demand for travel to Hong Kong. Li et al. (2004) used the long-run static and the short-run error correction-almost ideal demand system (EC-LAIDS) models to examine the demand for tourism in five European destinations by UK residents.

There are also a large number of studies focusing on tourism forecasting accuracy comparisons. Kulendran and King (1997) compared the forecasting performance of an error correction model (ECM), autoregressive (AR) model, autoregressive integrated moving average (ARIMA) model, a basic structural model and a regression based time series model. Their results demonstrated that the ECM performs poorly compared to the time series models. The reason why the ECM model performed badly may lie in the ways in which the non-stationary and seasonal data were used in the model specification. Kulendran and Witt (2003) examined seven forecasting models including an ARIMA model, ECM, and some structural time series models and found that the length of the forecasting horizon is highly related to a model's relative forecasting performance. Oh and Morzuch (2005) explored the performance of eight models in forecasting inbound tourism demand in Singapore and concluded that the selection of the performance measure and the forecasting horizon are the two main factors affecting performance. Song et al. (2000) generated *ex post* forecasts of the outbound tourism demand of UK residents to twelve destinations over a period of six years using an ECM and compared the forecasting performance of the ECM with that of AR, ARIMA and VAR models. Their results suggest that the ECM outperforms all the competitors. Witt et al. (2003) evaluated the forecasting performance of six econometric models and two univariate time series models using data on international tourism to Denmark. The results show that the time varying parameter (TVP) model performs consistently well in one-year-ahead forecasting, but the best model varies at longer forecast horizons. Smeral and Wüger (2005) showed that complex data adjustment procedures and adequate model structures also affect forecasting accuracy based on Austrian tourism demand data.

Although tourism forecasting has achieved much progress in terms of the use of modern modeling methodologies, one area that has attracted very little attention is forecast combination. The seminal work in the general forecasting literature on combination forecasts is attributed to Bates and Granger (1969); they examined the performance of combing two sets of forecasts of airline passenger data in which the weights are calculated based on the historic performance of each individual model. Their major finding is that the combined forecasts yielded much lower mean-square errors than either of the original individual forecasts. Clemen (1989) reviewed the development and applications of combination techniques in the various areas of forecasting before 1989. His key conclusion is that forecasting accuracy could be substantially improved through the combination of individual forecasts. The simple average combination method that attaches the same weight to each of the individual forecasts has been widely applied in the forecasting literature (for example Makridakis and Winkler 1983; Fang 2003; Hibon and Evgeniou 2005). However, many published studies have also used more advanced combination methods to achieve the optimal weights for combining the individual forecasts. In these procedures the past performance of the single forecast models is the key criterion for deciding the optimal weights for each of the individual models. Winkler and Makridakis (1983) applied several versions of the variance-covariance weighting method to examine the performance of the combined forecasts. The results show that most of the combined forecasts perform better than the individual forecasts. Walz and Walz (1989) examined the performance of a Bayesian method in comparison with an unconstrained regression procedure to combine forecasts through a study of four macroeconomic variables and drew the conclusion that the Bayesian procedure generates more accurate combined forecasts. Diebold and Pauly (1990) applied the shrinkage technique to incorporate prior information into the estimation of the combination weights and the empirical research in their study based on US GNP data found that the estimated combination weights are largely shrunken toward equality.

A number of studies suggest that combination techniques can outperform the best constituent single individual forecast based on empirical studies or simulation. However, Hibon and Evgeniou (2005) concluded that the best individual method and the

combination forecasts perform similarly based on an analysis of the 3003 series of the M3-competition. In their study only simple average combination is examined and the conclusion is that the advantage of combination is to decrease the forecasting risk but not significantly outperform the best single forecast. Koning et al. (2005) examined the performance of the combination of forecasts from three univariate forecasting models and concluded that the combined forecasts do not outperform the single forecasts.

The only published empirical study on combination forecasting in tourism since the 1980s is by Wong *et al* (2007). These authors examined the relative accuracy of combination and single model forecasts for one quarter ahead, but did not consider statistically significant differences. Wong *et al* (2007) concluded that combination forecasts are almost certain to outperform the worst single forecasts but only outperform the best single forecasts in less than 50 per cent of cases on average.

3. Data and Methodology

Data

The data on the demand for Hong Kong tourism by the top ten tourism generating countries/regions are used to estimate the forecasting models and these top ten major origin countries/regions comprise: Mainland China, Taiwan, Japan, USA, Macau, South Korea, Singapore, UK, Australia and Philippines. The demand variable is measured by tourist arrivals in Hong Kong from these origin markets. The price of Hong Kong tourism, tourism prices in substitute destinations and the income level in the tourist origin countries/regions are considered to be the explanatory factors which influence the demand for Hong Kong tourism (see Song et al. 2003).

The price of Hong Kong tourism can be represented by the relative consumer price index (CPI) between Hong Kong and the origin country/region. This variable is adjusted by the relevant exchange rate (EX). The specific variable is defined as follows:

$$P_{it} = \frac{CPI_{HK}/EX_{HK}}{CPI_i/EX_i} \qquad (1)$$

where *HK* denotes Hong Kong and *i* denotes the *i*th origin country/region.

In this study six countries/regions are considered as the substitute destinations of Hong Kong - Mainland China, Taiwan, Singapore, Thailand, Korea and Malaysia. The substitute price of Hong Kong tourism variable is calculated as the weighted average exchange rate adjusted CPI and is expressed as:

$$P_{it}^s = \sum_{j=1}^{n} \frac{CPI_j}{EX_j} w_{ij} \qquad (2)$$

where *j* denotes the *j*th substitute destination and *n*=6. The weights assigned to these six destinations are calculated based on their own inbound tourist arrivals from the studied origin countries/regions, and it can be written as:

$$w_{ij} = \frac{TA_{ij}}{\sum_{j=1}^{6} TA_{ij}} \qquad (3)$$

where TA_{ij} represents inbound tourist arrivals in substitute destination *j* from country/region *i*.

Tourists' income is measured by the GDP index (2000=100) in constant prices in these ten origin countries/regions. Seasonal dummies and one-off event dummies are also included in the modeling process to capture seasonal impacts and effects of some one-off events such as the hand-over of Hong Kong to China in quarter 3 of 1997.

In this research quarterly data are employed to generate *ex post* forecasts and the sample data starts from 1984q1. The SARS event which occurred in 2003q2 enormously

influenced Hong Kong inbound tourism. To avoid the effect of this outlier, the sample period ceases at 2003q1. The major data sources include *Visitor Arrivals Statistics* published monthly by the Hong Kong Tourism Board, *Tourism Statistical Yearbook* published by World Tourism Organization and the International Financial Statistics Online Service website of the International Monetary Fund.

Modeling methods

The seasonal ARIMA method can effectively handle the identification and modeling of seasonal time series. This method is developed from the standard Box-Jenkins model and incorporates both seasonal autoregressive and moving average factors into the modeling process. This method has been widely adopted in forecasting seasonal time series. Since quarterly data are analyzed, the seasonal ARIMA model is appropriate in this study.

The Autoregressive Distributed Lag (ADL) model is also included in this study. The general-to-specific approach suggests that the final forecasting model can be obtained through the simplification of a general ADL model, which can be written as:

$$y_t = \alpha + \sum_{i=1}^{p} \phi_i y_{t-i} + \sum_{i=0}^{q} \beta_i x_{t-i} + \varepsilon_t \tag{4}$$

where y_t denotes the dependent variable (tourist arrivals) and x_t denotes a vector of exogenous independent variables. p and q are lag lengths while α, ϕ_i and β_i are coefficient vectors to be estimated.

The third model is the two-stage ECM. Engle and Granger (1987) suggested that a linear combination of two or more non-stationary series may generate a stationary series. If this is the case, a cointegration relationship exists. The first stage of the EC modeling is to set up a long-run cointegration model, whereas in the second stage a short-run equilibrium model is built. Since quarterly data are employed in this paper, the HEGY (Hylleberg, Engle, Granger & Yoo, 1990) test is applied for testing for seasonal unit roots

and seasonal filters are used to eliminate the seasonal unit roots before an EC model is established.

The VAR model is a system estimation technique, and treats all variables as endogenous except the intercept, determinate time trend and dummies. The optimal lag length for the model is determined according to the Aikake Information Criterion (AIC). An excessive lag length will reduce the degrees of freedom for model estimation while a shorter than optimal lag length could lead to the misrepresentation of the data generating process.

Combination forecasting methods

In this study three combination methods are adopted to generate combined forecasts and they are the simple average combination, variance-covariance combination and discounted Mean Square Forecast Error (MSFE) combination. The simple average combination is a widely used method in practice because it is easy to comprehend and operate. The variance-covariance method takes the variance and covariance of single model forecast observations into account in order to generate the weights. The discounted MSFE method assigns heavier weights to more recent observations through a discount factor. Whereas the simple average combination method assigns equal weights to each of the individual forecasts, the latter two combination methods assign different weights to the individual forecasts.

The simple average combination: let f and w be the single forecast vector and weights vector respectively. The combined forecast f_c is:

$$f_c = wf \tag{5}$$

with $f = (f_1, f_2, ..., f_n)$, and $w = (w_1, w_2, ..., w_n)'$. In the simple average combination method, w is a $(n \times 1)$ vector with every element equal to $\frac{1}{n}$. Consequently f_c can be

expressed as $f_c = \dfrac{\sum_{i=1}^{n} f_i}{n}$. The simple average combination is the most commonly used combination forecasting technique in the existing literature due to its ease of implementation.

The variance-covariance combination: this assigns unequal weights to each single model forecast and the weights are determined by the historic performance of the single model forecasts. The specific weight vector is expressed as:

$$w' = u'\Sigma^{-1} / u'\Sigma^{-1}u \tag{6}$$

where Σ is the covariance matrix of the single forecasts, u is an $(n \times 1)$ vector of ones with constraints. $\sum_{i=1}^{n} w_i = 1$. Under the assumption that single model forecast errors are normally distributed with zero means, the estimate of Σ can be denoted as $(\hat{\Sigma})_{ij} = \sum e_i e_j$ where e_{it} is the difference between the actual and forecast series of the ith model, $e_{it} = y_t - f_{it}$. A detailed description of this method can be found in Clemen and Winkler (1986), Diebold and Pauly (1987), Granger and Newbold (1986), and Wong, et al. (2007).

The discounted MSFE combination: this method uses the mean square error to calculate the optimal weights and a discounting factor is used give more weight to the more recent forecasts (Diebold and Pauly, 1987, Bates and Granger, 1969, and Stock and Watson, 2004). The weighting scheme can be written as:

$$\Sigma_{ij} = \begin{cases} \sum_{t=1}^{T} \beta^{T-t+1} e_{it}^2 & i = j \\ 0 & i \neq j \end{cases} \tag{7}$$

where β is the discounting factor with $0 < \beta \le 1$, and e_{it} is the difference between the actual series and forecast series of the ith model. In this method, the off-diagonal elements of Σ are set to zero, which means that the effect of the covariance on the weights is ignored. The weight assigned to the ith single forecast can be represented as:

$$w_i = \left(\sum_{t=1}^{T} \beta^{T-t+1} e_{it}^2 \right)^{-1} \bigg/ \sum_{i=1}^{n} \left(\sum_{t=1}^{T} \beta^{T-t+1} e_{it}^2 \right)^{-1}$$

Measures of forecasting accuracy and statistical test of relative accuracy

Li et al. (2005) summarized the accuracy measures used in the tourism forecasting literature and they include MAPE (Mean Absolute Percentage Error), RMSE (Root Mean Square Error), MAE (Mean Absolute Error), RMSPE (Root Mean Square Percentage Error) and Theil's U statistic. Among these measures, MAPE is the most commonly used in tourism forecasting. The current study therefore also uses MAPE as the measure of forecast performance, which is denoted as:

$$MAPE = \frac{\sum_{t=1}^{n} \frac{|\hat{y}_t - y_t|}{y_t}}{n} \times 100 \tag{8}$$

where y_t and $\hat{y}_t$ are the actual and forecast values of tourist arrivals, respectively, and n is the length of the forecasting horizon. As expressed in Equation (8), MAPE removes the influence of the magnitudes of the variables; hence it can be used to compare the accuracy among different time series.

This study tests the hypothesis that combination forecasts and single model forecasts have the same accuracy. However, the traditional t test is not appropriate in some cases, as the normality test shows that some of the MAPE series are not normally distributed. Under this circumstance, a nonparametric technique would be more appropriate. In this study, the Wilcoxon signed rank test is used to test for forecasting accuracy differences between the single model forecasts and the combination forecasts in cases of non-

normality. The advantage of this method is that it can not only determine the direction of any difference in forecasting accuracy, but also can take account of the magnitude of any difference between the single model and combination forecasts.

For the variance-covariance combination and discounted MSFE combination, the historic performance of each single model forecast decides the weights assigned to the single model forecasts in order to calculate the combination forecasts. As mentioned above, the periods for estimation and forecasting are 1984q1-1994q4 and 1995q1-2003q1, respectively. As a result, 33 one-step-ahead forecasts, 32 two-step-ahead forecasts, 30 four-step-ahead forecasts and 26 eight-step-ahead forecasts are obtained. At every forecasting time horizon, the latest 15 observations are used for accuracy comparison while the rest are used for estimating the weights. Specifically, two methods are employed to assign the optimal weights, which are demonstrated using one-step-ahead forecasting as an example. As far as the first method is concerned (combination (a)), the 33 observations are separated into two parts. The weights are calculated based on the first 18 observations, and then these weights are assigned to the latter 15 observations to obtain the combination forecasts (see, for example, Granger and Ramanathan 1984; Diebold and Pauly 1990). The second method (combination (b)) calculates the weights recursively. Namely, the weights calculated from the first 18 observations are assigned to the 19th observations. Then the weights generated from the first 19 observations are assigned to the 20th observations. This process continues until all combined forecasts are obtained (see Clemen, 1986). The two-, four- and eight-steps-ahead combined forecasts can be obtained in a similar manner.

4. Empirical results

The calculated MAPEs of the individual forecasts and combined forecasts are shown in Table 1. The results show that no single forecast consistently outperforms all other forecasts across all horizons with the exception of the Philippines. This finding is consistent with past studies such as Witt et al. (2003). The asterisk symbol denotes the

situation in which the combined forecast is more accurate than the best constituent single forecast. It is demonstrated that from short-, medium- and long-run perspectives, the combined forecasts cannot always outperform the best single forecasts in all situations.

Furthermore, Table 2 shows the percentage of the combined forecasts which outperform the best individual forecasts. Forecast combination is superior to the best single model forecasts in only about 50% of all the cases. This result contradicts the findings of some previous studies that concluded that combination forecasts tend to outperform the best single model forecasts. It can also been seen that in the eight-step-ahead case, a larger proportion of combination forecasts perform better than the best single model forecasts. An explanation for this could be that the combined forecasts are usually more efficient when the single models are misspecified and in the longer term the models are more likely to be misspecified. As the forecasting horizon becomes longer, the additional uncertainty also causes the forecasts to be less accurate. From this viewpoint, forecast combination tends to be more suitable for long-run tourism forecasting. An empirical study by Lobo (1992) supports the above findings as it indicates that the difference between the average MAPEs of the individual models and the combined models decreases when the forecasting horizon gets shorter, which implies that forecast combination tends to be more useful for longer forecasting horizons. However, Lawrence, et al (1986) conclude that the greatest improvement in forecasting accuracy is often obtained in the short run rather than in the long run. They attribute this to the unrealistic assumption that the combination model structure remains constant over longer forecasting horizons. Further studies are necessary to confirm the benefits of forecast combination over different forecasting horizons in the tourism context.

Although forecast combination does not always beat the best single model forecasts, our empirical results do show that almost all the MAPE values of the combined forecasts are smaller than those of the worst single model forecasts for all countries/regions and across all forecasting horizons (see Table 3). This is to say that much worse *ex ante* forecasts could well be obtained if only one single forecasting method is used to generate tourism forecasts. Forecast combination can therefore reduce the risk of complete forecasting failure.

This study now compares the performance of combination forecasts with single model forecasts from a statistical point of view. As mentioned above, MAPE is a measure which allows the comparison not only among different models but also across countries/regions. The following explains the procedure of testing the difference between the combination and single model forecasts

Let $M_i^s = \{M_{ij}^s\}$ ($j=1, 2, ..., n$) represent the MAPEs of the corresponding n single forecasts of M_i^c. n is the number of single forecasts which are combined to generate M_i^c. The minimum series, maximum series and mean series of the MAPEs of the single model forecasts are defined as follows:

$$M_i^{\min} = \min_j(M_{ij}^s)$$

$$M_i^{\max} = \max_j(M_{ij}^s) \tag{9}$$

$$M_i^{mean} = \frac{1}{n}\sum_{j=1}^{n} M_{ij}^s$$

The aim is to test the following three null hypotheses: $M^c = M^{\min}$; $M^c = M^{\max}$ and $M^c = M^{mean}$.

The normality test shows that these series are not normally distributed and we have to employ the Wilcoxon matched-pairs signed ranks test, a nonparametric approach for which the tested series are not required to be normally distributed. The test results for hypothesis 1 (Table 4) show that, among the 28 cases, there are 7 cases in which the means of M^c are significantly lower than the means of $M^{\min}$; 6 cases in which the means of $M^{\min}$ are significantly lower than the means of M^c and 15 cases in which there is no significant difference between the means of M^c and $M^{\min}$ (significant level=0.05). This result demonstrates that, relative to the best individual model forecasts, the combined forecasts do not exhibit superior performance. However, the test results reject the null hypothesis that $M^c = M^{\max}$ for all the cases considered, which means that the combined forecasts significantly outperform the worst single forecasts. This verifies the findings discussed earlier in the paper.

The test results of the hypothesis $M^c = M^{mean}$ are shown in Table 5. The p-values demonstrate that the null hypothesis is rejected for all forecasting horizons and combination methods. This means that accuracy of the forecasts obtained by combining the forecasts generated by individual models is significantly higher than the average accuracy of these single model forecasts in all cases.

Another question one may ask is whether there is any difference in accuracy among the three combination methods. Statistical tests are carried out to determine whether there is any significant difference between the simple average method, variance-covariance (b) method and discounted MSFE (β=0.6 (b)) method and the results are given in Table 6. In the one-step-ahead forecasting case, the performance of the three combination methods is significantly different. The discounted MSFE method turns out to be the best and the variance-covariance method the worst. In the four-step-ahead case, the same is true. For two-step-ahead forecasting, however, no significant difference was found between the discounted MSFE and simple average methods, whereas the variance-covariance method performed significantly worse than the other two combination methods. For eight-step-ahead forecasting, the discounted MSFE method is more accurate than the other two

methods which are in turn not significantly different from each other. Overall, the accuracy of the variance-covariance method combination forecasts is the lowest and the accuracy of the discounted MSFE method combination forecasts is the highest.

The forecasting performance of the two, three and four single model combinations is also compared (see Table 7). The "Mean difference" columns in Table 7 indicate the accuracy difference between two, three and four single model combinations. It is shown that all the MAPE differences between two and three model combinations are positive, which suggests that the three-model combinations are more accurate than the two model combinations. Similarly, it is found that the differences between the two model and four model combinations are positive in 96% of cases and between the three model and four model combinations are positive in 86% of cases. These results lead to the conclusion that the forecast accuracy of two, three and four model combinations increases gradually as the number of models rises. This finding is consistent with the empirical study by Makridakis and Winkler (1983), which indicates that average forecasting accuracy improves along with the number of models included in the combination set.

This study also employs the t test to examine the hypothesis that the means of the combined forecasts of two, three and four single model forecasts are equal. The t test results in Table 7 indicate that this hypothesis cannot be rejected for virtually all combination methods and all forecasting horizons (99% of cases). This implies that, at least in this study, three and four model combinations are not statistically significantly more accurate than two model combinations and four model combinations are not statistically significantly more accurate than three model combinations. Due to the limitation of this study that only four single forecast models are included, this conclusion need to be treated with caution. Whether an "optimal" number of single forecasts to be included in the combinations exists or not is a matter for future research.

5. Concluding Remarks

The empirical study in this paper demonstrates that forecasting combination does not always improve forecasting performance, as only around 50% of combined forecasts outperform the best single model. This result is in line with Hibon and Evgeniou (2005). However, this percentage is higher for long-term forecasts (eight quarters ahead), suggesting that forecast combination may be more beneficial for long-run tourism forecasting. It is also found that almost all the combination forecasts outperform the worst single forecasts over all forecasting horizons, which implies that forecast combination does decrease the risk of complete forecasting failure.

A key contribution of this study is that the forecasting performance of the combination and single model forecasts based on the minimum, maximum and mean values of the MAPEs is compared statistically. The results show that not only are the combined forecasts significantly more accurate than the worst single model forecasts across all forecasting horizons and for all combination methods, but they are also significantly better than the average single model forecasts in all cases.

The implication of the above results for tourism practitioners is to provide a strong recommendation for forecast combination in order to improve forecasting accuracy if the magnitude of the tourism demand forecasting errors is the main concern of the decision makers, especially when the demand forecasts are used to assess the feasibility of long-term investment in tourism related infrastructures.

This study also compares the forecasting performance among the three combination methods in terms of statistically significant differences in forecasting accuracy. The results show that the variance-covariance method exhibits the worst performance and the discounted MSFE method the best. However, although the simple average method is less accurate than the discounted MSFE method, it may still be worthwhile to use the simple average combination sometimes as this combination method is easy to implement and requires fewer observations.

It is also shown that forecast accuracy does appear to increase gradually as the number of models included in the forecast combination increases – for example, three model combinations have lower MAPEs than two model combinations. However, when statistically significant differences are examined combinations with higher numbers of models do not outperform combinations with lower numbers. The implications of these empirical results are that combining forecasts reduces average forecast error, but it is not clear how many models to include in the combination to achieve an optimal result.

One possible area for further research is to include more forecasting models in the forecast combination evaluation. In this paper, only four models (three econometric and one time series) are considered when the accuracy of combination forecasts is examined. However, according to Granger and Newbold (1986, p. 273), forecast combination is expected to be most profitable when the individual forecasts are very dissimilar in nature. Attempts could therefore be made to combine the forecasts generated from, for example, artificial intelligence models and judgmental/expert predictions in addition to the extrapolation and causal models already considered. A second possible area for further research is to examine whether there exists an optimal number of single model forecasts to be included in forecast combinations. A third possible area for further research is to examine the effect of forecasting horizon on the improvement in accuracy resulting from combining forecasts.

References

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *Operational Research Quarterly,* 20, 451-468.

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliograghy. *International Journal of Forecasting,* 5, 559-583.

Clemen, R. T., & Winkler, R. L. (1986). Combining economic forecasts. *Journal of Business & Economic Statistics*, 4, 39-46.

Diebold, F. X., & Pauly, P. (1987). Structural change and the combination of forecasts. *Journal of Forecasting*, 6, 21-40.

Diebold, F. X., & Pauly, P. (1990). The use of prior information in forecast combination. *International Journal of Forecasting,* 6, 503-508.

Engle, R. F., & Granger, W. J. C. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica,* 55, 251-276.

Fang, Y. (2003). Forecasting combination and encompassing tests. *International Journal of Forecasting,* 19, 87-94.

Granger, C.W. J. & Newbold, P. (1986). Forecasting Economic Time Series, 2nd Edition, p273, Academic Press.

Hibon, M., & Evgeniou, T. (2005). To combine or not to combine: Selecting among forecasts and their combinations. *International Journal of Forecasting,* 21, 15-24.

Hylleberg, S., Engle, R. F., Granger, C. W., J., & Yoo, B. S. (1990). Seasonal Integration and Cointegration. *Journal of Econometrics,* 44, 215-238.

Kulendran, N., & Witt, S. F. (2003). Forecasting the Demand for International Business Tourism. *Journal of Travel Research,* 41, 265-271.

Koning, A. J., Franses, P. H., Hibon, M., & Stekler, H. O. (2005). The M3 competition: Statistical tests of the results. *International Journal of Forecasting,* 21, 397-409.

Kulendran, N., & King, M. L. (1997). Forecasting international quarterly tourist flows Using error-correction and time-series models. *International Journal of Forecasting,* 13, 319-327.

Law, R., and Au, N. (1999). A neural network model to forecast Japanese demand for travel to Hong Kong. *Tourism Management,* 20, 89-97.

Lawrence, M. J., Edmundson, R. H., & O'connor, M. J. (1986). The accuracy of combining judgmental and statistical forecasts. *Management Science,* 32, 1521-1532.

Li, G., Song, H., & Witt, S. F. (2004). Modeling tourism demand: A dynamic linear AIDS approach. *Journal of Travel Research,* 43, 141-150.

Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research,* 44, 82-99.

Lobo, G. J. (1992). Analysis and comparison of financial analysts', time series, and combined forecasts of annual earnings. *Journal of Business Research*, 24, 269-280.

Makridakis, S., & Winkler, R. L. (1983). Averages of forecasts: Some empirical results. *Management Science,* 29, 987-996.

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

Oh, C., & Morzuch, B. J. (2005). Evaluating time-series models to forecast the demand for tourism in Singapore: Comparing within-sample and postsample results. *Journal of Travel Research,* 43, 404-413.

Smeral, E., & Wüger, M. (2005). Does complexity matter? Methods for improving forecasting accuracy in tourism: The case of Austria. *Journal of Travel Research,* 44, 100-110.

Song, H., & Witt, S. F. (2003). Tourism forecasting: The general-to-specific approach. *Journal of Travel Research,* 42, 65-74.

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

Song, H., Romilly, P., & Liu, X. (2000). An empirical study of outbound tourism demand in the UK. *Applied Economics,* 32, 611-624.

Song, H., Wong, K. K. F., & Chon, K. K. S. (2003). Modelling and forecasting the demand for Hong Kong tourism. *International Journal of Hospitality Management,* 22, 435-451.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting,* 23, 405-430.

Walz, D. T., & Walz, D. B. (1989). Combining forecasts: Multiple regression versus a Bayesian approach. *Decision sciences,* 20, 77-89.

Winkler, R. L., & Makridakis, S. (1983). The combination of forecasts. *Journal of the Royal Statistical Society. Series A*, 146, 150-157.

Witt, S. F. & Song, H. (2001). Forecasting future tourism flows. In Lockwood, A. and Medlik, S. (eds.) *Tourism and Hospitality in the 21st Century*, Butterworth-Heinemann, Oxford, 106-118.

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

Wong, K. K. F., Song, H., Witt, S. F. and Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management*, 28, 1068-1078.

Table 1: MAPEs of Single and Combined Forecasts of Tourist Arrivals

		China				Taiwan			
		1 Step	2 Step	4 Step	8 Step	1 Step	2 Step	4 Step	8 Step
	A^1	6.55	9.29	14.45	22.90	4.37	6.73	8.21	12.28
	A^2	8.33	12.05	18.91	30.06	4.49	6.65	8.05	16.77
	E	8.28	11.46	15.59	20.16	6.96	11.16	16.25	27.07
	V	6.90	9.00	10.60	18.31	11.51	15.90	17.78	27.50
	A^1A^2	5.80*	8.53*	9.61*	9.95*	3.77*	5.48*	6.94*	11.98*
	A^1E	6.04*	8.65*	10.18*	11.26*	5.38	8.44	10.57	16.68
	A^1V	6.61	8.97*	12.53	19.89	6.99	10.69	12.59	18.30
	A^2E	8.22*	11.59	17.25	25.11	5.10	7.01	8.81	13.48*
Simple	A^2V	6.10*	8.94*	10.72	14.96*	6.79	9.44	12.51	21.24
Average	EV	6.57*	8.97*	10.17*	13.64*	8.37	12.64	15.78*	20.11*
Combination	A^1A^2E	6.05*	9.11*	10.43*	11.88*	4.61	6.81	8.32	12.09*
	A^1A^2V	5.76*	8.30*	9.86*	11.68*	5.43	7.87	10.44	16.47
	A^1EV	6.08*	8.44*	10.31*	12.74*	6.62	10.38	12.65	16.54
	A^2EV	6.61*	9.70	12.10	15.58*	6.44	9.24	11.50	15.95*
	A^1A^2EV	5.87*	8.57*	9.68*	11.35*	5.59	8.39	10.56	14.66
	A^1A^2	6.55*	8.62*	14.45*	11.58*	3.93*	5.69*	7.13*	11.70*
	A^1E	6.55*	8.44*	13.94*	20.16*	5.34	7.74	8.81	12.64
	A^1V	6.55*	9.29	13.21	21.50	4.42	6.73*	8.21*	12.28*
	A^2E	8.33	11.59	15.59*	20.16*	4.84	6.39	7.63*	11.73*
Variance-	A^2V	8.33	12.05	9.93*	22.05	4.74	6.79	9.91	20.00
covariance	EV	7.53	10.71	14.59	20.16	7.17	11.82	15.78*	20.75*
combination	A^1A^2E	6.55*	8.44*	13.94*	20.16*	4.29*	7.40	7.41*	11.25*
(a)	A^1A^2V	6.55*	8.62*	13.21	12.72*	3.93*	5.69*	7.13*	11.70*
	A^1EV	6.55*	8.44*	14.07	20.16	5.34	7.74	8.81	13.95
	A^2EV	8.33	11.59	14.59	20.16	4.98	7.81	9.46	15.81*
	A^1A^2EV	6.55*	8.44*	14.07	20.16	4.29*	7.40	7.41*	11.66*
	A^1A^2	7.52	8.89*	13.19*	11.40*	3.92*	6.15*	7.62*	11.51*
	A^1E	7.80	8.33*	9.62*	12.98*	5.22	7.61	8.98	14.02
	A^1V	6.87	10.02**	13.33	21.10	4.43	6.73*	8.21*	12.28*
	A^2E	8.54**	11.45*	15.59*	20.16*	5.11	6.92	7.80*	14.40*
Variance-	A^2V	7.90	8.93*	10.27*	17.44*	4.71	6.55*	9.71	18.83
covariance	EV	7.19	8.42*	11.05	13.96*	7.30	11.38	16.20*	21.93*
combination	A^1A^2E	7.55	8.62*	9.70*	12.85*	4.36*	7.30	7.96*	13.64
(b)	A^1A^2V	7.52	8.81*	12.97	11.49*	3.92*	6.15*	7.62*	11.51*
	A^1EV	7.81	8.24*	9.78*	12.94*	5.22	7.74	8.98	16.66
	A^2EV	7.94	8.59*	11.06	14.06*	5.30	7.30	9.08	16.16*
	A^1A^2EV	7.55	8.46*	9.84*	12.76*	4.36*	7.38	7.96*	13.65
	A^1A^2	5.74*	8.45*	9.98*	9.92*	3.83*	5.48*	6.97*	11.81*
	A^1E	5.95*	8.61*	10.81*	13.93*	5.45	8.39	9.90	12.84
	A^1V	6.59	8.96*	12.64	20.18	5.70	9.04	10.86	16.40
	A^2E	8.23*	11.59	16.38	21.77	5.04	6.75	8.02*	11.48*
Discount	A^2V	6.24*	9.32	10.31*	15.63*	5.42	7.18	10.61	20.02
MSFE (a)	EV	6.58*	9.16	12.08	16.00*	7.46	11.78	15.74*	20.65*

$\beta=0.9$	A^1A^2E	5.95*	8.95*	11.41*	16.00*	4.52	6.65	7.78*	11.26*
	A^1A^2V	5.69*	8.30*	10.18*	11.38*	4.63	6.13*	8.76	14.61
	A^1EV	6.02*	8.46*	9.98*	12.43*	6.12	9.34	11.47	15.67
	A^2EV	6.85*	10.02	13.09	18.03*	5.42	7.79	9.88	16.04*
	A^1A^2EV	5.78*	8.59*	10.47*	13.69*	4.90	7.34	9.12	13.92
	A^1A^2	5.76*	8.02*	9.07*	9.83*	3.87*	5.64*	7.12*	11.74*
	A^1E	6.04*	8.17*	8.87*	12.14*	5.25	7.85	9.41	13.66
	A^1V	6.68	9.14	12.49	20.11	4.95	7.92	10.44	15.70
	A^2E	8.25*	11.60	16.47	21.87	5.03	6.34*	7.27*	12.40*
Discount	A^2V	6.20*	8.49*	9.42*	13.51*	5.13	6.59*	9.94	19.54
MSFE (b)	EV	6.60*	8.44*	10.10*	13.09*	7.47	11.69	16.31	21.60*
$\beta=0.9$	A^1A^2E	5.94*	8.52*	10.28*	14.84*	4.46	6.26*	7.26*	11.66*
	A^1A^2V	5.76*	7.78*	9.52*	10.12*	4.45	5.76*	8.57	13.71
	A^1EV	6.16*	7.99*	8.10*	9.42*	5.61	8.61	10.79	14.56
	A^2EV	6.83*	9.28	11.55	15.38*	5.41	6.81	8.83	13.61*
	A^1A^2EV	5.83*	8.02*	8.97*	11.61*	4.78	6.64*	8.45	12.84

Notes: 1) A^1, A^2, E and V denote ADL model, ARIMA model, ECM and VAR model, respectively.

2) * denotes that the combined forecast is at least as good as the corresponding best single forecast.

3) ** denotes that the combination forecast is inferior to the worst single forecast.

4) The details of $\beta=0.6$ is not reported in Table 1 due to space limitations.

5) Sample period for the calculation of MAPEs: 1999q3-2003q1

Table1   (continued)

		Japan				USA			
		1 Step	**2 Step**	**4 Step**	**8 Step**	**1 Step**	**2 Step**	**4 Step**	**8 Step**
	A^1	6.10	9.56	10.10	16.81	7.07	7.67	10.45	21.19
	A^2	8.60	8.23	10.12	12.99	7.99	8.74	10.47	13.31
	E	6.64	7.07	8.31	15.25	6.64	6.89	9.55	27.78
	V	5.37	6.63	8.68	14.97	6.39	7.11	8.11	10.40
	A^1A^2	5.91*	6.40*	6.74*	7.87*	6.95*	6.46*	8.11*	9.31*
	A^1E	5.60*	6.13*	5.99*	10.92*	6.31*	6.39*	8.48*	22.88
	A^1V	5.43	7.38	8.36*	14.54*	6.32*	6.74*	8.52	12.52
	A^2E	5.85*	7.63	8.37	8.33*	6.58*	6.68*	6.42*	9.60*
Simple	A^2V	6.25	6.84	8.40*	10.48*	6.15*	7.19	8.31	8.83*
Average	EV	5.73	6.48*	7.60*	11.31*	6.20*	6.76*	7.89*	15.57
Combination	A^1A^2E	5.25*	5.76*	5.80*	7.54*	6.26*	5.83*	6.60*	11.78*
	A^1A^2V	5.53	5.95*	7.06*	9.82*	6.15*	6.36*	7.38*	9.10*
	A^1EV	5.47	5.90*	6.48*	11.15*	6.19*	6.36*	7.86*	16.38
	A^2EV	5.60	6.90	8.01*	8.95*	6.22*	6.56*	6.75*	8.82*
	A^1A^2EV	5.16*	5.80*	6.32*	9.24*	6.10*	6.09*	6.82*	10.10*
	A^1A^2	8.60	6.27*	10.12	12.99*	7.11	8.74	10.47	13.31*
	A^1E	6.27	7.07*	8.31*	15.25*	6.38*	7.01	10.45	21.19*
	A^1V	5.30*	8.30	8.23*	14.97*	6.35*	6.80*	7.87*	10.56
	A^2E	8.60	7.55	9.46	15.25	6.59*	8.01	9.77	13.31*
Variance-	A^2V	8.60	8.23	10.12	12.99*	6.38*	8.31	9.41	13.31
Covariance	EV	6.00	7.07	8.31*	15.25	6.33*	6.71*	8.11*	9.89*
Combination	A^1A^2E	8.60	7.55	9.46	15.25	6.59*	8.01	9.77	13.31*
(a)	A^1A^2V	8.60	6.27*	10.12	12.99*	6.40	8.31	9.41	13.31
	A^1EV	6.00	7.07	8.31*	15.25	6.31*	6.77*	7.87*	9.89*
	A^2EV	8.60	7.55	9.46	15.25	6.50	8.04	9.41	13.31
	A^1A^2EV	8.60	7.55	9.46	15.25	6.50	8.04	9.41	13.31
	A^1A^2	7.80	6.92*	9.59*	14.45	7.06*	8.10	10.25*	13.06*
	A^1E	6.31	7.58	8.69	15.36	6.43*	6.73*	9.96	20.27*
	A^1V	5.34*	7.35	7.12*	14.99	6.52	6.74*	8.17	11.15
	A^2E	8.24	7.25	9.70	11.02*	6.65	7.69	9.07*	9.59*
Variance-	A^2V	8.00	8.20	10.20**	14.28	6.14*	8.07	9.34	10.10*
Covariance	EV	5.91	7.11**	9.08**	14.98	6.83**	7.29**	8.09*	10.84
Combination	A^1A^2E	8.26	7.37	9.30	11.52*	6.65	7.88	9.28*	10.03*
(b)	A^1A^2V	8.04	6.85	9.59	14.45	6.38*	8.03	9.27	11.39
	A^1EV	5.70	7.58	8.80	15.69	6.83	6.85*	8.13	11.11
	A^2EV	8.43	7.30	9.65	11.03*	6.58	8.14	9.02	8.92*
	A^1A^2EV	8.39	7.37	9.30	11.52*	6.60	8.17	9.03	9.83*
	A^1A^2	6.19	6.27*	6.61*	10.10*	7.01*	7.43*	9.30*	12.33*
	A^1E	5.60*	6.05*	5.75*	13.11*	6.32*	6.44*	8.90*	22.28
	A^1V	5.43	7.44	8.35*	14.27*	6.31*	6.71*	8.31	10.46
	A^2E	6.09*	7.63	8.39	8.38*	6.59*	7.53	9.40*	12.23*
Discount	A^2V	6.53	6.91	8.59*	10.50*	6.19*	7.99	9.43	9.69*
MSFE (a)	EV	5.73	6.50*	7.64*	11.05*	6.20*	6.75*	7.82*	9.91*
β =0.9	A^1A^2E	5.44*	5.78*	6.35*	7.73*	6.29*	6.71*	8.52*	11.31*

	A^1A^2V	5.75	5.93*	6.87*	9.67*	6.15*	7.02*	8.66	9.19*
	A^1EV	5.48	5.88*	6.27*	10.56*	6.19*	6.35*	8.00*	10.23*
	A^2EV	5.69	6.95	8.10*	8.41*	6.23*	7.20	8.69	9.03*
	A^1A^2EV	5.31*	5.81*	6.55*	8.22*	6.11*	6.51*	8.07*	8.55*
Discount MSFE (b) β =0.9	A^1A^2	6.27	6.41*	6.62*	11.25*	7.05*	7.21*	9.33*	12.06*
	A^1E	5.65*	6.15*	5.82*	13.18*	6.30*	6.50*	8.39*	21.20
	A^1V	5.42	7.50	8.21*	14.25*	6.26*	6.75*	8.29	10.86
	A^2E	6.25*	7.57	8.50	8.55*	6.59*	7.27	8.82*	10.97*
	A^2V	6.53	6.91	8.60*	10.97*	6.13*	7.79	9.27	8.99*
	EV	5.74	6.48*	7.78*	11.59*	6.27*	6.83*	7.68*	9.66*
	A^1A^2E	5.70*	5.92*	6.49*	8.13*	6.30*	6.39*	7.92*	10.12*
	A^1A^2V	5.83	6.04*	6.85*	10.58*	6.14*	6.87*	8.43	9.23*
	A^1EV	5.52	5.98*	6.35*	11.58*	6.19*	6.45*	7.82*	10.41
	A^2EV	5.94	6.92	8.21*	8.80*	6.22*	7.00	8.21	8.43*
	A^1A^2EV	5.50	5.93*	6.67*	9.10*	6.10*	6.45*	7.53*	8.63*

Table1   (continued)

		Macau				Korea			
		1 Step	2 Step	4 Step	8 Step	1 Step	2 Step	4 Step	8 Step
	A^1	13.71	11.93	12.09	12.21	10.92	12.27	13.35	17.08
	A^2	7.86	11.03	11.43	16.54	7.41	8.65	16.48	31.91
	E	13.44	8.45	14.82	21.16	9.80	11.77	15.76	14.53
	V	8.85	7.44	8.80	20.90	9.66	10.28	15.80	35.71
	A^1A^2	7.92	9.08*	9.35*	11.27*	8.62	9.35	12.13*	22.30
	A^1E	11.38*	8.66	11.76*	12.49	9.87	11.74*	14.12	13.56*
	A^1V	10.67	8.78	8.61*	13.59	9.79	8.73*	11.55*	23.82
	A^2E	7.81*	8.05*	10.99*	15.55*	7.06*	7.91*	13.36*	17.26
Simple	A^2V	6.03*	6.46*	8.10*	17.18	7.91	7.74*	14.51*	29.75*
Average	EV	10.32	6.43*	9.25	18.33*	9.01*	8.40*	12.32*	20.72
Combination	A^1A^2E	8.08	7.83*	9.76*	11.81*	8.23	9.11	12.15*	16.98
	A^1A^2V	7.52*	7.51	7.87*	13.23	8.55	7.55*	12.15*	23.52
	A^1EV	10.04	7.59	8.78*	14.22	9.23*	9.14*	12.05*	18.51
	A^2EV	6.96*	6.24*	8.79*	16.87	7.57	6.93*	12.41*	20.83
	A^1A^2EV	7.59*	6.85*	8.63*	13.82	8.16	7.85*	11.56*	19.29
	A^1A^2	6.95*	9.02*	11.43*	16.54	10.92	12.27	13.35*	17.08*
	A^1E	11.40*	8.20*	14.70	20.72	10.92	12.27	13.35*	17.08
	A^1V	12.73	11.93	12.09	12.21*	10.11	12.27	13.35*	17.08*
	A^2E	7.27*	8.88	13.70	19.04	7.00*	8.63*	14.38*	12.25*
Variance-	A^2V	6.29*	9.74	11.43	16.54*	8.20	8.43*	16.48	31.91*
Covariance	EV	10.97	8.07	14.79	21.16	9.39*	8.41*	15.76*	13.60*
Combination	A^1A^2E	6.67*	9.02	13.70	19.04	10.92	12.27	13.35*	17.08
(a)	A^1A^2V	6.29*	9.42	11.43	16.54	10.11	12.27	13.35*	17.08*
	A^1EV	10.61	8.20	14.79	20.72	10.11	12.27	13.35*	17.08
	A^2EV	6.14*	9.74	14.72	19.04	8.20	7.39*	14.38*	12.25*
	A^1A^2EV	6.14*	9.42	14.72	19.04	10.11	12.27	13.35*	17.08
	A^1A^2	7.15*	9.60*	12.64**	13.06	10.23	11.20	12.94*	17.08*
	A^1E	12.48*	8.22*	13.76	14.04	11.04**	12.39**	13.35*	17.28**
	A^1V	9.67	9.21	9.83	12.46	9.97	11.16	12.89*	17.08*
	A^2E	7.69*	8.77	11.65	17.14	7.94	9.05	14.47*	12.12*
Variance-	A^2V	6.09*	8.57	11.16	15.66*	8.51	8.77	16.38	31.62*
Covariance	EV	10.64	6.98*	10.70	17.98*	9.73	8.13*	14.07*	13.78*
Combination	A^1A^2E	7.27*	9.44	11.84	13.26	10.30	11.20	12.94*	16.78
(b)	A^1A^2V	6.28*	8.95	11.22	13.08	9.54	11.14	12.89*	17.08*
	A^1EV	9.99	6.38*	10.89	13.62	10.20	11.16	12.89*	17.77
	A^2EV	6.98*	8.66	10.36	16.86	9.25	7.78*	14.22*	12.10*
	A^1A^2EV	7.13*	9.14	10.40	13.26	9.82	11.14	12.89*	16.76
	A^1A^2	6.87*	9.00*	10.77*	14.56	9.70	11.10	11.58*	17.67
	A^1E	11.40*	8.45*	13.90	20.31	10.25	11.88	13.80	15.46
	A^1V	11.43	11.44	12.05	12.21*	9.91	10.80	12.33*	17.33
	A^2E	7.03*	8.57	11.25*	18.06	7.08*	8.21*	13.02*	13.14*
Discount	A^2V	6.73*	10.69	11.42	16.54*	8.19	8.13*	14.64*	29.13*
MSFE (a)	EV	10.96	8.09	14.81	21.16	9.19*	8.40*	12.93*	13.93*
β =0.9	A^1A^2E	6.68*	8.11*	10.92*	17.49	9.34	10.86	12.27*	16.07

	A^1A^2V	6.41*	8.89	10.77	14.56	9.23	9.90	11.07*	17.83
	A^1EV	10.46	8.30	13.90	20.31	9.52*	10.66	12.99*	15.76
	A^2EV	6.36*	8.44	11.25	18.06	7.91	7.01*	12.13*	14.42*
	A^1A^2EV	6.63*	8.03	10.91	17.49	8.98	9.91	11.72*	16.25
	A^1A^2	6.83*	9.77*	11.33*	11.63*	9.10	10.10	10.71*	18.63
	A^1E	12.09*	8.51	13.07	13.83	10.21	12.17	13.98	14.88
	A^1V	10.29	9.18	9.55	11.41*	9.85	9.79*	11.53*	18.25
	A^2E	7.35*	8.77	10.45*	15.71*	7.21*	8.15*	11.98*	12.55*
Discount	A^2V	5.93*	9.01	10.51	15.29*	8.29	8.10*	14.77*	29.16*
MSFE (b)	EV	10.23	6.80*	10.78	15.81*	9.26*	8.02*	12.25*	13.41*
β =0.9	A^1A^2E	7.32*	8.81	10.84*	12.84	8.86	9.85	11.80*	15.14
	A^1A^2V	6.58*	9.01	10.33	11.49*	8.91	8.73	10.02*	19.20
	A^1EV	10.01	7.16*	10.27	12.73	9.50*	10.03*	12.56*	15.11
	A^2EV	6.66*	7.53	9.30	14.54*	7.96	7.13*	11.39*	13.56*
	A^1A^2EV	7.07*	7.77	9.48	12.28	8.63	8.76	11.07*	15.71

Table1　(continued)

		Singapore				UK			
		1 Step	2 Step	4 Step	8 Step	1 Step	2 Step	4 Step	8 Step
	A^1	10.63	10.33	15.12	31.26	5.31	5.01	11.48	29.27
	A^2	9.00	10.84	14.55	22.36	5.23	6.31	6.14	11.49
	E	8.99	9.61	13.59	22.82	5.69	5.19	8.13	19.57
	V	11.68	11.26	10.92	31.89	4.55	4.37	6.41	9.59
	A^1A^2	9.61	9.93*	14.23*	21.50*	4.93*	4.88*	6.94	12.25
	A^1E	9.52	9.91	14.35	22.77*	5.50	4.83*	9.81	24.42
	A^1V	9.74*	9.00*	9.86*	22.84*	4.62	4.05*	7.35	15.67
	A^2E	8.68*	9.51*	13.32*	20.71*	5.11*	4.53*	5.27*	8.87*
Simple	A^2V	9.61	9.48*	10.69*	20.43*	4.54*	4.75	6.08*	8.17*
Average	EV	8.98*	8.86*	8.72*	22.92	4.81	4.23*	5.56*	10.91
Combination	A^1A^2E	9.21	9.64	13.79	20.17*	5.00*	4.39*	7.02	13.92
	A^1A^2V	9.35	8.94*	11.41	17.99*	4.44*	4.20*	5.88*	8.98*
	A^1EV	9.30	8.61*	10.46*	19.88*	4.97	4.20*	7.44	16.97
	A^2EV	8.80*	8.73*	10.65*	19.85*	4.62	4.10*	5.04*	7.27*
	A^1A^2EV	9.12	8.54*	11.46	17.62*	4.71	3.99*	6.19	10.94
	A^1A^2	9.69	10.58	14.43*	22.36*	4.94*	4.82*	5.75*	11.49*
	A^1E	9.28	9.87	14.20	21.70*	5.31*	5.01*	11.39	29.27
	A^1V	10.35*	8.62*	9.42*	31.89	5.02	4.06*	6.41*	9.59*
	A^2E	8.99*	10.11	13.39*	20.47*	5.23*	4.84*	5.91*	11.49*
Variance-	A^2V	8.75*	10.84*	12.71	19.50*	4.71	4.67	6.06*	9.59*
Covariance	EV	8.64*	8.55*	9.15*	23.18	4.51*	4.15*	6.41*	9.59*
Combination	A^1A^2E	9.28	9.71	14.20	21.70*	4.94*	4.82*	5.75*	11.49*
(a)	A^1A^2V	9.65	10.58	12.78	19.50*	4.75	3.97*	6.06*	9.59*
	A^1EV	9.23	9.87	11.86	23.10	5.02	4.00*	6.41*	9.59*
	A^2EV	8.64*	10.11	10.01*	23.18	4.71	4.06*	6.06*	9.59*
	A^1A^2EV	9.23	9.71	11.86	23.10	4.75	3.97*	6.06*	9.59*
	A^1A^2	9.61	10.40	14.72	22.81	4.77*	5.10	5.92*	9.45*
	A^1E	9.17	9.74	14.22	21.72*	5.31*	5.02	9.51	25.30
	A^1V	10.38*	8.70*	9.86*	33.95**	4.49*	4.09*	6.38*	9.84
	A^2E	8.81*	10.33	13.81	21.35*	5.26	5.27	5.66*	10.66*
Variance-	A^2V	8.81*	10.75*	12.46	22.66	4.64	4.73	6.28	9.01*
Covariance	EV	8.78*	9.04*	9.45*	27.24	4.48*	4.25*	6.46	9.80
Combination	A^1A^2E	8.95*	9.98	13.57*	20.20*	4.77*	5.10	5.67*	9.35*
(b)	A^1A^2V	9.77	10.35	12.60	23.00	4.49*	4.31*	6.24	9.19*
	A^1EV	9.08	9.14*	10.60*	19.34*	4.49*	4.13*	6.40*	9.84
	A^2EV	8.78*	10.27	10.13*	25.41	4.67	4.48	6.07*	9.20*
	A^1A^2EV	9.07	9.87	10.60*	18.72*	4.49*	4.34*	6.08*	9.25*
	A^1A^2	9.65	10.01*	14.38*	20.43*	4.94*	4.92*	6.56	10.50*
	A^1E	9.37	9.85	14.07	21.14*	5.42	4.83*	9.78	25.11
	A^1V	10.35*	8.80*	9.62*	24.38*	4.67	4.01*	6.47	10.70
	A^2E	8.68*	9.54*	13.38*	20.54*	5.11*	4.68*	5.15*	7.40*
Discount	A^2V	8.78*	9.94*	11.91	19.70*	4.59	4.74	6.07*	8.23*
MSFE (a)	EV	8.58*	8.65*	9.71*	23.34	4.51*	4.15*	5.41*	8.72*
β =0.9	A^1A^2E	9.14	9.59*	13.53*	19.77*	4.97*	4.49*	6.60	11.83

	A^{1}A^{2}V	9.58	9.43*	12.30	18.02*	4.49*	4.24*	5.44*	7.44*
	A^{1}EV	9.32	8.65*	10.76*	21.62*	4.86	4.08*	6.79	11.68
	A^{2}EV	8.60*	9.10*	11.81	19.03*	4.49*	4.10*	5.17*	7.39*
	A^{1}A^{2}EV	9.12	8.90*	12.10	18.38*	4.59	4.01*	5.68*	8.16*
	A^{1}A^2	9.58	10.14*	14.58	20.27*	4.93*	4.99*	6.48	9.14*
	A^{1}E	9.37	9.86	14.08	20.85*	5.42	4.84*	9.64	24.13
	A^{1}V	10.12*	8.83*	9.79*	28.81*	4.51*	4.03*	6.22*	9.98
	A^{2}E	8.75*	9.70	13.42*	21.17*	5.14*	4.81*	5.18*	7.86*
Discount	**A^{2}V**	8.74*	9.89*	11.36	22.85	4.56	4.73	6.16	8.52*
MSFE (b)	**EV**	8.51*	8.71*	9.44*	26.43	4.45*	4.21*	5.48*	8.79*
β =0.9	**A^{1}A^{2}E**	9.15	9.77	13.85	20.74*	4.98*	4.59*	6.45	10.80*
	A^{1}A^{2}V	9.41	9.35*	11.99	21.43*	4.42*	4.27*	5.65*	7.75*
	A^{1}EV	9.26	8.83*	10.54*	25.23	4.73	4.03*	6.59	10.93
	A^{2}EV	8.61*	9.00*	11.41	21.90*	4.46*	4.18*	5.25*	7.53*
	A^{1}A^{2}EV	9.09	8.93*	11.88	21.50*	4.54*	4.08*	5.51*	7.74*

Table1　(continued)

		Australia				Philippines			
		1 Step	2 Step	4 Step	8 Step	1 Step	2 Step	4 Step	8 Step
	A^1	7.16	9.05	12.61	15.94	8.24	9.04	10.27	16.81
	A^2	6.52	8.56	11.06	11.53	6.68	7.60	10.18	13.85
	E	7.56	9.41	9.92	10.15	9.59	9.76	12.81	19.87
	V	7.71	9.41	9.36	8.96	8.79	10.74	19.85	36.10
	A^1A^2	6.84	7.89*	11.27	12.73	7.21	8.10	9.04*	10.63*
	A^1E	6.95*	7.80*	10.15	9.37*	7.84*	8.93*	10.44	15.24*
	A^1V	7.09*	7.96*	10.75	12.03	8.28	9.80	14.00	25.24
	A^2E	6.85	8.11*	9.37*	10.01*	7.48	7.90	11.20	14.78
Simple	A^2V	7.09	8.71	9.80	9.05	7.31	8.87	12.52	21.55
Average	EV	7.51*	8.95*	8.86*	7.86*	7.95*	10.10	15.39	27.26
Combination	A^1A^2E	6.68	7.63*	10.26	9.76*	7.13	8.15	10.17*	12.54*
	A^1A^2V	6.90	7.97*	10.55	10.85	7.49	8.92	11.09	19.12
	A^1EV	6.96*	8.00*	9.80	9.10	7.80*	9.37	13.01	21.87
	A^2EV	7.03	8.39*	9.29*	8.08*	7.23	8.69	12.02	20.57
	A^1A^2EV	6.76	7.79*	9.98	9.01	7.27	8.72	11.07	18.18
	A^1A^2	6.61	8.29*	11.06*	11.53*	7.74	8.52	10.27	16.81
	A^1E	7.01*	7.73*	10.96	8.80*	7.76*	8.94*	10.36	16.81*
	A^1V	7.14*	7.80*	11.05	15.13	8.03*	9.49	10.27*	16.81*
	A^2E	6.51*	8.44*	11.06	11.53	7.62	9.49	12.81	19.87
Variance-	A^2V	6.52*	8.56*	11.06	11.53	6.78	8.39	10.18*	13.85*
covariance	EV	7.48*	9.02*	8.83*	8.45*	8.42*	9.80	12.81*	19.87*
combination	A^1A^2E	6.61	8.29*	11.06	11.53	7.76	8.94	10.36	16.81
(a)	A^1A^2V	6.61	8.29*	11.06	11.53	7.82	9.49	10.27	16.81
	A^1EV	7.02*	7.83*	10.17	8.14*	7.76*	8.94*	10.36	16.81*
	A^2EV	6.51*	8.44*	11.06	11.53	7.63	9.67	12.81	19.87
	A^1A^2EV	6.61	8.29*	11.06	11.53	7.76	8.94	10.36	16.81
	A^1A^2	7.20**	8.50*	12.43	13.41	7.88	8.58*	10.49**	17.06**
	A^1E	6.98*	8.41*	10.70	9.59*	8.12*	8.87*	10.20*	17.95
	A^1V	7.32	8.15*	10.40	11.05	8.02*	9.58*	10.27*	16.81*
	A^2E	7.25	8.71	10.54	10.80	7.09	9.43	13.41**	20.36**
Variance-	A^2V	6.61	8.51*	11.17**	11.93**	6.89	8.65*	10.57	14.09
covariance	EV	7.74**	9.26*	9.43	8.97	8.29*	9.87*	12.91	19.87*
combination	A^1A^2E	7.56	8.70	11.56	10.24	8.01	9.02	10.36	18.36
(b)	A^1A^2V	7.23	8.50*	11.43	11.81	7.95	9.36*	10.49	17.06
	A^1EV	7.10*	8.44*	10.16	9.09	8.10*	8.88*	10.20*	17.95
	A^2EV	7.30	8.71	10.72	9.74	7.18	9.45	13.37	20.50
	A^1A^2EV	7.59	8.70	11.12	9.74	8.01	8.98	10.37	18.36
	A^1A^2	6.80	7.95*	11.22	12.06	7.29	8.21	9.22*	12.79*
	A^1E	6.96*	7.82*	10.15	8.62*	7.80*	8.93*	10.41	14.77*
	A^1V	7.10*	7.82*	10.86	13.09	8.20*	9.70	11.83	19.62
	A^2E	6.77	8.06*	9.50*	10.06*	7.44	8.06	11.34	16.46
Discount	A^2V	6.89	8.62	9.93	10.06	7.24	8.83	11.95	20.14
MSFE (a)	EV	7.47*	9.01*	8.93*	8.24*	7.97*	10.05	14.53	23.34
β =0.9	A^1A^2E	6.65	7.72*	10.28	9.60*	7.14	8.27	9.87*	13.59*

	A^1A^2V	6.85	7.95*	10.60	11.02	7.48	8.91	10.37	16.48
	A^1EV	6.96*	7.93*	9.83	8.55*	7.77*	9.30	11.57	17.40
	A^2EV	6.87	8.22*	9.39	8.90*	7.21	8.73	11.57	19.54
	A^1A^2EV	6.71	7.79*	10.03	9.20	7.26	8.69	10.43	15.99
	A^1A^2	6.94	8.03*	11.29	12.17	7.35	8.19	9.30*	12.61*
	A^1E	6.94*	8.02*	10.20	8.37*	8.04*	8.92*	10.37	15.23*
	A^1V	7.19	7.87*	10.57	11.23	8.21*	9.75	11.51	19.27
	A^2E	6.89	8.18*	9.49*	9.90*	7.32	7.96	11.43	16.57
Discount	A^2V	7.03	8.71	9.90	10.05	7.25	8.73	11.62	18.49
MSFE (b)	EV	7.50*	9.00*	9.04*	8.10*	8.16*	10.09	14.42	23.07
β =0.9	A^1A^2E	6.77	7.85*	10.29	8.85*	7.14	8.17	9.93*	13.90
	A^1A^2V	7.00	7.99*	10.45	10.97	7.40	8.85	10.29	16.17
	A^1EV	6.99*	8.11*	9.84	8.69*	7.97*	9.32	11.37	17.62
	A^2EV	7.01	8.36*	9.49	8.26*	7.18	8.63	11.47	19.12
	A^1A^2EV	6.84	7.93*	10.03	8.92*	7.29	8.63	10.40	16.07

Table 2:   Percentage of combined forecasts that outperform the best individual forecasts

	One step	Two steps	Four steps	Eight steps
Simple Average	47.27	66.36	59.09	54.55
Variance-covariance (a)	45.45	43.64	45.45	52.73
Variance-covariance (b)	33.64	38.18	41.82	46.36
Discounted MSFE β =0.9 (a)	50.91	56.36	51.82	60.00
Discounted MSFE β =0.6 (a)	53.64	55.45	47.27	56.36
Discounted MSFE β =0.9 (b)	50.91	60.91	50.00	62.73
Discounted MSFE β =0.6 (b)	43.64	66.36	66.36	74.55
Average	46.49	55.32	51.69	58.18

Table 3:   Percentage of combined forecasts that outperform the worst individual forecasts

	One step	Two steps	Three steps	Four steps	Eight steps
Simple Average	0.00	0.00	0.00	0.00	0.00
Variance-covariance(a)	0.00	0.00	0.00	0.00	0.00
Variance-covariance(b)	4.55	3.64	5.45	5.45	4.55
Discounted MSFE β =0.9(a)	0.00	0.00	0.00	0.00	0.00
Discounted MSFE β =0.6 (a)	0.00	0.00	0.00	0.00	0.00
Discounted MSFE β =0.9 (b)	0.00	0.00	0.00	0.00	0.00
Discounted MSFE β =0.6 (b)	0.00	0.91	0.91	0.00	0.00

Table 4:   Results of Wilcoxon matched-pairs signed rank test: difference between MAPEs of combined forecasts and minimum values of MAPEs of single forecasts

Combination method	one step			two steps			four steps			eight steps		
	D	z	p	D	z	p	D	z	p	D	z	p
Simple Average	-0.14	-1.38	0.17	0.17	-3.00	0.00*	0.19	-1.51	0.13	0.04	-0.18	0.85
Variance-covariance(a)	-0.41	-3.38	0.00*	-0.46	-3.34	0.00*	-0.83	-4.99	0.00*	-0.92	-3.86	0.00*
Variance-covariance(b)	-0.49	-5.04	0.00*	-0.32	-3.55	0.00*	-0.35	-3.39	0.00*	0.06	-0.48	0.63
Discounted MSFE β =0.9(a)	-0.09	-0.62	0.54	-0.04	-0.39	0.70	-0.09	-0.44	0.66	0.53	-1.78	0.07
Discounted MSFE β =0.6 (a)	-0.09	0.00	1.00	0.00	-1.23	0.22	-0.16	-0.69	0.49	0.19	-0.84	0.40
Discounted MSFE β =0.9 (b)	-0.07	-0.94	0.35	0.15	-2.43	0.02*	0.32	-1.26	0.21	0.98	-3.24	0.00*
Discounted MSFE β =0.6 (b)	-0.05	-1.19	0.24	0.26	-3.47	0.00*	0.74	-4.11	0.00*	1.98	-5.91	0.00*

Notes: 1) D denotes the difference between the means of MAPEs from the combined and
single forecasts.
2) z denotes Wilcoxon signed rank statistics.
3) p denotes the asymptotic significance level.
4) * significant at 0.05 level.

Table 5:   Results of Wilcoxon matched-pairs signed rank test: difference between MAPEs of combined forecasts and mean values of MAPEs of single forecasts

Combination method	One step			Two steps			Four steps			Eight steps		
	D	z	p	D	z	p	D	z	p	D	z	p
Simple Average	0.94	-9.02	0.00*	1.31	-9.10	0.00*	2.03	-9.05	0.00*	4.77	-9.02	0.00*
Variance-covariance(a)	0.66	-5.56	0.00*	0.68	-5.23	0.00*	1.00	-5.18	0.00*	3.81	-7.17	0.00*
Variance-covariance(b)	0.59	-5.07	0.00*	0.82	-7.17	0.00*	1.49	-7.28	0.00*	4.79	-8.48	0.00*
Discounted MSFE β =0.9(a)	0.99	-8.94	0.00*	1.10	-8.49	0.00*	1.75	-8.28	0.00*	5.26	-8.82	0.00*
Discounted MSFE β =0.6 (a)	0.99	-8.53	0.00*	1.14	-8.47	0.00*	1.67	-7.79	0.00*	4.93	-8.61	0.00*
Discounted MSFE β =0.9 (b)	1.01	-9.09	0.00*	1.29	-9.08	0.00*	2.16	-9.07	0.00*	5.71	-9.10	0.00*
Discounted MSFE β =0.6 (b)	1.02	-9.02	0.00*	1.40	-9.06	0.00*	2.58	-9.10	0.00*	6.71	-9.10	0.00*

Note: Same as Table 4.

Table 6   Accuracy comparison between three combination methods

	1 Step			2 Step			4 Step			8 Step		
	Mean Difference	t value	p-value	Mean Difference	t value	p-value	Mean Difference	t value	p-value	Mean Difference	z statistic	p-value
Simple Average-Variance-covariance(b)	-0.35	-3.70	0.00*	-0.49	-4.52	0.00*	-0.54	-3.41	0.00*	0.02	-.73(a)	0.47
Simple Average-Discounted MSFE β =0.6(b)	0.09	2.07	0.04*	0.09	1.27	0.21	0.55	5.31	0.00*	1.93	-6.49(b)	0.00*
Variance-covariance(b)-Discounted MSFE β =0.6(b)	0.44	5.41	0.00*	0.57	8.19	0.00*	1.09	9.05	0.00*	1.92	-7.22(b)	0.00*

Notes: 1) t test is used in 1-, 2- and 4- steps-ahead forecasts, whereas Wilcoxon Signed
Rank Test is used in 8 steps-ahead forecasts because of nonnormality of the series.
2) p values in 1-, 2-, and 4- steps-ahead forecasts are based on 2-tailed t statistics.
In the cases (a) and (b) of the 8-steps-ahead forecasts, the z statistics are Wilcoxon
signed rank statistics.
3) * denotes significant at 0.05 level.

Table 7 Forecasting performance among different numbers of individual forecasts

		1 Step			2 Step			4 Step			8 Step		
		Mean Difference	t-value	p-value	Mean Difference	t value	p-value	Mean Difference	t value	p-value	Mean Difference	t value	p-value
Simple	A	0.32	0.99	0.33	0.45	1.26	0.21	0.61	1.18	0.24	1.46	1.36	0.18
Average	B	0.52	0.93	0.36	0.72	1.18	0.24	0.97	1.07	0.29	2.28	1.21	0.23
Combination	C	0.20	0.39	0.70	0.27	0.48	0.64	0.36	0.45	0.66	0.81	0.52	0.60
Variance-	A	0.19	0.49	0.62	0.09	0.21	0.83	0.21	0.38	0.70	0.97	0.98	0.33
covariance	B	0.31	0.48	0.64	0.01	0.01	0.99	0.23	0.24	0.81	0.57	0.32	0.75
combination	C	0.12	0.19	0.85	-0.08	-0.11	0.91	0.02	0.02	0.99	-0.40	-0.28	0.78
Variance-	A	0.09	0.24	0.81	0.12	0.32	0.75	0.52	1.06	0.29	1.64	1.63	0.11
covariance	B	0.08	0.13	0.90	-0.08	-0.13	0.89	0.94	1.10	0.28	2.37	1.33	0.19
combination	C	-0.01	-0.01	0.99	-0.20	-0.32	0.75	0.42	0.57	0.57	0.72	0.51	0.61
Discount	A	0.36	1.04	0.30	0.41	1.11	0.27	0.58	1.13	0.26	1.53	1.57	0.12
MSFE (a)	B	0.59	0.99	0.33	0.61	0.97	0.34	0.95	1.08	0.29	2.25	1.31	0.20
β =0.9	C	0.23	0.41	0.69	0.20	0.33	0.75	0.37	0.47	0.64	0.71	0.49	0.63
Discount	A	0.34	0.93	0.36	0.44	1.23	0.22	0.56	1.04	0.30	1.51	1.57	0.12
MSFE (a)	B	0.62	1.00	0.32	0.71	1.16	0.25	1.04	1.13	0.26	2.13	1.26	0.21
β =0.6	C	0.28	0.46	0.65	0.27	0.43	0.67	0.48	0.55	0.59	0.61	0.40	0.69
Discount	A	0.32	0.93	0.35	0.44	1.27	0.21	0.69	1.42	0.16	1.76	1.72	0.09
MSFE (b)	B	0.52	0.89	0.38	0.68	1.15	0.25	1.10	1.29	0.20	2.45	1.36	0.18
β =0.9	C	0.20	0.37	0.71	0.24	0.42	0.67	0.41	0.58	0.57	0.69	0.45	0.66
Discount	A	0.32	0.95	0.34	0.47	1.37	0.18	0.84	1.80	0.08	2.03	2.16	0.03*
MSFE (b)	B	0.49	0.84	0.41	0.71	1.19	0.24	1.22	1.45	0.15	2.88	1.74	0.09
β =0.6	C	0.16	0.30	0.76	0.24	0.44	0.66	0.38	0.57	0.58	0.86	0.63	0.53

Notes: 1)A: statistical test of combination forecasts generated from two single model and three single model forecasts.

2) B: statistical test of combination forecasts generated from two single model forecasts and four single model forecasts.

3) C: statistical test of combination forecasts generated from three single model and four single model forecasts.

4) * denote significant at 0.05 level.