

An Assessment of Combining Tourism Demand Forecasts over Different Time Horizons

SHUJIE SHEN* GANG LI* AND HAIYAN SONG†¹

* University of Surrey, Guildford, UK

† The Hong Kong Polytechnic University, Hong Kong SAR, P. R. China

**Revised Version
February 2008**

¹ Corresponding author (hmsong@polyu.edu.hk). The second author would like to acknowledge the financial support of the Hong Kong UGC-CERG grant: B-Q04H.

An Assessment of Combining Tourism Demand Forecasts over Different Time Horizons

Abstract

This study investigates the performance of combination forecasts in comparison to individual forecasts. The empirical study focuses on the UK outbound leisure tourism demand for the USA. The combination forecasts are based on the competing forecasts generated from seven individual forecasting techniques. The three combination methods examined in this study are: the simple average combination method, the variance-covariance combination method and the discounted mean square forecast error method. Encompassing tests are applied to examine their effects on the forecasting performance of combination forecasts. The empirical results suggest that combination forecasts overall play an important role in the improvement of forecasting accuracy in that they are superior to the best of the individual forecasts over different forecasting horizons. Furthermore, the simple average combination method is outperformed by the other two more sophisticated combination methods. The variance-covariance combination method turns out to be the best among the three combination methods. Another finding of this study is that the encompassing test does not contribute significantly to the improved accuracy of combination forecasts. This study provides robust evidence of the efficiency of combination forecasts.

Keywords: *combination forecast; tourism demand, econometric model, forecast performance, encompassing test*

Introduction

Tourism demand forecasting plays an important role in tourism planning and management, due to the perishable nature of tourism products and the economic contribution of tourism activities to a destination's economy. The need for more accurate forecasts of tourism demand is driven by the desire to reduce risk and uncertainty (Yu and Schwartz, 2006). Much effort has been made in research on tourism demand forecasting accuracy over the past few decades. However, no single forecasting method has been found to outperform others in all situations (Li *et al.*, 2005). A new direction in tourism forecasting research has been to combine the forecasts produced by individual models, using various combination techniques. The favourable results in the general forecasting literature have inspired us to explore the usefulness of forecast combination in the tourism context. If forecasting accuracy could be improved through forecast combination, this would certainly be welcome by the decision makers in both the public and private sectors related to tourism.

The years since Bates and Granger's (1969) seminal work has seen the booming of research on forecast combination in various economics and business fields. However, the applications of forecast combination methodologies in the tourism context are still rare, with very few exceptions including Chu (1998), Fritz *et al.* (1984), Oh and Morzuch (2005) and Wong *et al.* (2007).

Chu (1998) and Fritz *et al.* (1984) both applied the variance-covariance combination method, but Chu (1998) only included two time series models in the combination. Fritz *et al.* (1984) focused on one time series model and one traditional econometric model. These two studies show empirical evidence in favour of forecast combination. Oh and Morzuch (2005) adopted the simple average combination method to combine the forecasts obtained from several competing time series models. They concluded that a combined forecast never performed less accurately than the poorest performing individual forecasts. Although Wong *et al.* (2007) combined the forecasts generated from three modern econometric models along with one time series model based on three combination methods. The latest forecasting techniques, such as the time-varying parameter (TVP) model, has not been included in their study.

Moreover, Wong *et al.* (2007) examined the combination efficiency of one-step-ahead forecasting only, and longer forecasting horizons have not been considered. Their empirical results show that the combined forecasts do not always outperform the best individual forecasts but are almost certain to outperform the worst individual forecasts.

This study aims to extend the previous studies on tourism forecast combination. Seven single forecasting techniques, including five econometric models and two time series models, are used to generate individual forecasts. The three combination methods used in the study include: the simple average combination method, the variance-covariance combination method and the discounted MSFE method. One- to four-quarter-ahead and eight-quarter-ahead individual forecasts are generated and the efficiency of forecast combination at multiple-step-ahead forecasting horizons is examined. Encompassing tests are also carried out in this study to examine whether they can contribute to the enhanced performance of forecast combination.

The empirical study of this paper is based on a dataset related to the demand for leisure tourism in the USA by the UK residents. UK is one of the five largest tourism generating countries in the world. During the period 1980-2006, the average annual growth rate of UK outbound tourism was above 5% and the total UK outbound tourists flows were almost tripled between 1980 and 2006. The substantial increase in international tourism demand in the UK is mainly due to the significant improvement in the residents' living standards together with the rapid development in communication, transportation and global economic integration. As the only non-European country on the top ten destinations for the UK tourists, USA remains the number one long-haul destination for the UK tourists. Meanwhile, according to the US Office of Travel and Tourism Industries, UK remains the third largest source market for the US inbound tourism and tourist arrivals from the UK account for over 8% of the total market share, with 60% of the visits were from the leisure travel market. Therefore, accurate forecasts of the UK outbound tourism demand for the USA plays an important role in the planning and management of the tourism businesses such as airlines in both countries.

Forecast Combination Methods

Simple Average Combination Method

A simple procedure for combining the forecasts is to take an arithmetic average of the forecasts. This procedure serves as a useful benchmark and has been shown to perform better than some complicated methods (see, for example, Makridakis and Winkler, 1983). The simple average combination method calculates the composite forecasts without taking the historical performance of the individual forecasts into account, as the combination weight is assigned equally to each of the individual forecasts. The simple average combination method can be expressed as:

$$f_{ct} = \sum_{i=1}^n f_{it} / n \quad (1)$$

where f_{ct} denotes the combined forecast; f_{it} is the i th forecast in time period t ; and n is the number of forecasts to be combined.

The simple average combination method (known as the ‘folk theorem’ in the forecast combination literature) assigns equal weight to each of the individual forecasts instead of using the optimal weights to minimize the variance of the combination forecasts. This method often works better in practice (e.g. Granger, 1989; Stock and Watson, 1999; Fildes and Ord, 2002). The reason why this is the case is that the estimation of the optimal combination weights can be very difficult in practice. Although forecast combinations with an equal weighting scheme may be biased, they could reduce the forecast error variance by not relying on estimated combination weights that depend on the second moments of forecast errors (Elliott and Timmermann, 2004). According to Palm and Zellner (1992), the advantages of simple average forecast combination are: First, the weights are known and do not have to be estimated—which is an important advantage if there is little evidence on the performance of individual forecasts or if the parameters of the model generating the forecasts are time varying; Secondly, in many situations a simple average of forecasts achieves substantial reductions in variance and bias through averaging out individual bias; and Thirdly, it is often superior to the optimal weighting scheme when the effect of sampling errors and model uncertainty on the estimates of the weights is taken into account..

Clemen (1989) drew a similar conclusion that simple average combination method has the virtues of impartiality, robustness and a good ‘track-record’ in economic and business forecasting. For this reason, it has been the common choice in many forecast combination studies. However, the shortcomings of the simple average method include the ignorance of potentially useful information about the precision of the forecasts and assumption that forecasts are exchangeable because each forecast receives the same weight (Clemen and Winkler, 1986).

Variance-Covariance Method

The variance-covariance method was proposed by Bates and Granger (1969). Normally the linear weights of individual forecasts are calculated to minimize the error variance of the combination forecasts (assuming un-biasedness for each individual forecast). The principle of the variance-covariance combination methods is illustrated using the case where two forecasting models are involved. Suppose the combined forecasts from the two unbiased forecasting models are given as:

$$\hat{y}_{t+k,t}^c = w\hat{y}_{t+k,t}^1 + (1-w)\hat{y}_{t+k,t}^2 \quad (2)$$

where $\hat{y}_{t+k,t}^c$ is the k -periods-ahead forecast combined from the k -period-ahead forecasts of $\hat{y}_{t+k,t}^1$ and $\hat{y}_{t+k,t}^2$ and w and $(1-w)$ are the weights assigned to $\hat{y}_{t+k,t}^1$ and $\hat{y}_{t+k,t}^2$, respectively. Since the weights sum to unity, the combined forecast will also be unbiased.

The error of the combined forecast is:

$$e_{t+k,t}^c = we_{t+k,t}^1 + (1-w)e_{t+k,t}^2 \quad (3)$$

and the variance of the error term is:

$$\sigma_c^2 = w^2\sigma_{11}^2 + (1-w)^2\sigma_{22}^2 + 2w(1-w)\sigma_{12} \quad (4)$$

where σ_{11}^2 and σ_{22}^2 are the unconditional forecast error variances and σ_{12} is the covariance. The weight that minimizes the combined forecast variance is

$$w^* = (\sigma_{22}^2 - \sigma_{12}) / (\sigma_{22}^2 + \sigma_{11}^2 - 2\sigma_{12}) \quad (5)$$

It can be easily seen that the forecast error variance from the optimal combination is lower than the individual variance of σ_{11}^2 or σ_{22}^2 . Therefore, combining the forecasts is beneficial. This example of two-model forecast combination can be easily extended to forecast combination of multiple-forecasting models.

In practice, $\sigma_{11}^2, \sigma_{22}^2$ and σ_{12} are unknown, but these parameters can be estimated from the data. Suppose that a set of T consecutive forecasts from model 1 is available, together with the corresponding forecasts from model 2, the forecast errors (e_{1t}, e_{2t}) , $t=1, 2, \dots, T$ can be determined. A possible estimator of the combination weight could be:

$$w_i = \frac{\sum_{t=1}^T e_{1t}^2 - \sum_{t=1}^T e_{1t} e_{2t}}{\sum_{t=1}^T e_{1t}^2 + \sum_{t=1}^T e_{2t}^2 - 2 \sum_{t=1}^T e_{1t} e_{2t}} \quad (6)$$

In fact, the combination methods that have been used in most of the practical studies are much simpler than (6). For example, Bates and Granger (1969) suggested that the weights attached to individual forecasts might be taken as inversely proportional to the mean squared forecast error in the recent past. Thus, instead of using (6) to combine the forecasts, Bates and Granger (1969) used (7), which entirely neglects the sample covariance term in (6).

$$w_i = \frac{\sum_{t=1}^T e_{it}^2}{\sum_{t=1}^T e_{1t}^2 + \sum_{t=1}^T e_{2t}^2} \quad (7)$$

Formula (7) could be easily extended to include more than two individual forecasts and the weights can be calculated, according to Fritz *et al.* (1984), by:

$$w_i = \frac{[\sum_{t=1}^T e_{it}^2]^{-1}}{\sum_{j=1}^m [\sum_{t=1}^T e_{jt}^2]^{-1}} \quad (8)$$

Discounted Mean Square Forecast Error (MSFE) Method

The discounted MSFE method was proposed by Bates and Granger (1969) for a two-individual forecast case and subsequently generalised by Newbold and Granger (1974) for n -individual-forecasts combination. The method uses the full sample but

weighs recent observations more heavily (Winkler and Makridakis, 1983; Diebold and Pauly, 1987). The combination of n -individual forecasts for period t is given as:

$$f_{ct} = \sum_{i=1}^n w_i f_{it} \quad (9)$$

where f_{it} is the forecast for period t from forecasting method i , w_i is the weight assigned to individual forecast i and n is the number of individual forecasts. Specifically, the weight of the discounted MSFE of the combined forecasts is defined as:

$$w_{it} = \phi_{it}^{-1} / \sum_{j=1}^n \phi_{jt}^{-1} \quad (10)$$

where $\phi_{it} = \sum_{s=1}^{t-1} \alpha^{t-1-s} (y_s - f_{is})^2$, α is a selected discounting factor with

$0 < \alpha \leq 1$ and y_s is the actual value for period s . A smaller value of α implies that more weight is given to the more recent observations. In practice, a few values of α close to 1 (such as 0.85, 0.90 and 0.95) are pre-selected to calculate the weights and the one that produces the most accurate combination forecasts would be selected (see Stock and Watson, 2004). When $\alpha = 1$ (no discounting) the discounted MSFE method corresponds to Bates and Granger's (1969) variance-covariance method. The difference between the two methods is that Equation (7) ignores the covariance information among the forecasting errors. Clemen and Winkler (1986) gave a justification for this: when the correlations among the forecast errors are high, the combination weights are likely to be more sensitive to changes in the correlations. To avoid the instability caused by interdependence between the combination weights and correlations in the forecasting errors, the covariance matrix should be ignored in Equation (10).

Encompassing Tests

An important procedure in forecast combination exercises is to carry out the encompassing tests. Forecast encompassing tests stem from forecast combination literature (see, *inter alia*, Bates and Granger, 1969; Clemen, 1989 and Granger, 1989). Encompassing tests seek to evaluate whether the competing forecasts may be

fruitfully combined to produce a forecast that is superior to the individual forecasts (Fang, 2003). According to Harvey and Newbold (2000), if the encompassing tests show that one forecast does not encompass its competitors, this would imply that the forecasts can be improved by linearly combining these forecasts with all the competing forecasts. Another purpose of this study is therefore to examine the effect of the encompassing tests on forecasting performance of the combined forecasts.

Harvey and Newbold (2000) proposed a multi-forecast encompassing test, based on the modified Diebold-Mariano approach, recommended by Harvey *et al.* (1998).

Regression-Based Multi-forecast Encompassing Test

The case of multi-forecast combination is discussed in Harvey and Newbold (2000). Let (f_{1t}, \dots, f_{kt}) be k competing forecasts of the actual value y_t , assuming that the forecasts are made one step ahead with non-autocorrelated errors, the test of the null-hypothesis that one forecast f_1 encompasses its competitors starts with:

$$f_{ct} = (1 - \lambda_1 - \lambda_2 - \dots - \lambda_{k-1})f_{1t} + \lambda_1 f_{2t} + \lambda_2 f_{3t} + \dots + \lambda_{k-1} f_{kt} \quad 0 \leq \lambda_i \leq 1 \quad (11)$$

which can be rewritten as:

$$e_{it} = \lambda_1(e_{1t} - e_{2t}) + \lambda_2(e_{1t} - e_{3t}) + \dots + \lambda_{k-1}(e_{1t} - e_{kt}) + \varepsilon_t \quad 0 \leq \lambda_i \leq 1 \quad (12)$$

where $e_{it} = y_t - f_{it}$ and ε_t is the error of the combined forecast. The null-hypothesis that Forecast 1 encompasses Forecast 2, ..., k is: $H_0 : \lambda_1 = \lambda_2 = \dots = \lambda_{k-1} = 0$ against the alternative that $H_1 : \lambda_1, \lambda_2, \dots, \lambda_{k-1}$ are not jointly equal to zero. Since the null-hypothesis $\lambda_1 = \lambda_2 = \dots = \lambda_{k-1} = 0$ is multi-dimensional, the F statistic should be used. This regression-based multi-forecast encompassing test can be applied to multi-step-ahead forecasts but it is subject to the problem of lacking robustness when the forecasting errors are not normally distributed (Harvey and Newbold, 2000).

Modified Diebold-Mariano-Type Multiple Encompassing Test

The modified Diebold-Mariano-Type test was proposed by Harvey *et al.* (1997, 1998) based on the test of Diebold and Mariano (1995).

According to Harvey *et al.* (1998), Regression (12) can be rewritten in the following general form:

$$y_t = X_t' \beta + \varepsilon_t, \quad y = X\beta + \varepsilon \quad (13)$$

where $y_t = e_{1t}$, $\beta = [\lambda_1 \lambda_2 \dots \lambda_{k-1}]'$ and $X_t = [(e_{1t} - e_{2t})(e_{1t} - e_{3t}) \dots (e_{1t} - e_{kt})]'$.

The null-hypothesis for multi-forecast encompassing in terms of Regression (13) is

$H_0 : \beta = 0$, or

$$H_0 : [E(X_t X_t')]^{-1} E(X_t y_t) = 0 \quad (14)$$

Equation (14) is true if and only if

$$H_0 : E(\Delta_t) = 0; \quad \Delta_t = [d_{1t} d_{2t} \dots d_{k-1,t}]'; \quad d_{it} = e_{1t}(e_{1t} - e_{i+1,t}) \quad (15)$$

The problem is now reduced to testing for the zero-mean of a vector of random variables, so the multivariate analogue of the Diebold-Mariano statistic, denoted as MS^* , takes the form of Hotelling's (1931) generalised T^2 -statistic (see, e.g. Anderson, 1958)

$$MS^* = (k-1)^{-1} (n-1)^{-1} (n-k+1) \bar{d}' \hat{V}^{-1} \bar{d} \quad (16)$$

where $\bar{d} = [\bar{d}_1 \bar{d}_2 \dots \bar{d}_{k-1}]'$, $\bar{d}_i = n^{-1} \sum d_{it}$ and \hat{V} is the sample covariance matrix. The construction of \hat{V} assumes $(h-1)$ -dependency and a corresponding rectangular kernel; as with the modified regression-based tests. The finite sample modification due to Harvey *et al.* (1997) applied directly to the sample variance (diagonal) terms of \hat{V} , and it is also straightforward to show that the same correction factor is appropriate when estimating the covariance terms. \hat{V} has (i, j) th element:

$$\hat{v}_{ij} = n^{-1} [n+1 - 2h + n^{-1}h(h-1)]^{-1} \times \left[\sum_{t=1}^n (d_{it} - \bar{d}_i)(d_{jt} - \bar{d}_j) + \sum_{m=1}^{h-1} \sum_{t=m+1}^n (d_{it-m} - \bar{d}_i)(d_{jt-m} - \bar{d}_j) + \sum_{m=1}^{h-1} \sum_{t=m+1}^n (d_{i,t-m} - \bar{d}_i)(d_{jt} - \bar{d}_j) \right] \quad (17)$$

In application the critical values of $F_{k-1, n-k+1}$ are used for statistic MS^* .

The modified Diebold-Mariano-type test is also subject to some under-sizing in smallest samples but is robust and provides a broadly reliable alternative to regression-based tests (Harvey *et al.*, 1998). The power of this test is weaker than the

F -test when the sample size is small, but the robustness of this test will increase significantly with the sample size increases.

The Data and Models

The empirical study focuses on the demand for outbound leisure tourism in USA by the UK residents. The tourism demand function can be written in the following general form:

$$TOU_t = f(Y_t, RRCP_t, RSUB_t, dummies) \quad (18)$$

where TOU_t is the UK outbound leisure tourism demand measured by quarterly tourist arrivals to the USA; Y_t is tourist income measured by real gross domestic product (GDP) of the UK in constant prices (1995=100); $RRCP_t$ represents the relative tourism price of the USA, which is calculated by dividing the price (measured by the consumer price index) in the USA (CPI_{USA}) by that of the UK (CPI_{UK}), then adjusted by the appropriate exchange rate (EX_{UK}):

$$RRCP = \frac{CPI_{USA}}{CPI_{UK} / EX_{UK}} \quad (19)$$

$RSUB_t$ represents the relative substitute price of the USA, measured by a weighted price index of the main alternative destinations relative to the tourism price of the UK, with shares of tourist arrivals in these potential substitute destinations being the weights. Three countries are chosen as substitutes to the USA. They are Canada (CA), Australia (AU) and New Zealand (NZ). The choice of these destinations is motivated by their relative significance for UK outbound tourism and the broad similarity of the tourism-related attributes among them (Divisekera, 2003). The substitute price is defined as:

$$RSUB = \frac{RRCP_{CA} \cdot TOU_{CA} + RRCP_{AU} \cdot TOU_{AU} + RRCP_{NZ} \cdot TOU_{NZ}}{TOU_{CA} + TOU_{AU} + TOU_{NZ}} \quad (20)$$

In this study, all of the above variables are transformed to logarithms, in line with most of the previous tourism demand studies.

Three dummy variables are included in the models to capture the effects of one-off events on the UK outbound tourism demand. Among them, DUM86 represents the severe decline in the world oil prices in 1986 (DUM86=1 in 1986Q2 and 1986Q3, 0 otherwise). The drastic decline in the world oil prices was due to a disagreement between the countries about how reductions in output should be allocated within the OPEC, which soon led to a collapse in the OPEC. (Trehan, 1986). The decline in the world oil prices should have a positive impact on the UK outbound tourism demand. DUM90 captures the effect of the invasion of Kuwait by Iraq in 1990 (DUM90=1 in 1990Q3 and 1990Q4, 0 otherwise). DUM91 is used to detect the effects of the Gulf War in 1991 (DUM91=1 in 1991Q1, 1991Q2 and 1991Q3, 0 otherwise). These two events may have negative effects on the UK outbound tourism demand. Seasonal dummies are also included in some of the models to capture seasonality. They are defined as $D_{i,t}$ ($i=2,3,\dots,s$), $D_{i,t}=1$ if time t corresponds to season s and $D_{i,t}=0$ otherwise ($s=4$ here as quarterly time series are used in the study) .

The data covers the period 1984Q1-2004Q4. The series on GDP, exchange rates and CPI are obtained from the *International Financial Statistics Yearbook* published by the International Monetary Fund (IMF). The tourist arrivals are obtained from the National Office of Statistics in the UK.

Individual Forecasting Models

In this study, five econometric models are used to generate individual *ex post* forecasts, and they are the reduced autoregressive distributed lag model (RE-ADLM), Wickens-Breusch error correction model (WB-ECM), Johansen Maximum Likelihood error correction model (JML-ECM), vector autoregressive model (VAR) and the TVP model.

RE-ADLM

Following the ‘general-to-specific’ modelling approach, estimation of an ADLM starts with a general specification which incorporates as many variables as possible supported by appropriate economic theory. The general model is estimated first and the least significant variable is removed from the equation. This process is repeated until all of the variables left in the equation are statistically significant.

ECMs

The ECM models were first developed by Engle and Granger (1987). ECM models are more appropriate usually when it is necessary to examine both the long-run equilibrium relationship and the short-run dynamics of the tourism demand model. Two estimation methods have been used in this study: the Wickens and Breusch (1988) one-stage approach (WB-ECM) and the Johansen (1988) maximum likelihood approach (JML-ECM). The difference between the two methods is that there might be more than one cointegration relationship identified in the JML-ECM. Detailed discussions of this method and its applications in the tourism context can be found in Kulendran and Witt (2001) and Lim and McAleer (2001).

VAR Model

Most of the traditional tourism demand models assume that the explanatory variables in the regression model are exogenous, while the VAR model is a system of equations in which all variables are treated as endogenous. It is used when the distinction between endogenous and exogenous variables is not clear or when the forecasters or practitioners are interested in the effects of policy 'shocks' on forecasting (Song and Witt, 2000).

TVP Model

Unlike traditional tourism demand analysis, which assumes the coefficients of variables to be constant over the sample period, the TVP model relaxes this restrictive assumption and can trace the behavioural changes of tourists over time. For instance, in a log-linear functional form, the TVP model can be used to analyse the evolution of tourism demand elasticities over time. The TVP model is normally specified in state space (SS) form and estimated using the Kalman filter algorithm (Kalman, 1960).

The specifications of these models follow those by Song, Witt and Jensen (2003). Due to space constraints and because single forecasting methods are not the focus of this study, the model specifications are omitted.

Time-Series Models

Two univariate time series models are included as benchmarks for comparison purpose: the seasonal Naive model and seasonal ARIMA (SARIMA) model (see Witt and Witt, 1992 and Pankratz, 1983 for model specifications).

Seasonality

As quarterly data are used in the study, it is likely that the time series exhibit seasonality. The HEGY test developed by Hylleberg *et al.* (1990) is used to test for seasonal and non-seasonal unit roots in the series. Two types of seasonality are relevant and they are the stochastic seasonality and deterministic seasonality. If the seasonal pattern of a data series evolves over time, so that “winter becomes summer,” the seasonality is stochastic. For stochastic seasonality, seasonal differencing is required. This applied to the WB-ECM and JML-ECM in this study. If seasonality is viewed as deterministic, i.e., the seasonal pattern is consistent over time, seasonal dummies should be used in estimating the models. ADLM, TVP and VAR specifications incorporate only the deterministic seasonality, in line with many previous studies. Different treatments of seasonality among various models also allow for the examination of their effects on the forecasting accuracy.

Empirical Results

Before the performance of various combination methods is evaluated, the performance of the individual forecasting approaches is examined.

Performance of Individual Forecasting Methods

The results of the HEGY test show that the UK outbound tourist arrivals series and the relative substitute price series exhibit trend and seasonality². Correspondingly, seasonal differences have been used in the WB-ECM and JML-ECM approaches before they are estimated..

The demand models are estimated based on the data from 1984Q1 to 1996Q4 and the *ex post* forecasts are generated for the period 1997Q1—2004Q4

² The results of the HEGY test are omitted here due to space constraints, but are available from the authors upon request.

³. Recursive forecasting techniques are used to generate forecasts, i.e., the models are estimated over the period 1984Q1—1996Q4 first, and the estimated models are used to forecast the tourist arrivals over the period 1997Q1—2004Q4. Subsequently the models are re-estimated using the data from 1984Q1 to 1997Q1 and forecasts are generated for the period 1997Q2—2004Q4. Such a procedure is repeated, each time with one more observation added to the estimation period, until all observations are exhausted. As a result, 32 one-quarter-ahead forecasts, 31 two-quarter-ahead forecasts, 30 three-quarter-ahead forecasts, 29 four-quarter-ahead forecasts and 25 eight-quarter-ahead forecasts are generated. As benchmarks, the seasonal naïve (no change) model and seasonal ARIMA (SARIMA) model are included in the comparison of forecasting performance. With regard to SARIMA model, a collection of SARIMA models with different orders of p , q , P , Q are estimated first, and one model is selected using information criteria such as Akaike Information Criterion (AIC) and Schwarz criterion (SC). The orders of p , q , P and Q are chosen from 0 to 2, according to Pankratz (1991), who stated that in practice, all the orders (p , d , q , P , D , Q) tend to be small, often no more than 1 or 2 (for SARIMA).

To be consistent with the previous tourism forecasting studies, the forecasting accuracy comparison is carried out based on the most frequently used error measure: the Mean Absolute Percentage Error (MAPE) (for detailed justification see Witt and Witt, 1992). The forecasting performances of the alternative models are ranked according to MAPE and the results are shown in Table 1.

(Insert Table 1 here)

The results show that for short-term forecasts (i.e. one and two quarters ahead) the WB-ECM and JML-ECM generate the most accurate forecasts, followed by the two time series models. With respect to medium-term forecasts (i.e. three and four quarters ahead), the JML-ECM is ranked top while the SARIMA performs the second best. As far as the long-run forecasts (i.e. eight quarters ahead) are concerned, the SARIMA model outperforms all of its counterparts, closely followed by the JML-

³ The model estimation results are omitted here due to space constraints but are available from the authors upon request.

ECM and the WB-ECM. The RE-ADLM, TVP and VAR models generate the least accurate forecasts across all horizons, with the VAR model being ranked bottom.

The aggregated error measures are calculated across all the forecasting horizons. Overall, the JML-ECM outperforms all the others followed by the SARIMA model. The WB-ECM and NAÏVE model share third and fourth places. The VAR model exhibits the poorest performance when forecasting the UK outbound tourism demand for the USA.

The superior performance of the JML-ECM and the seasonal time series models is likely to be associated with how the seasonality is dealt with. The specifications of the JML-ECM and the WB-ECM are based on the results of the seasonal unit root test. The seasonal Naïve model and the SARIMA models assume that there are seasonal unit roots at seasonal frequencies and this assumption is consistent with the results of the seasonal HEGY test which indicate that the UK outbound tourism series exhibits stochastic seasonality. However, the other three models use seasonal dummies to account for deterministic seasonality. According to Abeyasinghe (1994), the use of seasonal dummies in removing the seasonality in the data is likely to produce spurious regressions. Moreover, such a simplification is incapable of reflecting the dynamic nature of the seasonality inherent in the actual demand for tourism. This may explain the relatively poor forecasting performance of the RE-ADLM, VAR and TVP models. The conclusion that can be drawn from the analysis is that different assumptions about the effects of seasonality on forecasting performance of alternative models and the pre-testing seasonal unit roots can improve forecast accuracy. This conclusion is consistent with that drawn by Allyne (2006).

Performance of Combination Forecasts

Forecast combinations are carried out based on the individual forecasts generated from the seven single forecasting approaches, i.e. 32 one-quarter-ahead forecasts, 31 two-quarter-ahead forecasts, 30 three-quarter-ahead forecasts, 29 four-quarter-ahead forecasts and 25 eight-quarter-ahead forecasts. The individual out-of-sample forecasts at each horizon are divided into two sub-samples, with the first 20 observations used for the calculation of combination weights and the remainder for

the post-sample evaluation of combined forecasts. For example, for the 32 one-quarter-ahead forecasts, the weights using the data points from 1997Q1 to 2001Q4 (equivalently, forecast 1 through forecast 20) are obtained first. Then one more data point up to 2002Q1 is added, another weight is calculated from 1997Q1 to 2002Q1. This update is carried out recursively until the last weight is calculated from 1997Q1 through to 2004Q3. In line with the past studies such as Stock and Watson (2004), the discounted MSFE combination forecasts are computed using three values of the discounting factor- α : 0.95, 0.9, and 0.85.

Comparison between Combination Forecasts and Individual Forecasts

The performance of combination forecasts compared to that of the individual forecasts is the main focus of this study. All possible combinations among the seven individual models are considered. For a given number n of methods combined there are $\binom{7}{n}$ combinations, which is 21 for $n=2$ and 5; 35 for $n=3$ and 4; 7 for $n=6$ and 1 for $n=7$. So the total number of all possible combinations of the seven methods is 120. Thus, for each combination method applied, 120 combinations are carried out for each forecasting horizon.

In the forecasting performance evaluation, the percentage of superior combination forecasts is calculated. By the percentage of superior combination forecasts, we mean the proportion of the combination forecasts, which are more accurate than the best component forecasts based on the MAPE, to all of the 120 combinations. The performance of combination forecasts in comparison to that of the individual forecasts is evaluated across different forecasting horizons. Table 2 presents the percentages of superior combination forecasts across all forecasting horizons.

(Insert Table 2 here)

The results in Table 2 show that most of the combined forecasts outperform the component individual forecasts; i.e. the combined forecast is more accurate than the best component individual forecasts. For one-quarter-ahead forecasts, the combination methods outperform the best of the individual forecasts in more than 60% of cases.

For three-quarter-ahead forecasts, the variance-covariance combination method and the discounted MSFE methods generated more accurate forecasts than the most accurate individual forecasts in over 80% of cases. For longer forecasting horizons, the combination forecasts are superior to the best of individual forecasts in almost all as far as all combination methods are concerned. It indicates that the vast majority of the 120 combination forecasts generated by the three combination methods are more accurate than the best individual forecasts. It should be noted that, the superior performance of combination forecasts does not necessarily come from the combination of the best individual forecasts. For example, in the case of the simple average combination with one-quarter-ahead forecasts, 86% of the superior combination forecasts are from the combination of at least one of the worst three individual models. The formulas of the combination methods above indicate that there is no direct relationship between the accuracy of individual forecasts and that of the combined forecasts. This finding is inline with the previous studies such as Timmermann (2006).

Overall, the variance-covariance method and discounted MFSE methods outperform the best of single model forecasts in over 70% cases. The simple average method does not perform as well as the other two methods, but still generated more accurate forecasts than the best individual forecasts in over 60% of cases.

The above analysis shows that the superiority of the combination forecasts is relatively robust across different forecasting horizons except for two-quarter-ahead forecasts. The possible reason for this could be that the information included in each forecast tends to overlap (Wong *et al.*, 2007). The consistency of combination forecasting performance should be explored in future research using different datasets.

Comparison between Alternative Combination Methods

The comparison between the three combination methods is based on the percentage of superior combination forecasts as defined above. For one-quarter-ahead forecasts, the discounted MSFE method performs the best. There is little difference between the performances of the simple average method and variance-covariance method. The variance-covariance method outperforms its counterparts for two- to

four-quarter-ahead forecasts. With respect to eight-quarter-ahead forecasts, the discounted MSFE method ranked the best. The simple average method exhibits the poorest performance in most of the cases. In general, the variance-covariance method outperforms its counterparts, and the simple average method exhibits the poorest performance.

The results in this study show that more sophisticated combination methods, such as the discounted MSFE method and variance-covariance method, which take the historical performance of the individual forecasts into account, perform better than the simple average combination method. This finding is consistent with Holden and Peel (1986) and Bischoff (1989), but conflict with the findings of several other studies (e.g. Stock and Watson, 2004; Palm and Zellner, 1992). As far as the discounted MSFE method is concerned, there is no clear-cut answer to how the value of the discounting factor affects the performance of the combination method.

Comparison of Combination Forecasts with and without Encompassing Test

The regression-based F -test and the modified Diebold-Mariano-type test, MS^* , are considered in the study. Because the F -test is not robust when the forecast errors exhibit autocorrelations, which would be expected for the multi-step-ahead forecasts, it is only applied to one-quarter-ahead forecasts. The MS^* test is robust to conditional heteroscedasticity in the regression errors and also allows for forecast error autocorrelation in relatively large samples. The MS^* test is applied to one- to four-quarter-ahead and eight-quarter-ahead forecasts to examine whether its power is relatively low in small samples.

The two encompassing tests are conducted for every possible combination from the seven individual forecast series across different horizons. For a given set of forecasts at each forecasting horizon, one forecast must be chosen as the numeraire in the tests; with the interpretation being that this forecast encompasses the others. The encompassing tests are carried out, with each forecast used as the numeraire in turn, in order to allow for all possible results without pre-selection. It is found that the MS^* test results are severely mis-sized by over accepting the null-hypothesis, especially for medium to long-run forecasts due to the small sample size of this study. Therefore the examination of the effect of encompassing tests on the performance of combination

forecasts is based on the results from the regression-based F -test only, hence restricted to the one-quarter-ahead forecasting horizon⁴.

Based on the results of the regression-based F -test, one-quarter-ahead individual forecasts at each forecasting horizon are combined. For n competing individual forecasts, if the encompassing test results show that none of them encompasses its competitors, it suggests that the n individual forecasts should be combined and the forecast accuracy can be improved by combination. For example, f_{sa} , f_{re} and f_{iv} are used to represent one-quarter-ahead individual forecasts generated from the forecasting models: SARIMA model, RE-ADLM and TVP model. The encompassing test is carried out to test the null-hypothesis that f_{sa} encompasses f_{re} and f_{iv} ; f_{re} encompasses f_{sa} and f_{iv} ; and f_{iv} encompasses f_{re} and f_{sa} . The test statistics are 4.81, 5.41 and 10.27 with associated probability values of 0.0155, 0.0099 and 0.0004, respectively. The test results show that the hypotheses that the forecasts encompass their respective rivals are rejected at conventional significance levels. It implies that combination of the three individual forecasts would lead to an improvement in forecast performance.

Table 3 reports the results of the comparison between the performances of combination forecasts with and without encompassing pre-tests. For consistency with the above analysis, the evaluation is based on the measure of percentage of superior combination forecasts. It can be seen that the performance of combination methods based on the results of the encompassing test has improved—but only marginally (around 1%). The results suggest that the encompassing test does not contribute significantly to the improved accuracy of combination forecasts; and it may not be necessary to conduct encompassing test before individual forecasts are combined.

(Insert Table 3 here)

Conclusion

⁴ The results of both encompassing tests are not reported here but available from the author upon request.

This study investigated the accuracy of combination forecasts against individual forecasts. The individual forecasting models include the following econometric models: the RE-ADLM, two ECM approaches (WB-ECM and JML-ECM), unrestricted VAR model and the TVP model. These models represent the latest methodological developments in the tourism forecasting literature and have shown their advantages over the alternative forecasting models in previous studies. Three combination methods were employed and the performance comparison between alternative combination methods was carried out. The encompassing tests were applied and their effects on forecasting performance of combination methods were examined.

The empirical analysis in this study leads to the following main conclusions:

- This study has provided more robust evidence on the superiority of combination forecasts to individual forecasts over different forecasting horizons.
- The variance-covariance combination method turns out to be the best among the three combination methods investigated.
- More sophisticated combination forecasts, such as the discounted MSFE method and variance-covariance method, which take the historical performance of the individual forecasts into consideration, perform better than the simple average combination forecasts.
- The encompassing test does not contribute significantly to the improved accuracy of combination forecasts.

The empirical results provide further evidence on the efficiency of combination forecasts suggesting that combination forecasts are preferred to single model forecasts in tourism forecasting practices. Further research on the performance of combination forecasts in other situations such as different origin-destinations pairs, and different data frequencies would be useful. Moreover, to further enhance the conclusion of this study, other advanced combination methods could be used to evaluate the accuracy improvement of tourism forecast combinations.

References

- Abeyasinghe, T. (1994). Deterministic Seasonal Models and Spurious Regressions. *Journal of Econometrics*, 61: 259-272.
- Alleyne, D. (2006). Can Seasonal Unit Root Testing Improve the Forecasting Accuracy of Tourist Arrivals? *Tourism Economics*, 12, 45-64.
- Anderson, T.W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley: New York.
- Bates, J.M. and C.W.J. Granger, (1969). The combination of forecasts. *Operational Research Quarterly*, 20: 451-468.
- Bischoff, C.W. (1989). The Combination of Macroeconomic Forecasts. *Journal of Forecasting*, 8: 293-314.
- Chu, F.L. (1998). Forecasting Tourism: A Combined Approach. *Tourism Management*, 19: 515-520.
- Clemen, R.T. (1989). Combination Forecasts: A Review and Annotated Bibliography. *International Journal of Forecasting*, 5: 559-583.
- Clemen, R.T. and R.L. Winkler (1986). Combining Economic Forecasts. *Journal of Business and Economic Statistics*, 4: 39-46.
- Diebold F.X., R.S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*. 13: 253-263.
- Diebold, F.X. and P. Pauly (1987). Structural Change and the Combination of Forecasts. *Journal of Forecasting*, 6: 21-40.
- Divisekera, S. (2003). A Model of Demand for International Tourism. *Annals of Tourism Research*, 30: 31-49.
- Elliott, G. and A. Timmermann (2004). Optimal Forecast Combinations under General Loss Functions and Forecast Error Distributions. *Journal of Econometrics*, 122: 47-79.
- Engle, R.F. and C.W.J. Granger (1987). Cointegration 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.
- Fildes, R., and K. Ord (2002). Forecasting Competitions—Their Role in Improving Forecasting Practice and Research. In M. Clements and D. Hendry (Eds.), *A Companion to Economic Forecasting*. Oxford: Blackwell.
- Fritz, R.G., C. Brandon and J. Xander (1984). Combining Time-Series and Econometric Forecast of Tourism Activity. *Annals of Tourism Research*, 11: 219-229.
- Granger, C.W.J. (1989). Combining Forecasts- Twenty Years Later. *Journal of Forecasting*, 8: 167-173.
- Harvey, D.I. and P. Newbold (2000). Tests for Multiple Forecast Encompassing. *Journal of Applied Econometrics*, 15: 471-82.
- Harvey, D.I., S.J. Leybourne and P. Newbold (1997). Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*, 13: 281-291.
- Harvey, D.I., S.J. Leybourne and P. Newbold (1998). Tests for Forecast Encompassing. *Journal of business and Economic Statistics*, 16: 254-259.
- Holden, K. and D.A. Peel (1986). An Empirical Investigation of Combinations of Economic Forecasts. *Journal of Forecasting*, 5: 229-242.

- Hylleberg, S., R.F. Engle, C.W.J. Granger, B.S. Yoo (1990). Seasonal Integration and Cointegration. *Journal of Econometrics*, 44: 215-238.
- Johansen, S. (1988). A Statistical Analysis of Cointegration Vectors. *Journal of Economic Dynamics and Control*, 12: 231-254.
- Kalman, R.E. (1960). A New Approach to Linear Filtering and Prediction Problems, Transitions ASME. *Journal of Basic Engineering*, 82: 35-45.
- Kling, J.L. and D.A. Bessler (1985). A Comparison of Multivariate Forecasting Procedures for Economic Time Series. *International journal of Forecasting*, 1: 5-24.
- Kulendran, N. and S.F. Witt (2001). Cointegration versus Least Squares Regression. *Annals of Tourism Research*, 28: 291-311.
- Li, G., H. Song and S.F. Witt (2005). Recent Developments in Econometric Modelling and Forecasting. *Journal of Travel Research*, 44: 82-99.
- Lim, C., and M. McAleer (2001). Cointegration Analysis of Quarterly Tourism Demand by Hong Kong and Singapore for Australia. *Applied Economics*, 33: 1599-1619.
- Makridakis, S. and R.L. Winkler (1983). Averages of Forecasts: Some Empirical Results. *Management Science*, 29: 987-96.
- Newbold, P. and C.W.J. Granger (1974). Experience with Forecasting Univariate Time Series and the Combination of Forecasts. *Journal of the Royal Statistical Society, Series A*, 137: 131-146.
- Oh, C. and B.J. Morzuch (2005). Evaluating Time-Series Models to Forecast the Demand for Tourism in Singapore. *Journal of Travel Research*, 43: 404-413.
- Palm, F., and A. Zellner (1992). To Combine or Not to Combine? Issues of Combining Forecasts. *Journal of Forecasting*, 11: 687-701.
- Pankratz, A. (1983). *Forecasting with Univariate Box-Jenkins Models*. New York: Wiley.
- Song, H. and S.F. Witt (2000). *Tourism Demand Modelling and Forecasting: Modern Econometric Approaches*. Oxford, UK: Pergamon.
- Song, H., S.F. Witt and T.C. Jensen (2003). Tourism Forecasting: Accuracy of Alternative Econometric Models. *International Journal of Forecasting*, 19: 123-141.
- Stock, J.H. and M.W. Watson (1999). Forecasting Inflation. *Journal of Monetary Economics*, 44: 293-375.
- Stock, J.H. and M.W. Watson (2004). Combination Forecasts of Output growth in A Seven-Country Data Set. *Journal of Forecasting*, 23: 405-430.
- Timmermann, A. (2006). Forecast Combinations, in Elliot, G., C.W.J. Granger and A. Timmermann (eds), *Handbook of Economic Forecasting*, Amsterdam: Elsevier, pp. 135-196.
- Trehan, B. (1986). Oil Prices, Exchange Rates and the U. S. Economy: An Empirical Investigation. *Economic Review, Federal Reserve Bank of San Francisco*, 4: 25-43.
- Wickens, M.R. and T.S. Breusch (1988). Dynamic Specification, the Long-Run and the Estimation of Transformed Regression Model. *Economic Journal*, 98 (Conference): 189-205.
- Winkler, R.L. and S. Makridakis (1983). The Combination of Forecasts. *Journal of the Royal Statistical Society, Series A*, 146: 150-157.
- Witt, S.F., and C.A. Witt (1992). *Modelling and Forecasting Demand in Tourism*. Academic Press, London.
- Wong, K.K.F., H. Song, S.F. Witt and D.C. Wu (2007). Tourism Forecasting: To Combine or

Formatted: English (U.K.)

Not to Combine? *Tourism Management*, 28:1068-1078.

Yu, G., and Z. Schwartz (2006). Forecasting Short Time-Series Tourism Demand with Artificial Intelligence Models, *Journal of Travel Research*, 45: 194-203.

Table 1 Forecasting Accuracy over Different Forecasting Horizons

	Measure	NAÏVE	SARIMA	RE-ADLM	WB-ECM	JML-ECM	TVP	VAR
1-quarter-ahead	MAPE	2.045(5)	1.875(3)	2.008(4)	1.661(1)	1.697(2)	2.324(6)	2.389(7)
2-quarter-ahead	MAPE	2.051(4)	1.753(3)	2.302(5)	1.624(2)	1.540(1)	2.922(6)	3.232(7)
3-quarter-ahead	MAPE	2.058(4)	1.786(2)	2.245(5)	1.822(3)	1.729(1)	2.490(6)	3.341(7)
4-quarter-ahead	MAPE	2.113(3)	1.744(2)	2.164(5)	2.134(4)	1.562(1)	2.257(6)	3.736(7)
8-quarter-ahead	MAPE	3.040(4)	1.673(1)	3.705(5)	2.810(3)	1.745(2)	3.980(6)	6.996(7)
Overall	MAPE	2.261(4)	1.766(2)	2.485(5)	2.010(3)	1.655(1)	2.795(6)	3.939(7)

Note: Figures in brackets indicate the ranks of individual forecasting methods.

Table 2 Percentage of Superior Combination Forecasts over the Best Individual Forecasts

	Measure	SA	VACO	Dis0.85	Dis0.9	Dis0.95
1-quarter-ahead	MAPE	65.00%(3)	64.17%(4)	66.67%(1)	65.83%(2)	65.83%(2)
2-quarter-ahead	MAPE	12.50%(4)	24.17%(1)	11.67%(5)	14.17%(3)	17.50%(2)
3-quarter-ahead	MAPE	63.33%(3)	82.50%(1)	80.00%(2)	82.50%(1)	82.50%(1)
4-quarter-ahead	MAPE	96.67%(4)	99.17%(1)	97.50%(3)	98.33%(2)	99.17%(1)
8-quarter-ahead	MAPE	76.67%(4)	96.67%(3)	99.17%(1)	98.33%(2)	96.67%(3)
Overall	MAPE	62.83%(5)	73.33%(1)	71.00%(4)	71.83%(3)	72.33%(2)

Note: Figures in brackets indicate the ranks of combination methods in terms of their performance. Dis0.85, Dis0.9 and Dis0.95 denotes Discounted MSFE method ($\alpha = 0.85, 0.9, 0.95$, respectively)

Table 3 Percentages of Superior Combination Forecasts with and without Encompassing Test: One-Step-Ahead Forecasts

	MAPE
SA (without test)	65.00%
SA (with test)	65.25%
<i>improvement (percentage)</i>	0.39%
VACO (without test)	64.17%
VACO (with test)	64.41%
<i>improvement (percentage)</i>	0.37%
Discounted0.95 (without test)	65.83%
Discounted0.95 (with test)	66.10%
<i>improvement (percentage)</i>	0.41%
Discounted0.9 (without test)	65.83%
Discounted0.9 (with test)	66.10%
<i>improvement (percentage)</i>	0.41%
Discounted0.85 (without test)	66.67%
Discounted0.85 (with test)	66.95%
<i>improvement (percentage)</i>	0.42%