Does the Combination of Models with Different Explanatory Variables Improve Tourism Demand Forecasting Performance?

Journal:	*Tourism Economics*
Manuscript ID	TEU-21-0493.R3
Manuscript Type:	Empirical Article
Keywords:	combination forecasting, tourism demand forecasting, climate, diversification
Abstract:	The aim of this study is to assess whether combining econometric models with different explanatory variables can contribute to better tourism demand forecasts. Inbound tourism demand to the UK from seven leading markets is forecast respectively based on quarterly data using both individual and combination models. Causal econometric models that serve as constituents in combination take two specifications which are different in identified influencing factors. The empirical results show that generally including different explanatory variables in combination can produce better predictions according to both predictive accuracy measures and statistical tests. It suggests that the combination forecasting approach is superior to the individual one, and diversified information embedded in different explanatory variables should be integrated to improve tourism demand forecasting performance.

Does the Combination of Models with Different Explanatory Variables Improve Tourism Demand Forecasting Performance?

Abstract:

The aim of this study is to assess whether combining econometric models with different explanatory variables can contribute to better tourism demand forecasts. Inbound tourism demand to the UK from seven leading markets is forecast respectively based on quarterly data using both individual and combination models. Causal econometric models that serve as constituents in combination take two specifications which are different in identified influencing factors. The empirical results show that generally including different explanatory variables in combination can produce better predictions according to both predictive accuracy measures and statistical tests. It suggests that the combination forecasting approach is superior to the individual one, and diversified information embedded in different explanatory variables should be integrated to improve tourism demand forecasting performance.

Keywords

combination forecasting, tourism demand forecasting, climate, diversification

Introduction

The combination forecasting approach has been extensively applied in many areas such as meteorology, economics and insurance as a robust and powerful method to improve forecasting ability (Qian et al., 2022; Stock and Watson, 1998, 2003, 2004). However, in the context of tourism demand forecasting, combining individual forecasts is not the mainstream method. Some studies apply the combination, and constituent models considered include time series, econometric and artificial intelligence (AI) models with causal models being confined to the ones that identify the same set of economic variables as influencing factors. When forecasting tourism demand, no combination has been applied to causal models with different explanatory variables. It means that a type of information has been neglected, which origins from distinct independent variables included in different single models. Bates and Granger (1969) pointed out that to make as good a forecast as possible, combining single forecasts based on different variables was a wise procedure. This paper evaluates whether combining econometric models with different explanatory variable can improve tourism demand forecasting performance.

In this paper, econometric models serving as constituents in combination take two model specifications, which are different in identified influencing factors: either only include mainstream economic determinants or introduce the climate factor as a demand determinant. Climate is an important influencing factor in choice of destination, time of departure and length of stay and can hence affect tourism demand. In the context of climate change, to which tourism is both a significant contributor and substantially affected by (Scott et al., 2019), it is important to understand the value of climate variables when projecting the geographical and seasonal shifts in tourism demand. Besides, the ongoing COVID-19 crisis holds important messages regarding the interwoven nature between tourism and the environment, which highlights the need to consider the role of climate variables in forecasting tourism demand. In the current literature, however, mainstream causal models treat climate variables as fixed with time, as seasonal dummies, or as fixed effects in panel data studies, which ignores variation as well as long-term change in climate conditions. Tourism demand forecasts generated from such models neglect the relationship between climate condition and tourism demand. This study fills this gap through incorporating climate factors when combining tourism demand forecasts.

The rest of this paper is structured as follows. Section 2 reviews the literature on combination forecasting and its application in the tourism demand context. Section 3 focuses on the research method by introducing the individual and combination forecasting models, the variables and data as well as the forecasting procedure. Section 4 presents the empirical results with discussion and section 5 concludes the study.

Literature Review

Why to Combine?

There are two distinct forecasting approaches: the individual forecasting approach, which produces direct forecasts from single models; and the combination forecasting approach, which generates composite forecasts by combining constituent forecasts yielded by single models. When the individual approach is followed, forecasts are generated relying on only one model, and the forecasting performance of rival models are always compared to identify the best-performing one with the inferior forecasts being discarded. The disadvantages of such a procedure include: firstly, the discarded predictions may contain useful independent information, and secondly, the identification of the 'best single model' is like a moving target. Many existing studies show that the forecasting performance of different individual models depends on the accuracy measure used, the forecasting horizon under consideration and the origin-destination pair under study (Gunter and Önder, 2015; Hassani et al., 2017). There are no clear-cut evidences showing which single model is superior to others under all situations, hence there exists no principles regarding the selection of the best single forecasting method. Rather than trying to choose the best single model, the combination forecasting approach pools a range of constituent forecasts together. The rationale is that single forecasts from diverse models based on competing theories, functional forms and specifications contain independent information, the combination of which can achieve diversification gain. As a result, the performance of combination forecasting is supposed to be more stable than the individual one.

The idea of combining multiple forecasts of the same event dates to the 1960s. Bates and Granger (1969) published the seminal work in 1969, which showed that better predictions can be obtained by combining two forecasts yielded by different models. Since then, the general forecasting literature has seen considerable studies on combination forecasts with contributions from many disciplines such as forecasting, statistics and management (Clemen, 1989). The constituent forecasts have been extended from two to multiple ones with various combining methods being presented and tested and different forecasting horizons and accuracy measures being considered. The empirical results support the conclusion that combining alternative forecasts together can reduce uncertainty and increase accuracy (Diebold and Pauly, 1990; Makridakis et al., 2020; 2018; Zhang and Yu, 2018).

Weighting Schemes

One important step of combination is to identify the optimal weights that are assigned to each constituent projection. Main combination approaches differ in the way they use historical information to compute the weights. The simplest weighting scheme is the simple average (SA) method, which assigns equal weights to all individuals. The SA method has been found to be a robust, stable and easy-to-use way, often outperforming more sophisticated weighting schemes and hence is always used as a benchmark in combination forecasting studies (Makridakis and Winkler, 1983; Makridakis et al., 2020; Henry and Clements, 2004; Wu et al., 2020).

The variance-covariance (VACO) method was presented by Bates and Granger (1969) and extended by Fritz et al. (1984) to multiple constituents. To minimize the combined forecasts variance, the VACO scheme assigns larger weights to individual forecasts with smaller forecasting errors, which links the weighting scheme to the historical performance of constituent forecasts. The VACO method is a common choice in forecasting studies (Baumeister and Kilian, 2015; Fritz et al., 1984; Stock and Watson, 2004; Wong et al., 2007; Wu et al., 2020).

A similar weighting method is the discounted mean square forecast error (DMSFE) method, which is proposed by Bates and Granger (1969) and generalized by Newbold and Granger (1974). Weights in DMSFE are inversely related to the individual forecasting accuracy, which is measured by the forecasting error, and the recent forecasts are weighed more heavily by applying a discounting factor. The discounting factor lies between 0 and 1, and in practice, 0.95, 0.9, 0.85, 0.8 are all common choices (Diebold and Pauly, 1987; Shen et al., 2008, 2011; Stock and Watson, 2004).

All the above-mentioned methods share one common feature which is that the weights add up to unity. Granger and Ramanathan (1984) presented the regression method which does not require the weights to add up to unity. The proposed method regresses the actual values on each constituent forecasts and a constant term with the estimated parameters to be the corresponding weights. Some applications of the regression method have demonstrated its satisfactory performance (Guerard, 1987; Holmen, 1987; MacDonald and Marsh, 1994), while others showed evidence of its unstable predicting ability (Lobo, 1991; Shen et al., 2011). The limitation of the regression method is obvious: when the number of the constituent forecasts are large compared to the sample size, it is inappropriate as the regression for working out the weights is invalid. In the regression-based framework, Diebold and Pauly (1990) applied Bayesian shrinkage techniques to incorporate prior information into the weighting scheme and concluded that shrinkage improved the accuracy of the regression-based combination forecasts.

Another extension of the regression method is the time-varying parameter (TVP) method, which uses Kalman filter algorithm to estimate the weights that are allowed to vary with time if the data suggests so. Applications of the time-varying combination method include LeSage and Magura (1992), Sessions and Chatterjee (1989), Shen et al. (2011) and Stock and Watson (2004). Sun et al. (2021) further developed the TVP weighting method through introducing the nonparametric estimation and proposed the TV jackknife model averaging (TVJMA), which can handle structural changes and nonstationary trends in tourism data.

Applications of Combination Forecasting in the Tourism Demand Literature

An increasing number of studies on combination forecasts have appeared in the tourism demand literature, among which differences can be found in weighting schemes, individual model inputs and accuracy measures with one common finding being that combination forecasts are generally superior to individual ones (Kourentzes et al., 2021; Li et al., 2019; Liu et al., 2021; Song and Li, 2021; Wu et al., 2020). The SA, VACO and DMSFE methods are popular weighting schemes (Shen et al., 2008, 2011; Song, Witt, Wong, and Wu, 2009; Wong et al., 2007). Other methods including the stochastic frontier analysis (SFA), the cumulative sum control chart (CUSUM) method, the management-oriented approach, AI techniques and nonparametric techniques have also been explored to determine the optimal weights (Andrawis et al., 2011; Chan et al., 2010; Claveria et al., 2016; Coshall and Charlesworth, 2011; Qiu et al., 2021; Sun et al., 2021).

When it comes to individual model inputs, the most popular ones are time-series and econometric models. Regarding time series techniques, autoregressive integrated moving average (ARIMA) and exponential smoothing (ETS) models are widely chosen; and for causal models, popular candidates are autoregressive distributed lag (ADL), error correction (EC) and TVP models (Cang et al., 2011, 2014; Chan et al., 2010; Li et al., 2019; Shen et al., 2008, 2011). However, independent information embedded in different explanatory variables in econometric models has never been combined. Causal models that generate component forecasts in existing combination studies are the same in identifying the same influencing factors: the origins' real income, the relative price between destination and origin, the substitute price in competing markets as well as seasonal and one-off events dummies (Chan et al., 2010; Li et al., 2019; Shen et al., 2008, 2011; Wong et al., 2007). AI techniques and grey models are also considered in more recent studies (Hu et al., 2021; Liu et al., 2021; Qiu et al., 2021).

When evaluating forecasting performance, widely employed accuracy measures include mean absolute percentage error (MAPE), mean absolute error (MAE), root mean squared error (RMSE) and root mean squared percentage error (RMSPE). Besides descriptive accuracy measures, statistical tests such as the Diebold and Mariano (D-M) test, the Harvey, Leybourne, Newbold (HLN) test and the Mann-Whiteney test are also applied to assess whether the combination forecasting approach is significantly better than the single method (Bangwayo-Skeete and Skeete, 2015; Cang, 2014; Coshall, 2009).

Research Method

Variables and Data

UK's inbound tourism demand from its seven leading markets: France, Germany, Irish Republic, Italy, the Netherlands, Spain and the US are studied with tourism demand being proxied by tourist arrivals. For traditional econometric models, income, own price and substitute price are chosen as demand determinants, and seasonal and event dummies are included to capture the seasonal and event effects. To compute substitute prices, Germany and France are chosen as competing destinations to the UK, as France, the UK and Germany are the top three most visited destinations in northwestern Europe in 2017 (UNWTO, 2018). For climate econometric models, besides economic influencing factors, climate condition in the UK, which is measured by Tourism Climatic Index (TCI) of the UK, is introduced to represent the climate attribute of the destination, which is considered as another determinant of tourism demand.

TCI is a human-oriented, synthetic evaluation of climate attractiveness to tourists, which takes the most relevant climate elements to tourism experience into account. It comprises five sub-indices. The composition of $\mathrm{TCI}_{\mathrm{UK,t}}$ follows Mieczkowski (1985):

$$TCI_{UK,t} = 2 \times [4(cid_{UK,t}) + cia_{UK,t} + 2(P_{UK,t}) + 2(S_{UK,t}) + W_{UK,t}] \quad 1$$

where $cid_{UK,t}$ is the daytime comfort index composed of the maximum daily temperature and the minimum daily relative humidity; $cia_{UK,t}$ is the daily comfort index composed of daily temperature and daily relative humidity; $P_{UK,t}$ denotes the rating for precipitation; $S_{UK,t}$ signifies the rating for sunshine duration; and $W_{UK,t}$ represents the rating for wind speed.

Detailed introduction of the TCI can be found in (Mieczkowski, 1985). Data for the UK's

TCI are from the Met Office, which provides monthly meteorological data from many weather stations across the UK. Only data from the Heathrow weather station are taken into account mainly for two reasons. First and most importantly, it is dictated by data availability. Data on seven meteorological variables are required to construct TCI which include the maximum daily temperature, the minimum daily relative humidity, the mean daily temperature, the mean daily relative humidity, the mean monthly precipitation, the mean hours sunshine per day and the wind speed. Only the Heathrow station provides all data needed for the whole sample period under study, while other stations fail to provide some of the data either for the whole sample period or for some seasons. Secondly, most international holiday-makers tend to take the weather condition of their first stop into consideration when making travel decisions. In 2017, 76% of inbound tourists reached the UK by air, and as the busiest airport in the UK, Heathrow Airport is the first stop for most of them. Table 1 summarizes variables and data sources.

The general form of traditional and climate econometric models are specified in equation 2 and equation 3 respectively:

$$y_t = f(GDP_t, rp_t, sp_t, Q_1, Q_2, Q_3, D_{DS}, D_{911}, D_{BM}, D_{2008}, D_{OL}) \qquad 2$$

$$y_t = f(GDP_t, rp_t, sp_t, TCI_{UK,t}, D_{DS}, D_{911}, D_{BM}, D_{2008}, D_{OL}) \qquad 3$$

Table 1: Variables and Data Sources

Individual Forecasting Models and Three Combination Groups

Individual models selected as constituents in combination serve not only as forecasting tools but also as sources of diversification. If we see forecasts as information, combining forecasts is an aggregation of information. According to Bates and Granger (1969), the combination of models that contain independent information is most likely to improve forecasting performance. To ensure that constituent models contain as much independent information as possible, a variety of individual models including causal econometric and non-causal time series techniques, which are different in modelling techniques, assumptions and explanatory variables are selected. The seasonal naive no-change forecasts serve as benchmarks.

In the current tourism demand literature, popular econometric models based on time series data include cointegration techniques, ADL, leading indicator (LI), vector autoregressive (VAR), TVP and simple dynamic (SD) models (Song et al., 2019), all of which are applied to generate individual forecasts. All econometric techniques are applied with two specifications which are different in identified explanatory variables. The ones that follow the mainstream specification in the current literature by considering economic influencing factors and dummy variables are referred to as ***traditional econometric models***. And the others that introduce the climate factor as a demand determinant are called ***climate econometric models***. The employed time series techniques consist of seasonal Naïve no-change, SARIMA, ETS and state space ETS methods. Except the seasonal naive no-change model, all other individual forecasting models are considered as candidate components in combinations. Therefore, there are altogether 15 individual forecasting models, all of which are summarized in Table 2.

According to the PP (Phillips and Perron, 1988) unit root test results, all model variables are integrated of order zero or order one, based on which the bounds test cointegration approach is selected in this study as it is robust no matter whether the model variables are integrated of the same order (not integrated of order $d > 2$) (Song et al., 2012; Wang, 2009). The lag lengths in econometric models are selected based on Akaike Information Criteria (AIC) with maximum lags of four. The specific forms of SARIMA models are also selected based on AIC and the types of the error, trend and seasonal components in state space ETS models are automatically selected based on AIC, according to which the smoothing methods of ETS models are chosen. Three residual diagnostic tests including serial correlation LM test (Breusch, 1978; Godfrey, 1978), Harvey heteroscedasticity test (Harvey, 1976) and autoregressive conditional heteroscedasticity (ARCH) test (Engle, 1982) are run after estimation. The detailed discussion of the bounds test cointegration approach can be found in Pesaran et al. (2001), those of other econometric models can be found in Song, Witt, and Li (2009) and that for SARIMA, ETS and state space ETS methods can be found in Box et al. (2015), Holt (2004) and Hyndman et al. (2002) respectively.

Table 2: Summary of Individual Forecasting Models

The 15 individual models available to combine are categorized into three combination groups: ***Group A*** consists of all models; ***Group B*** comprises six traditional econometric models and three time series models; and ***Group C*** contains six climate econometric models and three time series models. Table 3 shows the component models in each combination group. For each group, we need to consider all possible combinations for every weighting scheme, origin and forecasting horizon, the number of which can be specified as $N = \sum_{r=2}^{n} C_n^r$ with $C_n^r = \frac{n!}{r!(n-r)!}$, where n is the number of individual models in each group. The number of all possible combinations is 32752 for Group A with $n = 15$ and 502 for Group B and Group C with $n = 9$ respectively.

Table 3: The Component Models of Three Combination Groups

Combination Forecasting Methods

This paper evaluates the most popular statistical weighting schemes in the current tourism demand literature including SA, VACO and DMSFE. The regression-based methods are excluded from this study as they are inappropriate because of the large number of constituent forecasts in the combination panel relative to the small training sample size. In addition, a new weighting scheme, which is referred to as ***the inverse-MAE method***, is introduced and tested with weights being expressed as:

$$w_i = \frac{[MAE_i]^{-1}}{sum_{j=1}^{n}[MAE_j]^{-1}} \quad 4$$

where $MAE_i = \frac{1}{T}\left|\sum_{t=1}^{T} f_{t+h} - \hat{f}_{t+h}^{i}\right|$ signifies the mean absolute error of the ith individual forecast $\hat{f}_{t+h}^{i}$. Rather than the mean squared error (MSE), which is used in VACO and DMSFE, MAE is selected as the measure of historical forecasting performance. With the same individual forecasts, the better performing constituent forecasts are weighted lighter in such a scheme as the forecasting error is not amplified by the square operator.

Three popular accuracy measures including MAE, MAPE and RMSE are applied to evaluate the forecasting performance, and the D-M test (Diebold and Mariano, 1995) is conducted to check whether the difference in the accuracy of competing models is statistically significant. Applications of the D-M test in the tourism demand literature include Athanasopoulos and Silva (2012), Gil-Alana (2010) and Álvarez-Díaz and Rosselló-Nadal (2010).

Forecasting Procedure

The sample covers the 1994Q1-2017Q4 period and different time spans are considered for different origins given data availability. The whole sample is divided into three periods as illustrated by Figure 1. Observations from 1994Q1 to 2012Q4 are used for model estimation, and the individual out-of-sample forecasts are generated from 2013Q1 to 2017Q4, with forecasts from 2013Q1 to 2015Q4 used to determine the combining weights, and the ones from 2016Q1 to 2017Q4 retained for comparison. The out-of-sample combination forecasts are generated from 2016Q1 to 2017Q4. The sample from 2013Q1 to 2015Q4 is the training sample, and that from 2016Q1 to 2017Q4 is the comparison sample.

Figure 1: Illustration of Data Sample

This study follows the recursive forecasting procedure which is popular in the tourism demand forecasting literature (Song et al., 2019). One- to four-step-ahead out-of-sample forecasts are generated from every 15 individual forecasting model for combination and comparison with the seasonal naive no-change forecasts serving as benchmarks. Three groups of individual forecasts are combined respectively with different weighting schemes. The weights determined by different combination methods (except SA) are time-varying by applying the recursive weighting procedure, which is illustrated in Figure 2. For example, for composite forecasts in 2016Q1, the historical performance of single forecasts from 2013Q1 to 2015Q4 are considered to construct the weights. For combination forecasts in 2016Q2, the historical performance of individual forecasts from 2013Q1 to 2016Q1 are taken into account to decide the weights. Such a procedure is repeated with one more single forecast added to the training sample each time, updating the weights each period according to the historical individual forecasting performance. The codes for computing combination forecasts as well as conducting forecasting comparison and statistical tests are written in Matlab 2018a.

Figure 2 The Recursive Weighting Procedure for One-Step-Ahead Combination Forecasts

Results and Discussion

Performance Comparison Based on Descriptive Measures

For seven origins and four forecasting horizons, the performance of every single and combination model is evaluated and compared based on MAE, MAPE and RMSE. For the comparison between the individual and the combination forecasting approach, the percentages of the superior combination forecasts compared to the best single ones (referred to as ***superior percentages*** hereafter) and the percentages of the inferior combination projections compared to the worst single ones judged by MAE, MAPE and RMSE respectively are generated for each case. The results show that combining individual forecasts can improve forecasting accuracy in all cases regardless of origin country, forecasting length, accuracy measure, combination group or combination method. The most accurate forecasts are always produced by combination methods, and the worst projections are always generated through individual models. To evaluate whether including econometric models with different explanatory variables in combination is superior to only considering causal models of the same independent variables to combine, composite forecasts from combining different models in three groups are compared.

For every weighting scheme, the average superior percentages for each origin over four forecasting horizons judged by MAE, MAPE and RMSE are presented in Table 4 to Table 6 respectively with the highest superior percentages highlighted. For example, according to Table 5 (evaluated by MAPE), for the German origin, averagely 9616 ($29.36\% \times 32752$) SA combination forecasts are more accurate than the best single prediction when combining all models in Group A. The results show that the performance ranking of three combination groups varies according to origin market and accuracy measure. In general, combining all models in Group A is the best, which achieves the highest superior percentages for at least five out of seven markets despite three different accuracy measures and six distinct weighting methods.

It also shows that for one market, the difference in the forecasting ability of different weighting schemes given the same combination group is small (refer to each column of Table 4 to Table 6). For instance, according to Table 4 (judged by MAE), the greatest difference in the superior percentages achieved by the best and the worst weighting methods to combine Group A is seen from the American case between SA and DMSFE (.85), which is 3.7%. It means that there are 1212 more superior combination forecasts if the weights are obtained based on SA instead of DMSFE (.85).

Table 4: Superior Percentages of Three Combination Groups for Each Origin (MAE)

Table 5: Superior Percentages of Three Combination Groups for Each Origin (MAPE)

Table 6: Superior Percentages of Three Combination Groups for Each Origin (RMSE)

For every weighting scheme, the average superior percentages for each forecasting horizon over seven origins based on MAE, MAPE and RMSE are presented in Table 7 to Table 9 respectively with the highest superior percentages highlighted. It shows that combining all models in Group A performs the best in all cases. Similarly, the difference in the forecasting ability of different weighting schemes for one forecasting length is not great given the same combination group (refer to each column of Table 7 to Table 9).

Table 7: Superior Percentages of Three Combination Groups for Each Forecasting Horizon (MAE)

Table 8: Superior Percentages of Three Combination Groups for Each Forecasting Horizon (MAPE)

Table 9: Superior Percentages of Three Combination Groups for Each Forecasting Horizon (RMSE)

Table 10 demonstrates the results of the general comparison among three combination groups for each weighting method evaluated by three measures by taking averages across seven origins and four horizons. It is clear that combing all models achieves the best results for all cases. The values in Table 10 confirms previous observation, which is that the forecasting ability of different combination methods is similar considering the same combination group (refer to each row of Table 10).

Table 10: Superior Percentages of Three Combination Groups for Each Weighting Scheme

Performance Comparison Based on the D-M Test

For a given weighting scheme, 32752 combination models and 15 single models are available to generate forecasts for seven origins at four forecasting horizons when combining Group A. (Regarding Group B and Group C, there are 502 combination forecasting models and nine single forecasting models respectively with one weighting scheme.) To conduct the D-M test for each forecasting horizon, every combination model is compared with the best single model for each combination group with each weighting scheme. Considering seven origins and the application of the recursive forecasting procedure, each forecasting model generates 140 (7×20) one-step-ahead forecasts, 133 (7×19) two-step-ahead forecasts, 126 (7×18) three-step-ahead forecasts and 119 (7×17) four-step-ahead forecasts.

The null hypothesis of the D-M test is that there exists no difference in the forecast accuracy between the combination model and the best single model. When the null is rejected with a negative statistic, it means that forecasts from the combination model are significantly more accurate than those from the best single model; and when the null is rejected with a positive statistic, it means that the best single model forecasts significantly better than the combination model. The tests are carried out based on the quadratic loss function, and two percentages are provided in Table 11 for three combination groups, four forecasting horizons and six weighting schemes. The first percentage is the proportion of times of a negative significant value for the tests, and the second percentage is the percentage of times of a positive significant value for the tests. The highest negative significant value percentage and the lowest positive significant value percentage for each case are highlighted.

According to Table 11, the combination forecasting approach is statistically better than the single one. For each case, the percentage of the times that the D-M test is rejected with a negative statistic, which means that the combination model forecasts statistically more accurate than the best single one, is much higher than the percentage of the times that the D-M test is rejected with a positive statistic, which signifies that the best single model performs significantly better than the combination model. The results imply that when generating one- to four-step-ahead forecasts, forecasting performance can be improved significantly if combination is applied. For example, for the case of combining Group A using SA for one-step-ahead forecasts, the percentage of times of a negative significant statistic for the tests is 22.66%, and that of times of a positive significant statistic is 6.02%. It means that 7422 ($22.66\% \times 32752$) out of 32752 combination models from combining Group A using SA generate significantly more accurate forecasts than the best single model, and the best single model forecasts significantly better than 1971 ($6.02\% \times 32752$) of the 32752 sets of 140 one-step-ahead forecasts.

Table 11 also shows that combining econometric models with different explanatory variables, i.e., combining models in Group A, is better than integrating econometric models with the same set of independent variables. For four forecasting horizons, combining all models in Group A always yields the highest percentage of negative significant statistic and the lowest percentage of positive significant statistic no matter which combination method is applied. For instance, for the D-M tests between the VACO combination models and the best single model for four-step-ahead forecasts, the percentage of times of a negative significant statistic is 22.47% when combining Group A, 21.12% if combining Group B and 21.51% when Group C is combined. It means that if VACO weighting is applied with Group A, 7359 ($22.47\% \times 32752$) out of 32752 combination models generate significantly more accurate forecasts than the best single model; when Group B is combined, 106 ($21.12\% \times 502$) out of 502 combination models perform significantly better than the best single model; and if combination is applied with Group C, 108 ($21.51\% \times 502$) out of 502 combination models forecast significantly more accurate than the best single method.

Table 11 further indicates that for four forecasting horizons, the difference in the forecasting ability of different weighting schemes given the same combination group is small. For instance, the greatest difference in the percentage of times of a negative significant statistic achieved by the best and the worst weighting methods to combine Group A is seen from producing one-step-ahead forecasts between DMSFE (.95) and VACO, which is 1.6%. It means that there are 524 ($1.6\% \times 32752$) more significantly better combination models if the weights are obtained based on DMSFE (.95) instead of VACO. In addition, the performance rankings of six weighting schemes to combine Group A are relatively consistent across four forecasting horizons judged by the D-M test results, with DMSFE (.95), inverse-MAE and SA always ranking top three and VACO staying at the bottom. Another interesting observation is that when combining all models in Group A, six weighting schemes consistently behave the best for the fourth horizon (one year), which is of direct interest to policy-makers.

Table 11: The D-M Test Results

Conclusion

This paper evaluates forecasting combination of econometric models with different influencing factors through an empirical study on UK's inbound tourism demand with results showing that such a combination is statistically better than just integrating econometric models with the same explanatory variables. It also demonstrates that for a given combination panel, the forecasting ability of different combination methods is similar with the presented inverse-MAE scheme showing good performance.

The findings of this paper have implications for future research and for stakeholders in the tourism industry. Most importantly, this paper paves the way for further empirical investigations on combination forecasting including component models with distinct explanatory variables. In addition, it suggests more research attention to the combination forecasting approach. This study shows the general forecasting superiority of combination models compared to individual ones, which is in agreement with the conclusions of many existing studies (Li et al., 2019; Liu et al., 2021; Qiu et al., 2021; Wu et al., 2020). In the current literature, too much attention is paid to improving the forecasting ability of individual models and to identifying the best single model. If improving forecasting accuracy is the aim, combination forecasting deserves more study and should be included in forecasting comparisons. Besides, government and destination managers can improve the efficiency of their planning exercises by taking into account additional information on climate trends.

Several future research directions have been identified. Firstly, combining econometric models with different explanatory variables deserves more exploration in the future. Climate variables such as the origin's climate condition, the difference in the climate condition between the destination and the origin, and the relative climate condition of the main destination to the alternative competitors can be considered as tourism demand influencing factors. And the value of other variables such as Air Quality Index (AQI) and search engine data in improving combination forecasting accuracy can be explored.

Secondly, a user-friendly software which can produce combination forecasts easily should be made available considering the powerful forecasting ability of the combination forecasting approach. The biggest obstacle of popularizing combination forecasting is the cost of applying it. It is extremely time-consuming under the current condition to generate combination forecasts as it requires different programs for different tasks. With the help of the software, combination forecasts should be included in forecasting comparisons and can be used as benchmarks for forecasting evaluation.

Moreover, the findings of this paper suggest the application of combination forecasting to assess the impact of the COVID-19 crisis on tourism demand. In the current literature, one important way to evaluate the impact of crises on tourism demand is to compare actual values of demand with reference demand forecasts which are generated on the assumption that the crisis had not happened (Page et al., 2012). Combination can be applied to improve the accuracy of the reference forecasts and hence to improve the performance of such evaluations.

This paper, like other studies, is not without limitations. Only one- to four-step-ahead forecasts are generated and other forecasting horizons are omitted. Besides, nonparametric techniques such as AI-based individual and combination forecasting methods are not included. In addition, important determinants may be missing from the causal models. For instance, no variables for travel costs are included due to unavailability of suitable data.

References

Andrawis RR, Atiya AF and El-Shishiny H (2011) Combination of long term and short term forecasts with application to tourism demand forecasting. *International Journal of Forecasting 27* (3): 870-886.

Athanasopoulos G, Hyndman RJ, Song H, and Wu DC (2011) The tourism forecasting competition. *International Journal of Forecasting 27* (3): 822-844.

Athanasopoulos G and Silva AD (2012) Multivariate exponential smoothing for forecasting tourist arrivals. *Journal of Travel Research*, 51 (5): 640-652.

Bates JM and Granger CWJ (1969) The Combination of Forecasts. *Operational Research Society 20* (4): 451-468.

Bangwayo-Skeete PF and Skeete RW (2015) Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46: 454-464.

Baumeister C and Kilian L (2015) Forecasting the real price of oil in a changing world: a forecast combination approach. *Journal of Business & Economic Statistics 33* (3): 338-351.

Box GD, Jenkins GM, Reinsel GC and Ljung GM (2015) *Time series analysis: forecasting and control*. John Wiley & Sons.

Breusch TS (1978) Testing for autocorrelation in dynamic linear models. Australian Economic Papers 17 (31): 334-355.

Cang S (2011) A non-linear tourism demand forecast combination model. *Tourism Economics 17* (1): 5-20.

Cang S (2011) A Comparative Analysis of Three Types of Tourism Demand Forecasting Models: Individual, Linear Combination and Non-linear Combination. *International Journal of Tourism Research 16* (6): 596-607.

Chan CK, Witt SF, Lee Y and Song H (2010) Tourism forecast combination using the CUSUM technique. *Tourism Management 31* (6): 891-897.

Claveria O, Monte E and Torra S (2016) Combination forecasts of tourism demand with machine learning models. *Applied Economics Letters 23* (6): 428-431.

Clemen RT (1989) Combining forecasts: A review and annotated bibliography. *International journal of forecasting 5* (4): 559-583.

Clemen RT and Winkler RL (1986) Combining economic forecasts. *Journal of Business & Economic Statistics 4* (1): 39-46.

Coshall JT (2009) Combining volatility and smoothing forecasts of UK demand for international tourism. *Tourism Management 30* (4): 495-511.

Coshall JT and Charlesworth R (2011) A management orientated approach to combination forecasting of tourism demand. *Tourism Management 32* (4): 759-769.

Dickey DA and Fuller WA (1981) Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*: 1057-1072.

Diebold FX and Mariano RS (1995) Comparing predictive accuracy. *Journal of Business and Economic Statistics 3*: 253-263.

Diebold FX and Pauly P (1987) Structural change and the combination of forecasts. *Journal of Forecasting 6* (1): 21-40.

Diebold FX and Pauly P (1990) The use of prior information in forecast combination. *International Journal of Forecasting 6* (4): 503-508.

Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*: 987-1007.

Fritz RG, Brandon C and Xander J (1984) Combining time-series and econometric forecast of tourism activity. *Annals of Tourism Research 11* (2): 219-229.

Gil-Alana LA (2010) International arrivals in the Canary Islands: persistence, long memory, seasonality and other implicit dynamics. *Tourism Economics*, 16 (2): 287-302.

Godfrey LG (1978) Testing against general autoregressive and moving average error models when the regressors include lagged dependent variables. *Econometrica: Journal of the Econometric Society:* 1293-1301.

Granger CWJ and Ramanathan R (1984) Improved Methods of Combining Forecasts. *Journal of Forecasting 3* (2): 197-204.

Guerard Jr JB (1987) Linear constraints, robust-weighting and efficient composite modeling. *Journal of Forecasting 6* (3): 193-199.

Gunter U and Önder I (2015) Forecasting international city tourism demand for Paris: accuracy of uni- and multivariate models employing monthly data. *Tourism Management 46*: 123-135.

Harvey AC (1976) Estimating regression models with multiplicative heteroscedasticity. *Econometrica: Journal of the Econometric Society:* 461-465.

Hassani H, Silva ES, Antonakakis N, Filis G and Gupta R (2017) Forecasting accuracy evaluation of tourist arrivals. *Annals of Tourism Research 63*: 112-127.

Hendry DF and Clements MP (2004) Pooling of forecasts. *The Econometrics Journal 7(1):* 1–31.

Holmen JS (1987) A note on the value of combining short-term earnings forecasts: A test of Granger and Ramanathan. *International Journal of Forecasting 3* (2): 239-243.

Holt CC (2004) Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting 20* (1): 5-10.

Hu YC, Wu G and Jiang P (2021). Tourism demand forecasting using nonadditive forecast combinations. *Journal of Hospitality & Tourism Research*. doi: 10963480211047857.

Hyndman RJ, Koehler AB, Snyder RD and Grose S (2002) A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting 18*: 439-454.

Kourentzes N, Saayman A, Jean-Pierre P, Provenzano D, Sahli M, Seetaram N and Volo S (2021) Visitor arrivals forecasts amid COVID-19: A perspective from the Africa team. *Annals of Tourism*

Research 88: 103197.

LeSage JP and Magura M (1992) A mixture-model approach to combining forecasts. *Journal of Business & Economic Statistics 10* (4): 445-452.

Li G, Wu DC, Zhou M and Liu A (2019) The combination of interval forecasts in tourism. *Annals of Tourism Research 75*: 363-378.

Liu A, Vici L, Ramos V, Giannoni S and Blake A (2021) Visitor arrivals forecasts amid COVID-19: A perspective from the Europe team. *Annals of Tourism Research 88*: 103182.

Lobo GJ (1991) Alternative methods of combining security analysts' and statistical forecasts of annual corporate earnings. *International Journal of Forecasting 7* (1): 57-63.

MacDonald R and Marsh IW (1994) Combining exchange rate forecasts: What is the optimal consensus measure? *Journal of Forecasting 13* (3): 313-332.

Makridakis S and Winkler RL (1983) Averages of forecasts: Some empirical results. *Management Science 29* (9): 987-996.

Makridakis S, Spiliotis E and Assimakopoulos V (2018). The M4 Competition: Results, findings, conclusion and way forward. *International Journal of Forecasting 34(4):* 802–808.

Makridakis S, Spiliotis E and Assimakopoulos V (2020) The M4 Competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting 36(1):* 54–74.

Mieczkowski Z (1985) The tourism climatic index: a method of evaluating world climates for tourism. *The Canadian Geographer/Le Géographe canadien 29* (3): 220-233.

Newbold P and Granger CWJ (1974). Experience with forecasting univariate time series and the combination of forecasts. *Journal of the Royal Statistical Society: Series A (General) 137* (2): 131-146.

Page S, Song H and Wu DC (2012) Assessing the impacts of the global economic crisis and swine flu on inbound tourism demand in the United Kingdom. *Journal of Travel Research 51*(2):142-153.

Pesaran MH, Shin Y and Smith RJ (2001) Bounds testing approaches to the analysis of level relationships. *Journal of applied econometrics 16* (3): 289-326.

Phillips PC and Perron P (1988) Testing for a unit root in time series regression. *Biometrika 75 (2)*: 335-346.

Qian W, Rolling CA, Cheng G and Yang Y (2022) Combining forecasts for universally optimal performance. *International Journal of Forecasting 38(1):* 193-208.

Qiu RT, Wu DC, Dropsy V, Petit S, Pratt S and Ohe Y (2021) Visitor arrivals forecasts amid COVID-19: A perspective from the Asia and Pacific team. *Annals of Tourism Research 88:* 103155.

Scott D, Hall CM and Gossling S (2019) Global tourism vulnerability to climate change. *Annals of Tourism Research 77*: 49–61.

Sessions DN and Chatterjee S (1989) The combining of forecasts using recursive techniques with non-stationary weights. *Journal of Forecasting 8* (3): 239-251.

Shen S, Li G and Song H (2008) An assessment of combining tourism demand forecasts over different time horizons. *Journal of Travel Research 47* (2): 197-207.

Shen S, Li G and Song H (2011) Combination forecasts of international tourism demand. *Annals of Tourism Research 38* (1): 72-89.

Song H, Gartner WC and Tasci AD (2012) Visa restrictions and their adverse economic and marketing implications–Evidence from China. *Tourism Management 33 (2):* 397-412.

Song H, Qiu RT and Park J (2019) A review of research on tourism demand forecasting: Launching the Annals of Tourism Research Curated Collection on tourism demand forecasting. *Annals of Tourism Research 75:* 338–362.

Song H, Witt SF and Li G (2009) *The Advanced Econometrics of Tourism Demand*. New York: Routledge.

Song H, Witt SF, Wong KF and Wu DC (2009) An empirical study of forecast combination in tourism. *Journal of Hospitality & Tourism research 33* (1): 3-29.

Stock J and Watson M (1998) A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. *National Bureau of Economic Research Working Papers*. doi: 10.3386/w6607.

Stock J and Watson M (2003) How did leading indicator forecasts perform during the 2001 recession? *FRB Richmond Economic Quarterly 89* (3): 71-90.

Stock J and Watson M (2004) Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting 23* (6): 405-430.

Sun Y, Zhang J, Li X and Wang S (2021). Forecasting tourism demand with a new time-varying forecast averaging approach. *Journal of Travel Research*. doi: 00472875211061206.

The World Tourism Organization (UNWTO) (2018) UNWTO Tourism Highlights 2018 Edition. Available at: www.unwto.org/global/publication/unwto-tourism-highlights-2018-edition (accessed 10 July 2020)

Wan SK and Song H (2018) Forecasting turning points in tourism growth. *Annals of Tourism Research 72*: 156-167.

Wang Y (2009) The impact of crisis events and macroeconomic activity on Taiwan's international inbound tourism demand. *Tourism Management 30 (1):* 75-82.

Wong KKF, Song H, Witt SF and Wu DC (2007) Tourism forecasting: To combine or not to combine? *Tourism Management 28* (4): 1068-1078.

Wu J, Cheng X and Liao S (2020) Tourism forecast combination using the stochastic frontier analysis technique. *Tourism Economics 26* (7): 1086-1107.

Zhang X and Yu J (2018) Spatial weights matrix selection and model averaging for spatial autoregressive models. *Journal of Econometrics 203(1):* 1–18.

Álvarez-Díaz M and Rosselló-Nadal J (2010) Forecasting British tourist arrivals in the Balearic Islands using meteorological variables. *Tourism Economics* 16 (1):153-168.

Table 1: Variables and Data Sources

Variables	Explanation	Transformation / Formula	Data source
y_t	dependent variable: tourist arrivals for holiday purpose from one origin to the UK	logarithm	Travelpac
GDP_t	independent variable: tourists' income: real GDP in the origin	logarithm	Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org)
rp_t	independent variable: own price/relative price: exchange rate adjusted CPI of the UK to that of the origin country (OG)	logarithm, $lnrp_t = ln\left(\frac{\frac{CPI_{UK,t}}{EX_{UK,t}}}{\frac{CPI_{OG,t}}{EX_{OG,t}}}\right)$	Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org)
sp_t	independent variable: substitute price: the market share adjusted relative prices of alternative destinations	logarithm, $lnsp_t = ln\left(\frac{\sum_k^n rp_{k,t} * K_{k,t}}{\sum_k^n K_{k,t}}\right)$ $K_{k,t}$ represents the market share of competing destination k.	Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org); German National Tourist Board (GNTB); Directorate General for Enterprise

			(DGE)
Q_1, Q_2,Q_3	independent variable: seasonal dummies	To capture the instantaneous and delayed effects of these events, the dummies take values of 1 in the quarter the event happened and the following quarter, and they take values of 0 otherwise.	NA
D_{DS}	independent variable: one-off event dummy: the outbreak of foot-and-mouth disease at the beginning of 2001		
D_{911}	independent variable: one-off event dummy: the terrorist attack on 11^{th} of September, 2001		
D_{BM}	independent variable: one-off event dummy: the terrorist bombing in London on 7^{th} of June, 2005		
D_{2008}	independent variable: one-off event dummy: the global financial crisis in 2008		

D_{OL}		independent variable: one-off event dummy: the Olympic Games held in London in 2012.			
$TCI_{UK,t}$	$cid_{UK,t}$	independent variable: The UK's climate condition (the TCI of the UK)	daytime comfort index	logarithm, $TCI_{UK,t} = 2 \times \begin{bmatrix} 4(cid_{UK,t}) + cia_{UK,t} + 2(P_{UK,t}) + \\ 2(S_{UK,t}) + W_{UK,t} \end{bmatrix}$	Met Office
	$cia_{UK,t}$		daily comfort index		
	$P_{UK,t}$		index for precipitation		
	$S_{UK,t}$		index for sunshine duration		
	$W_{UK,t}$		index for wind speed		

Table 2: Summary of Individual Forecasting Models

Econometric Models		
Bounds Test Cointegration	$\Delta y_t = \alpha + \sum_{j=1}^{n}\sum_{i=0}^{p_j-1}\beta_{ji}\Delta x_{j,t-i} + \sum_{i=1}^{p-1}\phi_i \Delta y_{t-i} + \lambda_0 y_{t-1} + \sum_{j=1}^{n}\lambda_j x_{j,t-1} + u_t$	The climate factor is introduced as an explanatory variable in climate econometric models (represented by $x_{j,t}$ or x_t)
ADL	$y_t = \alpha + \sum_{j=1}^{n}\sum_{i=0}^{p_j}\beta_{j,i} x_{j,t-i} + \sum_{i=1}^{p}\phi_i y_{t-i} + \varepsilon_t$	
Leading Indicator	$y_t = \sum_{j=1}^{n}\sum_{i=1}^{p_j}\beta_{ji} x_{j,t-i} + \varepsilon_t$	
VAR	$y_t = \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \dots + \Pi_p y_{t-p} + H x_t + U_t$	
TVP	$y_t = x_t'\beta_t + \varepsilon_t, \quad \varepsilon_t \sim i.i.d.(0, H_t)$ $\beta_{t+1} = T_t\beta_t + \eta_t, \quad \beta_1 \sim N(b_1, P_1), \eta_t \sim i.i.d.(0, Q_t)$	
Simple Dynamic	$y_t = \sum_{j=1}^{n}\beta_{j0} x_{j,t} + \phi_1 y_{t-s} + \varepsilon_t$	
Time Series Techniques		

<table>
<tr><td>SARIMA</td><td>$\Phi_P(B^s)\phi_p(B)(1-B^s)^D(1-B)^d y_t = \mu + \Theta_Q(B^s)\theta_q(B)\varepsilon_t$</td><td rowspan="3">Non-causal models</td></tr>
<tr><td>ETS</td><td>$\hat{y}_{t+k} = (L_t + kT_t)S_{t+k-p}$
$L_t = \alpha\frac{y_t}{S_{t-p}} + (1-\alpha)(L_{t-1} + T_{t-1})$
$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$
$S_t = \delta\frac{y_t}{L_t} + (1-\delta)S_{t-p}$
or,
$\hat{y}_{t+k} = L_t + kT_t + S_{t+k-p}$
$L_t = \alpha(y_t - S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1})$
$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$
$S_t = \delta(y_t - L_t) + (1-\delta)S_{t-p}$</td></tr>
<tr><td>State Space ETS</td><td>$y_t = w'X_{t-1} + \varepsilon_t$
$X_t = FX_{t-1} + g\varepsilon_t$</td></tr>
</table>

Notes: Only the estimation equations for each model are provided here.

Table 3: The Component Models of Three Combination Groups

	Group A	Group B	Group C
6 traditional econometric models	✓	✓	
6 climate econometric models	✓		✓
3 time series models	✓	✓	✓

Table 4: Superior Percentages of Three Combination Groups for Each Origin (MAE)

	France	Germany	Irish Republic	Italy	the Netherlands	Spain	the US
SA							
Group A	***27.93%***	***29.67%***	29.50%	***18.40%***	***19.23%***	***29.65%***	***27.80%***
Group B	25.10%	28.54%	28.34%	17.18%	11.11%	26.25%	19.57%
Group C	14.39%	27.29%	***29.63%***	13.35%	18.73%	28.34%	12.25%
VACO							
Group A	***29.05%***	***29.71%***	***29.58%***	19.19%	22.08%	***29.70%***	***25.48%***
Group B	24.25%	28.83%	28.24%	***23.61%***	15.29%	27.04%	17.98%
Group C	16.98%	27.24%	29.38%	12.45%	***23.16%***	28.54%	12.40%
DMSFE (.85)							
Group A	***29.24%***	***29.64%***	***29.60%***	20.84%	21.43%	***29.78%***	***24.10%***
Group B	24.55%	28.74%	27.84%	***24.95%***	13.89%	27.29%	17.73%
Group C	17.93%	26.34%	29.33%	14.64%	***22.96%***	28.19%	11.70%
DMSFE (.90)							
Group A	***29.19%***	***29.66%***	***29.59%***	20.29%	21.66%	***29.76%***	***24.53%***
Group B	24.55%	28.74%	28.04%	***24.40%***	14.44%	27.29%	17.73%
Group C	17.58%	26.69%	29.38%	13.84%	***23.01%***	28.39%	11.95%
DMSFE (.95)							
Group A	***29.12%***	***29.69%***	***29.58%***	19.75%	21.87%	***29.73%***	***25.00%***
Group B	24.40%	28.74%	28.14%	***24.05%***	14.94%	27.14%	17.93%
Group C	17.33%	26.99%	29.33%	13.05%	***23.01%***	28.44%	12.15%
inverse-MAE							
Group A	***28.92%***	***29.76%***	29.65%	19.44%	***20.14%***	***29.67%***	***27.35%***
Group B	25.10%	28.93%	28.59%	***22.16%***	13.20%	26.64%	18.87%
Group C	15.29%	27.24%	***29.68%***	12.55%	19.72%	28.59%	12.20%

Note: Two decimal places are retained for all percentages for neatness of presentation and the superior percentages of the best-performing combination group for each combination method and each origin are highlighted.

Table 5: Superior Percentages of Three Combination Groups for Each Origin (MAPE)

	France	Germany	Irish Republic	Italy	the Netherlands	Spain	the US
SA							
Group A	26.39%	***29.36%***	***29.68%***	***19.79%***	***28.21%***	***29.68%***	19.24%
Group B	***27.04%***	27.14%	28.83%	17.08%	7.47%/	26.10%	***28.09%***
Group C	15.64%	24.15%	29.43%	14.94%	25.90%	28.09%	17.48%
VACO							
Group A	***27.87%***	***29.55%***	***29.76%***	18.61%	***28.32%***	***29.69%***	16.41%
Group B	26.49%	27.64%	29.33%	***22.71%***	8.22%/	26.94%	***27.84%***
Group C	20.32%	25.80%	29.13%	12.95%	26.69%	27.84%	15.74%
DMSFE (.85)							
Group A	***28.22%***	***29.44%***	***29.78%***	20.22%	***27.95%***	***29.79%***	17.15%
Group B	27.04%	27.59%	28.98%	***24.15%***	6.62%/	27.24%	***27.59%***
Group C	21.22%	25.20%	29.03%	14.99%	26.39%	27.59%	15.54%
DMSFE (.90)							
Group A	***28.14%***	***29.47%***	***29.78%***	19.69%	***28.06%***	***29.76%***	16.94%
Group B	26.89%	27.59%	29.08%	***23.56%***	7.02%/	27.14%	***27.69%***
Group C	21.02%	25.35%	29.03%	14.34%	26.54%	27.69%	15.59%
DMSFE (.95)							
Group A	***28.02%***	***29.51%***	***29.77%***	19.13%	***28.19%***	***29.73%***	16.69%
Group B	26.64%	27.59%	29.18%	***23.16%***	7.62%/	27.09%	***27.79%***
Group C	20.72%	25.50%	29.03%	13.65%	26.59%	27.79%	15.69%
inverse-MAE							
Group A	***27.81%***	***29.62%***	***29.81%***	19.67%	***28.44%***	***29.68%***	17.84%
Group B	27.04%	27.69%	29.33%	***21.71%***	8.72%/	26.49%	***28.29%***
Group C	17.53%	25.70%	29.48%	13.94%	26.34%	28.29%	16.24%

Note: Two decimal places are retained for all percentages for neatness of presentation and the superior percentages of the best-performing combination group for each combination method and each origin are highlighted.

Table 6: Superior Percentages of Three Combination Groups for Each Origin (RMSE)

	France	Germany	Irish Republic	Italy	the Netherlands	Spain	the US
SA							
Group A	26.02%	***29.82%***	29.76%	***27.95%***	***17.74%***	***29.86%***	***28.69%***
Group B	***27.94%***	27.29%	28.54%	26.49%	17.18%	26.64%	25.55%
Group C	19.47%	29.58%	***30.13%***	25.55%	14.99%	29.38%	13.45%
VACO							
Group A	***27.71%***	***29.78%***	29.84%	28.41%	***23.09%***	***29.81%***	***26.21%***
Group B	27.49%	27.34%	28.09%	***28.54%***	22.31%	27.34%	24.20%
Group C	22.36%	29.28%	***30.13%***	26.05%	21.41%	29.38%	12.70%
DMSFE (.85)							
Group A	***28.16%***	***29.77%***	29.83%	28.38%	***22.97%***	***29.83%***	***24.79%***
Group B	27.79%	27.09%	27.69%	***28.78%***	21.76%	27.64%	23.75%
Group C	23.85%	28.93%	***30.03%***	26.49%	21.17%	29.18%	12.20%
DMSFE (.90)							
Group A	***28.03%***	***29.77%***	29.83%	28.43%	***22.99%***	***29.82%***	***25.22%***
Group B	27.69%	27.24%	27.79%	***28.64%***	21.96%	27.54%	24.00%
Group C	23.41%	29.08%	***30.13%***	26.39%	21.31%	29.28%	12.30%
DMSFE (.95)							
Group A	***27.88%***	***29.78%***	29.84%	28.45%	***23.03%***	***29.81%***	***25.68%***
Group B	27.64%	27.29%	27.84%	***28.59%***	22.21%	27.34%	24.10%
Group C	22.96%	29.23%	***30.13%***	26.29%	21.36%	29.28%	12.50%
inverse-MAE							
Group A	27.58%	***29.81%***	29.85%	***28.37%***	***19.72%***	***29.84%***	***28.33%***
Group B	***27.99%***	27.54%	28.69%	28.19%	18.92%	27.19%	25.20%
Group C	*21.66%*	29.43%	***30.13%***	26.05%	17.08%	29.43%	13.40%

Note: Two decimal places are retained for all percentages for neatness of presentation and the superior percentages of the best-performing combination group for each combination method and each origin are highlighted.

Table 7: Superior Percentages of Three Combination Groups for Each Forecasting Horizon (MAE)

	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead
SA				
Group A	***23.68%***	***26.23%***	***27.16%***	***27.04%***
Group B	22.96%	20.94%	22.22%	23.05%
Group C	21.74%	20.89%	19.41%	20.23%
VACO				
Group A	***23.97%***	***26.66%***	***27.48%***	***27.48%***
Group B	23.02%	22.37%	23.16%	25.87%
Group C	22.42%	22.11%	20.06%	21.20%
DMSFE (.85)				
Group A	***23.51%***	***26.47%***	***27.64%***	***27.88%***
Group B	22.68%	22.43%	23.16%	26.01%
Group C	22.28%	22.37%	20.18%	21.51%
DMSFE (.90)				
Group A	***23.68%***	***26.55%***	***27.58%***	***27.72%***
Group B	22.77%	22.43%	23.13%	26.07%
Group C	22.31%	22.31%	20.12%	21.46%
DMSFE (.95)				
Group A	***23.83%***	***26.62%***	***27.53%***	***27.59%***
Group B	22.88%	22.43%	23.22%	25.95%
Group C	22.31%	22.20%	20.06%	21.31%
inverse MAE				
Group A	***24.32%***	***26.44%***	***27.23%***	***27.68%***
Group B	23.79%	21.77%	22.68%	25.18%
Group C	21.94%	21.46%	19.10%	20.52%

Note: Two decimal places are retained for all percentages for neatness of presentation and the superior percentages of the best-performing combination group for each combination method and each forecasting horizon are highlighted.

Table 8: Superior Percentages of Three Combination Groups for Each Forecasting Horizon (MAPE)

	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead
SA				
Group A	***25.83%***	***25.69%***	***25.90%***	***26.78%***
Group B	23.39%	22.88%	23.93%	22.22%
Group C	21.80%	22.68%	22.14%	22.31%
VACO				
Group A	***25.30%***	***25.49%***	***25.45%***	***26.74%***
Group B	23.85%	23.59%	24.62%	24.62%
Group C	22.23%	22.97%	22.51%	22.85%
DMSFE (.85)				
Group A	***25.30%***	***25.67%***	***26.06%***	***27.28%***
Group B	23.68%	23.65%	24.59%	24.79%
Group C	22.11%	23.22%	22.85%	23.22%
DMSFE (.90)				
Group A	***25.32%***	***25.63%***	***25.87%***	***27.09%***
Group B	23.73%	23.62%	24.50%	24.70%
Group C	22.23%	23.16%	22.77%	23.02%
DMSFE (.95)				
Group A	***25.31%***	***25.57%***	***25.66%***	***26.91%***
Group B	23.79%	23.59%	24.56%	24.67%
Group C	22.23%	23.08%	22.65%	22.88%
Inverse-MAE				
Group A	***26.34%***	***25.56%***	***25.48%***	***27.12%***
Group B	24.30%	23.56%	24.76%	24.10%
Group C	22.57%	22.94%	21.86%	22.65%

Note: Two decimal places are retained for all percentages for neatness of presentation and the superior percentages of the best-performing combination group for each combination method and each forecasting horizon are highlighted.

Table 9: Superior Percentages of Three Combination Groups for Each Forecasting Horizon (RMSE)

	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead
SA				
Group A	***25.28%***	***27.26%***	***27.48%***	***28.47%***
Group B	25.18%	24.90%	26.78%	25.78%
Group C	24.10%	23.45%	21.51%	23.82%
VACO				
Group A	***25.56%***	***28.01%***	***28.51%***	***29.26%***
Group B	25.18%	25.58%	27.92%	27.21%
Group C	24.59%	24.62%	23.39%	25.30%
DMSFE (.85)				
Group A	***25.37%***	***27.68%***	***28.45%***	***29.20%***
Group B	24.87%	25.61%	27.86%	27.09%
Group C	24.53%	24.59%	23.68%	25.41%
DMSFE (.90)				
Group A	***25.43%***	***27.79%***	***28.48%***	***29.22%***
Group B	25.04%	25.61%	27.83%	27.15%
Group C	24.56%	24.64%	23.65%	25.38%
DMSFE (.95)				
Group A	***25.49%***	***27.90%***	***28.50%***	***29.24%***
Group B	25.07%	25.61%	27.89%	27.15%
Group C	24.62%	24.64%	23.56%	25.33%
inverse-MAE				
Group A	***26.07%***	***27.73%***	***27.82%***	***28.94%***
Group B	25.58%	25.41%	27.15%	26.84%
Group C	24.70%	24.05%	22.08%	24.70%

Note: Two decimal places are retained for all percentages for neatness of presentation and the superior percentages of the best-performing combination group for each combination method and each forecasting horizon are highlighted.

Table 10: Superior Percentages of Three Combination Groups for Each Weighting Scheme

		SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE
MAE	Group A	***26.03%***	***26.40%***	***26.38%***	***26.38%***	***26.39%***	***26.42%***
	Group B	22.30%	23.61%	23.57%	23.60%	23.62%	23.36%
	Group C	20.57%	21.45%	21.59%	21.55%	21.47%	20.75%
MAPE	Group A	***26.05%***	***25.74%***	***26.08%***	***25.98%***	***25.86%***	***26.12%***
	Group B	23.11%	24.17%	24.17%	24.14%	24.15%	24.18%
	Group C	22.23%	22.64%	22.85%	22.79%	22.71%	22.50%
RMSE	Group A	***27.12%***	***27.84%***	***27.67%***	***27.73%***	***27.78%***	***27.64%***
	Group B	25.66%	26.47%	26.36%	26.41%	26.43%	26.25%
	Group C	23.22%	24.47%	24.55%	24.56%	24.54%	23.88%

Note: Two decimal places are retained for all percentages for neatness of presentation and the superior percentages of the best-performing combination group for each combination method are highlighted.

Table 11: The D-M Test Results

	1-step-ahead	2-step-ahead	3-step-ahead	4-step-ahead
SA				
Group A	*22.66%* *6.02%*	*22.75%* *5.75%*	*22.64%* *6.07%*	*23.40%* *5.71%*
Group B	21.31% 7.17%	21.51% 6.97%	21.71% 6.17%	21.91% 6.77%
Group C	20.91% 6.57%	21.91% 6.77%	21.51% 6.97%	22.31% 6.57%
VACO				
Group A	*21.81%* *6.79%*	*21.70%* *6.10%*	*21.89%* *6.57%*	*22.47%* *6.26%*
Group B	20.91% 6.77%	21.51% 7.37%	21.12% 6.57%	21.12% 6.77%
Group C	20.91% 6.97%	21.71% 7.77%	19.72% 6.97%	21.51% 6.57%
DMSFE (.85)				
Group A	*22.41%* *6.09%*	*22.50%* *6.02%*	*22.13%* *5.48%*	*22.98%* *4.94%*
Group B	22.50% 6.57%	22.70% 6.97%	21.91% 6.77%	22.90% 6.57%
Group C	20.91% 6.57%	20.31% 6.77%	21.12% 6.97%	21.51% 6.17%
DMSFE (.90)				
Group A	*22.29%* *6.30%*	*21.86%* *6.71%*	*22.52%* *6.32%*	*22.70%* *5.46%*
Group B	21.12% 7.17%	20.91% 6.77%	21.12% 6.57%	21.51% 6.57%
Group C	20.71% 7.17%	21.51% 7.17%	21.91% 6.97%	22.11% 7.17%

DMSFE (.95)				
Group A	*23.41%* *5.94%*	*22.77%* *6.00%*	*22.65%* *5.96%*	*23.61%* *5.60%*
Group B	20.52% 7.37%	20.71% 6.77%	21.91% 6.77%	22.31% 6.57%
Group C	21.31% 6.57%	21.91% 6.18%	20.71% 6.57%	22.11% 6.37%
inverse-MAE				
Group A	*22.99%* *6.04%*	*22.61%* *5.93%*	*22.85%* *6.17%*	*23.23%* *5.99%*
Group B	21.51% 6.77%	21.71% 6.77%	21.31% 7.17%	21.91% 6.37%
Group C	21.51% 6.97%	21.51% 6.37%	21.71% 6.77%	21.91% 6.57%

Note: Each cell of the table comprises two entries. The first entry is the percentage of times the Null of equal forecast accuracy between the combination model and the best single model is rejected at a 5% level of significance with a negative statistic, i.e., forecasts from the combination model are significantly more accurate than the best single model. The second entry is the percentage of times the Null of equal forecast accuracy between the combination model and the best single model is rejected at a 5% level of significance with a positive statistic, i.e., forecasts from the combination model are significantly less accurate than the best single model.

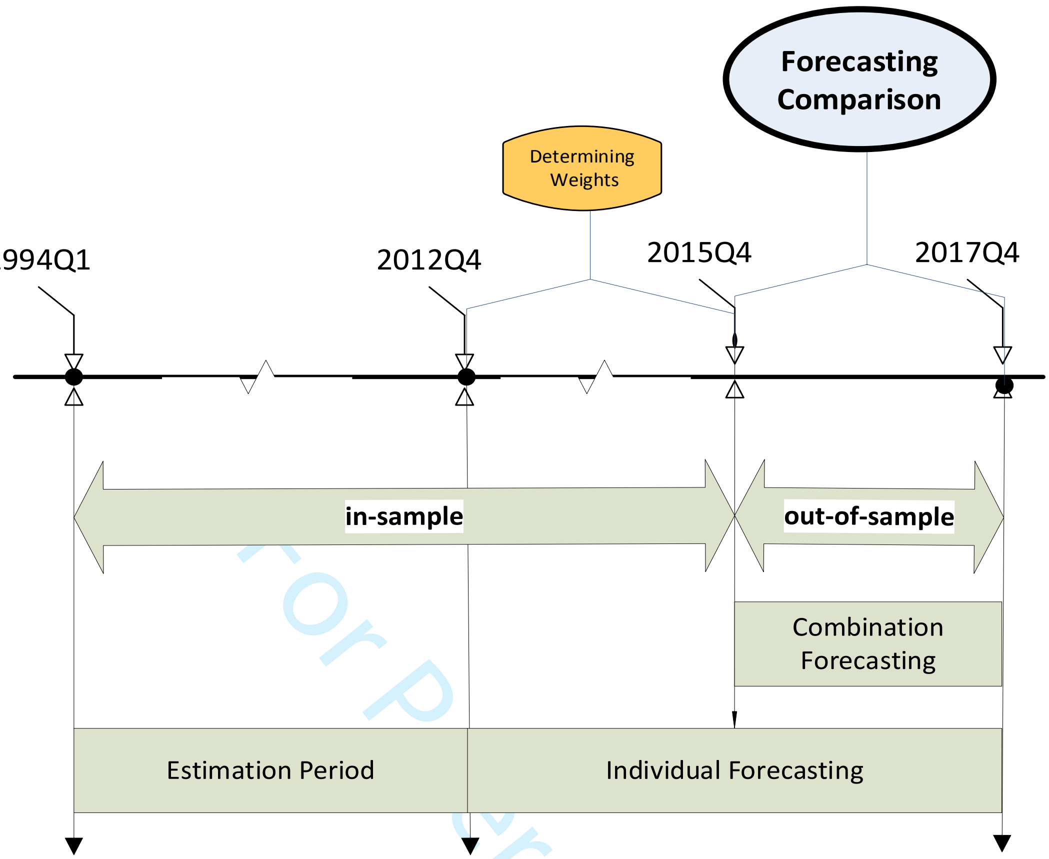

Figure 1: Illustration of Data Sample

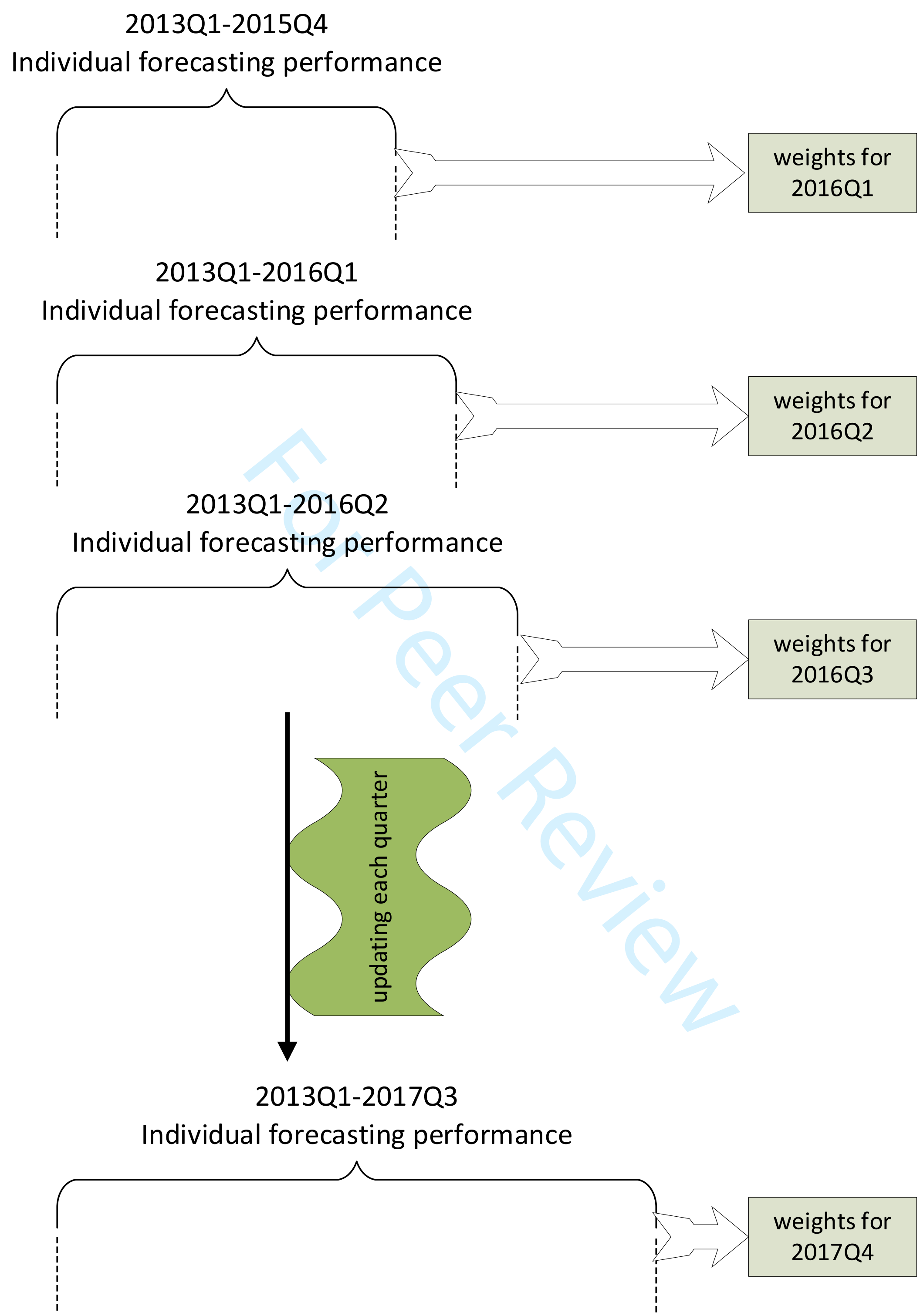

Figure 2 The Recursive Weighting Procedure for One-Step-Ahead Combination Forecasts

Notes: The weights for two- to four-step-ahead combination forecasts are generated iteratively in the same way.

Unit Root Test Results

Table S1 Unit Root Test Results

		France	Germany	Irish Republic	Italy	The Netherlands	Spain	The US
lny_t	Level	-6.25***	-8.53***	-8.32***	-5.37***	-9.35***	-4.64***	-9.37***
	1st difference	-22.15***	-16.79***	-20.40***	-31.21***	-21.75***	-25.70***	-14.85***
$lnGDP_t$	Level	-1.46	-1.10	-0.68	-5.51***	-1.65	-2.06	-1.15
	1st difference	-18.06***	-10.92***	-10.44***	-30.73***	-21.08***	-25.15***	-11.09***
$lnrp_t$	Level	-2.05	-2.06	-1.57	-1.69	-1.88	-1.48	-1.77
	1st difference	-9.61***	-9.80***	-9.89***	-9.71***	-9.56***	-9.42***	-8.29***
$lnsp_t$	Level	-1.79	-1.80	-1.71	-1.81	-1.61	-1.60	-1.63
	1st difference	-11.84***	-11.85***	-7.08***	-7.52***	-9.00***	-14.55***	-6.81***
$lnTCI_{UK,t}$	Level				-14.45***			
	1st difference				-18.07***			

Notes: Two decimal places are retained for all values. The superscript * means significant at the 10% level, ** means significant at the 5% level, and *** means significant at the 1% level.

Weights in VACO and DMSFE Combination Methods

Weights in the VACO method are calculated to minimise the error variance of the combination forecasts (assuming unbiasedness of each single forecast), which can be expressed as:

$$w_{it} = \frac{\left[\sum_{t=1}^{T} e_{it}^2\right]^{-1}}{\mathrm{sum}_{j=1}^{n}\left[\sum_{t=1}^{T} e_{jt}^2\right]^{-1}} \quad (1)$$

where w_{it} is the weight of the ith individual forecast, e_{it} is the estimated individual forecast error for the ith individual forecast, e_{jt} denotes the estimated individual forecast error for any individual forecasts, T is the sample size, and n is the number of single constituents in the combination panel. Because there exist correlations among the forecast errors, negative weights may appear in some cases.

Weights in the DMSFE scheme can be written as:

$$w_{it} = \frac{\phi_{it}^{-1}}{\sum_{j=1}^{n}\phi_{jt}^{-1}}, \; \phi_{it} = \sum_{s=1}^{T-1} \alpha^{t-1-s}\left(f_{s+h} - \hat{f}_{s+h,s}^{i}\right)^2 \quad (2)$$

where $\hat{f}_{s+h,s}^{i}$ is the individual h-step-ahead forecast generated at time s from single model i for the actual value f_{s+h}, α is a selected discounting factor with $0 < \alpha \ll 1$. The smaller the value of α is, the heavier the more recent forecasts are weighted. Following the current literature, the DMSFE combination forecasts are computed for three values of α: $\alpha = 0.95, 0.90, 0.85$ (Shen et al., 2008, 2011).

Forecasting Performance of Individual Models

All MAE, MAPE and RMSE ratios of the 15 individual models have been normalized relative to the seasonal naive no-change forecasts with ratios below one indicating predictive gains relative to the benchmark forecast. The comparisons are conducted in the market-specific, the forecasting-horizon-specific and the general ways and the results are presented in Table S2, S3 and S4 respectively

Table S2 Forecasting Performance of Individual Models for Seven Origins

origin	bounds©	ETS	SARIMA	ADLM©	State ETS	LI©	VAR©	TVP©	SD©	bounds	ADLM	LI	VAR	TVP	SD
MAE ratios															
France	0.648	0.881	0.697	0.848	0.714	1.380	0.916	0.895	1.009	0.784	0.785	0.936	0.742	0.784	0.876
	1	10	2	8	3	15	12	11	14	6	7	13	4	5	9
Germany	0.646	0.480	0.720	0.817	0.489	0.959	0.889	0.727	0.679	0.570	0.663	0.540	0.531	0.649	0.594
	7	1	11	13	2	15	14	12	10	5	9	4	3	8	6
Irish Republic	0.567	0.628	0.646	0.746	0.629	0.414	0.617	0.711	0.486	0.751	0.905	0.650	0.564	0.563	0.663

	5	7	9	13	8	1	6	12	2	14	15	10	4	3	11
Italy	0.870	0.619	0.677	0.669	0.606	0.796	0.784	0.863	0.988	0.811	0.794	1.270	0.904	0.860	0.865
	12	2	4	3	1	7	5	10	14	8	6	15	13	9	11
the	0.986	0.718	0.723	0.800	0.719	1.128	0.939	1.040	0.998	0.931	0.891	0.745	0.958	0.922	1.039
Netherlands	11	1	3	5	2	15	9	14	12	8	6	4	10	7	13
Spain	0.859	0.764	0.641	0.669	0.618	0.744	0.745	0.756	0.890	0.796	0.820	0.875	0.714	0.561	0.712
	13	10	3	4	2	7	8	9	15	11	12	14	6	1	5
US	0.855	0.648	0.609	0.604	0.716	0.976	0.810	0.863	0.836	0.870	0.661	0.801	0.815	0.830	0.711
	12	3	2	1	6	15	8	13	11	14	4	7	9	10	5
MAPE ratios															
France	0.684	0.954	0.753	0.840	0.778	1.421	0.955	0.956	1.007	0.857	0.800	1.029	0.789	0.897	0.864
	1	10	2	6	3	15	11	12	13	7	5	14	4	9	8
Germany	0.598	0.439	0.648	0.752	0.449	1.018	0.770	0.620	0.614	0.511	0.587	0.475	0.501	0.542	0.517

	9	1	12	13	2	15	14	11	10	5	8	3	4	7	6
Irish Republic	0.577	0.690	0.750	0.821	0.698	0.339	0.754	0.939	0.454	0.692	1.109	0.663	0.534	0.538	0.631
	5	8	11	13	10	1	12	14	2	9	15	7	3	4	6
Italy	0.955	0.675	0.727	0.727	0.646	0.891	0.827	0.911	1.075	0.872	0.845	1.356	0.994	0.943	0.950
	12	2	3	4	1	8	5	9	14	7	6	15	13	10	11
the Netherlands	0.985	0.833	0.840	0.840	0.803	1.333	0.973	1.126	1.014	0.952	0.867	0.814	1.034	1.035	1.059
	9	3	5	4	1	15	8	14	10	7	6	2	11	12	13
Spain	0.975	0.873	0.750	0.746	0.736	0.818	0.856	0.858	1.029	0.934	0.930	0.961	0.831	0.693	0.847
	14	10	4	3	2	5	8	9	15	12	11	13	6	1	7
US	0.838	0.688	0.626	0.668	0.719	0.873	0.801	0.859	0.805	0.909	0.649	0.736	0.766	0.773	0.659
	12	5	1	4	6	14	10	13	11	15	2	7	8	9	3
RMSE ratios															
France	0.665	0.910	0.743	0.971	0.760	1.463	1.005	0.938	1.163	0.814	0.909	0.944	0.759	0.798	0.983

	1	8	2	11	4	15	13	9	14	6	7	10	3	5	12
Germany	0.691	0.612	0.805	0.821	0.621	0.958	0.928	0.783	0.751	0.682	0.738	0.733	0.651	0.780	0.711
	5	1	12	13	2	15	14	11	9	4	8	7	3	10	6
Irish Republic	0.587	0.654	0.674	0.803	0.655	0.479	0.562	0.665	0.519	0.814	1.010	0.699	0.532	0.541	0.705
	6	7	10	13	8	1	5	9	2	14	15	11	3	4	12
Italy	1.268	0.924	0.992	1.053	0.911	1.305	1.262	1.284	1.452	1.149	1.257	1.763	1.257	1.208	1.273
	10	2	3	4	1	13	9	12	14	5	7	15	8	6	11
the	0.957	0.671	0.671	0.774	0.668	1.012	0.908	0.991	0.997	0.916	0.883	0.758	0.898	0.828	1.023
Netherlands	11	2	3	5	1	14	9	12	13	10	7	4	8	6	15
Spain	0.886	0.850	0.731	0.740	0.700	0.848	0.868	0.877	0.948	0.889	0.871	0.930	0.756	0.601	0.819
	12	8	3	4	2	7	9	11	15	13	10	14	5	1	6
US	1.073	0.736	0.657	0.660	0.802	1.092	1.053	1.184	0.869	1.154	0.782	0.935	0.957	1.014	0.852
	12	3	1	2	5	13	11	15	7	14	4	8	9	10	6

Notes: There are two entries in each cell with the first ones being accuracy measure ratios and the second ones being ranks. Three decimal places are retained for all ratios and the ratios and ranks of the top three individual models are highlighted for each comparison.

Table S3 Forecasting Performance of Individual Models for Four Forecasting Horizons

horizon	bounds©	ETS	SARIMA	ADLM©	State ETS	LI©	VAR©	TVP©	SD©	bounds	ADLM	LI	VAR	TVP	SD
	MAE ratios														
1	0.750 8	0.675 3	0.635 1	0.697 4	0.639 2	0.911 15	0.761 11	0.846 14	0.816 13	0.754 9	0.744 7	0.790 12	0.731 5	0.739 6	0.758 10
2	0.806 10	0.683 3	0.676 2	0.728 4	0.635 1	0.914 15	0.847 12	0.850 13	0.857 14	0.787 9	0.776 7	0.820 11	0.767 6	0.745 5	0.779 8
3	0.777 7	0.677 2	0.687 3	0.750 6	0.664 1	0.909 15	0.825 11	0.832 12	0.845 13	0.798 9	0.803 10	0.853 14	0.745 5	0.736 4	0.790 8
4	0.770 7	0.673 2	0.695 3	0.768 6	0.628 1	0.921 15	0.824 11	0.818 10	0.845 13	0.811 9	0.830 12	0.860 14	0.744 5	0.734 4	0.792 8
	MAPE ratios														
1	0.783	0.729	0.691	0.751	0.683	0.964	0.798	0.909	0.839	0.785	0.802	0.833	0.777	0.785	0.779

	7	3	2	4	1	15	10	14	13	8	11	12	5	9	6
2	0.831	0.753	0.730	0.759	0.696	0.948	0.880	0.904	0.870	0.802	0.813	0.850	0.790	0.776	0.785
	10	3	2	4	1	15	13	14	12	8	9	11	7	5	6
3	0.789	0.737	0.739	0.776	0.706	0.946	0.857	0.889	0.856	0.818	0.833	0.875	0.773	0.768	0.793
	7	2	3	6	1	15	12	14	11	9	10	13	5	4	8
4	0.804	0.725	0.768	0.795	0.673	0.967	0.859	0.880	0.863	0.868	0.859	0.889	0.774	0.751	0.802
	8	2	4	6	1	15	9	13	11	12	10	14	5	3	7
	RMSE ratios														
1	0.863	0.767	0.708	0.794	0.741	1.017	0.917	1.011	0.916	0.898	0.870	0.927	0.827	0.850	0.874
	7	3	1	4	2	15	12	14	11	10	8	13	5	6	9
2	0.905	0.771	0.764	0.820	0.705	1.015	0.972	0.982	0.960	0.919	0.902	0.948	0.850	0.822	0.898
	9	3	2	4	1	15	13	14	12	10	8	11	6	5	7
3	0.863	0.747	0.766	0.849	0.758	1.024	0.945	0.937	0.965	0.910	0.941	0.987	0.820	0.814	0.925

	7	1	3	6	2	15	12	10	13	8	11	14	5	4	9
4	0.870	0.775	0.775	0.865	0.720	1.033	0.929	0.911	0.987	0.940	0.972	1.002	0.823	0.811	0.941
	7	3	2	6	1	15	9	8	13	10	12	14	5	4	11

Notes: There are two entries in each cell with the first ones being the accuracy measure ratios and the second ones being the ranks. Three decimal places are retained for all ratios and the ratios and ranks of the top three individual models are highlighted for each comparison.

Table S4 General Forecasting Performance of Individual Models

measure	bounds©	ETS	SARIMA	ADLM©	State ETS	LI©	VAR©	TVP©	SD©	bounds	ADL	LI	VAR	TVP	SD
MAE ratios	0.776/7	0.677/3	0.673/2	0.736/4	0.641/1	0.914/15	0.814/11	0.836/13	0.841/14	0.788/9	0.789/10	0.831/12	0.747/6	0.738/5	0.780/8
MAPE ratios	0.802/8	0.736/3	0.728/2	0.770/4	0.690/1	0.956/15	0.848/11	0.896/14	0.857/12	0.818/9	0.827/10	0.862/13	0.778/6	0.774/5	0.790/7
RMSE ratios	0.875/7	0.765/3	0.753/2	0.832/6	0.731/1	1.022/15	0.941/11	0.960/13	0.957/12	0.917/9	0.921/10	0.966/14	0.830/5	0.824/4	0.909/8

Notes: There are two entries in each cell with the first ones being the accuracy measure ratios and the second ones being the ranks. Three decimal places are retained for all ratios and the ratios and ranks of the top three individual models are highlighted for each comparison.