Does The Inclusion of Climate Variables Improve Tourism Demand Forecasting Performance?

XI WU

Faculty of Management

July 2019

Does The Inclusion of Climate Variables Improve Tourism Demand Forecasting Performance?

by

XI WU

Abstract

The aim of this study is to assess whether incorporating climate variables in econometric and combination forecasting models can improve tourism demand forecasting performance. Climate conditions are important tourism resources which can influence tourists' decision as to when and where to travel, however, our understanding of the value of climate variables in forecasting tourism demand is limited. The current research fills this gap through empirical studies on UK's international tourism demand.

Inbound tourism demand to the UK from seven leading markets, namely, France, Germany, Irish Republic, Italy, the Netherlands, Spain and the US are studied respectively based on quarterly time series data from 1994Q1 to 2017Q4. The bounds test cointegration approach is applied to assess the long-run relationships between tourism demand and its influencing factors and to evaluate the impact of climate on tourism demand.

Individual tourism demand forecasts are generated through both causal econometric and non-causal time series models, which are popular in the current tourism demand literature. Causal econometric models, which consist of the bounds test cointegration approach, the autoregressive distributed lag model (ADLM), the leading indicator (LI) model, the vector autoregressive (VAR) model, the time-varying parameter (TVP) model and the simple dynamic (SD) model, take two model specifications, which are different in identified influencing factors. Econometric models that only consider economic factors as demand determinants are named as traditional econometric models, and the others that include the climate factor as a demand determinant are called climate econometric models. Non-causal time series techniques consist of the seasonal naïve no-change model, the seasonal autoregressive integrated moving average (SARIMA) model, the exponential smoothing (ETS) model and the state space ETS model. One- to four-step-ahead out-of-sample single forecasts are generated from every individual forecasting model through the recursive forecasting procedure with the seasonal naïve no-change model serving as the benchmark.

Except the naïve model, all other individual forecasting models are selected as candidate constituents for combination. For combination forecasting, the 15 selected individual models are categorized into three groups. The first group includes all individual models; the second one contains traditional econometric and time series models; and the third category consists of climate econometric and time series models. Combination is conducted for each group respectively, resulting in three sets of combination forecasts: the first set is generated through combining 15 individual models; the second and third ones are produced from integrating 9 individual models. Different combination methods are applied including the simple average (SA) method, the variance-covariance (VACO) method, the discounted mean square forecast error (DMSFE) ($\alpha = 0.85/0.90/0.95$) methods, as well as the newly-introduced inverse-MAE and the two-stage combination approaches.

Comprehensive comparisons of the predictive powers of the individual and combination forecasting approaches for seven origins and four forecasting horizons are conducted based on three accuracy measures including mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean square error (RMSE). The results show that individual model's forecasting performance varies greatly according to the origin market under consideration. No single model can perform the best in all cases, and in most cases, more advanced individual models forecast better than the naïve benchmark. In general, non-causal time series techniques are superior to causal econometric models. Whether including the climate factor can improve the forecasting accuracy of econometric models should be evaluated case-by-case.

With respect to the forecasting ability of the combination approach, it is demonstrated that combining individual forecasts is beneficial regardless of origin country under study, forecasting horizon under consideration, accuracy measure used, combination methods applied or combination group under analysis. In all cases, there are always a portion of combination forecasts that are more accurate than the best single projections, and the worst forecasts are always produced by individual forecasting models. It means that the combination forecasting approach is superior to the individual one, as it can improve forecasting accuracy and reduce forecasting failure. Comparisons among alternative combination methods show that no single combination method can provide the best composite forecasts in all situations. The newly-introduced inverse-MAE scheme performs quite well, but the two-stage combination methods behave unsatisfactorily.

Comparisons among three combination groups reveal that, generally, combining all individual models, which include traditional econometric models, climate econometric models and time series techniques produce the best combination forecasts, which means that combining econometric models with different influencing factors and introducing climate variables into combination can contribute to more accurate projections. It implies that through combining, diversity gain can be achieved not only by incorporating different modelling techniques but also by integrating different model specifications.

Regarding which and how many models to combine, it is shown that individual models' frequencies to constitute the superior combination forecasts are irrelevant to their forecasting abilities. More accurate individual forecasts do not have higher opportunities to construct superior composite projections. The number of single constituents in the best forecasts range from two to six, and for most origins, combining two individual models can bring about the most accurate projections.

To the best of my knowledge, this research represents the first effort to evaluate the combination forecasting approach which consider econometric models with different explanatory variables as candidate constituents, and climate variables have been, for the first time, introduced to the combination forecasts. It proves that better combination forecasts can be obtained by integrating econometric models with different influencing factors, and the value of non-economic explanatory variables in combination forecasting deserves more attention. It is suggested that a user-friendly software for combination forecasting should be made available and combination forecasts should be included in forecasting comparisons considering the general superiority of the combination forecasting approach compared to the single forecasting method.

Acknowledgement

At the final stage of my PhD journey, which is one of the most challenging, rewarding and memorable explorations in my life, I feel proud of myself and extremely grateful to others.

I would not be able to get to this stage without the support from my family. Thank you to my mum and dad, who, although do not think it is a good idea for me to pursue a PhD, always come to my rescue without any hesitation. I could not have done it without you and am very grateful.

Thank you to Yong, my husband, who is tolerant, kind and considerate most of the time and works very hard to support our family. It means more to me than I can possibly express.

Thank you to Alex, my big boy, who, although can be a nightmare to me sometimes, has always been here for me throughout the whole journey and warms my heart when I am lonely and desperate. Thank you to George, my little boy, who, although has never failed to wake me up in the middle of the nights since he was born, cures any negative emotions with his innocent smiles and big hugs. Alex and George, thank you for coming to my life. The deep love I feel towards and from you motivates and inspires me every day.

I would like to thank my aunt, who has taken the trouble to help me to look after George. Without her kindness, I would not be able to survive the first year.

I would like to express my great gratitude to my supervisors, Professor Adam Blake and Professor Shuang Cang. They have always supported and encouraged me throughout my PhD journey, offering fundamental guidance to me, making valuable suggestions for my research and providing important feedback on my thesis.

I would also like to thank Dr. Neelu Seetaram for her kindness and encouragement to me.

And thank you to Chris Pitt from IT department, who helped me a lot with the formatting of my thesis.

Finally, I acknowledge the financial support provided by Erasmus Mundus, as this research is mainly funded by Erasmus Mundus Fusion program (545831-EM-1-2013-1-IT-ERAMUNDUSEMA21), and I would like to extend my special thanks to those who worked on the program.

Contents

List of Tables

List of Figures

List of Abbreviations

2SLS: TWO STAGE LEAST SQUARES REGRESSION

ADE: ADAPTIVE DIFFERENTIAL EVOLUTION ALGORITHM

ADF: AUGMENTED DICKEY FULLER

ADLM: AUTOREGRESSIVE DISTRIBUTED LAG MODEL

AGARCH: ASYMMETRIC GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTIC

AI: ARTIFICIAL INTELLIGENCE

AIC: AKAIKE INFORMATION CRITERION

AIDS: ALMOST IDEAL DEMAND SYSTEM

ANFIS: ADAPTIVE NETWORK-BASED FUZZY INFERENCE SYSTEM

ANN: ARTIFICIAL NEURAL NETWORK

APE: ABSOLUTE PERCENTAGE ERROR

AR: AUTOREGRESSIVE

ARAR: AUTOREGRESSIVE AUTOREGRESSIVE MOVING AVERAGE

ARCH: AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTICITY

ARFIMA: AUTOREGRESSIVE FRACTIONAL INTEGRATED MOVING AVERAGE

ARIMA: AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

ARIMAX: AUTOREGRESSIVE INTEGRATED MOVING AVERAGE WITH EXPLANATORY VARIABLES

ARMA: AUTOREGRESSIVE MOVING AVERAGE

ARMAX: AUTOREGRESSIVE MOVING AVERAGE WITH EXPLANATORY VARIABLES

BPNN: BACK-PROPAGATION NEURAL NETWORK

BSM: NON-CAUSAL BASIC STRUCTURAL MODEL

BVAR: BAYESIAN VECTOR AUTOREGRESSIVE

CCC: CONSTANT CONDITIONAL CORRELATION

CCE-MG: COMMON CORRELATED EFFECT MEAN GROUP

CGE: COMPUTABLE GENERAL EQUILIBRIUM MODEL

CI: COINTEGRATION

CLSDV: CORRECTED LEAST SQUARE DUMMY VARIABLES

CMCSGM: CUCKOO-MARKOV CHAIN-SEGMENT GREY MODEL

CP: CUBIC POLYNOMIAL MODEL

CPI: CONSUMER PRICE INDEX

CSM: CAUSAL STRUCTURAL MODEL

CUSUM: CUMULATIVE SUM CONTROL CHART

DBT: DRY-BULB TEMPERATURE

DC: DIRECTION OF CHANGE

DCA TEST: DIRECTIONAL CHANGE ACCURACY TEST

DCC: DYNAMIC CONDITIONAL CORRELATION

DF: DICKEY FULLER

DFE: DYNAMIC FIXED EFFECTS

DGP: DATA GENERATING PROCESS

D-M TEST: DIEBOLD-MARIANO TEST

DMSFE: DISCOUNTED MEAN SQUARE FORECAST ERROR COMBINATION

DOLS: DYNAMIC ORDINARY LEAST SQUARES

ECM: ERROR CORRECTION MODEL

EG-ECM: ENGLE-GRANGER ECM

EGARCH: EXPONENTIAL GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTIC

ETS: EXPONENTIAL SMOOTHING

FE: FIXED EFFECT

FMOLS: PANEL FULLY MODIFIED ORDINARY LEAST SQUARES

FTS: FUZZY TIME SERIES

GARCH: GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTIC

GDFM: GENERALIZED DYNAMIC FACTOR MODEL

GDP: GROSS DOMESTIC PRODUCT

GMM: GENERAL METHOD OF MOMENTS

GPR: GAUSSIAN PROCESS REGRESSION

GFS: GENETIC FUZZY SYSTEM

GJR: GLOSTEN, JAGANNATHAN & RUNKLE

GNP: GROSS NATIONAL PRODUCT

GRM: GROWTH RATE MODEL

GVAR: GLOBAL VECTOR AUTOREGRESSIVE

HIER: HIERARCHICAL FORECAST COMBINATION

HLN TEST: HARVEY, LEYBOURNE, NEWBOLD TEST

IID: INDEPENDENT AND IDENTICALLY DISTRIBUTED

INV-MES: INVERSE OF THE MEAN SQUARE ERROR COMBINATION

JML-ECM: JOHANSEN MAXIMUM LIKELIHOOD ECM

LAIDS: LINEAR ALMOST IDEAL DEMAND SYSTEM

LI: LEADING INDICATOR

MA: MOVING AVERAGE

MAE: MEAN ABSOLUTE ERROR

MAFE: MEAN ABSOLUTE FORECASTING ERROR

MAPE: MEAN ABSOLUTE PERCENTAGE ERROR

MASE: MEAN ABSOLUTE SCALED ERROR

MASSFE: MEAN ABSOLUTE SUM OF SQUARES FORECASTING ERROR

MDASE: MEDIAN ABSOLUTE SCALED ERROR

MEDAPE: MEDIAN ABSOLUTE PERCENTAGE ERROR

MFE: MEAN FORECASTING ERROR

MGARCH: MULTIVARIATE GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTIC

MIDAS: MIXED DATA SAMPLING

ML: MAXIMUM LIKELIHOOD

MLP: MULTI-LAYER PERCEPTION

MSDR: MARKOV SWITCHING DYNAMIC REGRESSION

MSE: MEAN SQUARED ERROR

MSSA: MULTIVARIATE SINGULAR SPECTRUM ANALYSIS

NMSE: NORMALIZED MEAN SQUARE ERROR

NN: NEURAL NETWORK

OLS: ORDINARY LEAST SQUARES

ONS: OFFICE FOR NATIONAL STATISTICS

PB: PERCENTAGE BETTER

PCA: PRINCIPAL COMPONENT ANALYSIS

PGR: POSTGRADUATE RESEARCHER

PMG: POOLED MEAN GROUP ESTIMATOR

POLS: POOLED ORDINARY LEAST SQUARE

PP: PHILIPS-PERRON

CQR: (COUNT) QUANTILE REGRESSION

QADL: QUANTILE AUTOREGRESSIVE DISTRIBUTED LAG MODEL

RANK: RANK BASED WEIGHTING COMBINATION

RBF: RADIAL BASIS FUNCTION

RE: RANDOM EFFECT

RMSE: ROOT MEAN SQUARED ERROR

RMSFE: ROOT MEAN SQUARED FORECASTING ERROR

RMSPE: ROOT MEAN SQUARED PERCENTAGE ERROR

SA: SIMPLE AVERAGE COMBINATION

SARFIMA: SEASONAL AUTOREGRESSIVE FRACTIONAL INTEGRATED MOVING AVERAGE

SARIMA: SEASONAL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE

SIC: SCHWARZ INFORMATION CRITERION

SD: SIMPLE DYNAMIC

SETAR: SELF-EXCITING THRESHOLD AR

SR: STATIC REGRESSION

STAR: SMOOTH TRANSITION AUTOREGRESSIVE MODEL

STSM: STRUCTURAL TIME SERIES MODEL

SUR: SEEMINGLY UNRELATED REGRESSION

SVR: SUPPORT VECTOR REGRESSION

SW: SINE WAVE NONLINEAR REGRESSION

TBATS: TRIGONOMETRIC BOX-COX ARMA TREND SEASONAL MODEL

TCI: TOURISM CLIMATIC INDEX

TFM: TRANSFER FUNCTION MODEL

TGARCH: THRESHOLD GENERALIZED AUTOREGRESSIVE CONDITIONAL HETEROSKEDASTIC

TS: THE TWO-STAGE COMBINATION METHOD

TVP: TIME-VARYING PARAMETER

VACO: VARIANCE-COVARIANCE METHOD COMBINATION

(S) (G) VAR: (STRUCTURAL) (GLOBAL) VECTOR AUTOREGRESSION

VECM: VECTOR ERROR CORRECTION MODEL

VLTS: VECTOR LOCAL TREND SEASONAL MODEL

WB-ECM: WICKENS-BREUSCH ECM

WMA: WEIGHTED MOVING AVERAGE

Chapter 1 Introduction

1.1 Research Background and Motivation

Accurate forecasts of tourism demand can help marketers, managers and planners reduce risks, and therefore are of great importance for both the public and private sectors. In the last few decades, a wide range of forecasting techniques have been applied to predict international tourism demand with the purpose of improving forecasting performance. In the current literature, great attention is paid to the methodological advances of individual forecasting models, while the following research areas are generally neglected.

Firstly, the value of climate factors in forecasting tourism demand is understudied.

Tourism is an industry that is highly dependent on climate resources, and potentially sensitive to changes in climate conditions. Climate not only determines the suitability of one location for different types of recreation activities, but also affects the seasonality of one destination. Many tourists are in search of an enjoyable climate, and they prefer destinations that can reputably provide good climate with a certain reliability. Being an important factor in choice of destination, time of departure and length of stay, climate can potentially influence tourist flow.

In the current tourism demand literature, mainstream causal econometric models only consider economic variables as demand determinants. Climate variables are treated as fixed with time, as seasonal dummies, or as fixed effects in panel data studies, which ignores variation in climate conditions as well as long-term change, with the effects of climate therefore being subsumed into error terms, dummy variables or with variables that change over time. Tourism demand forecasts generated from such models ignore the relationship between climate condition and tourism demand, or travellers' attitudes towards climate change. Understanding tourists' reactions to changes in climate conditions is essential when projecting the potential geographical and seasonal shifts in tourism demand. And whether the forecasting performance of econometric models can be improved by including climate determinants should be examined.

Secondly, combination forecasts of tourism demand do not receive enough attention as they deserve.

The combination forecasting approach generates indirect forecasts through taking weights of individual projections produced by single models. It has been extensively applied in many areas such as meteorology, economics, finance and insurance as a robust and powerful approach to improve forecasting ability (Stock and Watson 1998; 2003; 2004). In contrast to the individual forecasting approach, which employs only one forecasting model or chooses the best performing one by discarding other forecasting methods, the combination forecasting approach makes use of a range of forecasting models. Individual forecasts are pooled together and assigned corresponding weights which are obtained through different weighting schemes. The weighted averages of all possible combinations of the individual forecasts are the combination forecasts. By aggregating information embedded in different individual models, which are based on competing theories, functional forms and specifications, diversification gains are achieved and forecasting ability can be improved. Although the performance of the combination forecasting approach is proved to be more stable than that of the individual one, in the context of tourism demand forecasting, only a handful of studies apply combination forecasting methods.

Furthermore, the value of climate factors in combination forecasting has never been evaluated.

In the current tourism demand literature, the individual constituent models that are considered in combination forecasts include causal econometric models and non-causal time series techniques. And the causal econometric constituent models are confined to the ones that only consider economic variables as influencing factors. No combination approach has been applied with causal econometric models incorporating climate variables as demand determinants. Econometric models that identify different influencing factors contain diverse information, which can potentially contribute to better combination forecasts. 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. The role played by climate factors in combination forecasts of tourism demand should be assessed. And the forecasting performance of combining econometric models with different sets of explanatory variables should be evaluated.

With these research gaps, several questions remain unanswered:

1. Whether including climate variables can improve the forecasting performance of econometric models?
2. Whether combining single econometric models which include climate factors can result in better combination forecasts?
3. Whether combining single econometric models with different explanatory variables are beneficial?
4. Whether the combination forecasting approach is consistently superior to the individual one?
5. Which individual models to combine to achieve better forecasts?
6. How many individual models to combine to generate the best forecasts?

To answer these questions, the current research comprehensively compares the performance of combination and individual forecasts of tourism demand. The econometric models included in the combination take two different specifications: with and without the climate determinant. Combination forecasts are generated through different weighting approaches. Comparisons among combination and individual forecasts are evaluated based on measures of accuracy commonly used in the tourism demand forecasting literature. The roles played by combination and climate factors in improving forecasting ability are evaluated, and the impact of climate on tourism demand is assessed.

1.2 Research Aim and Objectives

The aim of this study is to explore whether tourism demand forecasting accuracy can be improved through including the climate factor in econometric and combination forecasting models.

The following objectives are intended to be achieved:

1. quantify the long-run impacts of climate conditions on UK inbound tourism demand from seven leading markets;
2. generate one- to four-step-ahead individual forecasts of UK inbound tourism demand from seven leading markets;
3. introduce new weighting schemes to the tourism demand forecasting literature;
4. produce one- to four-step-ahead combination forecasts of UK inbound tourism demand from seven leading markets;
5. compare the forecasting performance of different individual and combination models based on accuracy measures commonly used in the tourism demand forecasting literature;

6. investigate which and how many individual models to combine to achieve more accurate forecasts.

1.3 Overview of International Demand for UK Tourism

The UK is one of the world's most popular international tourism destinations. It ranks number seven in terms of international tourist arrivals, and number five in terms of international tourism receipts in 2017 (UNWTO Tourism Highlights 2018 Edition). Figure 1-1 to Figure 1-3 show the trends and year on year growth rates for inbound tourist arrivals, inbound tourism expenditure and inbound tourist nights spend for the last 13 years (2005-2017) (VisitBritain, 2017). Inbound tourist flows have experienced several years of growth since 2011 after three years' decline since 2008. In 2017, the number of tourist arrivals grows 4.27% to a record 39.2 million. The value of spending, according to figure 1-2, has increased every year since 2005 except a slight decrease in 2007. £24.5 billion is the new record value achieved in 2017 with an 8.71% raise. Figure 1-3 demonstrates that the number of inbound tourist nights spent in the UK has got significant rises for the last five years after several years of fluctuations. It raised by 2.70% to about 285 million in 2017.

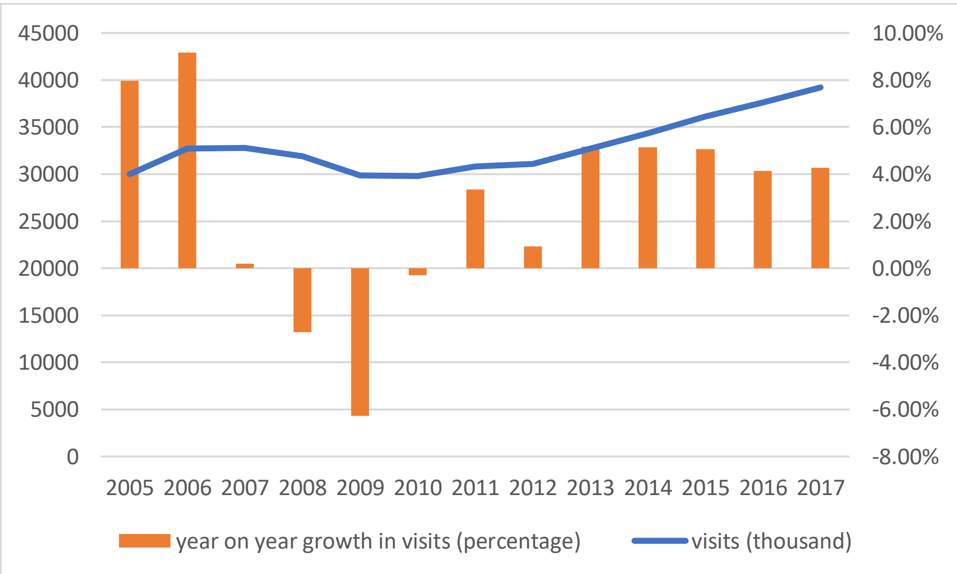

Figure 1-1 UK Inbound Tourist Arrivals from 2005 to 2017

Source: adapted from VisitBritain 2017 Snapshot

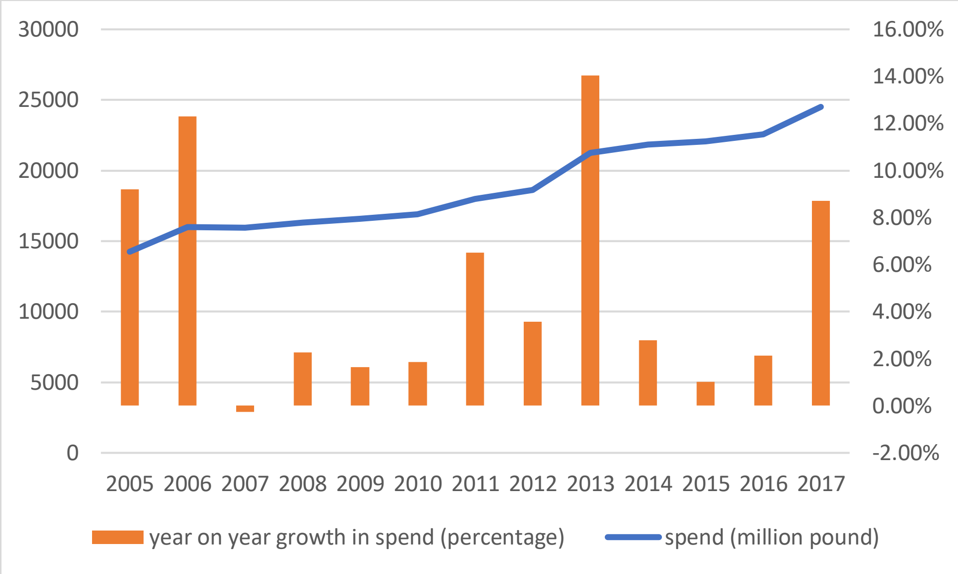

Figure 1-2 UK Inbound Tourism Expenditure from 2005 to 2017

Source: adapted from VisitBritain 2017 Snapshot

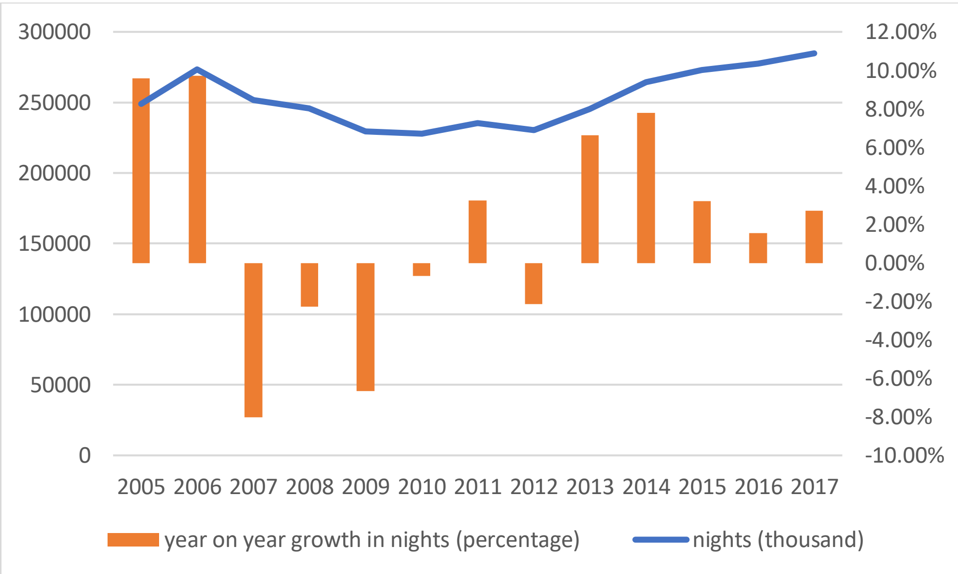

Figure 1-3 UK Inbound Tourist Nights from 2005 to 2017

Source: adapted from VisitBritain 2017 Snapshot

The top seven inbound markets for the UK in terms of number of arrivals during the period from 2005 to 2017 are France, USA, Germany, Irish Republic, Spain, the Netherlands and Italy, which accounted for 52.54% of the total visits in 2017 (see figure 1-5). The top seven markets have been the same every year since 2005 to 2016; the only change in their ranking was that in 2015 the USA, which is the only long-haul market in the top seven, overtook Germany to claim second place. In 2017, Poland, which ranked number eight in 2016, replaced Italy to rank number seven. But in terms of total arrivals for the last 13 years (2005 to 2017), Italy is the seventh largest origin. Figure 1-4 shows the total number of visits from the top seven markets for the period from 2005 to 2017.

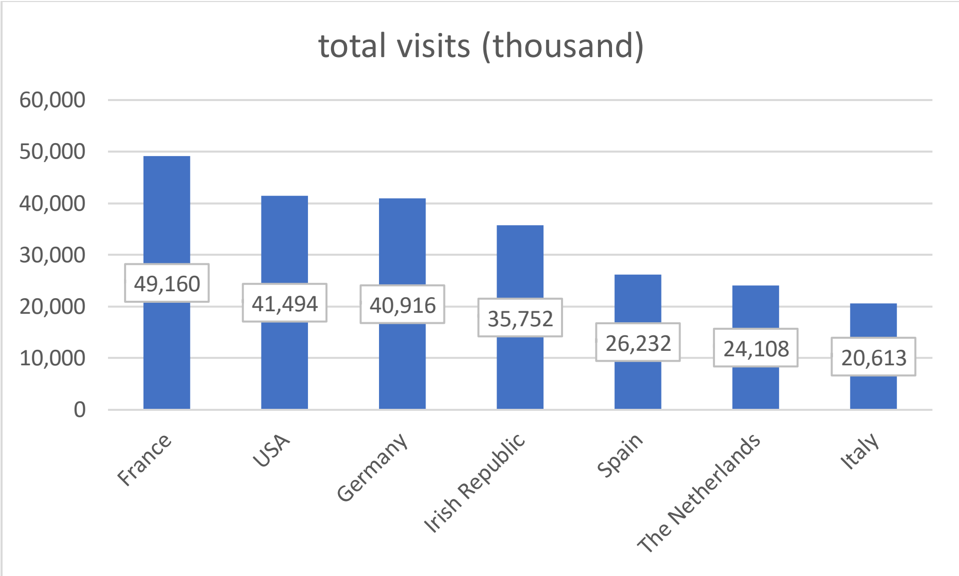

Figure 1-4 Total Visits from The Top Seven Markets to the UK from 2005 to 2017

Source: adapted from VisitBritain 2017 Snapshot

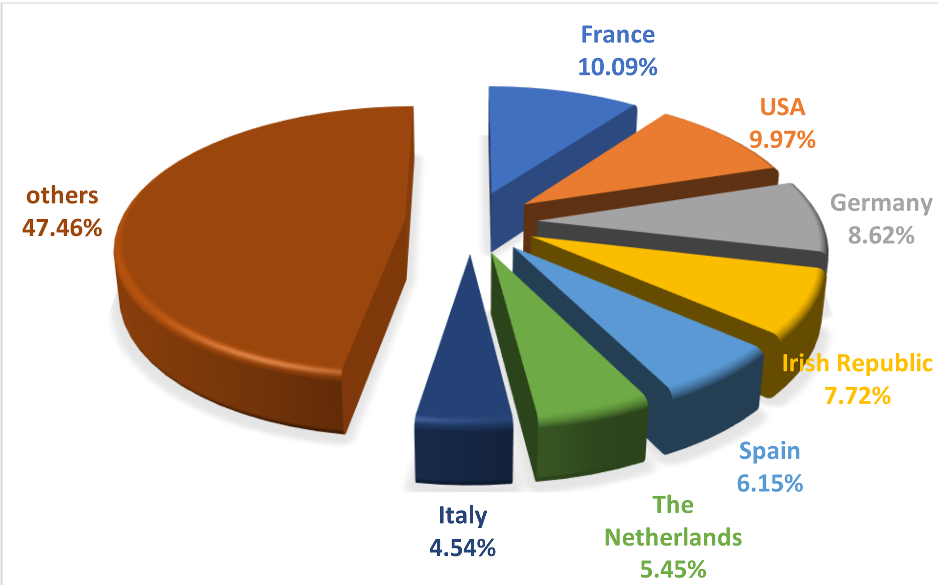

Figure 1-5 Market Shares of UK Inbound Tourist Arrivals in 2017

Source: adapted from VisitBritain 2017 Snapshot

1.4 Structure of This Research

The rest of this study is organized into five chapters. Chapter 2 is literature review, which covers publications up to 2018 with focuses on studies applying non-causal time series techniques, causal econometric models and the combination forecasting approach, which are the main methodologies of this research. A detailed review on time series studies comes first, followed by a systematic account of econometric researches. Afterwards, the benefits, methods and applications of the combination forecasting approach is illustrated, and the characteristics of data used in the current tourism demand literature is summarized.

Chapter 3 presents the research method of this study. Firstly, the research plan is illustrated by a chart featuring the details of how the objectives of this study are fulfilled, which is followed by an explanation of model variables, data sources and data sample. Next, the bounds test cointegration technique that is used to evaluate the long-run cointegration relationship between climate conditions and tourism demand is discussed. Several diagnostic tests including residual diagnostics and stability diagnostics are run after model estimation, and these tests are presented in this chapter. Afterwards, the individual forecasting models that are considered as candidate constituents in combinations are described with a justification on model selection being provided, and the combination forecasting methods that are employed in this study are interpreted. The forecasting procedure and the accuracy measures are examined next. And an explanation of the programs for the combination forecasting concludes this chapter.

The empirical results are reported and discussed in chapter 4 and chapter 5. In chapter 4, the integration orders of model variables are tested, and the results are presented, based on which, the bounds test cointegration technique is chosen to evaluate the cointegration relationship between tourism demand and its influencing factors. The model parameter estimates, and the diagnostic test results are summarized and the impact of different influencing factors including the climate factor on tourism demand are evaluated.

Chapter 5 covers the results of forecasting comparisons, which are conducted for one- to four-step-ahead forecasts for seven destination-origin pairs based on three accuracy measures. Firstly, the forecasting performance of individual models are assessed and compared. The country-specific evaluation, the forecasting-horizon-specific assessment and the general comparison are conducted among 15 individual models including six traditional econometric models, six climate econometric models and three time series techniques with the seasonal naive no-change forecasts serving as benchmarks. The ratios of the accuracy measures of the 15 individual models compared to those of the benchmark forecasts are provided. Except the seasonal naive no-change model, all other individual forecasting models are considered as constituents in combinations. A whole section is devoted to evaluating whether including the climate determinant can improve the forecasting performance of econometric models.

The comparisons between combination and individual forecasts come next. Individual models are categorized into three groups: the first group includes all 15 individual models; the second group contains three time series models and six traditional econometric models; and the third group consists of three time series models and six climate econometric models. The three groups of individual models are combined respectively, and the combination methods applied consist of two new schemes. The numbers of all possible combination forecasts for the first group are 32752 and 5007 (for the two-stage combination method), and that for the second and third groups is 502. The percentages of the superior combination forecasts compared to the best single ones are worked out and reported for each country-specific evaluation, forecasting-horizon-specific assessment and general comparison. A comparison among the three combination groups is conducted to assess whether including econometric models with the climate determinant and combining econometric models with different influencing factors can contribute to better forecasts. In addition, the frequencies of each individual constituent model in the superior combination forecasts and the number of the components in the best forecasts in every comparison are also reported in this chapter.

The core empirical results of the current research are all included in chapter 5, which are the percentages of the superior combination forecasts obtained through different combination methods and combination groups and the comparisons among three combination groups. These results answer the research questions as whether including the climate factor in the combination forecasting approach is beneficial and whether combination forecasts are consistently superior to individual ones.

Chapter 6 concludes this study by outlining the major findings, contributions, limitations of the current research, and making some recommendations for future forecasting practice and future research.

Chapter 2 Literature Review

2.1 Introduction

Parallel with the fast development of tourism industry is the rapid advancement of tourism demand modeling and forecasting techniques, and the impressive increase in the number of studies published in this area. Since the 1960s, an accumulating body of tourism demand researches have emerged in the literature with a wide range of analytical approaches being applied. Some notable early works include Artus (1972), Gray (1966) and Kwack (1972). Most studies employ macroeconomic quantitative methodologies, which are the focus of this review. Several review articles have published (e.g. Crouch 1994; Witt and Witt 1995; Lim 1997; 1999; Li et al. 2005; Song and Li, 2008; Wu et al. 2017). This chapter does not intend to duplicate the existing work. It provides a systematic review on quantitative tourism demand studies focusing on publications from 2008 to 2018. The purpose is to identify new research trends, present valuable research questions, and discover potential research gaps.

After an extensive search on a variety of databases including Google Scholar, citations from publications, and *MySearch*, an academic search engine provided by the library of Bournemouth University, more than 300 studies have been identified. Table A-1 in Appendix 1 provides a summary of publications from 2008 to 2018. It is generally acknowledged that the dominant methodologies are non-causal time series techniques and causal econometric models, with other quantitative methods such as artificial intelligence (AI) and the rough set approach having been introduced since the turn of this century. This review emphasizes on studies applying econometric and time series models, as well as the combination forecasting approach, which are the core methodologies of this study.

This chapter is organized into six sections, with the first one being introduction and the last one being summary. Section 2.2 focuses on different time series techniques, with detailed introduction of the most popular ones in the tourism demand literature including the naïve model, the autoregressive integrated moving average (ARIMA) model, the exponential smoothing (ETS) model, the state space ETS model and the structural time series model (STSM). The newly introduced Singular Spectrum Analysis (SSA) is also covered.

Section 2.3 goes over various econometric models which are categorized into two groups based on the data type used: the ones based on pure time series data and the ones based on panel data. Econometric studies based on pure time series data take two different modelling approaches: the single-equation approach and the system-of-equations approach, which are addressed in section 2.3.2 and 2.3.3. Panel data analysis, which was rare before 2008, has become increasingly popular and deserves attention. It is presented in section 2.3.4. In addition to being employed as forecasting tools, econometric models can also be used to evaluate causal relationships between tourism demand and its influencing factors. A subsection (section 2.3.5) is devoted to studies on the roles played by climate factors in modelling and forecasting tourism demand.

Section 2.4 presents the combination forecasting approach, where the benefits of combining, different combination methods and the applications of combination forecasting in the tourism demand literature are discussed.

Section 2.5 focuses on the characteristics of data used in the current literature, including data type and data frequency. And section 2.6 summarizes the main findings of this chapter.

2.2 Non-causal Time Series Techniques

2.2.1 Overview

Non-causal time series models have longer history and wider application in tourism demand studies compared to causal econometric models. The assumption underlying time series techniques is that tourism demand can be modelled and forecasted based on its own past values, which is justified by the belief that historic pattern of a time series can evolve into the future. As a result, the emphasis is put on revealing the historic trends and patterns (such as cycle and seasonality) of the series and predicting the future value of the series based on the properties identified (Song and Li 2008). When forecasting tourism demand, there is no need to take the roles of explanatory variables into account. Instead, the intrinsic evolution of tourism demand series is captured.

A general time series model can be expressed as:

$$y_t = f(\beta_i y_{t-i}, \varepsilon_t) \quad 2.1$$

where y_t denotes a measure of tourism demand, y_{t-i} are the lags of the dependent variable with a lag length of i and corresponding parameters of β_i, ε_t is the error term and $f(\cdot)$ represents the functional form determined by the relationship between y_t and its own lags.

Time series techniques are advantageous for they are cost-effective in terms of data collection and forecasting. When forecasts are generated by time series techniques, there is no need to obtain predictions of other variables, which is a prerequisite of causal econometric models. Therefore, the use of time series techniques is closely associated with forecasting practice, and different time series models are popular candidates in forecasting competitions. On the other hand, the limitation of time series techniques is obvious. They are not capable of modelling causal relationships, and hence can only serve the objective of forecasting instead of explanation.

The time series techniques that are popular in the 1970s belong to the decomposition approach, which include simple exponential smoothing (ETS) models, moving average (MA) models and autoregressive (AR) models (Witt and Witt 1995). In the current literature, there are mainly five popular time series techniques: the naïve model, Box-Jenkins' autoregressive integrated moving average (ARIMA) models, exponential smoothing (ETS) models, the state space ETS model and the STSM, all of which frequently appear in tourism demand analysis. The SSA has been introduced to the tourism demand literature in the 2010s and has been proven to possess robust predictive power.

2.2.2 The Naïve Model

The naïve model is the simplest form of the time series technique. It assumes that historic pattern evolves based on static growth rate, and therefore future is just a repeat of history. In the current tourism demand literature, there are basically two types of naïve models: the naïve 1 model and the naïve 2 model.

The naïve 1 model is also called the no-change model, which assumes that the future value equals the most recent available value, i.e., $\hat{y}_t = y_{t-s}$, where s represents the number of seasons under consideration. When annual data is used, $s = 1$; with seasonal data, $s = 4$, and with monthly data, $s = 12$. The naïve 2 model assumes that the value of a variable grows at a constant rate, so the naïve 2 model is also known as the constant growth rate model. Forecasts can be generated based on the constant growth rate: $\hat{y}_t = y_{t-s}[1 + (y_{t-s} - y_{t-2s})/y_{t-2s}]$, again, s represents the number of seasons under consideration. When annual data is used, $s = 1$; with seasonal data, $s = 4$, and with monthly data, $s = 12$. The naïve 1 model is a special case of the naïve 2 model, where the constant growth rate is assumed to be zero.

The naïve models often serve as benchmarks in forecasting comparisons (Chu 2008a; Cang and Hemmington 2010; Moore 2010; Fildes et al. 2011; Song et al. 2011a; Lee 2011b) and are sometimes included in combination forecasts (Shen et al. 2008; 2009; 2011; Cang 2011; 2014).

2.2.3 The Autoregressive Integrated Moving Average Model

The ARIMA model was presented by Box and Jenkins in the 1970s, and it has become the most popular time series technique ever since (Goh and Law 2011; Song and Li 2008). It is renowned for its wide applicability, as it can handle any stationary or non-stationary time series, both with or without seasonality (Lim and McAleer 2002). The ARIMA model integrates the AR process, which specifies that the dependent variable depends linearly on its past values, and the MA process, which signifies that the current value of the dependent variable is a liner combination of current and previous white noise error terms.

A non-seasonal $\mathrm{ARIMA}(\mathrm{p}, \mathrm{d}, \mathrm{q})$ process is given by:

$$\phi_p(B)(1-B)^d y_t = \mu + \theta_q(B)\varepsilon_t \qquad 2.2$$

where y_t denotes a measure of tourism demand, ε_t is the error term with $\varepsilon_t \sim \mathrm{IID}\,(0, \sigma^2)$, B is the backshift operator with $By_t = y_{t-1}$, $B^d y_t = y_{t-d}$, and μ is the overall mean of the series, which is constant. $\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is a polynomial of order p representing the autoregressive (AR) part; $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is a polynomial of order q representing the moving average (MA) part. d is the differencing integer which equals the number of times the variable under consideration (y_t) needs to be differenced to achieve stationary.

Besides the basic ARIMA model, there are different variations and extensions of the ARIMA model, among which the seasonal ARIMA (SARIMA) is the most popular one.

A $\mathrm{SARIMA}(\mathrm{p}, \mathrm{d}, \mathrm{q})(\mathrm{P}, \mathrm{D}, \mathrm{Q})_s$ process can be specified as:

$$\Phi_P(B^s)\phi_p(B)(1-B^s)^D(1-B)^d y_t = \mu + \Theta_Q(B^s)\theta_q(B)\varepsilon_t \qquad 2.3$$

where $\Phi_P(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{s2} - \cdots - \Phi_p B^{Sp}$ is a seasonal AR of order P; $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \theta_2 B^{s2} - \cdots - \theta_Q B^{Sq}$ is a seasonal MA of order Q. (Box et al. 2015).

Other variations of the ARIMA model that have been successfully applied in the tourism demand literature include the Autoregressive Fractional Integrated Moving Average (ARFIMA) model and the Autoregressive Autoregressive Moving Average (ARAR) model. The ARFIMA model integrates long-range dependencies with the ARIMA model and allows the differencing parameter d to be a non-integer to represent the fractional order of the integration (Chu 2009). The ARAR model assumes that a time series is transformed from a long-memory AR filter to a short-memory filter. It challenges the practice of differencing to achieve stationary and can make use of the information contained in the level data, which is usually lost in differencing (Chu 2009).

The ARIMA model can also be expanded by including explanatory variables, which contribute to the Autoregressive Integrated Moving Average with Explanatory Variables (ARIMAX) model. The ARIMAX model represents the integration of the time series technique and the econometric model and has been suggested by several studies to be an improvement over the time series technique (Morley 2009; Yang et al. 2015; Chatziantoniou et al. 2016; Tsui and Balli 2017).

The ARIMAX model can be specified as:

$$\Phi_P(B^s)\phi_p(B)(1-B^s)^D(1-B)^d y_t = \mu + \sum_{j=1}^{n}\sum_{i=0}^{p_j}\beta_{j,i}\, x_{j,t-i} + \Theta_Q(B^s)\theta_q(B)\varepsilon_t \qquad 2.4$$

where y_t denotes a measure of tourism demand, ε_t is the error term with $\varepsilon_t \sim \mathrm{IID}\,(0,\sigma^2)$, B is the backshift operator with $By_t = y_{t-1}$, $B^d y_t = y_{t-d}$, and μ is the overall mean of the series, which is constant. $\Phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$ is a polynomial of order p representing the autoregressive (AR) part; $\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$ is a polynomial of order q representing the moving average (MA) part. $X_{j,t-i}$ are the identified explanatory variables with the number of n and corresponding parameters of $\beta_{j,i}$, p_j are the lag lengths of the independent variables.

Among post-2008 studies, the applications of the ARIMA and SARIMA models cover various empirical studies focusing on wide-ranging destination-origin pairs (e.g. Assaf et al. 2011; Brida and Risso 2011; Chatziantoniou et al. 2016; Ma et al. 2016; Apergis et al. 2017; Saayman and Botha 2017). Interval forecasts can also be produced by AR or ARIMA models. Kim et al. (2010) proposed the use of the bias-corrected bootstrap for interval forecasting of AR models and introduced a new stationary-correction method, based on stable spectrum factorization, as an alternative to Kilian's method (Kilian 1998). Through employing this method in forecasting HK inbound tourism demand, they found that the presented approach had desirable small sample properties and the generated interval predictions were tighter and more stable than those based on Kilian's stationary correction. In a later study, Kim et al. (2011) explored alternative interval forecasts generated by the AR model, the AR model using the bias-corrected bootstrap, the SARIMA model, STSM and the state space ETS model. They investigated HK and Australia inbound tourism demand respectively and concluded that all models except for the AR model, produced satisfactory prediction intervals, and the AR model based on the bias-corrected bootstrap performed the best generally.

The applications of the ARFIMA and ARAR models are still limited in the current literature, examples include Chu (2008a; 2008b; 2009; 2011), Nowman and Dellen (2012), Apergis et al. (2017) and Hassani et al. (2017). Chu (2008b) employed the ARAR model to forecast tourism demand of nine destinations in the Asian-Pacific region and concluded that the ARAR model could be used as a reliable forecasting method after comparing the forecasting accuracy of the ARAR model and the SARIMA model. Nowman and Dellen (2012) assessed the forecasting ability of continuous time models with discrete data when forecasting UK inbound tourism demand and found that discrete time models outperformed the continuous time counterparts after comparing continuous models with the ARIMA and ARFIMA models. Apergis et al. (2017) assessed the forecasting performance of four univariate seasonal time series models: the SARIMA model, the SARIMA model with Fourier transformation, the ARAR model and the SARFIMA model based on monthly data of total tourist arrivals to 20 Croatian counties. They demonstrated that the SARIMA model with Fourier transformation provided the best forecasts for the destinations investigated.

2.2.4 The Exponential Smoothing Model

Compared to (S)ARIMA models, which can be complicated to identify and estimate in many business environments, the ETS model, which has been a popular technique for more than half a century, is simple to implement when forecasting data with seasonal patterns. The ETS model was developed based on the MA technique and uses weighted values of past observations to generate forecasts with the weights decaying exponentially over time justified by the belief that the most recent information is considered to be more influential on forecasts than older ones.

In the single form of the ETS model, the forecast in period t is based on weighting the observations in period t by a smoothing factor α, and the most recent forecast by $(1-\alpha)$. The single ETS mode is only suitable for non-seasonal stationary time series with no structural change, and more advance algorithms, such as double exponential smoothing, Holts's method (Holt 2004) and Holt-Winter's method (Holt 2004; Winter 1960) have been developed and applied in the tourism demand literature, with the most popular one being Holt-Winter's additive and multiplicative seasonal models (the three parameters ETS model).

The three parameters ETS model, which can reflect level L_t , trend T_t and seasonality S_t can be expressed in a multiplicative form (equation 2.5-2.8) or an additive form (equation 2.9-2.12):

$$\hat{y}_{t+k} = (L_t + kT_t)S_{t+k-p} \quad 2.5$$

$$L_t = \alpha\frac{y_t}{S_{t-p}} + (1-\alpha)(L_{t-1} + T_{t-1}) \quad 2.6$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \quad 2.7$$

$$S_t = \delta\frac{y_t}{L_t} + (1-\delta)S_{t-p} \quad 2.8$$

$$\hat{y}_{t+k} = L_t + kT_t + S_{t+k-p} \quad 2.9$$

$$L_t = \alpha\left(y_t - S_{t-p}\right) + (1-\alpha)(L_{t-1} + T_{t-1}) \quad 2.10$$

$$T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1} \quad 2.11$$

$$S_t = \delta(y_t - L_t) + (1-\delta)S_{t-p} \quad 2.12$$

where $\hat{y}_{t+k}$ is the k-period ahead forecasting value, p denotes the number of seasons per year with $p = 4$ for seasonal data and $p = 12$ for monthly data, and α, γ and δ are the three smoothing parameters with constant values between 0 and 1.

The ETS model is widely employed in tourism demand forecasting studies (e.g. Lim and McAleer 2001b; Cho 2003; Chen 2011; Gounopoulos et al. 2012; Untong et al. 2015).

2.2.5 The State Space ETS Model

A recent development of the ETS model is the state space ETS model, which was proposed by Hyndman et al. (2002). The state space ETS model encapsulates the notion of exponential smoothing in a state space form by including an observation equation for the forecast variable and a number of state equations for the components such as trend, level and seasonality which cannot be observed. The state space ETS model is estimated by maximum likelihood and can generate interval forecasts.

The general form of state space ETS can be expressed as:

$$y_t = w'X_{t-1} + \varepsilon_t \quad 2.13$$

$$X_t = FX_{t-1} + g\varepsilon_t \quad 2.14$$

where y_t denotes an observation of tourism demand at time t, X_t represents a vector of unobserved components which can be a mixture of level, trend and season, $\varepsilon_t \sim \mathrm{IID}\,(0, \sigma^2)$, w is a vector consisting of elements of zeros and ones, F is a transition matrix of zeros, ones and model parameters, and g is a vector of unknown parameters which determine the impact of the random noise on the unobserved components.

Equation 2.13 is the observation (or measurement) equation, which describes the relationship between unobserved parts and the observation y_t, and equation 2.14 is the state (or transition) equation, which depicts the evolution of the states over time.

The state space ETS model is more popular among post-2008 tourism demand studies compared to its traditional counterparts, and it has been proved by many studies to have more powerful forecasting ability (e.g. Athanasopoulos and Hyndman 2008; Athanasopoulos et al. 2009; Athanasopoulos et al. 2011; Kim et al. 2011; Gunter and Önder 2015; Hassani et al. 2015).

In addition, the ETS model can be extended to the multivariate framework. Athanasopoulos and Silva (2012) presented a new set of multivariate stochastic models within the vector innovations structural time-series (VISTS) framework, where multiple variables can be modeled simultaneously. Through comparing among univariate and multivariate time series models in the context of forecasting Australia and New Zealand inbound tourism demand from 11 source markets, they concluded that by pooling information together using a multivariate approach, more accurate forecasts can be generated.

2.2.6 The Structural Time Series Model

A more common form of the state space model, which is known as the structural time series model (STSM), was presented by Harvey (1990). The STSM has also been successfully utilized in the current tourism demand literature.

The basic STSM, which is also called basic structural model (BSM), was introduced to the tourism demand literature in the middle 1990s by Gonzalez and Moral (1995). It decomposes an observed time series into various unobserved components, which can be forecast individually and combined to generate a forecast for the observed series.

The BSM can be expressed as:

$$y_t = \mu_t + \gamma_t + \psi_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{IID}(0, H_t) \tag{2.15}$$

$$\mu_t = \mu_{t-1} + \delta_{t-1} + \eta_{t,} \qquad \delta_t = \delta_{t-1} + \zeta_t \tag{2.16}$$

$$\eta_t \sim i.i.d.\left(0, \sigma_\eta^2\right), \qquad \zeta_t \sim i.i.d.\left(0, \sigma_\zeta^2\right),$$

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + w_t, \qquad w_t \sim i.i.d.(0, \sigma_w^2) \tag{2.17}$$

$$\begin{cases} \psi_t = \rho \cos \lambda_c \psi_{t-1} + \rho \sin \lambda_c \psi_{t-1} + \kappa_t \\ \psi_t^* = -\rho \sin \lambda_c \psi_{t-1}^* + \rho \cos \lambda_c \psi_{t-1}^* + \kappa_t^* \end{cases}, \qquad \kappa_t, \kappa_t^* \sim i.i.d.(0, \sigma_\kappa^2) \tag{2.18}$$

where y_t is the observed time series comprised of a stochastic trend (μ_t), a seasonal term (γ_t), a cyclical component (ψ_t) and a random term (ε_t). The trend μ_t is assumed to be a random walk with stochastic random drift δ_t (equation 2.16). The seasonal component γ_t is assumed to follow a stochastic dummy specification where s depends on the data frequency (for quarterly data,$s = 4$). And when $\sigma_w^2 = 0$, γ_t reduces to be a deterministic seasonal component (equation 2.17). The cyclical component is specified in the trigonometric form, where ρ is a damping factor between zero and one, λ_c is the frequency of the cycle measured in radians (equation 2.18). $\eta_{t,}\ \zeta_t, w_t, and\ \kappa_t/\kappa_t^*$ are mutually uncorrelated white-noise Gaussian errors with corresponding variances of σ_η^2, $\sigma_\zeta^2, \sigma_\omega^2\ and\ \sigma_\kappa^2$.

The difference between the state space ETS model and the STSM lies in that the state space ETS model only involves single source of error, while the STSM covers multiple sources of errors. The STSM is also called multiple source of error model, for each equation of a STSM system carries its own independent error term.

Exogenous variables can be readily included in a STSM specification to form a model called CSM, which falls into the category of causal econometric models. Furthermore, modern econometric techniques such as ECM and TVP can be easily incorporated with CSM. Applications of BSM and CSM include Ouerfelli 2008, Algieri and Kanellopoulou 2009, Shen et al. 2009; 2011, Guizzardi and Mazzocchi 2010, Cortes-Jimenez and Blake 2011, and Guizzardi and Stacchini 2015.

2.2.7 The Singular Spectrum Analysis

The Singular Spectrum Analysis (SSA) is a filtering technique, and it generates nonparametric forecasts. The classical time series techniques such as ARIMA and ETS models forecast both the signal and noise, assuming that a time series consists of signal and noise. On the other hand, the spectrum analysis aims at filtering the noise and forecasting the signal. The noise in time series is filtered and forecasts are generated using the newly constructed less noisy time series. The expressions of SSA is beyond the scope of this study and those who are interested in the detailed description of SSA are directed to Hassani (2007).

There are two major benefits of SSA. Firstly, as a nonparametric technique, it requires no assumptions regarding the data generating process. ARIMA and ETS models all belong to parametric techniques, which rely on assumptions such as normality and stationary. Such assumptions are both likely to be violated in tourism demand series, and data transformation would be required. Secondly, it does not require large historical data, and a minimum of three observations are enough to generate forecasts (Hassani et al. 2017)

Despite the advantages of spectrum analysis, SSA does not gain popularity in the current tourism demand literature. There are a few studies applying the spectrum analysis (Chan and Lim 2011; Kožić 2014; Hassani et al. 2015; 2017; Saayman and Botha 2017; Silva et al. 2017). For instance, Saayman and Botha (2017) considered both linear and nonlinear methods when forecasting tourist arrivals to South Africa. The comparison of the seasonal naïve model, the SARIMA model, BSM, SSA and smooth transition autoregressive model revealed that the non-linear forecasts were better than the linear ones. Hassani et al. (2017) applied several parametric and nonparametric techniques including MA, weighted MA (WMA), SARIMA, ETS, ARFIMA, Trigonometric box-cox ARMA trend seasonal model (TBATS), artificial neural network (ANN) and SSA to forecast tourist arrivals to selected European countries. They concluded that there was no single best model for any of the countries in the short-, medium- and long-run, and SSA performed most consistently across all countries and all forecasting horizons.

The superior forecasting ability of spectrum analysis compared to other time series techniques has been proven by some empirical studies. For example, Hassani et al. (2015) applied SSA to project US inbound tourism demand and compared it with other time series models including the ARIMA and ETS models, as well as neural network techniques. SSA's superiority in forecasting tourism demand was verified by the empirical study, and the authors concluded that SSA performed significantly better and was worthy of consideration in future studies. Silva et al. (2017) introduced multivariate singular spectrum analysis (MSSA) to the tourism demand literature, which was used to identify leading indicators as well as forecast tourism demand. They showed evidence that tourist arrivals in a European country can serve as the leading indicator of tourist arrival to another European country. The predicting ability of MSSA was found to be superior to the SARIMA model, ETS and SSA when forecasting monthly tourist arrivals to 10 European countries. The forecasting ability of the spectrum analysis deserves more attention.

2.3 Causal Econometric Models

2.3.1 Overview

The rationale behind econometric modeling is that tourism demand is determined by influencing factors such as disposable income, price levels and transportation costs, hence tourism demand can be specified in an equation consisting of all the influencing factors and an error term. The major benefits of econometric models are their abilities to assess causal relationships between tourism demand and its influencing factors, and therefore to conduct elasticity estimation, impact analysis and policy evaluation. As Clements and Hendry (1998, p16) state, the advantage of econometric analysis is that it '*fulfills many useful roles other than just being a device for generating forecasts; for example, such models consolidate existing empirical and theoretical knowledge of how economies function, provide a framework for a progressive research strategy, and help to explain their own failures*.'

Tourism demand models may be built on different assumptions or theories, be rich in dynamic specification or be static, be based on time series or panel data, employ a linear functional form or other sophisticated ones. According to the data type utilized, main econometric studies can be classified into two groups: the ones based on pure time series data and the ones based on panel data.

Time series data are single-dimensional data, which are indexed at successive equally spaced points in time. They are discrete-time observations of variables for one particular entity over a period of time. An example of time series data is the quarterly international tourist arrivals to the UK from the first quarter of 1994 to the last quarter of 2017. Panel data are multi-dimensional data, which provide observations of variables for a number of entities over a period of time. An example of panel data is the quarterly international tourist arrivals to each European country from the first quarter of 1994 to the last quarter of 2017. The combination of the temporal and cross-sectional dimension results in more observations.

Most econometric models rely on pure time series data, with panel data models having become increasingly popular in impact analysis. Single equation approaches with dynamic specifications dominate in the first group, while system-of-equations approaches are unpopular. Table 2-1 summarizes the main econometric modelling techniques applied among the post-2008 studies, which shows that the most popular ones are the cointegration (CI) approaches and the panel data analysis.

Table 2-1 Summary of Main Econometric Modeling Techniques

		Time Series Data		**Panel Data**
		Single Equation Approaches	**System-of–equations Approaches**	
	Static Models	static regression	static AIDS	static panel
	Dynamic Models	ADLM; Cointegration analysis & ECM; simple dynamic; leading indicator; TVP; TVP-ECM; TVP-STSM	EC-AIDS; VAR & VECM; Bayesian VAR; Global VAR; Bayesian Global VAR	panel ADLM; panel cointegration analysis & panel ECM
Novel		TVP-STSM	Bayesian Global VAR	Dynamic Spatial Panel Data Model
Most Popular		Cointegration analysis & Panel data analysis		

Source: the author

2.3.2 The Single-Equation Approach

In the tourism demand literature, most econometric models are estimated based on time series data, which basically follow two approaches: the single-equation approach, which involves one equation about the observed part in a model[1], and the system-of-equations approach, which makes use of a system of equations regarding the observed part in a model. The single-equation approach dominates in the tourism demand literature with extensive empirical studies covering a wide range of destination/origin pairs.

The general form of a single equation econometric model can be specified as (Song, Witt and Li 2009):

$$y_t = f(\beta_{j,i} x_{j,t-i}, \varepsilon_t) \quad 2.19$$

where y_t represents a measure of tourism demand, $x_{j,t-i}$ are the identified explanatory variables, which may contain different independent variables and their lags, as well as the lags of the dependent variable, with the corresponding parameters of $\beta_{j,i}$, and ε_t is the error term which is assumed to be an independent and identically distributed (IID) variable of zero mean and a variance of σ^2.

In the current literature, most studies apply a linear functional form with y_t and $x_{j,t-i}$ specified in log form (also called double-log linear form), which transforms the data to a smaller scale. The benefits of such a functional form include: firstly, it smooths the fluctuations of the data, and therefore may reduce the integration order of model variables, which facilitates to standard cointegration analysis (Li et al. 2005); secondly, the residual variance produced by models estimated in the double-log linear is relatively low compared to that generated by models estimated in other forms (Goh and Law 2011); and most importantly, the estimated parameter can be interpreted as the demand elasticity of the corresponding variable directly.

---

[1] The time varying parameter model is classified as a single-equation model as there is only one equation regarding the observed part in the model.

The autoregressive distributed lag model (ADLM) represents a general form of model specification, and a variety of models are available by imposing different assumptions regarding the relationship between the dependent variable and the independent variables. Examples include the static model and various dynamic models such as the growth rate model, the partial adjustment model, the cointegration (CI) analysis and error correction model (ECM) and the leading indicator (LI) model. Another popular dynamic modelling technique is the time-varying parameter (TVP) model, which allows the parameters of the explanatory variables to change over time.

Different single-equation models have their benefits and drawbacks and have always been chosen based on the aims and objectives of the studies or the expertise of the researchers. However, due to their ability to conducting both long-run and short-run demand elasticities analysis, the cointegration analysis and ECM techniques have gained much popularity with many applications in tourism research since the 1990s (Song, Witt and Li 2009, p. 31). Among all these CI techniques, the bounds test cointegration approach has been extensively chosen among post-2008 studies, because it is applicable no matter whether the model variables are integrated of order zero, or order one, or mutually cointegrated (e.g. Lee 2011a; Onafowora and Owoye 2012; Seetaram et al. 2014; Li et al. 2015; Saayman and Saayman 2015).

2.3.2.1 The Autoregressive Distributed Lag Model

The ADLM, which considers lagged values of the dependent and independent variables as well as current values of independent variables as potential influencing factors (Song and Witt 2000), can be specified as:

$$y_t = \alpha + \sum_{j=1}^{n} \sum_{i=0}^{p_j} \beta_{j,i}\, x_{j,t-i} + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \qquad 2.20$$

where y_t represents a measure of tourism demand, $x_{j,t-i}$ are the identified explanatory variables with the number of n and corresponding parameters of $\beta_{j,i}$, $p_j\ and\ p$ are the lag lengths of independent and dependent variables respectively, ϕ_i is the coefficient on y_{t-i}, which needs to be estimated and ε_t is the error term which is assumed to be an IID variable of zero mean and a variance of σ^2 (Morley 2009; Song, Witt and Li 2009, p47).

A general guide for the maximum lag length $p_j\ or\ p$ is that $p_j = 1$ for annual data, $p_j = 4$ for quarterly data, and $p_j = 12$ for monthly data, and Akaike Information Criteria (AIC) and Schwarz Information Criteria (SIC) are used to determine the lag length (Song, Witt and Li 2009, p42).

The ADLM allows the nature of data suggest the lag lengths instead of imposing restrictions. By allowing for sufficient lag lengths, the disturbance term is well-behaved, and the statistics calculated from the estimation are asymptotically standard normal. Applications of the ADLM can be found in many post-2008 studies (e.g. Athanasopoulos et al. 2011; Song et al. 2011a; 2013).

The ADLM encompasses a range of econometric models including the static model and different dynamic models, which are achieved by imposing different assumptions.

2.3.2.2 The Static Model

Being the most traditional models in the tourism demand literature, the single-equation static models are extensively used among tourism demand studies before the 1980s (e.g. Gray 1966; Artus 1972; Barry and O'Hagan 1972; Kwack 1972; Loeb 1982). It assumes that the current value of tourism demand is only related to the current values of the explanatory variables. The static model can be expressed as:

$$y_t = \alpha + \sum_{j=1}^{n} \beta_j \, x_{j,t} + \varepsilon_t \tag{2.21}$$

where y_t represents a measure of tourism demand, $x_{j,t}$ are the identified explanatory variables with the number of n and corresponding parameters of β_j, and ε_t is the error term, which is assumed to be an IID variable of zero mean and a variance of σ^2.

Since tourism demand variables are always non-stationary, the error terms in static models are always found to be highly autocorrelated, which indicates the existence of spurious regression relationships (Song, Witt and Li 2009 p48). Autocorrelation means that there is correlation between values of the same series at different times, which results in biased t-statistics and other important statistical indicators. To address these problems, dynamic specifications have been introduced and static models have become unpopular since the 1990s. Only a few of post-2008 studies utilize static models (Goh et al. 2008; Stepchenkova and Eales 2011; Schiff and Becken 2011; Kim et al. 2012; Smeral 2012; Falk, 2014; Untong et al. 2015), the first of which deserves attention. Goh et al. (2008) made use of rough sets data mining technique to construct leisure time and climate index and found that leisure time and climate had stronger impacts on tourist arrivals than economic factors.

2.3.2.3 The Growth Rate Model

The growth rate model, which makes use of data differencing, represents an early attempt to address the problem of spurious regression relationship resulted from trended variables in static models (e.g. Witt 1980a). It regresses the first difference of the dependent variable on the first differences of the explanatory variables. The growth rate model can be specified as:

$$\Delta y_t = \sum_{j=1}^{n} \beta_j \Delta x_{j,t} + \varepsilon_t \tag{2.22}$$

where Δy_t signifies the first difference of the dependent variable with $\Delta y_t = y_t - y_{t-1}$, $\Delta x_{j,t}$ represents the first differences of the explanatory variables with $\Delta x_{j,t} = x_{j,t} - x_{j,t-1}$, n is the number of the independent variables and β_j is the corresponding parameter. ε_t is the error term, which is assumed to be an IID variable of zero mean and a variance of σ^2.

Although data differencing can solve the problem caused by trended economic variables, valuable information on long-run properties are lost. As a result, the growth rate model pays its attention to short-term dynamics of demand variations. In some pre-2008 studies, the growth rate model was proved to be an appropriate functional form (e.g. Li et al. 2002; Song and Witt 2003). But its utilization among post-2008 studies is quite limited, as it is unable to reveal the long-run relationship between tourism demand and its influencing factors. Falk (2014) applied the growth rate model and found that it was unreliable when evaluating climate’s impact on international and domestic tourism demand of Australia.

2.3.2.4 The Partial Adjustment Model

Dynamics was introduced to tourism demand models in the early 1980s in a simple form: partial adjustment (Kliman 1981; Witt 1980a; 1980b). The partial adjustment model assumes that only lagged values of dependent variables and current values of independent variables have influences on dependent variable. Since the 1980s, the partial adjustment model has become a popular dynamic specification, and has been extensively used when modelling habit persistence, word-of-mouth (WOM) effect or supply constraints (e.g. Marin and Witt 1988; Song et al. 2003b). It can be specified as:

$$y_t = \sum_{j=1}^{n} \beta_j x_{j,t} + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \qquad 2.23$$

where y_t represents a measure of tourism demand, $x_{j,t}$ are the identified explanatory variables with the number of n and corresponding parameters of β_j, p is the lag length of the dependent variable, ϕ_i is the coefficient on y_{t-i}, which needs to be estimated and ε_t is the error term, which is assumed to be an IID variable of zero mean and a variance of σ^2.

A special case of the partial adjustment model receives great attention, which is the simple dynamic (SD) model. It is achieved by only including one lagged dependent variable (y_{t-s}). In the tourism demand literature, simple dynamic specification is widely chosen, especially in panel data analysis (e.g. Morley 2009; Seetaram 2010; Lorde et al. 2015; Albaladejo et al. 2016; Habibi 2016). Besides, it is also popular with time series data studies. For example, the simple dynamic model was applied to forecast US commercial air travel in Carson et al. (2011), which focused on 179 busiest airports in US and used the ratio of the number of passengers originating from an airport to the population of the Metropolitan Statistical Area served by the airport as the dependent variable.

The simple dynamic model can be specified as:

$$y_t = \sum_{j=1}^{n} \beta_j x_{j,t} + \phi_1 y_{t-s} + \varepsilon_t \qquad 2.24$$

where s represents the number of seasons under consideration. When annual data is used, $\mathrm{s} = 1$; with seasonal data, $\mathrm{s} = 4$, and with monthly data, $\mathrm{s} = 12$.

2.3.2.5 Cointegration Analysis and Error Correction Models

The cointegration and error correction models were presented by Engle and Granger (1987). A time series variable is stationary when its mean, variance and covariance remain constant over time (Song, Witt and Li 2009). Tourism demand variables, which are often non-stationary, may belong to the same economic system, and a linear combination of them can be stationary. If it is the case, the collection is said to be cointegrated. According to Engle and Granger (1987), if a pair of non-stationary economic variables, x_t and y_t, belongs to the same economic system, there should be an attractor or cointegration relationship that prevents these two time series from drifting away from each other; that is, there exists a force of equilibrium that keeps the two variables moving together in the long run.

Cointegration, which represents the long-run equilibrium relationship between a collection of non-stationary time series variables, has been widely used in modelling and forecasting tourism demand for decades since the late 1990s. The early applications include Kulendran (1996), Kulendran and King (1997) and Kim and Song (1998). In the current literature, considerable attention has been paid to testing the existence of cointegration relationships in levels between variables, and various cointegration estimation techniques such as Engle-Granger two-step approach (EG-ECM) (Engle and Granger 1987), Wickens-Breusch one-step approach (WB-ECM) (Wickens and Breusch 1988), the bounds test cointegration approach (Pesaran et al. 2001) and the Johansen maximum likelihood (JML) approach (Johansen 1988) have been widely applied.

Unit Root Tests

Before testing for the existence of cointegration relationship, unit root tests are necessary, as different cointegration techniques have specific requirements on the integration orders of model variables. A non-stationary time series variable, which is also referred to as an integrated process, has as many unit roots as the number of times that the variable needs to be differenced to achieve stationary. For instance, if a time series variable becomes stationary after taking first differences, the variable is said to be integrated of order one, and it has one unit root. Except the bounds test cointegration approach, which does not require that the variables under consideration are integrated of the same order, the other three techniques all demand that all variables are integrated of order one. All techniques are not valid for variables that are integrated of order two or above.

There are different unit root tests available including the Dickey-Fuller (DF) test, the augmented Dickey-Fuller (ADF) test (Dickey and Fuller 1981) and the Phillips-Perron (PP) test (Phillips and Perron 1988), and the latter two are generally used in the tourism demand literature.

The Dickey-Fuller Test

The DF test is based on the following auxiliary equation, assuming that the time series under consideration (y_t) can be modeled by an autoregressive process of order 1 $(\mathrm{AR}(1))$:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \varepsilon_t \qquad 2.25$$

where y_t is the time series under consideration, α_0 is the constant form, α_1 is the autocorrelation coefficient, and ε_t is an error term.

The null hypothesis of y_t is a non-stationary process or there is at least one unit root is defined as: $H_0: \alpha_1 = 1$, and the alternative hypothesis of y_t is a stationary process is defined as: $H_1: \alpha_1 < 1$. Instead of testing $\alpha_1 = 1$ directly, Dickey and Fuller (1979) transformed equation 2.25 to the following form:

$$\Delta y_t = \alpha_0 + \phi y_{t-1} + \varepsilon_t \qquad 2.26$$

where $\Delta y_t = y_t - y_{t-1}$, $\phi = \alpha_1 - 1$, and the null hypothesis is $H_0: \phi = 0$ against the alternative hypothesis $H_0: \phi < 0$.

The commonly used statistic for $\phi = 0$ is the t ratio: $\hat{\phi}/[SE(\hat{\phi})]$, where $\hat{\phi}$ stands for the estimated value of ϕ with $SE(\hat{\phi})$ as the standard error of the estimation. Under the null hypothesis of non-stationarity, the t ratio has a non-standard distribution, and the conventional values for the t statistic are not applicable. Based on Monte Carlo simulation, Dickey and Fuller (1979) obtained appropriate critical values. If the calculated t value is lower than the critical value, the null hypothesis of non-stationarity should be rejected. Otherwise, if the calculated t value is greater than the critical value, the null hypothesis should be accepted, which means that time series y_t has at least one unit root. The test is then repeated on the differenced series of y_t:

$$\Delta^2 y_t = \alpha_0 + \phi \Delta y_{t-1} + \varepsilon_t \qquad 2.27$$

where $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$.

If the null hypothesis of $H_0: \phi = 0$ is rejected in equation 2.27, the time series y_t is considered as integrated of order one, which is denoted by $I(1)$. If the null hypothesis of $H_0: \phi = 0$ is accepted in equation 2.27, the DF test should be conducted again on higher differenced series of y_t. The process will be repeated until the null hypothesis of non-stationarity is rejected, and the integration order or the number of unit roots can be confirmed.

The Augmented Dickey-Fuller Test

One major problem of the DF test is that the residuals in the auxiliary equation (equation 2.26) may have serial correlation, which can be augmented by including lagged dependent variables. The pth order ADF statistic is based on the following equation:

$$\varDelta y_t = \alpha_0 + \phi y_{t-1} + \sum_{i=1}^{p} \varphi_i \, \varDelta y_{t-i} + \varepsilon_t \qquad 2.28$$

where p is the lag length for the dependent variable, and φ_i is the parameter of the lagged dependent variable y_{t-i}.

The null hypothesis of non-stationarity is defined as: $H_0\colon \phi = 0$, and the alternative hypothesis of stationarity is defined as: $H_0\colon \phi < 0$. Similar to the DF test, the t statistic ($\hat{\phi}/[SE(\hat{\phi})]$, where $\hat{\phi}$ stands for the estimated value of ϕ with $SE(\hat{\phi})$ as the standard error of the estimation) is calculated and compared to the critical values. When the calculated t value is lower than the critical value, the null hypothesis of non-stationarity is rejected, which would be accepted when the t statistic is higher than the critical value.

To test the null hypothesis of non-stationarity, one important step is to select the appropriate lag length for the dependent variable. Too few lags may result in over-rejection of the null hypothesis when it is true, and too many lags may reduce the power of the test due to loss of degrees of freedom. The criteria for selecting the lag length is to minimize both the AIC and the SIC.

The Philips-Perron Test

The assumptions of the DF test and the ADF test include that the error terms in the auxiliary regression is IID, which are very restrictive. The PP test is a generalization procedure of the ADF test which relaxes the assumptions. It is nonparametric with respect to nuisance parameters and thereby allows for a very wide class of weakly dependent and possibly heterogeneously distributed data (Phillips and Perron 1988). The expressions of the PP test are extremely complex and beyond the scope of this study.

EG-ECM

According to Engle and Granger (1987), to test whether there exists cointegration relationship between a set of variables which are integrated of order one, a long-run static regression is run and estimated by ordinary least square (OLS) firstly. If the estimated residual is stationary, then the cointegration relationship is confirmed. The long-run static regression takes the following form:

$$y_t = k_0 + \sum_{j=1}^{n} k_j \, x_{j,t} + \varepsilon_t \qquad 2.29$$

where y_t is the dependent variable, and x_{jt} are the independent variables of number n with corresponding parameters of $\mathrm{k_j}$, and ε_t is the error term.

The estimated residual, $\hat{\varepsilon}_t$ is assumed to follow an autoregressive process of order p (AR(p)):

$$\hat{\varepsilon}_t = \sum_{i=1}^{p} \vartheta_i \hat{\varepsilon}_{t-i} + e_t \qquad 2.30$$

The unit root test is then employed to test whether the estimated residual, $\hat{\varepsilon}_t$, is stationary or not based on:

$$\Delta\hat{\varepsilon}_t = \phi^* \hat{\varepsilon}_{t-1} + \sum_{i=1}^{p-1} \gamma^* \Delta\hat{\varepsilon}_{t-i} + e_t \qquad 2.31$$

The test statistic is the t values of the estimated coefficient ϕ^*. The null hypothesis is that the estimated error term in the long-run static model (equation 2.29) is a non-stationary process, i.e., there exists no cointegration relationship between y_t and $x_{j,t}$. If $\hat{\varepsilon}_t$ is found to be stationary, the null hypothesis is rejected, which means that $\mathrm{y_t}$ and $\mathrm{x_{j,t}}$ are cointegrated.

After confirming the cointegration relationship, the second step is to estimate the error correction equation:

$$\Delta y_t = \sum_{j=1}^{n}\sum_{i=0}^{p_j} \beta_{j,i}\,\Delta x_{j,t-i} + \sum_{i=1}^{p} \phi_i\,\Delta y_{t-i} - \lambda\hat{\varepsilon}_{t-1} + u_t \qquad 2.32$$

where Δy_t signifies the first difference of the dependent variable with $\Delta y_t = y_t - y_{t-1}$, $\Delta x_{j,t}$ represents the first differences of the explanatory variables with $\Delta x_{j,t} = x_{j,t} - x_{j,t-1}$, n is the number of the independent variables, $p_j\ and\ p$ are the lag lengths of independent and dependent variables respectively, $\hat{\varepsilon}_{t-1} = y_{t-1} - \widehat{k_0} - \sum_{j=1}^{n} \hat{k}_j\, x_{j,t-1}$ is the OLS residuals from equation 2.29, and the lag structure of the differenced variables ($\Delta y_{t-i}\ and\ \Delta x_{j,t-i}$) is determined by experimentation.

Equation 2.32 integrates both the long-run and short-run relationships between model variables: the estimated coefficients on the level terms, i.e., $\widehat{k_0}$ and $\hat{\mathrm{k}}_\mathrm{j}$ can be interpreted as the long-run demand elasticities; and the estimated coefficients on the first differenced terms ($\Delta y_{t-i}\ and\ \Delta x_{j,t-i}$), i.e., $\hat{\beta}_{j,i}$ and $\hat{\phi}_i$ represent the short-run dynamics. The term $-\lambda\hat{\varepsilon}_{t-1}$ represents the error correction mechanism, and the coefficient $-\lambda$ is called the adjustment speed, whose value is between -1 and 0. The adjustment speed shows the extent to which the variables have the tendency to converge towards the long-run equilibrium value following shocks, or more specific, the system will adjust itself towards equilibrium by removing λ of a unit from the error made from the previous period (Smeral 2010; Song, Witt and Li 2009, p89). So, the larger the value of λ is, the faster the adjustment is.

The assumption of EG-ECM is that y_t and $x_{j,t}$ are integrated of order one. Stationary variables except event dummies or time trend cannot be included, and any variables that are integrated of higher orders need to be differenced accordingly to reduce their integration orders.

EG-ECM has some major drawbacks. Firstly, the cointegration regression (equation 2.32) requires all the variables to be integrated of order one (denoted by $\mathrm{I}(1)$), so higher order of integration should be differenced, which results in loss of information. Secondly, since the variables (x_t and y_t) are not stationary, OLS estimation should not be chosen in the first step (equation 2.29), which means that EG-ECM is unreliable, especially in small samples, as it chooses the wrong estimation method in the first place. As a result, EG-ECM is unpopular among post-2008 studies with only a few applications (Song et al. 2009; Song 2010; Lee 2011b; Cheng 2012). Other approaches including the WB-ECM and the bounds test cointegration approach have emerged in the tourism demand literature to address these problems (Choyakh 2008; Shen et al. 2008; 2009; 2011; Song et al. 2011b; Liu 2014; Lin et al. 2015).

WB-ECM

Wickens and Breusch (1988) presented a method to test for cointegration as well as estimate the long-run and short-run parameters in a single step by re-parameterizing ADLM. Equation 2.20 can be re-parameterized into the following form on which the Wickens-Breusch estimation is based:

$$\Delta y_t = \alpha + \sum_{j=1}^{n}\sum_{i=0}^{p_j-1} \beta_{ji}\,\Delta x_{j,t-i} + \sum_{i=1}^{p-1} \phi_i\,\Delta y_{t-i} + \lambda_0 y_{t-1} + \sum_{j=1}^{n} \lambda_j x_{j,t-1} + u_t \qquad 2.33$$

where Δy_t signifies the first difference of the dependent variable with $\Delta y_t = y_t - y_{t-1}$, $\Delta x_{j,t}$ represents the first differences of the explanatory variables with $\Delta x_{j,t} = x_{j,t} - x_{j,t-1}$, n is the number of the independent variables, p_j and p are the lag lengths of independent and dependent variables respectively, the lagged level variables (y_{t-1} and $x_{j,t-1}$) represent the cointegration relationship, and u_t is the error term.

The cointegration relationship, which is represented by the lagged level variables, can be tested by the coefficient Wald test. The test is run regarding all the lagged variables in the model, and the null hypothesis of no cointegration is defined as:

$$\mathrm{H_0}: \lambda_0 = \lambda_1 = \lambda_2 = \cdots = \lambda_\mathrm{n} = 0 \qquad 2.34$$

When the null hypothesis is rejected, the cointegration relationship is confirmed and the long-run cointegration parameters are estimated by OLS, which can be derived from the following estimation:

$$y_t = k_0^* + k_j^* x_{jt} \qquad 2.35$$

where $\mathrm{k}_0^* = -\frac{\hat{\alpha}}{\widehat{\lambda_0}}$ and $\mathrm{k}_\mathrm{j}^* = -\frac{\widehat{\lambda_\mathrm{j}}}{\widehat{\lambda_0}}$.

This approach also requires that only variables that are integrated of order one can be included and pre-testing of the integration order is necessary. Examples of the application of WB-ECM among post-2008 studies include Shen et al. (2008; 2009; 2011) and Liu (2014).

The Bounds Test Cointegration Approach

As a more-recently developed technique, the bounds test cointegration approach (Pesaran et al. 2001) is more flexible as it can be conducted irrespective of whether the underlying variables are integrated of order zero or one, or whether they are mutually co-integrated. It has become the most popular cointegration technique in the tourism demand literature with a number of applications among post-2008 studies (e.g. Onafowora and Owoye 2012; Otero-Giráldez et al. 2012; Seetaram et al. 2014; Lin et al. 2015; Saayman and Saayman 2015).

The bounds test cointegration technique can also be used to generate interval estimates and hence has the ability to produce interval forecasts. For example, Song and Lin (2010) applied the bounds test approach to evaluate the impacts of financial and economic crisis on Asia tourism. They used the delta method to generate interval estimates of demand elasticities and produced interval forecasts. Arguing that the delta method is inappropriate to generate confidence intervals when the statistic of interest does not follow a normal distribution, Song et al. (2010a) introduced Kilian's (Kilian 1998) bias-corrected bootstrap, a statistical method which had been proved to provide accurate and reliable confidence intervals for demand elasticities, to tourism demand analysis when modelling Hong Kong tourism based on the bounds test approach. A similar method was applied in Otero-Giráldez et al. (2012) to evaluate the long-run effects of socioeconomic and meteorological factors on tourism demand. Although these two studies did not generate interval forecasts, they have provided evidences showing the successful applications of (bias-corrected) bootstraps in generating interval estimates of tourism demand elasticities.

The detailed discussion of the bounds test cointegration approach is provided in section 3.4.

The JML Approach

The JML approach (Johansen 1988) is based on the vector autoregressive (VAR) framework, which belongs to the system-of-equations approach. The underlying cointegration test is an extension of the DF unit root test. The JML approach is invalid if the variables are not integrated of order one. It is preferable when there is more than one cointegration relationship existing, as it can detect multiple cointegration relationships (Song et al. 2003b). Examples of the application of the JML model include McAleer (2001a), Saayman and Saayman (2008), Shen et al. (2008; 2009; 2011), Bonham et al. (2009), Lim and Goh (2012) and Vanegas, Sr. (2013). Detailed discussion of the JML approach is provided in section 2.3.3.1.

2.3.2.6 The Leading Indicator Model

Another type of dynamic specification is the leading indicator model, which is a useful tool for macroeconomic forecasting (Song, Witt and Li 2009). It was introduced to the tourism demand literature in the 1990s, and early applications include Turner et al. (1997) and Kulendran and Witt (2003). The LI model assumes that only lagged independent variables have impacts on current values of dependent variable, and can be specified as:

$$y_t = \sum_{j=1}^{n} \sum_{i=1}^{p_j} \beta_{j,i}\, x_{jt-i} + \varepsilon_t \qquad 2.36$$

where y_t represents a measure of tourism demand, $x_{j,t-i}$ are the identified explanatory variables with the number of n and corresponding parameters of $\beta_{j,i}$, p_j signifies the lag length of independent variables, and ε_t is the error term which is assumed to be an IID variable of zero mean and a variance of σ^2

For the LI model to behave well, it is crucial to identify the appropriate indictors, which is usually done by trial and error as there lacks theoretical justification. Besides, it is required that the coefficients of the leading indicators to be constant in the period under study, because '*there is no point in building a leading-indicator model if the link between what happens today and the previous values of the indicator will not persist over the relevant forecast horizon* (Hendry 1995, p252). Applications of the leading indicator model among the post-2008 studies include Kulendran and Wong (2009; 2011) and Yap and Allen (2011).

2.3.2.7 The Time Varying Parameter Model

Traditional regression analysis assumes that the parameters of explanatory variables stay unchanging over the time span under consideration, while theTVP model relaxes this restriction and allows for temporal changes in parameters, therefore possesses the ability of capturing dynamics of demand elasticity and assessing whether tourists' taste preferences evolve over time.

TVP is normally specified in a state space form (SSF) and estimated by Kalman filter algorithm (Kalman 1960). A linear SSF can be written as:

$$y_t = x_t'\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.(0, H_t) \qquad 2.37$$

$$\beta_{t+1} = T_t\beta_t + \eta_t, \beta_1 \sim N(b_1, P_1), \eta_t \sim i.i.d.(0, Q_t) \qquad 2.38$$

where y_t represents dependent variable; x_t is a row vector of k explanatory variables with β_t being the corresponding column vector of k coefficients called the state vector, T_t is a $\mathrm{k} \times \mathrm{k}$ matrix, ε_t is a vector of residuals with zero mean and constant variance matrix H_t, and η_t is a vector of residuals with zero mean and constant variance matrix Q_t. And ε_t refers to the temporary disturbance and η_t is the permanent disturbance, which are serially independent and independent of each other at all-time points. The matrices x_t, H_t and Q_t are initially assumed to be known.

Equation 2.37 is called the *observation equation* and Equation 2.38 is called the *state equation*, which explains the evolution of the unobserved part of the model. In most economic applications, it is assumed that $T_t = I$, where I is the identity matrix. In this sense, β_t follows a multivariate random walk. The initial vector of β_t, i.e. β_1, has a mean of b_1, which can be estimated by maximum likelihood from the first few observations of y_t *and* x_t, and P_1 is its variance (Durbin and Koopman 2012; Harvey 1990).

The TVP model was firstly applied in the tourism context in the early 2000s, and the pioneer works include Song et al. (2003a), Song and Witt (2000) and Song and Wong (2003). It has become one of the most popular econometric models with many successful applications among post-2008 studies (e.g. Shen et al. 2011; Page et al. 2012; Wu et al. 2012; Untong et al. 2014; Zhou-Grundy and Turner 2014).

The latest development of the TVP model is by Song *et al.* (2011), who incorporated TVP estimation of the explanatory variable coefficients and a time series technique called structural time series model (STSM) to present the time varying parameter- structural time series model (TVP-STSM). The TVP-STSM was employed to model and forecast quarterly tourist arrivals to Hong Kong from China, South Korea, UK and US, and was proved to outperform other seven competitors including the STSM, the causal structural model (CSM), the TVP model, the ADLM, the SARIMA model, the Naïve 1 and Naïve 2 models.

The specification of the TVP-STSM follows:

$$y_t = \mu_t + \gamma_t + \psi_t + x_t\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim \mathrm{IID}(0, H_t) \tag{2.39}$$

$$\beta_{t+1} = T_t\beta_t + \eta_t, \qquad \beta_1 \sim N(b_1, P_1), \qquad \eta_t \sim i.i.d.(0, Q_t) \tag{2.40}$$

$$\begin{gathered}\mu_t = \mu_{t-1} + \delta_{t-1} + \varrho_t, \qquad \delta_t = \delta_{t-1} + \zeta_t, \\ \varrho_t \sim i.i.d.\left(0, \sigma_\varrho^2\right), \qquad \zeta_t \sim i.i.d.\left(0, \sigma_\zeta^2\right)\end{gathered} \tag{2.41}$$

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + w_t, \qquad w_t \sim i.i.d.(0, \sigma_w^2) \tag{2.42}$$

$$\begin{cases} \psi_t = \rho\cos\lambda_c\psi_{t-1} + \rho\sin\lambda_c\psi_{t-1} + \kappa_t \\ \psi_t^* = -\rho\sin\lambda_c\psi_{t-1}^* + \rho\cos\lambda_c\psi_{t-1}^* + \kappa_t^* \end{cases}, \qquad \kappa_t, \kappa_t^* \sim i.i.d.(0, \sigma_\kappa^2) \tag{2.43}$$

where y_t is the observed time series comprised of a stochastic trend (μ_t), a seasonal term (γ_t), a cyclical component (ψ_t) and a random term (ε_t). y_t is influenced by a vector of explanatory variables x_t with β_t being the corresponding vector of coefficients called the state vector. Equation 2.39 is called the *observation equation* and Equation 2.40 is called the *state equation*. In most economic applications, it is assumed that $T_t = I$, where I is the identity matrix. The temporary disturbance ε_t and the permanent disturbance η_t are assumed to be serially independent and independent of each other at all-time points with zero means and variance matrix as H_t and Q_t respectively.

The trend μ_t is assumed to be a random walk with stochastic random drift δ_t. The seasonal component γ_t is assumed to follow a stochastic dummy specification where s depends on the data frequency (for quarterly data, $s = 4$ and for monthly data, $s = 12$), and when $\sigma_w^2 = 0$, γ_t reduces to be a deterministic seasonal component. The cyclical component is specified in the trigonometric form, where ρ is a damping factor between zero and one, λ_c is the frequency of the cycle measured in radians. $\varrho_t, \zeta_t, w_t, and\ \kappa_t/\kappa_t^*$ are mutually uncorrelated white-noise Gaussian errors with corresponding variances of σ_ϱ^2, $\sigma_\zeta^2, \sigma_\omega^2\ and\ \sigma_\kappa^2$ respectively.

When dropping the time subscript of β_t, the model reduces to a standard CSM, which takes the following form:

$$y_t = \mu_t + \gamma_t + \psi_t + x_t\beta + \varepsilon_t, \qquad \varepsilon_t \sim IID(0, H_t) \tag{2.44}$$

$$\mu_t = \mu_{t-1} + \delta_{t-1} + \varrho_t, \qquad \delta_t = \delta_{t-1} + \zeta_t \tag{2.45}$$

$$\varrho_t \sim i.i.d.\left(0, \sigma_\varrho^2\right), \qquad \zeta_t \sim i.i.d.\left(0, \sigma_\zeta^2\right),$$

$$\gamma_t = -\sum_{j=1}^{s-1} \gamma_{t-j} + w_t, \qquad w_t \sim i.i.d.(0, \sigma_w^2) \qquad 2.46$$

$$\begin{cases} \psi_t = \rho \cos \lambda_c \psi_{t-1} + \rho \sin \lambda_c \psi_{t-1} + \kappa_t \\ \psi_t^* = -\rho \sin \lambda_c \psi_{t-1}^* + \rho \cos \lambda_c \psi_{t-1}^* + \kappa_t^* \end{cases}, \qquad \kappa_t, \kappa_t^* \sim i.i.d.(0, \sigma_\kappa^2) \qquad 2.47$$

The explanation of the CSM is similar to the TVP-STSM, except that the vector of parameters (β) of the identified influencing factors (x_t) is assumed to be stable over time.

2.3.3 The System-of-Equations Approach

The system-of-equations approach is adopted mainly to address the limitations of the single-equation approach. There are two major types of system-of-equations models: the vector autoregressive (VAR) model and the almost ideal demand system (AIDS).

2.3.3.1 The Vector Autoregressive Model

The VAR Model

One major restriction of the single equation models is the exogeneity assumption, which requires that all independent variables are exogenous of the model. Validation of the exogeneity assumption results in invalid OLS estimation. To relax this assumption, Sims (1980) presented the unrestricted VAR model. The VAR model assumes that, except the deterministic variables such as time trend, constant and dummy variables, all variables are endogenous (Song and Witt 2006).

The VAR model regresses the current values of all the variables on all the lagged values of the same set of variables in the system, and therefore provides a multivariate framework where changes in a particular variable are related to changes in its own lags and in other variables in the system. In forecasting practice, one major benefit of this approach is that the system can generate forecasts for all variables simultaneously, which saves the trouble of obtaining projections for influencing factors separately.

A general VAR(p) with m variables can be expressed in a matrix form:

$$y_t = \Pi_1 y_{t-1} + \Pi_2 y_{t-2} + \cdots + \Pi_p y_{t-p} + H x_t + U_t, \qquad U_t \sim IID(0, \Sigma) \qquad 2.48$$

where $y_t (m \times 1)$ are vectors of m variables in the system which are all treated endogenous, Π_p $(m \times m)$ are the corresponding matrices of parameters, $x_t (d \times 1)$ represents a vector of deterministic variables including intercept, trend or seasonal dummies, and $U_t (\mathrm{m} \times 1)$ is a vector of regression errors which are assumed to be contemporaneously correlated but not autocorrelated with Σ $(\mathrm{m} \times \mathrm{m})$ as the covariance matrix.

In the tourism demand literature, most VAR models identify tourism demand of one destination-origin pair and its influencing factors such as disposable income, own price and substitute price as endogenous variables (e.g. Athanasopoulos et al. 2011; Shen et al. 2011; Daniel and Rodrigues 2012; Gunter and Önder 2015). Valadkhani and O'Mahony (2015) chose tourism demands of different destination-origin pairs as endogenous variables in the system and analyzed the dynamics of Australian demand in a multimarket context. Since there were no mainstream influencing factors of tourism demand included in the model, the VAR model applied was classified as a multivariate time series technique.

The latest development of the VAR model is the introduction of the Bayesian VAR (BVAR) model and the global VAR (GVAR) model. The BVAR model is estimated based on Bayesian statistical theory, which combines priors with sample information in model estimation, and it has been proved to be a robust forecaster in the literature relating to macroeconomic forecasting (e.g. Ramos 2003). The BVAR model was firstly introduced to the tourism demand literature by Wong et al. (2006), who compared the forecasting accuracy of various VAR models. The information prior used in the BVAR model was the Minnesota prior, which imposed restrictions on the more distant lags of a VAR rather than eliminating them. He found that introducing the Bayesian priors to the unrestricted VAR process lead to more accurate forecasts. In a later study (Gunter and Önder 2015), the BVAR model was found to possess superior forecasting ability among a range of causal models and time series techniques.

The GVAR model, which was first proposed by Pesaran et al. (2004), links up a range of regional systems into a unified global system. It was initially applied to macroeconomics studies on global economic linkages and is also proper for tourism demand studies on a global setting. The GVAR model can overcome the common issues of tourism demand models, which are endogeneity and over-parameterization.

Cao et al. (2017) introduced the GVAR model to the tourism demand literature to quantify the co-movements of tourism demand at a global level. They showed that the negative shocks to China's real income and tourism price variables would induce fluctuations in international tourism demand and tourism prices in almost all the 24 countries under consideration in the short run and would affect the developed countries more in the long run. Assaf et al. (2018) applied the Bayesian Global VAR (BGVAR) to forecast tourism demand of nine countries in Southeast Asia. They showed that the BGVAR model could capture the spillover effects of international tourism demand in the region and had superior forecasting ability compared to three alternative VAR models: the VAR, BVAR and GVAR models when one- to four-step-ahead forecasting horizons are considered. The applications of BVAR, GVAR and BGVAR are still limited in the tourism demand literature.

The JML Cointegration Approach

Based on the VAR framework, Johansen (1988) presented the Johansen maximum likelihood approach (JML), which has become a popular estimation procedure for cointegration. It relaxes the assumption of other cointegration techniques that there is only one cointegration relationship among model variables, and hence is more reliable when there are more than one co-integrating vectors existing. The EG-ECM, WB-ECM and bounds test cointegration approach are all based on the single-equation approach, which implicitly assumes that there is only one cointegration relationship among a set of economic variables. '*However, this assumption is too restrictive and sometimes is unrealistic. In reality, there may be more than one cointegration relationship if the long-run model involves more than two variables*' (Song, Witt and Li 2009, p127).

The JML cointegration approach extends the multivariate Dicken-Fuller unit root test in the VAR framework. By taking differencing, equation 2.48 can be transformed to the following form:

$$\Delta y_t = \sum_{i=1}^{p-1} \Phi_i \Delta y_{t-i} + \Phi y_{t-p} + Hx_t + U_t \qquad 2.49$$

where $\Phi_i = -(I - \Pi_1 - \Pi_2 - \cdots - \Pi_i)$ and $\Phi = -\left(I - \Pi_1 - \Pi_2 - \cdots - \Pi_p\right)$.

Equation 2.49 is known as a vector error correction model (VECM), and the error correction term is embodied in the term Φy_{t-p}. The JML approach generates maximum likelihood estimators rather than applying OLS estimation, and the parameter matrices Φ and Φ_i represent the long-run and short-run adjustments to the changes in y_t respectively.

The cointegration test is to detect the rank of the matrix Φ, as the number of cointegrating vectors equals the number of characteristic roots of the matrix Φ that are different from zero. The number of characteristic roots can be tested by two tests. The first one is the trace test, which has the null hypothesis as there are at most r cointegrating vectors, that is, the rank of the matrix Φ is less than or equal to r. The alternative hypothesis is that there are more than r cointegrating vectors, that is, the rank of the matrix Φ is greater than r.

The second test is the maximal eigenvalue test. The null hypothesis of the maximal eigenvalue test is that the rank of Φ is r, which means that the number of the cointegration vectors is r. And the alternative hypothesis of the maximal eigenvalue test is that the rank of Φ is $\mathrm{r}+1$, which means that the number of the cointegration vectors is $\mathrm{r}+1$.

Johansen and Juselius (1990) provided the critical values for these two statistics. The test begins with $\mathrm{r}=0$. When the null hypothesis of there is zero ($\mathrm{r}=0$) cointegration vector (no cointegration) rejected, r is reset to be the next higher integer value (here is $\mathrm{r}=1$). Such a procedure is followed until the null hypothesis is accepted and the alternative hypothesis is rejected, i.e. it is confirmed that there are r cointegration vectors.

2.3.3.2 The Almost Ideal Demand System

The AIDS can address another major limitation of the single-equation approach, which is that the single-equation models cannot capture the interdependence of budget allocations to different consumer goods and services. The AIDS was presented by Deaton and Muellbauer (1980), who provided solid theoretic underpinning as to which explanatory variables should be included in the model based on consumer demand theory.

According to consumer demand theory, consumers face the choice among a number of products under a budget constraint. The consumption of one product interacts with the consumption of others, and the changes in the price of one product influences the whole system. AIDS can adequately model the interaction among the consumption of different products. In the context of tourism demand, different products can be competing destinations or various tourism services and products.

The static AIDS is specified as:

$$w_i = a_i + \sum_j \gamma_{ij} logp_j + b_i \log\left(\frac{z}{P}\right) + v_i, \quad (i,j = 1,2,\dots,n) \tag{2.50}$$

$$\log\mathrm{P} = a_0 + \textstyle\sum_i \alpha_i logp_i + \frac{1}{2}\sum_i \sum_j \gamma_{ij} logp_i\, logp_j \tag{2.51}$$

where w_i is the budget share of the ith good, p_j is the price of the jth good, z is the total expenditure on all goods in the system, P is the aggregate price index for the system, v_i is the disturbance term with $v_i \sim i.i.d.\left(0,\sigma_i^2\right)$ and n is the number of the goods (or equations) in the system.

When the aggregate price P index is replaced by P^* with $\log P^* = \sum_i w_i logp_i$, the AIDS reduces to a linear AIDS (LAIDS), which is generally used in empirical studies. The static AIDS can also be combined with other modelling techniques such as the ECM and TVP to explore the dynamics in tourists' behavior.

The AIDS was firstly introduced to the tourism demand literature in the 1980s, but it did not attract much attention until the late 1990s (e.g. Sinclair et al. 2003; Li et al. 2005). Compared to other econometric techniques, the AIDS remain unpopular among post-2008 tourism demand studies with only a handful applications. One major reason is that researchers' interest mainly lies in exploring the determinants of tourism demand and the impacts of them, which can be addressed by the single-equation models, other than tourists' budget allocation.

In the current tourism demand literature, goods in the AIDS can be various tourism products and services including accommodation, dining, sightseeing and shopping, and the aim of such studies is to explore tourists' consumption pattern in one destination (e.g. Divisekera 2010; Wu et al. 2011; 2012). There are also studies considering various destinations including competing international ones for one origin, or domestic and foreign destinations for one market as goods in the system with the purpose of analyzing the substitution effect in a multi-market context (e.g. Mangion et al. 2012; Athanasopoulos et al. 2014). However, this technique has not gained popularity in forecasting studies, with only one exception (Cortes-Jimenez et al. 2009) seen among post-2008 publications, which applied the long run AIDS model and the EC-AIDS model to evaluate Italian outbound tourism to four main European destinations and found that the EC-AIDS model outperformed the long run model in forecasting accuracy.

2.3.4 Studies Based on Panel Data

A new trend in the current tourism demand literature is the increasing popularity of panel data analysis, with a large number of applications (e.g. Khadaroo and Seetanah 2008; Kuo et al. 2009; Falk 2010; Albaladejo et al. 2016; Balli et al. 2016; Yazdi and Khanalizadeh 2016; Li et al. 2016; Dogru et al. 2017; Buigut 2018; Long et al. 2018). The superiority of panel data models over pure time series ones lies in the relatively large number of observations and the consequent increase in degrees of freedom, which reduces collinearity and improves estimate efficiency (Song, Witt and LI 2009, P149).

The research aim of panel data studies lies mainly in assessing the impact of a particular influencing factor on tourism demand or identifying the determinants of tourism demand and estimating the corresponding elasticities. For instance, Seeteram (2012b) explored the relationship between immigration and inbound tourism demand of Australia from 15 main markets using simple dynamic panel data models. Resident population born overseas was chosen as the proxy for immigration, and the data was only available in census years of 1981, 1986, 1991, 1996, 2001 and 2006. The method of White (2007) was applied to calculate a time-consistent data series. The finding suggested that immigration helped to explain tourist arrivals to Australia from 1980-2008. Su and Lin (2014) employed static panel data models to study world heritage sites' impact on tourism demand of 66 destinations and showed that positive relationships did exist between number of WHSs and tourism demand. Massidda and Etzo (2012) implemented the system generalized method of moments (GMM) estimation procedure in a simple dynamic panel data model to investigate the main determinants of Italian domestic tourism. The influencing factors they considered were chosen based on an extended gravity model.

There are three papers utilizing panel data when forecasting tourism demand (Moore 2010; Fildes et al. 2011; Long et al. 2018). Moore (2010) utilized panel ECM to investigate the impact of climate change on Caribbean tourism demand and produced forecasts of tourism arrivals based on different scenarios for future climatic conditions. The forecasting performance of the dynamic panel data model and two naïve models were evaluated with the result showing that the panel ECM was superior to two naïve counterparts over both short and long horizons. Fildes et al. (2011) concentrated on airline traffic prediction and evaluated the forecasting ability of the pooled ADLM, the TVP model, the VAR model, and an automatic method for econometric model specification. After an empirical study of UK inbound and outbound tourism demand, they concluded that the pooled ADLM including the world trade variable was hard to beat. Long et al. (2018) evaluated whether pooling can improve tourism demand forecasting performance through an empirical study on domestic tourism demand of 341 cities in China based on yearly data from 2005 to 2013. They found that when both spatial and temporal effects were incorporated, the pooled OLS model outperformed the OLS and naïve models.

2.3.4.1 General Model Specification

A general panel data model takes the form of:

$$y_{it} = \alpha + x'_{it}\beta_{it} + \delta_i + \gamma_t + \epsilon_{it} \qquad 2.52$$

where y_{it} denotes the dependent variable, and x_{it} is a k-vector of regressors with coefficients β_{it}, and ϵ_{it} is the error term for $i = 1,2, \dots., M$ cross-section units during periods $t = 1,2, \dots., T$. The parameter α is the overall constant in the model, while δ_i represents the cross-section-specific effect, and γ_t stands for the period-specific effect, which both can be random or fixed.

The coefficients β_{it} can be divided into sets of common (β), cross-section-specific (β_i), as well as period-specific (β_t) regressor parameters. Most panel data empirical studies in the current tourism demand literature apply the assumption of common parameters of β, which means that equation 2.52 reduces to the following form:

$$y_{it} = \alpha + x'_{it}\beta + \delta_i + \gamma_t + \epsilon_{it} \qquad 2.53$$

Equation 2.53 can be further classified into three categories based on the assumptions regarding the cross-section-specific effect δ_i, and the period-specific effect γ_t: the pooled ordinary least square (POLS) model, the fixed effect (FE) model and random effect (RE) model.

In the POLS model, the intercept is regarded as a constant across all cross-section units for each time period, which means that the cross-section-specific (δ_i) and period-specific (γ_t) effects are both assumed to be zero. On the other hand, the FE and RE models both allow the intercept to vary between cross-section units and/or periods. The difference between the FE and RE models lies in that the FE model treats the cross-section-specific (δ_i) or the period-specific intercept (γ_t) as fixed, while the RE model assumes that the variation in the effects is randomly determined (Song, Witt and Li 2009, p151).

Most tourism demand studies disregard the period-specific intercept (γ_t) and the corresponding model takes the following form:

$$y_{it} = \alpha + x'_{it}\beta + \delta_i + \epsilon_{it} \qquad 2.54$$

The cross-section-specific effect is denoted by δ_i, and if it is assumed to be correlated with x_{it}, we have the FE model. In contrast, if δ_i is assumed to be randomly determined, hence uncorrelated with x_{it}, we have the RE model. Statistical tests must be conducted regarding the choice between these two models, as the appropriate estimation approaches are different under these opposite assumptions.

2.3.4.2 The Simple Dynamic Panel Data Model

Panel data models also varies in terms of dynamic specification. There are static panel data models applied in post-2008 studies which regress the current value of tourism demand on the current values of influencing factors (e.g. Fourie and Santana-Gallego 2011; Granvorka and Strobl 2013; Ridderstaat et al. 2014; Saha and Yap 2014; Su and Lin 2014). More empirical studies chose different dynamic specifications, among which the simple dynamic panel data model has been the most popular one.

The Model

The simple dynamic panel data model takes the following form:

$$y_{it} = \alpha + \gamma y_{i,t-1} + x'_{it}\beta + \delta_i + \epsilon_{it} \qquad 2.55$$

where y_{it} denotes the dependent variable, and x_{it} is a k-vector of regressors with the coefficient β, $y_{i,t-1}$ is the lagged value of the dependent variable with the coefficient γ, and ϵ_{it} is the error term for $i = 1,2, \ldots, M$ cross-section units during periods $t = 1,2, \ldots, T$. The parameter α is the overall constant in the model, while δ_i represents the cross-section-specific effect, which can be random or fixed.

The dynamic is modelled in a simple way, which is embedded in the lagged dependent variable $y_{i,t-1}$. Similar to simple dynamic models which estimate on pure time series data, the lagged dependent variable $y_{i,t-1}$ is included to account for habit persistence or word-of-mouth effect. The simple dynamic panel data model is very popular in the tourism demand literature with many applications (e.g. Seetaram 2010; 2012a; 2012b; Lorde et al. 2015; Albaladejo et al. 2016; Ghaderi et al. 2016; Habibi 2016; Li et al. 2016).

Model Estimation

The OLS estimator is inappropriate for the simple dynamic panel data model and will be inconsistent and biased, as the lagged dependent variable is correlated with the cross-section-specific effect δ_i, be it fixed or random. Besides, the within groups and random effects estimators do not eliminate the bias and are also biased and inconsistent. The difference GMM, which was presented by Arellano and Bond (1991), and the system GMM, which was proposed by Blundell and Bond (1998) are popular approaches to deal with such problems.

The difference GMM involves first-differencing the model to remove the unobserved fixed effect δ_i, and then using lagged levels of the regressors as instrument variables (IV) to solve the problem of the endogeneity of the differenced lagged dependent variables. The system GMM builds a system of equations in both first differences and levels forms and uses lagged first-differences as IV for equations in levels and the usual lagged levels as IV for equations in first-differences. These two techniques are both extensively applied in the tourism demand literature when estimating dynamic panel data models, and examples of applications include Kuo et al. 2008, Massidda and Etzo (2012), Bento (2014), Albaladejo et al. (2016), Balli et al. (2016), Ghaderi et al. (2016), Habibi (2016) and Li et al. (2016).

However, they are not without limitations. Kiviet (1995) has demonstrated that in samples where T is small (smaller than 30), the original bias persists in GMM estimators, and proposed the corrected least square dummy variable (CLSDV) method, which only applies to balanced panel, to correct it. Studies utilizing the CLSDV method include Seetaram (2010; 2012a; 2012b).

2.3.4.3 The Panel Autoregressive Distributed Lag Model

The dynamic in the simple dynamic panel data model is modelled in a very simple way by only including one lagged dependent variable, which may be inadequate to capture the data generating process (DGP). A more general dynamic specification, i.e., the panel ADLM is also available in the current tourism demand literature.

The Model

The panel ADLM model[2] is specified in the following form:

[2] It reduces to a simple dynamic model when $\mathrm{p} = 1, p_i = 0$.

$$y_{it} = \alpha + \sum_{j=1}^{p} \gamma_{ij} y_{i,t-j} + \sum_{j=0}^{p_i} x'_{i,t-j}\beta_{ij} + \delta_i + \epsilon_{it} \qquad 2.56$$

where y_{it} denotes the dependent variable, and $x_{i,t-j}$ is a k-vector of regressors with coefficients β_{ij}, $y_{i,t-j}$ is the lagged dependent variable with the coefficients γ_{ij}, p_i and p are the lag lengths of independent and dependent variables respectively, and ϵ_{it} is the error term for $i = 1,2, \ldots, M$ cross-section units during periods $t = 1,2, \ldots, T$. The parameter α is the overall constant in the model, while δ_i represents the cross-section-specific effect, which can be random or fixed.

When the variables are cointegrated, equation 2.56 can be re-parameterized as a panel ECM:

$$\Delta y_{it} = \sum_{j=1}^{p} \gamma_{ij} \Delta y_{i,t-j} + \sum_{j=0}^{p_i} \Delta x'_{i,t-j}\beta_{ij} + \phi_i\left[y_{i,t-1} - \alpha - x'_{it}\eta_i - \delta_i\right] + \varepsilon_{it} \qquad 2.57$$

where Δy_{it} and $\Delta x_{i,t-j}$ are the first differences of the dependent and independent variables. $\left[y_{i,t-1} - \alpha - x'_{it}\eta_i - \delta_i\right]$ is the error correction term and ϕ_i is the error correction coefficient, which is expected to be negative to adjust deviations from the long-run equilibrium. The estimated coefficients on the level terms, i.e., $\hat{\eta}_i$ can be interpreted as the long-run demand elasticities; and the estimated coefficients on the first differenced terms, i.e., $\hat{\gamma}_j$ and $\hat{\beta}_j$ represent the short-run dynamics.

The panel ADLM is less popular compared to the simple dynamic panel data model with fewer applications (e.g. Moore 2010; Falk 2010; 2015; Massidda and Piras 2015; Yazdi and Khanalizadeh 2016).

Model Estimation

There are three standard estimation procedures available to estimate the panel ECM, namely the dynamic fixed effects (DFE), the pool mean group (PMG) and the mean group (MG) estimation procedure. The choice of the appropriate estimation procedure is based on the assumption regarding how the data is pooled.

The DFE estimator pools cross-sectional observations assuming common long- and short- run dynamics, convergence coefficient and error variances. And only the intercepts are allowed to differ across units. In other words, it is assumed that $\gamma_{ij} = \gamma_j$, $\beta_{ij} = \beta_j, \eta_i = \eta$.

The PMG estimator, which was proposed by Pesaran et al. (1999), allows for cross-section heterogeneity in the short-run parameters ($\delta_i, \gamma_{ij}, \beta_{ij}, \phi_i$), and assumes common long-run coefficients across units ($\eta_i = \eta$).

By imposing no restrictions, the MG estimator allows both the intercepts and the slope coefficients to vary across units, which means that the MG estimates separate equations for each cross-section unit and then averages the estimated coefficients (Massidda and Piras 2015).

Most panel ECM studies apply the PMG estimator (e.g. Moore 2010; Falk 2010; 2015). Massidda and Piras (2015) utilized all of them when exploring the relationship between migration and Italian domestic tourism demand based on panel data. They found that there was a strong positive relationship between domestic tourism nights and internal migration shock, supporting the migration-tourism nexus in the context of domestic tourism.

2.3.5 Climate and Tourism Demand

Tourism demand is determined by a wide range of factors, which can be broadly categorized into two groups: economic factors and non-economic factors. Mainstream econometric models only consider economic factors including income, prices, exchange rates and travel costs. The traditional omission of non-economic factors could be the result of the tendency to follow the existing literature, the difficulty of quantifying the qualitative factors, and/or the lack of available data. Besides, many researchers' and planners' interests lie in estimating income elasticities and/or price elasticities and evaluating the consequences of taxes or exchange rate policies. Therefore, many econometric analyses of tourism demand focus on economic factors by summarizing the effects of non-economic ones into the error terms or dummy variables.

Such simple treatment is questionable, and efforts have been made to incorporate other economic and non-economic factors into econometric models. There are studies aiming at exploring whether introducing new explanatory variables can improve the forecasting ability, and the new factors considered range from search engine data (e.g. Yang et al. 2014b; 2015; Bangwayo-Skeete and Skeete 2015; Li et al. 2017; Park et al. 2017; Önder 2017; Camacho and Pacce 2018; Dergiades et al. 2018) to climate conditions (e.g. Lise and Tol 2002; Agnew and Palutikof 2006; Falk 2014; Lorde et al. 2015). For instance, to explore the relationship between business and tourism demand cycles, Croes et al. (2017) applied two stage lease squares regression methods based on annual data from 1970 to 2015. The inbound tourism demand of two small islands: Aruba and Barbados were examined, and the study indicated that business cycles explained nearly 49% of inbound tourism demand to Aruba and nearly 91% to Barbados. Yang et al. (2015) employed search query volume data from Baidu and Google to forecast monthly tourism arrivals to Hainan, China. They pointed out that two types of search engine data both helped to greatly decrease forecasting errors. Bangwayo-Skeete and Skeete (2015) used Google Trends' search query data-based mixed data sampling (MIDAS) when predicting monthly visitor arrivals to five Caribbean destinations from three main source markets. They revealed that the AR-MIDAS model outperformed the AR and the SARIMA models in terms of 12-month-ahead forecasts. Álvarez-Díaz and Rosselló-Nadal (2010) included meteorological variables to model tourism demand from UK to Balearic Islands. Lee (2011b) explored the role played by permanent income and asset wealth in modelling and forecasting Hong Kong inbound tourism demand on the basis of permanent income life cycle hypothesis and found that permanent income was the most important influencing factor.

The study of the relationships between climate and tourism demand has gained renewed attention under a global background of increasing recognitions of the impacts of climate change (e.g. Maddison 2001; Álvarez-Díaz and Rosselló-Nadal 2010; Eugenio-Martin and Campos-Soria 2010; Amelung and Moreno 2012). Climate conditions are important time-variant tourism resources and should be taken into consideration in empirical studies. Changes in climate conditions can lead to large shifts in tourist flows with significant economic implications. Understanding the impact of climate is valuable in not only better interpreting tourism demand, but also obtaining more accurate forecasts.

2.3.5.1 Measures of Climate Conditions

Climate refers to the conditions of the atmosphere over relatively long periods of time, while weather is how the atmosphere behaves over a short period of time. Climate change is represented by the changes in long-term averages of daily weather. In the tourism demand literature, various meteorological measurements including temperature, precipitation, sunshine duration, wind speed, days of air frost and snow depth are widely considered as proxies of climate conditions (e.g. Saayman and Saayman 2008; Falk 2013; 2014; Ridderstaat et al. 2014). Besides, composite climate indices, which are derived from a range of meteorological measures, have been chosen by many studies to represent climate conditions with the most popular one being Mieczkowski's Tourism Climatic Index (TCI) (1985) (e.g. Goh et al. 2008; Moore 2010; Goh 2012; Lorde et al. 2015). Weather phenomenon is also used as a measure of climate conditions. When assessing the determinants and their impacts on Galician domestic tourism demand, Otero-Giráldez et al. (2012) chose North Atlantic Oscillation (NAO), a weather phenomenon in the North Atlantic Ocean to account for meteorological effects and revealed a significant positive connection between NAO and tourist nights.

Composed of five sub-indices of daytime thermal comfort, daily thermal comfort, precipitation, hours of sunshine, and wind speed, TCI was designed based on bio-meteorological literature on human comfort and has been adapted to tourism to reflect average tourism wellbeing. Being a synthetic evaluation of the climatic elements that most affect the quality of the tourism experience, TCI is therefore a good choice to represent the climate conditions in tourism demand impact analysis, especially when there is no need to distinguish tourist activities. However, the weights of different elements in TCI were decided subjectively based on expert judgements and meteorological literature, which lacks empirical justification. This is the main drawback of TCI.

To provide empirical validation, Morgan et al. (2000) presented beach climate index (BCI) which based the rating and weighting schemes on information regarding tourists' preferences collected from questionnaires to 1354 beach users. Similar to TCI, BCI consists of sub-indices, but the weights are quite different from that proposed by Mieczkowski (1985). Applying BCI, Moreno and Amelung (2009) assessed the impacts of climate change on Europe's beach tourism in summer and found relatively modest shifts in attractiveness.

One important sub-index of the climate indices is the thermal one, which is represented by objective climate measures such as air temperature and relative humidity in TCI, and skin temperature in BCI. However, some researchers prefer Physically Equivalent Temperature (PET) when establishing the influences of climate on tourism demand (Cegnar and Matzarakis 2004; Morabito et al. 2004; Lin and Matzarakis 2011). Because they believe that it is human's thermal perception which determines their behavior, and such perception is different from the objective measurements. PET was developed by Matzarakis et al. (1999) to assess human comfort in general, but this index disregards the other two important climate aspects: the physical and aesthetic aspect.

2.3.5.2 Impact Studies

Tourism demand literature has seen a number of studies on the impact of climate on tourism demand (e.g. Hamilton et al. 2005a; 2005b; Hamilton and Tol 2007; Eugenio-Martin and Campos-Soria 2010; Rosselló-Nadal et al. 2011; Amelung and Moreno 2012; Falk 2010; 2013; 2014; Pintassilgo et al. 2016). A group of these studies incorporate time series data regarding climate conditions into econometric models to assess the effects of climate conditions/change on tourism demand. Both static and dynamic modelling techniques are used.

Most studies utilize dynamic models to capture the interaction between climate factors and tourism demand. For instance, Falk (2014) investigated the relationship between climate conditions in the peak summer season and domestic and foreign over-night stays in Austria for the period from 1960 to 2012. Both the static regression model and the ECM were employed, and average sunshine duration, temperature and precipitation were considered as measures of climate conditions. He found that short-run (annual) weather variations had significant impact on tourism demand, while the long-run impact over the considered 50 years had been quite modest. Goh (2012) applied monthly data from August 1984 to December 2011 in an ECM with destination's TCI as one of the model inputs to study Hong Kong inbound tourism demand from US, UK, China and Japan respectively. He concluded that, as an important socio-psychological variable, climate had significant impact on tourism demand and suggested that tourism demand models should take climate variables into account.

The other group of impact studies considers climate conditions both spatially and temporally by utilizing panel data on climate variables in model estimation. To assess the impact of climate on seasonal tourism demand, Li et al. (2018a) constructed a relative climate index based on TCI, which represented the ratio between the difference of destination and origin TCI and the origin TCI, and an interannual relative climate index, which took the deviation of the relative index from its long-term average into account. Hong Kong outbound tourism demand to 13 mainland China cities were examined in a panel data model, and the empirical results indicated that the relative climate conditions had significant effect on tourism demand. Li et al. (2016) assessed the impact of climate on seasonal tourism demand from Hong Kong to 19 major cities in China through simple dynamic panel data models covering data from 2006Q1 to 2011Q4. Lorde et al. (2015) employed an augmented gravity model which included climate distance (the gap between climate conditions in origin and destination) to study tourism demand from USA, Europe, Canada and the Caribbean to 18 Caribbean countries, and concluded that climate distance was an important demand determinant. Ridderstaat et al. (2014) analyzed seasonal patterns of climate's impact on seasonal variations in Aruba's inbound tourism demand from USA and Venezuela. Climate variables both in the destination and in the origin were considered in the model, and the data range from 1986M1 to 2011M12. They showed that both pull and push seasonal factors of climate played roles in tourism demand fluctuations.

2.3.5.3 Forecasting Studies

The papers assessing whether incorporating climate variables can improve the predictive power of econometric models all reached the same conclusion that meteorological variables can contribute to better tourism demand forecasts (Alvarez-Diaz and Rosselló-Nadal 2010; Moore 2010; Kulendran and Dwyer 2012; Zhang and Kulendran 2016). The forecasts of the seasonal variation in tourist numbers, which is measured as the fluctuation from season to season from the mean value, have attracted some attention. Kulendran and Dwyer (2012) and Zhang and Kulendran (2016) both evaluated the effect of climate variables on seasonal variations in tourism demand using Euclidean Distance statistics, and generated forecasts through the autoregressive conditional heteroscedasticity (ARCH) model and the ADLM respectively. As an example of the forecasting study based on panel data, Moore (2010) showed the superiority of the panel ECM, which incorporates the climate factor, compared to naïve models in forecasting Caribbean tourism demand.

2.4 The Combination Forecasting Approach

2.4.1 Why to Combine?

There are two different 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, the forecasting performance of rival models are compared to identify the best-performing one, and the inferior forecasts are discarded. The disadvantages of the individual forecasting approach include: firstly, the discarded predictions may contain some useful independent information, and secondly, the identification of the ‘best single model’ is like a moving target, as many empirical studies have shown that the forecasting performance of different individual models depends on the accuracy measure used, the forecasting horizon under consideration and the origin-destination pairs under study (e.g. Shen et al. 2009; Athanasopoulos et al. 2011; 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 among a wide range of competitors.

Many studies have reported empirical evidences showing the unstable performances of individual models. Gunter and Önder (2015) compared causal models with time series techniques when forecasting tourist arrivals to Paris, and they concluded that the Bayesian VAR model was the best according to root mean squared error (RMSE) in six-step-ahead forecasts, which was beaten by the VAR model when mean absolute error (MAE) was used as an accuracy measure. But as far as 24-steps ahead forecasts were concerned, the TVP model ranked number one no matter which measure was utilized. Shen et al. (2009) explored the relationship between seasonality treatment and forecasting performance when forecasting the outbound leisure tourism demand from UK to seven destinations. They conducted comparison among models including Naïve 1, BSM, SARIMA, ADLM, WB-ECM, JML-ECM, VAR, TVP and CSM based on mean absolute percentage error (MAPE) and root mean squared percentage error (RMSPE) and concluded that pre-test for seasonal unit root was necessary, which could improve forecasting accuracy. Regarding the forecasting abilities, they found that the rankings did vary dramatically across different destinations and forecast horizons, and MAPE and RMSPE could give controversial results in some cases. Athanasopoulos et al. (2011) conducted a comprehensive comparison of the forecasting performance among various methods including three fully automated time series algorithms (Forecast Pro, ARIMA and EP based algorithms), two method-specific approaches (the Theta method and the damped trend), and five causal models (static and dynamic regression, ADLM, TVP, and VAR). Monthly, quarterly and yearly data were all used and percentage better (PB), MAPE, mean absolute scaled error (MASE) as well as median absolute scaled error (MdASE) were considered as accuracy measures. They showed that there was no single best model in all situations, but generally speaking, pure time series approaches forecasted more accurately than causal models, and the TVP model possessed the most consistent performance among all the causal modes. They also pointed out that the difficulty in forecasting the explanatory variables and possible model misspecifications were two major reasons of the inferiority of causal models.

Rather than trying to choose the best single model, the combination forecasting approach pools all available constituent forecasts together. The rationale is that forecasts from diverse models based on competing theories, functional forms and specifications contain independent information, and combination forecasts can achieve diversification gains by aggregating information. As a result, the performance of the combination forecasts is more stable than the constituent ones.

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, management, science, operations research and psychology, and applications in many fields including meteorology, economics, finance, insurance, sales and price (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 (e.g.

Granger and Ramanathan 1984; Diebold and Pauly 1990; Stock and Watson 2004).

2.4.2 Weighting Schemes

The 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 the included 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 (e.g. Makridakis and Winkler 1983; Stock and Watson 1998; 2003; 2004).

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 the 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 (e.g. Fritz et al. 1984; Diebold and Fauly 1987; Stock and Watson 2004; Wong et al. 2007; Cang 2011; 2014; Baumeister and Kilian 2015).

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. They explored three different regression specifications to work out the optimal weights, and after testing them with data on quarterly forecasts of hog prices, they showed that the best one is to add a constant term and not to constrain the weights to add 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 the prior information into the weighting scheme. The presented shrinkage method uses empirical Bayes procedures to estimate prior precision from the data, which coaxes the weights toward equality without forcing them to be exact equal, and the least squares and simple average weights are the polar cases for the posterior mean. They tested the method by forecasting US GNP and concluded that shrinkage improved the accuracy of the regression-based combination forecasts.

Another extension of the regression method is the time-varying parameter method, which relaxes the assumption that the parameters in the combined regression are fixed. The weights are estimated using Kalman filter algorithm and are varying with time if the data suggests so. Applications of the time-varying combination method include Sessions and Chatterjee (1989), LeSage and Magura (1992), Stock and Watson (2004) and Shen et al. (2011).

2.4.3 Applications in The Tourism Demand Literature

There are a handful of studies on combination forecasts in the tourism demand literature, among which differences can be found in terms of weighting schemes, individual model inputs and accuracy measures with one common finding being that combination forecasts are generally superior to individual ones (e.g. Fritz et al. 1984; Shen et al. 2008, 2011; Song et al. 2009; Andrwais et al. 2011; Cang 2011; 2014). The SA, VACO and DMSFE methods are popular weighting schemes (Wong et al. 2007; Shen et al. 2008; 2011; Song et al. 2009; Chan 2010; Cang 2011; 2014; Coshall and Charlesworth 2011).

Wong et al. (2007) compared the predicting ability of four single models with three combination methods when forecasting tourism arrivals to Hong Kong. The individual models included the SARIMA model, the ADLM, the EG-ECM and the VAR models, and the weighting schemes were the SA, VACO and DMSFE methods. They showed that forecast combination could considerably reduce the risk of forecasting failure when one-step-ahead forecasts were considered and pointed out that the relative forecasting accuracy of single versus combination models depended on the origin under consideration and the weighting scheme used. This research was extended by Song et al. (2009) by taking different forecasting horizons into account. The same single and combination methods were considered, and similar conclusions were drawn: combination forecasts are superior to the average single ones across all horizons. They also explored whether more accurate forecasts could be obtained as the number of the constituents increased in the combination and found that forecasting accuracy did not increase as the number of the constituents in the combination forecasts increased.

Shen et al. (2008) conducted comparisons among popular econometric models, time series techniques and combination methods when forecasting UK outbound tourism demand to seven countries. Seven individual models were chosen: ADLM, WB-ECM, JML-ECM, VAR, TVP, season naïve and SARIMA and three combination methods were included: SA, VACO and DMSFE. They showed that combination played a significant role in improving forecasting performance across all horizons, and the VACO method was the best weighting method. In a latter study (Shen et al. 2011), three more weighting schemes were introduced: the regression method, the shrinkage method and the TVP combination method. 120 possible composite forecasts for seven origins, five forecasting horizons and six weighting schemes were generated and compared with the single ones, which demonstrated that combination was generally beneficial in improving forecasting accuracy, and the VACO method and the DMSFE method with a discounting factor of 0.85 provided the most consistent performance. They also revealed that introducing the best individual model in the composite forecast did not always contribute to better predictions and including all the single models in one combination forecast always performed poorly.

Other combination methods including non-linear weighting schemes, cumulative sum control chart (CUSUM) method and management-oriented approach have also been explored to determine the optimal combining weights in the tourism demand forecasting literature (Chan et al. 2010; Andrawis et al. 2011; Coshall and Charlesworth 2011). Chan et al. (2010) focused on combination forecasts using cumulative sum control chart (CUSUM) techniques and they employed a quadratic programming approach to determine the combination weights for individual forecasts. Their conclusion was that the controlled weighting method both saved time in updating the combination weights and improved the overall performance of the combined forecasts. Although this research explored novel combination methods, the statistical combining approaches included in the comparison were confined to the SA and VACO methods, and more comprehensive comparisons need to be conducted before reaching a decisive conclusion. Andrawis et al. (2011) examined the forecasting ability of both linear and non-linear statistical combination approaches. They found that the top four combination methods were two linear ones of the regression method and the inverse of the mean square error (INV-MSE) method followed by two non-linear ones of the weighted geometric mean (GEMO-WTD) and the weighted harmonic mean (HARM-WTD) methods. Coshall and Charlesworth (2011) explored a management-oriented approach to combine forecasting models and concluded that Goal Programming (GP) methods provided a flexible management-oriented focus for combining forecasts. But the single forecasting methods involved did not include modern econometric models.

In addition, ANN combination methods such as multi-layer perception (MLP), radial basis function (RBF) and support vector regression (SVR) have drawn some attention (Cang 2011; 2014; Claveria et al. 2016). Cang (2011) presented a non-linear combination model based on MLP and compared it with nine individual models and three linear combination models. She found that the MLP model was robust, powerful and performed better. In a later study (Cang 2014), she proposed another two non-linear combination methods using RBF and SVR, and similar conclusion was drawn which advocated non-linear ANN combination approaches over other methods. This research can be extended in two directions: firstly, more advanced statistical combining schemes can be included in the comparison; secondly, econometric modelling techniques can be introduced as individual forecasting methods.

For example, Wan and Song (2018) utilized logistic models to forecast the growth of Hong Kong inbound tourism, which was evaluated by the hit rates and quadratic probability score. They also examined whether combining probability forecasts can improve the forecasting accuracy and concluded that the performance of the constituent forecasts affected the predictive ability of the combination ones to a large extent.

When it comes to individual model inputs, both time-series and econometric models are considered. Regarding time series techniques, the (S)ARIMA and ETS models are widely chosen, and for causal models, popular candidates are the ADLM, the ECM and TVP models (e.g. Shen et al. 2008; 2011; Chan et al. 2010; Cang et al. 2011; 2014). These modelling techniques have different assumptions regarding the form of the relationships between the variables and the dynamics in the system, hence possess different information, which can be pooled together by combining.

In the current tourism demand literature, a type of information has been neglected, which origins from different explanatory variables included in the 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. When modelling and forecasting tourism demand, causal econometric models are different in identified explanatory variables, which can be economic or non-economic. The various model specifications in terms of identified demand determinants are based on different theories and contain useful independent information. Current combination forecasting studies only consider econometric models based on economic influencing factors. Chan et al. (2010), Shen et al. (2011; 2008); Song et al. (2009) and Wong et al. (2007) all include modern econometric models as constituents in the combination forecasts, and the econometric models are the same in identified 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 are considered as demand determinants. Studies need to be conducted to explore whether forecasting accuracy can be improved by combining causal econometric models with different influencing factors.

When evaluating forecasting performance, the most widely employed accuracy measure is MAPE, (e.g. Shen et al. 2008; 2011; Coshall 2009; Song et al. 2009; Chan et al. 2010; Coshall and Charlesworth 2011; Cang 2011; 2014). Other popular measures include MAE, RMSE and RMSPE. Besides the accuracy measures, which are purely descriptive, formal statistical tests are also applied to assess whether one forecasting method is significantly better than the other. Coshall (2009) utilized Diebold-Mariano (D-M) test (Diebold and Mariano 2002) and Harvey, Leybourne

Newbold (HLN) test to check whether the difference in the accuracy of competing models was statistically significant. Cang (2014) applied Mann-Whiteney test and proved that the forecasting ability of combination methods is statistically superior to the individual ones.

2.5 Data

In terms of data sources, types and frequencies, post-2008 tourism demand studies are more diversified than previous researches. Main data sources include tourism administration departments in different countries, which provide information on various measures of tourism demand, as well as a wide range of national or international organizations which release data on a variety of economic and non-economic variables.

2.5.1 Data Type

Secondary data in the form of time series still dominate in tourism demand studies. There are merely a handful of papers making use of primary data in quantitative studies. Lee et al. (2008), Song et al. (2008; 2013) and Lin et al. (2014) applied an integration approach, which made adjustments to forecasts yielded through quantitative techniques based on predictions obtained from qualitative analysis. They integrated qualitative forecasts, data on which were collected through Delphi methods, with quantitative ones, which were generated based on secondary data. For instance, Song et al. (2013) presented a web-based tourism demand forecasting system where statistical and judgmental forecasts were integrated. The variables of interest included tourist arrivals, total and sectional tourist expenditures, as well as hotel room demands, and the destination considered was HK. After generating forecasts using the ADLM, judgmental adjustments were made by 21 postgraduate researchers (PGRs) and five staffs from The Hong Kong Polytechnic University. The result of the case study showed that integration of statistical and judgmental forecasts improved forecasting accuracy.

Panel data analysis has gained great popularity since 2008, which is a new trend in the tourism demand literature. Panel data models have also been used to forecast tourism demand (Moore 2010; Fildes et al. 2011; Long et al. 2018). Panel data contain observations on a number of cross-sectional units over time, and compared to pure time series data, which are observed on one unit for a time period, they can address many problems in model estimation procedure such as collinearity and the lack of degree of freedom. In the context of tourism analysis, different destinations or origins can be considered as the cross-section units, and time-insensitive variables such as the length of coastline, age structure and education level can therefore be included in the models to explore their impacts on tourism demand.

2.5.2 Data Frequency

Regarding data frequency, higher frequency data studies, especially monthly data studies, have become more popular among post-2008 publications. Traditionally, causal econometric models mainly rely on yearly data and non-causal time series models generally utilize higher-frequency ones (quarterly and monthly). Since 2008, it has become more flexible in terms of model selection when forecasting higher frequency time series data. When conducting impact analysis or policy evaluation, more econometric models choose monthly data (Otero-Giráldez et al. 2012; Chatziantoniou et al. 2013; Untong et al. 2015). When forecasting tourism demand, a large number of econometric models utilize quarterly (Athanasopoulos et al. 2017; Zhu et al. 2018) or monthly data (Goh et al. 2008; Bangwayo-Skeete and Skeete 2015; Folgieri et al. 2017), and almost all time series techniques utilize quarterly or monthly data (Gil-Alana et al. 2014; Ma et al. 2016; Apergis et al. 2017; Hassani et al. 2017; Saayman and Botha 2017; Vergori 2017). Guizzardi and Stacchini (2015) employs 4-monthly data to forecast inbound tourism demand to Rimini, Italy. The reason why it chooses 4-monthly data is that this paper included business sentiment indicators (BIS) from business surveys as an influencing factor and these surveys were conducted on a 4-monthly basis.

With increasing interests in exploring the value of internet data in forecasting tourism demand, weekly data are utilized in some studies. Li et al. (2017) constructed weekly composite search index data based on generalized dynamic factor model (GDFM), which was used to forecast domestic tourism demand of Beijing, China. The data on tourism demand variables were transformed from monthly to weekly and the forecasting model chosen was the ADLM. Pan and Yang (2017) predicted weekly hotel demand in Charleston, South Carolina, US using ARIMAX models and the Markov switching dynamic regression (MSDR) model. They incorporated several tourism big data sources including search engine queries, website traffic, and weekly weather information in the forecasting models and showed that ARMAX models with both search engine queries and website traffic data generated the most accurate forecasts. Yang et al. (2014b) utilized weekly web traffic volume data of a destination marketing organization to predict hotel demand for Charleston, and the forecasting models chosen are the ARMAX and ARMA models. They found significantly improved forecasting accuracy of the ARMAX model over the ARMA model and concluded that website traffic data had great value in forecasting. There is one paper using daily data (Divino and McAleer 2010), which is quite unique in the tourism demand literature, to model the growth rate and volatility in daily international tourist arrivals to Peru.

2.6 Summary

Tourism demand has been, as always, one of the most popular topics in the current literature. Since 2008, an accumulating body of tourism demand studies have emerged in literature with a wide range of analytical approaches being applied. Advanced modelling techniques have been widely accepted, and a large number of destinations/origins are covered in empirical studies with the most popular ones being Australia, Hong Kong and the UK. Although the trend of introducing new techniques into tourism demand analysis shows no sign of downward, the most popular methodologies are still non-causal time series and causal econometric models.

Various time series techniques such as naïve, ARIMA, ETS and state space models are widely-accepted forecasting devices and are usually selected as candidates or benchmarks when forecasting tourism demand. The spectrum analysis, which has been introduced to the tourism demand literature in the 2010s, has been proven to possess strong predictive ability (Chan and Lim 2011; Kožić 2014; Hassani et al. 2015; 2017; Saayman and Botha 2017; Silva et al. 2017). This type of time series method deserves more attention in the future.

Econometric models have been extensively applied in both impact and forecasting studies. Different econometric methods are utilized to estimate tourism demand elasticities (Seetaram 2012a; Fuleky et al. 2014; Gatt and Falzon 2014; Untong et al. 2014), evaluate the impact of a particular factor on tourism demand (Kuo et al. 2008; 2009; Falk 2010; Tveteras 2014; Chen and Haynes 2015; Balli et al. 2016; Li et al. 2016; Ongan and Gozgor 2018) and conduct policy analysis (Mangion et al. 2012; Li and Song 2013; Chou et al. 2014). Such studies rely heavily on single-equation approaches, while the utilization of one type of system-of-equations approaches, i.e., the AIDS are limited. Examples of the application of the AIDS include Cortes-Jimenez et al. (2009), Divisekera (2009; 2010), Wu et al. (2011; 2012), Mangion et al. (2012), Athanasopoulos et al. (2014) and Gatt and Falzon (2014). The AIDS can model the allocation of tourism demand among a range of competing destinations and the allocation of expenditure on various tourism products or services in one trip. These aspects of tourism demand should be explored more in the future.

Panel data studies have gained popularity since 2008 with a number of applications in impact analysis. However, the forecasting performance of panel data models are seriously under-studied with only three studies generating forecasts based on panel data (Moore 2010; Fildes et al. 2011; Long et al. 2018). Long et al. (2018) introduced the dynamic spatial panel data model to the tourism demand forecasting literature. The dynamic spatial panel data model deserves more attention in the future, as it can incorporate both the spatial and the temporal effects when forecasting tourism demand.

In the context of tourism demand forecasting, mainstream econometric models only consider economic factors including tourists' income, own price and substitute price as demand determinants ignoring the effect of other influencing factors. The value of other economic and non-economic factors such as permanent income, business cycle, climate variables and internet data in improving forecasting performance has been explored in some studies (Falk 2014; Lorde et al. 2015; Yang et al. 2015; Croes et al. 2017). It is worth noting that internet data has received great interest when seeking additional explanation of tourism demand evolution (Yang et al. 2014b; 2015; Bangwayo-Skeete and Skeete 2015; Li et al. 2017; Park et al. 2017; Önder 2017; Camacho and Pacce 2018; Dergiades et al. 2018).

There are studies paying their attention to further advance methodologies by introducing or presenting new models, addressing potential problems of existing methods and discovering the best-performing forecasting approach through making comparisons (Morley 2009; Chan et al. 2010; Athanasopoulos et al. 2011; Hadavandi et al. 2011; Song et al. 2011a; Wan et al. 2013; Akin 2015; Yang et al. 2015; Sun et al. 2016). For example, ever since the turn of this century, several quantitative forecasting approaches, which are popular in other areas, have been introduced into the tourism demand literature, among which AI techniques have received great attention. Various artificial neural networks (ANN) methods are applied to forecast tourism demand either as single forecasting models or as combination methods (Cang 2011; 2014; Claveria and Torra 2014; Akin, 2015; Claveria et al. 2015; 2016). Besides, the rough set approach (Goh, 2008) and the fuzzy time series method (Hadavandi et al. 2011) have both been employed in tourism demand analysis.

With further advancement in methodologies, which can be found in different areas such as econometric techniques (Song et al. 2011a; Wu et al. 2012; Smeral 2014; Assaf et al. 2018), time series approaches (Coshall 2009; Nowman and Dellen 2012; Hassani et al. 2015; 2017; Apergis et al. 2017; Saayman and Botha 2017; Silva et al. 2017) and AI methods (Hadavandi et al. 2011; Cang 2011; 2014; Li et al. 2018b), there are still some research gaps identified.

Firstly, the forecasting of the growth cycle of tourism demand, which is of vital importance both to businesses and governments, is seriously under-researched. Most studies pay their attention to modelling and forecasting the amount of tourism demand for a specified time span, ignoring predicting the turning points and directional changes. There are only a few studies concerning tourism demand growth cycles, and even fewer generate forecasts (e.g. Andraz et al. 2009; Kulendran and Wong 2011; Kožić 2014; Wan and Song 2018).

Besides, most tourism demand forecasting studies only generate point forecasts, neglecting interval forecasts with only a few exceptions (e.g. Song and Lin 2010; Kim et al. 2010; 2011; Song et al. 2011b; Athanasopoulos et al. 2011; 2017). Point forecasts are estimates of the unknown true future value, which cannot give information to the variability associated with the forecasts. While interval forecasts provide a range of possible future outcomes with a prescribed level of confidence, which are of greater value to decision-makers, as they provide estimates of uncertainty and allow for contingency planning. The utilization of the bias-corrected bootstrap for interval forecasting of AR models has been presented and tested (Kim et al. 2010; 2011). In the context of econometric modelling, interval forecasts are produced based on interval estimates of demand elasticities, where the delta method and Kilian's (Kilian 1998) bias-corrected bootstrap have been utilized (Song and Lin 2010; Song et al. 2010a; Otero-Giráldez et al. 2012).

In addition, studies on forecast combination is in its infancy stage in the current tourism demand literature. Although empirical studies that evaluate the combination forecasting approach all provide positive evidence supporting combining individual forecasts, the mainstream forecasting methods still belong to the individual forecasting approach. This is due heavily to the difficulty of computing combination forecasts without an easy-to-use software. The combination forecasting approach deserves more attention, and researches on combination forecasts can be extended in different ways. For example, econometric models with different explanatory variables, which include economic and non-economic influencing factors, should be integrated in the combination panel to serve as constituent models, recently-introduced or novel single forecasting techniques such as the BGVAR model and SSA should be considered as single inputs, combination of interval forecasts should be studied, and the forecasting ability of other combination methods such as information-criteria-based weighting schemes should be evaluated and compared with AI-based weighting schemes.

When it comes to evaluating the forecasting performance, most studies make comparisons among diverse forecasting methods based on different accuracy measures, with the most popular ones in the tourism demand literature being MAPE, RMSE and RMSPE. However, the comparison results do change according to which accuracy measure is used. Considering the conflicting results based on different accuracy measures, there are doubts on whether the forecasting rankings generated are reliable. Besides, the accuracy measures are purely descriptive. To assess whether one method is significantly better than the other, formal statistical tests need to be utilized. There are some papers applying D-M test to check whether the difference in the accuracy of competing models is statistically significant (Álvarez-Díaz and Rosselló-Nadal 2010; Gil-Alana 2010; Bangwayo-Skeete and Skeete 2015). HLN test and Clements & Harvey test are also utilized (Bonham et al. 2009; Guizzardi and Stacchini 2015). Besides, when forecasting tourism demand growth cycle, directional change error or turning point error should be evaluated. For example, Hassani et al. (2015; 2017) utilized the direction of change (DC) criterion proposed by Hassani et al. (2013), and Chen (2011) chose the directional change accuracy test (DCA) presented by Pesaran and Timmermann (1992). Future studies should be more critical when it comes to choosing the proper forecast accuracy measures, and statistical tests concerning relative forecasting abilities are suggested.

At last, studies on model selection is rare in the current literature. With more and more research methods available, it becomes even harder to choose the appropriate technique as there exists no rule regarding model selection. Empirical findings are controversial in terms of forecasting accuracy rankings, on which forecasting horizons, accuracy measures as well as origin-destination pairs are all found to have influences. Why a particular model is chosen is mainly decided based on the purpose of the study and the researches' expertise. It is true that there is no single model which can behave the best in all situations, but there should be some guidelines on which types of techniques should be better under particular conditions. There is only one study on model selection (Akin, 2015) and more studies in this area are welcomed in the future.

Chapter 3 Research Method

3.1 Introduction

This chapter is about the research method of this study.

The economic analysis of tourism appeared in the literature at the end of the 1950s, and it has undergone significant development ever since the 1970s (Crouch 1992). The methods used can be broadly classified into two categories: quantitative and qualitative, and quantitative tourism demand studies dominate in the current literature in terms of the amount of applications (Song and Li 2008). Quantitative methods are widely used to evaluate the impacts of specific policies/events, assess the effects of particular factors, and forecast the future flows of tourism demand. According to Makridakis et al. (2008), quantitative forecasting methods are suitable given the following three conditions: first, sufficient information about the past is available; second, this information can be quantified in the form of numerical data; and third, it can be assumed that some aspects of the past pattern will repeat in the future. On the other hand, if little or no quantitative information is available, but sufficient qualitative knowledge exists, qualitative methods are appropriate. Qualitative forecasting methods require no mathematical rules but rely on intuition thinking and experts' judgement and experience.

Based on the research aim, data availability and the researcher's expertise, this study applies quantitative methods which are based on secondary data. Basically, this study is an econometric methodology research with the application on tourism demand forecasting.

The rest of this chapter is organized into seven sections. An illustration of the research plan is provided in section 3.2, which explains in detail how the aim and objectives of this research are achieved. Section 3.3 delineates the variables and data used in this study. The dependent variable, i.e. the proxy of tourism demand is introduced firstly, followed by the independent variables including economic determinants and dummy variables which are considered by the traditional econometric models, as well as the climate factor which is included by the climate econometric models. The highlight is on the explanation of the climate determinant, which is the UK's TCI. The data sources are addressed next, followed by an illustration of the data sample.

Section 3.4 introduces the bounds test cointegration approach, which is used to evaluate the impact of climate on tourism demand. Unit root tests are essential before cointegration analysis is run, as different cointegration techniques have specific requirements on the integration orders of model variables. The choice of the bounds test cointegration approach is based on the integration orders of the model variables. An introduction of diagnostic tests consisting of residual diagnostics and stability diagnostics, which are necessary for econometric techniques, is presented in section 3.5.

Section 3.6 and section 3.7 are concerned with the forecasting methodologies. The forecasting methods including the individual and the combination ones are discussed in section 3.6. The individual forecasting models applied in this study include causal econometric models and non-causal time series techniques with the seasonal naive no-change model being used as the forecasting benchmark. Except the naïve model, all other individual models are considered as potential constituents in combinations, which are summarized, and a justification of model selection is provided. Afterwards, different combination methods are investigated. The combination methods applied to generate combination forecasts consist of SA, VACO, DMSFE as well as the newly-introduced two methods including the inverse-MAE and the two-stage combination methods. Section 3.7 covers the forecasting procedures and the accuracy measures. The recursive individual forecasting procedure and the recursive weighting procedure are explained firstly, which is followed by a description of the three forecasting accuracy measures.

Lastly, section 3.8 puts its effort on illustrating the programs for computing combination forecasts and conducting forecasting comparison and analysis.

3.2 Research Plan

To achieve the research, aim and objectives as presented in section 1.2, the plan illustrated by figure 3-1 is followed. Inbound tourism demand to the UK from seven leading markets: France, Germany, Irish Republic, Italy, the Netherlands, Spain and the US are studied respectively based on quarterly time series data. The bounds test cointegration approach is utilized to assess the long-run relationships between tourism demand and its influencing factors, and to evaluate the impact of climate on tourism demand. Individual forecasting models considered include causal econometric models, which consist of the bounds test cointegration, the ADLM, LI, VAR TVP and SD models, as well as non-causal time series techniques, which are comprised of the SARIMA, ETS and state space ETS models. Besides, each econometric model takes two model specifications, which are different in identified influencing factors. Traditional econometric models only consider economic factors and dummy variables, and climate econometric models include the climate determinant. 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.

To generate combination forecasts, the individual models to be combined should be determined first.

Except the seasonal naive no-change model, all individual forecasting models are considered as potential constituents in combinations, which are categorized into three groups: the first group which includes all individual models; the second one which contains traditional econometric and time series models; and the third category which consists of climate econometric and time series models. Combination is conducted for each group respectively, resulting in three sets of combination forecasts: the first set is generated through combining 15 individual models; the second and third ones are produced from integrating 9 individual models.

After identifying constituent forecasts, the methods for computing the weights should be chosen. Each group of individual forecasts are combined respectively with different weighting schemes. The SA, VACO, DMSFE ($\alpha = 0.85/0.90/0.95$) methods, as well as the newly-introduced inverse-MAE and the two-stage combination approaches are applied for the first group. The SA, VACO, DMSFE ($\alpha = 0.85/0.90/0.95$) and the inverse-MAE weighting schemes are utilized for the second and third groups, and as the two-stage combination approaches perform unsatisfactorily, they are excluded.

The comparisons of forecasting performance are conducted in several levels and ways. Firstly, single forecasts from different individual models are compared to evaluate which individual models have stronger forecasting abilities, and whether introducing the climate determinant can improve the forecasting accuracy of causal models. Secondly, to assess whether combining is beneficial, individual and combination predictions are compared. And to reveal which weighting schemes have satisfactory and consistent performance, comparisons are made among different combining methods. Thirdly, to evaluate whether including models of different explanatory variables in combination is superior to only considering models of the same variables to combine, composite forecasts from the three groups are compared. Each origin is studied respectively based on three accuracy measures including MAE, MAPE and RMSE. Furthermore, analysis is carried out to address such questions as whether combining better single forecasts results in more accurate composite predictions; and what is the optimal number of constituents in the best forecast.

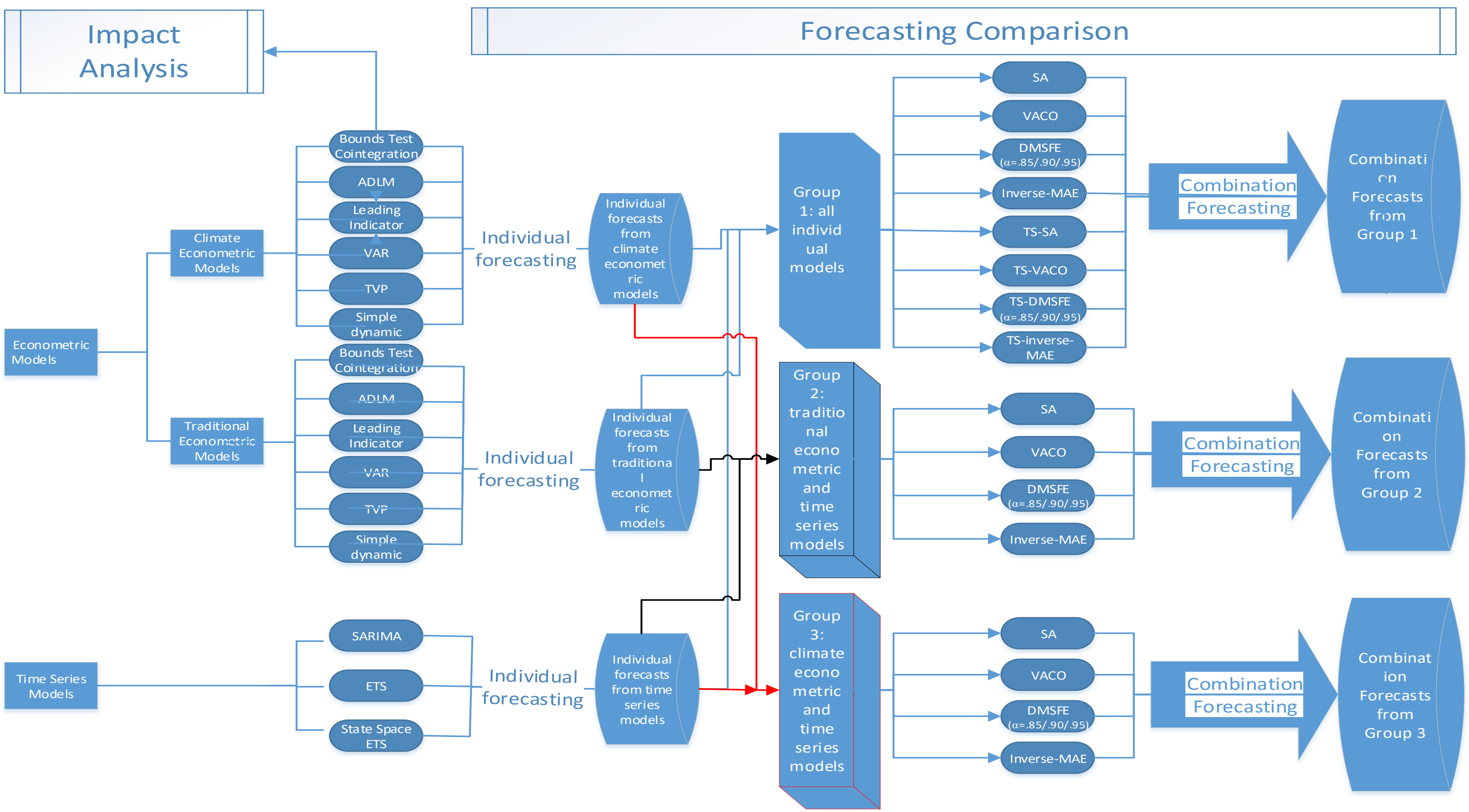

Figure 3-1 Research Plan

Source: the author

Notes: The seasonal naïve no-change model serves as the benchmark for individual forecasting comparisons and is also applied to generate individual forecasts. But it is not considered as a candidate constituent for forecast combination and hence is not included in this figure.

3.3 Variables and Data

3.3.1 The Dependent Variable

The term 'tourism demand' can be defined as the quantity of the tourism product, which is a combination of tourism goods and services, that consumers are willing and able to purchase during a specified period for a particular destination. (Song, Witt and Li 2009). The choice of the right measurement for tourism demand is critically important for empirical studies. Song et al. (2010) summarized the four measurement criteria for all types of travel and tourism demand: (1) a doer criterion: for instance the number of tourist arrivals, the number of tourist visits and the visit rates; (2) a pecuniary criterion: such as the level of tourism expenditure/receipt, and the share of tourism expenditure/receipt in income; (3) a time-consumed criterion: for example tourist-days, tourist-nights; and (4) a distance-traveled criterion: such as the distance traveled in miles or kilometers. In the current tourism demand literature, the first three criteria, i.e. tourist arrivals, tourist expenditure/receipt and length of stay are commonly accepted with tourist arrivals being the most popular one. (Lim 1999; Witt and Song 2000; Li et al. 2005; Song and Li 2008).

Data on tourist arrivals can be collected on the aggregated level, i.e. total number of tourist arrivals, or on the disaggregated level in terms of purpose of travel. Based on travel purposes, tourism can be classified into different categories including travel for holiday, travel for business, travel for visiting friends and family, travel for study, travel for transit and travel for miscellaneous purposes. Different types of tourism are determined by varying influencing factors and associated with different decision-making process. As holiday makers are the most sensitive ones to climate conditions among all the travelers, tourist arrivals for holiday purpose (lny_t) is used as the measure of tourism demand in this study.

3.3.2 Explanatory Variables

3.3.2.1 Economic Determinants and Dummy Variables

In the current literature, economic variables including tourists' income, relative prices between destination and origin, substitute prices in the competing markets and exchange rates are commonly identified as the influencing factors (Lim 1999; Witt and Song 2000; Li et al. 2005; Song and Li 2008). The traditional econometric models follow the mainstream literature and take the following influencing factors into account: income, own price/relative price, substitute price and dummy variables.

Income

Income is a key determinant for tourism demand. According to Lim (1997), income ranked number one among all explanatory variables identified in the current tourism demand literature, with 84 out of 100 studies using it. Tourism demand elasticity with respect to income, or income elasticity, which is defined as the percentage change in quantity of tourism demand induced by one percentage change in income of tourists, is always the question of concern. Tourism demand, especially international tourism demand, is generally regarded as income elastic, i.e., one percentage change in income can induce more than one percentage change in tourism demand, which means income elasticity of tourism demand is greater than one. Popular measures/proxies of income include private consumption, personal disposable income, gross domestic product (GDP) and gross national product (GNP) in real or nominal terms. This study chooses real GDP in the origin country ($lnGDP_t$) as the income measure.

Own Price

Tourism prices, which include travel cost from the origin to the destination and living cost at the destination are important influencing factors of demand. However, because of the potential multicollinearity problems and lack of available data, researchers often choose to omit the travel cost variable (Witt and Witt 1995). This study follows this practice and considers costs of products and services at the destination such as accommodation, local transportation, dining and entertainment as tourism prices.

The ideal proxy of tourism prices would be tourists price indices, which represent the weighted average of the prices of a basket of goods and services consumed by tourists. But there are no continuous and consistent data for such indices. This study follows the common practice in the current literature to choose the relative price, which is the destination's exchange rate adjusted consumer price index (CPI) relative to that of the origin country, as the proxy of tourism prices. Relative price combines the impacts of exchange rate fluctuations as well as price movements in both the origin and the destination, and is also called own price, as it denotes the price level of the destination. Own price has been widely chosen in empirical studies (e.g. Shen et al. 2001; 2009; Seetaram 2010). It is constructed as:

$$lnrp_t = ln\left(\frac{\frac{CPI_{UK,t}}{EX_{UK,t}}}{\frac{CPI_{OG,t}}{EX_{OG,t}}}\right) \tag{3.1}$$

where $\mathrm{CPI}_{\mathrm{UK,t}}$ is CPI of the UK, $\mathrm{CPI}_{\mathrm{OG,t}}$ is CPI of the origin country, and $\mathrm{EX}_{\mathrm{UK,t}}$ and $\mathrm{EX}_{\mathrm{OG,t}}$ represent the exchange rate of the UK and the origin country respectively, which are quarterly national currency to US dollar exchange rates and are not seasonally adjusted.

Tourism demand elasticity with respect to own price, or own price elasticity, is defined as the percentage change in quantity of tourism demand caused by one percentage change in own price. Tourism demand normally has negative own price elasticity, and is regarded as price-elastic if the absolute value of own price elasticity is greater than one, which means that one percentage change in own price can lead to more than one percentage change in tourism demand; or as price-inelastic if the absolute value of own price elasticity is smaller than one, which shows that one percentage change in own price can induce less than one percentage change in tourism demand. The magnitude of own price elasticity is an important influencing factor of price policies.

Substitute Price

Besides the relative price between destination and origin, price levels in alternative destinations are also influential demand determinants. It is assumed that after deciding whether to travel, and whether to travel domestic or abroad, tourists will choose the destination among a range of alternatives, and the relative price levels of these competing destinations influences the destination choice. In 2017, France, the UK and Germany are the top three most visited destinations in northwestern Europe (UNWTO Tourism Highlights 2018 Edition). As a result, Germany and France are chosen as competing destinations to the UK in this study.

The market share adjusted relative prices of the competing destinations are used as the proxy of substitute price, which is constructed as:

$$lnsp_t = ln\left(\frac{\sum_k^n rp_{k,t} * K_{k,t}}{\sum_k^n K_{k,t}}\right) \qquad 3.2$$

where $rp_{k,t}$ represents the relative price of alternative destination k, $K_{k,t}$ represents the market share of destination k, which is measured by tourist arrivals.

Tourism demand elasticity with respect to substitute price, or substitute price elasticity, which is defined as the percentage change in quantity of tourism demand caused by one percentage change in substitute price, can be positive or negative. When it is positive, it means that tourism demand in the main destination is positively related to the price changes in alternative destinations, implying that the alternative destinations substitute the main one. On the other hand, when substitute price elasticity is negative, tourism demand in the main destination is negatively related to the price changes in alternative destinations, so the alternatives are complements to the main destination.

Dummy Variables

To capture the impacts of seasonality and special events, traditional econometric models include seasonal dummies (Q_1, Q_2, Q_3) and one-off events dummies. Generally, the special events considered include major economic or political changes, mega events and natural or man-made disasters. During the sample period which ranges from the first quarter of 1994 to the last quarter of 2017, the following events are considered: the outbreak of foot-and-mouth disease at the beginning of 2001, the terrorist attack on 11th of September, 2001, the terrorist bombing in London on 7th of June, 2005, the global financial crisis in 2008 and the Olympic Games held in London in 2012, which are represented by $D_{DS,}D_{911,}D_{BM,}D_{2008}$ and D_{OL} respectively. To capture the instantaneous and delayed effects of these events, the impact in two quarters are considered. The dummies take values of 1 in the quarters the events happen and the next following quarter and take values of 0 otherwise.

The general form of the traditional econometric models can be specified as:

$$\ln y_t = f(\ln GDP_t, \ln rp_t, \ln sp_t, Q_1, Q_2, Q_3, D_{DS,}D_{911,}D_{BM,}D_{2008,}D_{OL}) \qquad 3.3$$

3.3.2.2 The Tourism Climatic Index

To evaluate the impact of climate on tourism demand and assess whether introducing climate determinants can improve the forecasting performance of causal models, the climate econometric model is presented, which introduces the climate condition in the destination as a demand determinant. The climate condition is measured by the TCI of the UK.

There are three distinct aspects of climate that have roles to play in tourism: thermal, physical and aesthetic. The thermal component, which mainly refers to temperature and humidity, determines tourists' physiological comfort. The physical aspect includes features such as rain and wind that may cause physical annoyance to tourists. And the aesthetic component represents features like sunshine and cloud coverage which may benefit or spoil the experiences of sight-seeing. When establishing the impact of climate on tourism demand, the three aspects should all be taken into consideration. TCI is a human-oriented, synthetic evaluation of climate attractiveness to tourists, which takes the most relevant climate elements to tourism experience into account. TCI is comprised of five sub-indices: the daytime comfort index ($cid_{UK,t}$), the daily comfort index ($cia_{UK,t}$), the index for precipitation ($P_{THE\ UK,t}$), the index for sunshine duration $(S_{UK,t})$ and the index for wind speed ($W_{UK,t}$) (see figure 3-2).

The composition of $TCI_{UK,t}$ follows Mieczkowski (1985):

$$\mathrm{TCI_{UK,t}} = 2 \times \left[4\left(\mathrm{cid_{UK,t}}\right) + \mathrm{cia_{UK,t}} + 2\left(\mathrm{P_{UK,t}}\right) + 2\left(\mathrm{S_{UK,t}}\right) + \mathrm{W_{UK,t}}\right] \qquad 3.4$$

where $\mathrm{cid_{UK,t}}$ is the daytime comfort index of the UK, composed of the maximum daily temperature in degree Celsius and minimum daily relative humidity stated as a percentage; $\mathrm{cia_{UK,t}}$ represents the daily comfort index of the UK, composed of daily temperature in degree Celsius and daily relative humidity stated as a percentage; $\mathrm{P_{UK,t}}$ means the rating for precipitation in the UK, measured in mm; $\mathrm{S_{UK,t}}$ is the rating for duration of sunshine in the UK, measured in hours per day; and $\mathrm{W_{UK,t}}$ represents the rating for wind speed in the UK, measured in kilometers per hour.

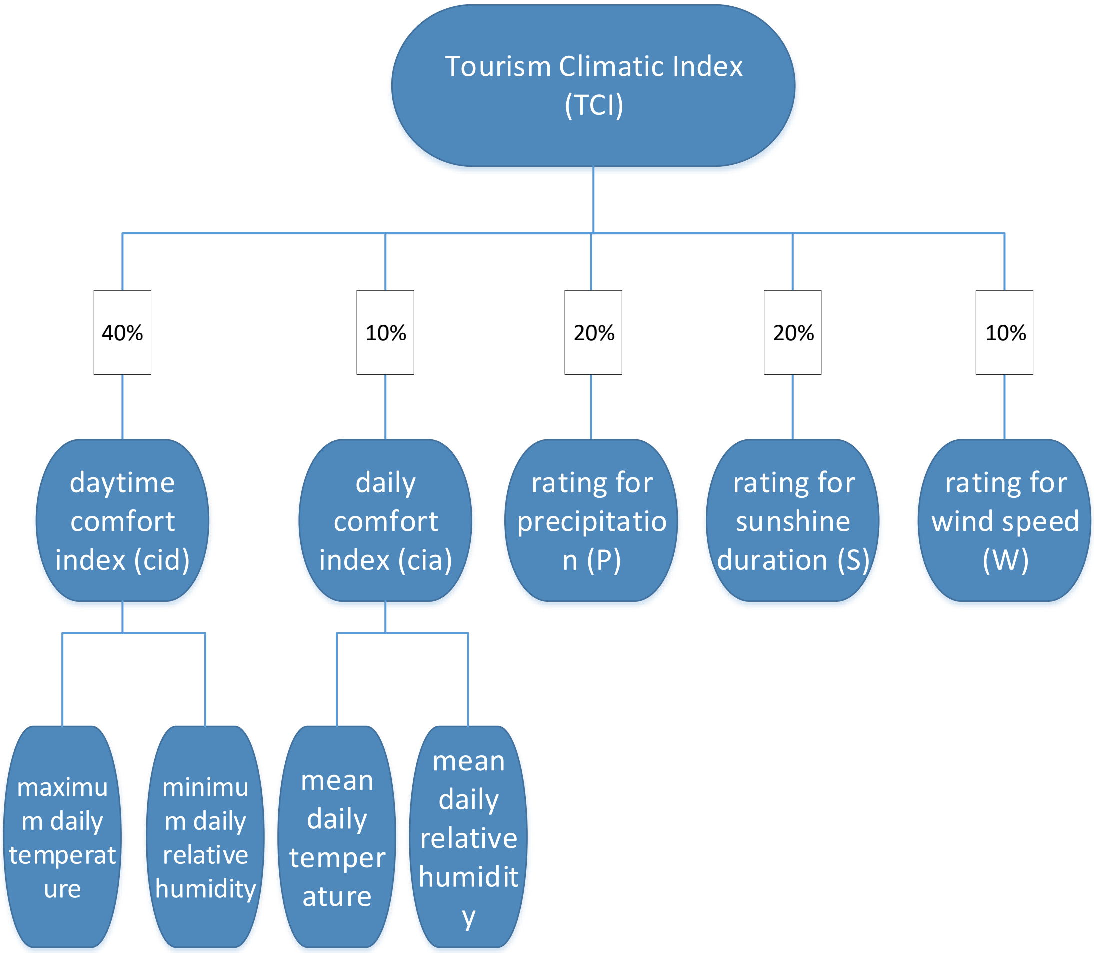

Figure 3-2 The Composition of Tourism Climatic Index

Source: adapted from Mieczkowski (1985)

The Daytime Comfort Index

The first two sub-indices represent the assessment of tourists' thermal comfort based on effective temperature, which is drawn from physiological literature. Simply speaking, effective temperature is a single index of combining dry-bulb temperature (DBT) and relative humidity, which can be shown on a psychrometric chart (Mieczkowski 1985). The daytime comfort index considers the variables of the maximum daily DBT and minimum daily relative humidity, which usually occur between 12:00 pm and 16:00 pm when tourists are most active outdoors. Therefore, the daytime comfort index is considered the most important indicator of tourists' wellbeing, and it is assigned the heaviest weights of 40 percent in the TCI formula. To obtain the index, the values of the pairs of maximum daily DBT and minimum daily relative humidity are plotted on the psychrometric chart given by Mieczkowski (1985). For example, the combination of a maximum daily DBT of 26 degree Celsius and a minimum daily relative humidity of 35% gives the optimum index of 5.0, and the same DBT with a relative humidity of 80% results in an index score of 4.0.

The Daily Comfort Index

The daily comfort index reflects tourists' thermal comfort of the full 24 hours based on two different variables: mean daily DBT and mean daily relative humidity. It is considered less important compared to the daytime index for it includes the temperature of the night, when most tourists stay indoors. It carries 10 percent in the TCI formula. The index is obtained in the same way as the daytime index by plotting the values of the mean daily DBT and the mean daily relative humidity on the psychrometric chart. For instance, the combination of a mean daily DBT of 20 degree Celsius and a mean daily relative humidity of 50% gives the optimum index of 5.0.

The Index for Precipitation

Precipitation plays an important role in tourists' wellbeing, both through the total amounts and through their distribution in time. Because of data unavailability, only the amount of precipitation is considered, and the distribution of precipitation is excluded from the construction of TCI. Mieczkowski (1985) provides a table of the rating scales which is based on mean monthly precipitation amount. Since this study is based on quarterly data, adjustments to Mieczkowski's rating scale is necessary. Table 3-1 summarizes the rating scores based on mean quarterly precipitation following Mieczkowski (Mieczkowski 1985). For example, the mean quarterly precipitation ranges from 90.0 to 134.9 millimeters gives the index score of 4.0, and that ranges from 270.0 to 314.9 millimeters only gives a score of 2.0. The index also takes the progressive disadvantage of increasing precipitation into consideration by assigning a value of -1.0 for each additional 180 millimeters of precipitation when the quarterly mean is above 450 millimeters. The sub-index is assigned a weight of 20% considering the importance of precipitation in affecting tourists' experience.

Table 3-1 Rating Scores for the Precipitation Index in TCI

Rates	Mean quarterly precipitation (millimeter)
5.0	0.0-44.9
4.5	45-89.9
4.0	90-134.9
3.5	135-179.9
3.0	180-224.9
2.5	225-269.9
2.0	270-314.9
1.5	315-359.9
1.0	360-404.9
0.5	405-449.9
0.0	445 or more

Source: adapted from Mieczkowski (1985)

The Index for Sunshine Duration

For most tourists, sunshine is identified as a positive factor because they can improve the results of photography, provide health benefits of moderate ultraviolet radiation and support sunbathing. Generally, more sunshine has better rating, except for regions, which are mainly located in desert climate where high amounts of sunshine are usually associated with high DBT. Mieczkowski (1985) presents a rating scheme which is based on mean monthly hours of sunshine per day. For example, the variable ranging from 9 hours to 9 hours 59 minutes gives a score of 4.5, and that from 4 hours to 4 hours 59 minutes only gives a score of 2.0. To account for negative effects of sunshine in hot weathers, 10 points is deducted from the total index score when the temperature is above 33 degree Celsius, and 20 points is deducted from the total index when the temperature is above 36 degree Celsius. The sunshine factor carries a weight of 20%, reflecting its substantial effect on tourists' wellbeing.

The Index for Wind Speed

The effect of wind speed on tourists' comfort is associated with the temperature. For example, wind increases the chill sensation by removing the heated layer of air near the skin at cooler temperatures and cools the body by the same action at warmer temperatures (Mieczkowski 1985). To thoroughly reflect the impact of wind speed, Mieczkowski (1985) provides four separate rating schemes according to the temperature range. The normal system is recognized when the daily maximum temperature is between 15 degree Celsius and 24 degree Celsius. Wind is considered to be a negative factor, and the score decreases as the wind speed increases. For example, a score of 5.0 is assigned when the mean wind speed is below 2.88 kilometers per hour, and the mean wind speed from 9.04 kilometers per hour to 12.23 kilometers per hour is set a score of 3.5. The trade wind system is identified when the daily maximum temperature ranges from 24 degree Celsius to 33 degree Celsius. The best score is set at a moderate wind speed of 12.24 kilometers per hour to 19.79 kilometers per hour to reflect the positive evaporative cooling effect of the wind. When the daily maximum temperature exceeds 33 degree Celsius, any wind is considered to be unfavorable, as wind can increase the heat load by adding convective heat to human body. The lowest mean wind speed has the highest score, but the value is only 2.0, and 0 point is set to any mean wind speed that exceeds 19.79 kilometers per hour in this hot climate system. To reflect the negative effect of wind in cold season when the daily maximum temperature is below 15 degree Celsius and the mean wind speed is greater than 8 kilometers per hour, a wind chill nomogram is provided (Mieczkowski 1985). When the mean wind speed is below 8 kilometers per hour, the normal system still applies. The wind factor is assigned a weight of 10% as it is less important compared to the precipitation and the sunshine duration.

Each sub-index is rated to take values ranging from 5 representing optimal to -3 standing for extremely unfavorable. And the TCIs are obtained by aggregating every sub-index, resulting in values ranging from 100 for ideal to -20 for impossible.

The general form of the climate econometric models can be specified as:

$$\mathrm{lny_t} = \mathrm{f^c}\left(\mathrm{lnGDP_t}, \mathrm{lnrp_t}, \mathrm{lnsp_t}, \mathrm{lnTCI_{UK,t}}, \mathrm{D_{DS}}, \mathrm{D_{911}}, \mathrm{D_{BM}}, \mathrm{D_{2008}}, \mathrm{D_{OL}}\right) \quad 3.5$$

where $\mathrm{TCI_{UK,t}}$ is the quarterly tourism climatic index of the UK representing the overall climate attractiveness of the UK to tourists.

3.3.3 Data Sources

Data for the dependent variable, i.e., the quarterly tourist arrivals for holiday purposes from seven leading origins to the UK are from Travelpac. Travelpac consists of series of quarterly data files derived from the International Passenger Survey (IPS), which is conducted by the Office for National Statistics (ONS) to passengers entering and leaving the UK from 1994, and provides information such as demographics, travel mode, travel purpose, length of stay, number of visits, nights spent and tourists' expenditure. The total number of tourist arrivals for each season from

1994Q1 to 2017Q4 from different countries of origin can be extracted from Travelpac using Excel. To compute substitute price, market shares of potential alternative destinations including Germany and France are needed. Due to data unavailability, market share is represented by yearly number of overnight stays in one destination. Data for inbound tourism demand from France, Irish Republic, Italy, the Netherlands and the US to Germany are collected from German National Tourist Board (GNTB), and that for tourist arrivals from Germany, Irish Republic, Italy, the Netherlands and the US to France are derived from The Directorate General for Enterprise (DGE).

Data for the economic explanatory variables are from the website of Federal Reserve Bank of St. Louis[3]. The website gathers thousands of economic data regarding a variety of countries from 87 sources including Bank of England, Bank for International Settlements, Eurostat, US Bureau of Economic Analysis, etc. Real GDP data for European origins are based on millions of chained 2010 Euros, and that for the US are based on millions of chained 2012 dollars. Since there exists no price index for the tourists, CPI of all items (2010=100), which can reflect the general living costs in one country are chosen. National currencies to US Dollar exchange rates are measured by the averages of daily rates, and they are extracted for each country. Different time spans are considered for each origin giving data availability: for the French and German markets, data are from 1994Q1 to 2017Q4; for the Irish, Dutch and Spanish cases, data are from 1995Q1 to 2017Q4, for the Italian origin, it is from 1996Q1 TO 1017Q4, and for the American market, it is from 2002Q1 to 2017Q4. All data are quarterly in frequency and they are without seasonal adjustment.

Data for the climate determinant, i.e. the UK's TCI are from the Met Office. Met Office provides monthly meteorological data from many weather stations across the UK. Only data from the Heathrow weather station are taken into account for the following reasons. Firstly, to construct TCI, data on seven meteorological variables including 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 are needed. Heathrow stations provides all the data required for the longest time span, while, others fail to provide some of the data either for the whole time period under consideration (1994Q1-2017Q4), or for some seasons. Secondly, there exist regional variations in climate conditions across the UK. If we sum up data from different stations located in various regions, the characteristics of the climate conditions for each season may be covered. For example, if weather stations from Scotland, middle England and Southeast England are considered, and the data are pooled together by taking averages, the true nature of the climate condition in each region cannot be revealed because of the offset effect. Thirdly, most tourists tend to take the weather condition of their first stop into consideration when they make travel decisions. The UK enjoys excellent global connectivity, with over 100 countries having direct air connections to the UK in 2017, and 76% of inbound visitors reached the UK by air in 2017. Being the busiest airport in the UK, Heathrow Airport is the first stop for most international travelers. So even there were data from several stations available, it is better to choose data from the Heathrow station to construct TCI in this study. Quarterly mean values are generated for the maximum daily temperature, the minimum daily relative humidity, the mean daily temperature, the mean daily relative humidity, the mean hours sunshine per day and the wind speed by taking averages of the monthly data, and the quarterly precipitation is got by taking sum of the monthly variables. After getting the original data, the rating scores for each sub-index in TCI are obtained following Mieczkowski (1985) before calculating TCI. Table 3-2 summarizes the variables and data sources.

[3] The website of Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org

Table 3-2 Variables and Data Sources

Variables	**Explanation**	**Transformation / Formula**	**Data source**
$\ln y_t$	dependent variable: tourist arrivals for holiday purpose from one origin to the UK	logarithm	Travelpac
$\ln GDP_t$	independent variable: tourists' income: the real GDP in the origin	logarithm	Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org)
$\ln rp_t$	independent variable: own price/relative price: exchange rate adjusted CPI of the UK to that of the origin country	$\ln rp_t = \ln\left(\frac{\frac{CPI_{UK,t}}{EX_{UK,t}}}{\frac{CPI_{OG,t}}{EX_{OG,t}}}\right)$	Federal Reserve Bank of St. Louis (https://fred.stlouisfed.org)
$\ln sp_t$	independent variable: substitute price: the market share adjusted relative prices of alternative destinations	$\ln sp_t = \ln\left(\frac{\sum_k^n rp_{k,t} * K_{k,t}}{\sum_k^n K_{k,t}}\right)$ $K_{k,t}$ 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 11th of September, 2001		
D_{BM}	independent variable: one-off event dummy: the terrorist bombing in London on 7th 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.			
$LnTCI_{UK,t}$	$cid_{UK,t}$	Independent variable: The UK's climate conditions: the TCI of the UK	the daytime comfort index	$TCI_{UK,t} = 2 \times \left[4\left(cid_{UK,t}\right) + cia_{UK,t} + 2\left(P_{UK,t}\right) + 2\left(S_{UK,t}\right) + W_{UK,t}\right]$	Met Office
	$cia_{UK,t}$		the daily comfort index		
	$P_{UK,t}$		the index for precipitation		
	$S_{UK,t}$		The index for sunshine duration		
	$W_{UK,t}$		The index for wind speed		

Source: the author

3.3.4 Data Sample

The whole sample is divided into three periods as illustrated by figure 3-3. The observations ranging from 1994Q1 to 2012Q4 are used for model estimation. The individual out-of-sample forecasts are generated from 2013Q1 to 2017Q4, with forecasts from 2013Q1 to 2015Q4 being used to determine the combining weights, and the ones from 2016Q1 to 2017Q4 being retained for comparison. The out-of-sample combination forecasts are generated from 2016Q1 to 2017Q4. The sample from 1994Q1 to 2012Q4 is called the estimation sample, that from 2013Q1 to 2015Q4 is named the training sample, and that from 2016Q1 to 2017Q4 is referred to as the comparison sample.

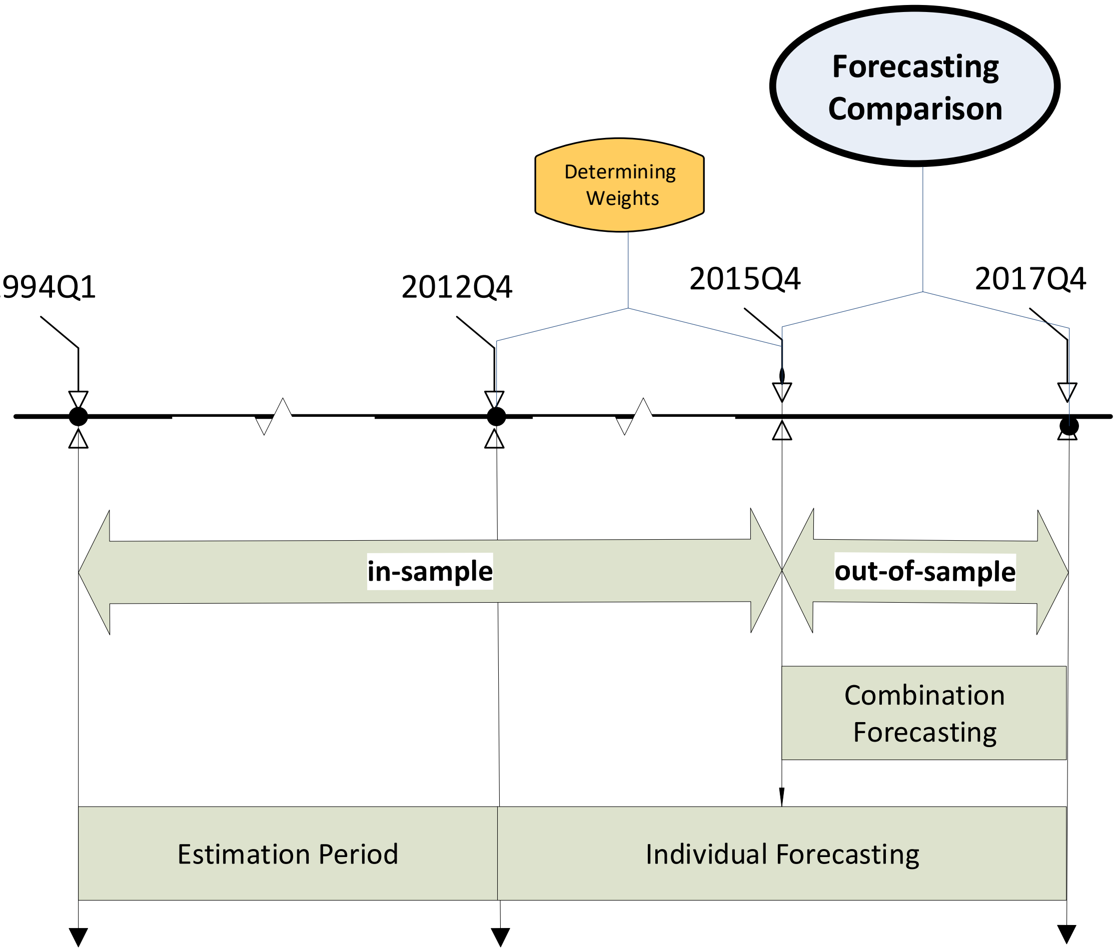

Figure 3-3 Illustration of Data Sample

Source: the author

3.4 The Bounds Test Cointegration Approach

The bounds test cointegration approach is based on the general ADLM of the following form:

$$y_t = \alpha + \sum_{j=1}^{n} \sum_{i=0}^{p_j} \beta_{j,i}\, x_{j,t-i} + \sum_{i=1}^{p} \phi_i y_{t-i} + \varepsilon_t \qquad 3.6$$

where y_t represents a measure of tourism demand, $x_{j,t-i}$ are the identified explanatory variables with the number of n and corresponding parameters of $\beta_{j,i}$, p_j and p are the lag lengths of independent and dependent variables respectively, ϕ_i is the coefficient on y_{t-i}, which needs to be estimated and ε_t is the error term which is assumed to be an IID variable of zero mean and a variance of σ^2 (Morley 2009; Song, Witt and Li 2009, p47).

To conduct the bounds test, equation 3.6 is re-parameterized to the following form:

$$\Delta y_t = \alpha + \sum_{j=1}^{n} \sum_{i=0}^{p_j - 1} \beta_{j,i} \Delta x_{j,t-i} + \sum_{i=1}^{p-1} \phi_i \Delta y_{t-i} + \lambda_0 y_{t-1} + \sum_{j=1}^{n} \lambda_j x_{j,t-1} + u_t \qquad 3.7$$

where Δy_t signifies the first difference of the dependent variable with $\Delta y_t = y_t - y_{t-1}$, $\Delta x_{j,t}$ represents the first differences of the explanatory variables with $\Delta x_{j,t} = x_{j,t} - x_{j,t-1}$. The lagged level variables (y_{t-1} and $x_{j,t-1}$) represent the cointegration relationship, and the estimated coefficients on the first differenced terms (Δy_{t-i} and $\Delta x_{j,t-i}$), i.e., $\hat{\beta}_{j,i}$ and $\hat{\phi}_i$ represent the short-run dynamics.

Bounds test for cointegration is applied by testing the joint significance of the lagged levels of all the variables under consideration in equation 3.7 using the F-test. The null hypothesis of no cointegration is defined as:

$$H_0: \lambda_0 = \lambda_1 = \lambda_2 = \cdots = \lambda_n = 0 \qquad 3.8$$

And the alternative hypothesis of cointegration is stated as at least one of the λ_j is non zero.

The computed F-statistic is compared to the critical values which are provided by Pesaran et al. (2001). The critical value bounds can classify the variables into purely $I(1)$, purely $I(0)$ or mutually cointegrated. The null hypothesis of 'no long-run relationship' is rejected if the computed F-statistic is greater than the upper critical bound value, and it is accepted if the F-statistic is less than the lower bound value. When the F-statistic falls within the lower and the upper bounds, the test is inconclusive (Halicioglu 2010).

After the cointegration relationship is confirmed, the long-run parameters are derived from the following estimation:

$$y_t = k_0^* + k_j^* x_{jt} \qquad 3.9$$

where $k_0^* = \frac{\hat{\alpha}}{1 - \sum_{i=1}^{p} \widehat{\phi_i}}$ and $k_j^* = \frac{\sum_{i=0}^{p_j} \widehat{\beta_{ji}}}{1 - \sum_{i=1}^{p} \widehat{\phi_i}}$.

The underlying assumption of the bounds test cointegration technique is that the variables included in the model are stationary or integrated of order one. When variables that are integrated of order two or above are present, the computed statistics are invalid. To avoid spurious results in the bounds test, the unit root test is necessary.

Compared to other cointegration techniques, the bounds test cointegration approach is advantageous as it is able to detect cointegration relationship and solve the small sample problem no matter whether the model variables are purely $I(1)$, purely $I(0)$ or a mixture of both (Song et al. 2012; Wang 2009). It is applicable when the model variables are not integrated of the same order (order zero or order one).

3.5 Diagnostic Tests

Diagnostic tests including residual diagnostics and stability diagnostics are used to check whether essential assumptions of econometric techniques are met and whether the model is correctly specified and stable over time. The necessary tests such as testing for normality, testing for autocorrelation, testing for heteroscedasticity and testing for misspecification are applied in this study.

3.5.1 The Jarque-Bera Normality Test

It is assumed that the residuals of econometric models should be normally distributed, and validation of this assumption will result in invalid t and F statistics. The normality test was presented by Jarque and Bera (1980), and it is also known as the J-B test. The J-B test statistic measures the difference between the skewness and kurtosis of the series under study and those of the normal distribution, and is computed as follows:

$$\mathrm{n}\left[\frac{\mu_3^2}{6\mu_2^3}+\frac{\left(\frac{\mu_4}{\mu_2^2}-3\right)^2}{24}\right] \qquad 3.10$$

where $\hat{\varepsilon}_t$ is the residual from the estimated model, $\mu_2=\sum_{t=1}^{n}\hat{\varepsilon}_t^2/n$, $\mu_3=\sum_{t=1}^{n}\hat{\varepsilon}_t^3/n$ and $\mu_4=\sum_{t=1}^{n}\hat{\varepsilon}_t^4/n$ are the second, third and fourth moments of the residuals respectively.

The J-B statistic has a χ^2 distribution with two degrees of freedom (denoted as $\chi^2(2)$). When the calculated J-B statistic is greater than the critical value of $\chi^2(2)$, the null hypothesis of normally distributed residuals is rejected, and when the J-B statistic is smaller than the critical value, the null hypothesis is accepted.

3.5.2 The Breusch-Godfrey Lagrange Multiplier Test

The B-G LM test was presented by Breusch (1978) and Godfrey (1978), and it has become the most popular tests for autocorrelation in the tourism demand literature. It can be used to test for higher orders autocorrelation and is valid when there is lagged dependent variables existing on the right-hand side of the model. The test is based on the following auxiliary equation:

$$\hat{\varepsilon}_t=\alpha+\sum_{i=1}^{k}\beta_i x_{it}+\sum_{j=1}^{p}\rho_j\hat{\varepsilon}_{t-j}+u_t \qquad 3.11$$

where x_{it} are the explanatory variables of the model of number k, $\hat{\varepsilon}_t$ is the estimated residual of the model, p is the lag length of the lagged estimated residual, which is pre-specified, and β_i and ρ_j are the parameters.

The null hypothesis is that there is no serial correlation up to lag p, i.e., $H_0\colon\rho_1=\rho_2=\cdots=\rho_p=0$, where p is a pre-specified integer, and the alternative hypothesis is that there exists ARMA(r,q) errors, where the number of lag terms $p=\max(r,q)$. It means that the B-G LM test can be used to test for higher order ARMA errors. It may have power against a variety of autocorrelation structures (Godfrey, 1991).

Under the null hypothesis, the test statistic is nR^2, where n is the sample size and R^2 is the coefficient of determination of equation 3.11. The test statistic has an asymptotical χ^2 distribution with p degrees of freedom (denoted as $\chi^2(p)$). When the value of nR^2 is greater than the critical value of $\chi^2(p)$, the null hypothesis of no autocorrelation is rejected, and the null hypothesis is accepted when the value of nR^2 is smaller than the critical value of $\chi^2(p)$.

The B-G LM test is popular in the current tourism demand literature because of its general applicability and is applied in this study. It overcomes the limitations of the Durbin-Watson (DW) test, which was developed by Durbin and Watson (1950; 1951). The DW test can only be used to test for first order autocorrelation. The DW statistic is defined as:

$$DW = \frac{\sum_{t=2}^{n}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^{n}\hat{\varepsilon}_t^2} \qquad 3.12$$

where $\hat{\varepsilon}_t$ is the residual from the estimated model.

The null hypothesis of the test is that the errors are serially uncorrelated, and the alternative hypothesis is that the errors follow a first order autoregressive process. The values of the DW statistic ranges from 0 to 4. When the value is around 2, the null hypothesis of no autocorrelation is accepted, a DW statistic of value 0 would suggest perfect positive autocorrelation, and a DW statistic of value 4 would indicate perfect negative autocorrelation. The main limitations of the DW test are as follows: firstly, there are inclusive region, which is determined by the sample size; secondly, it cannot be used to detect higher orders of autocorrelation; and lastly, it is invalid when there is lagged dependent variables as influencing factors in the model.

3.5.3 Testing for Heteroscedasticity

It is assumed that the variance of the errors from an econometric model is constant over time, which is called homoscedasticity. When the assumption of homoscedasticity does not hold, the ordinary least squares estimates are still unbiased, but efficiency is lost, and the conventional computed standard errors are no longer valid. When there is evidence of heteroscedasticity, either the heteroscedasticity-consistent standard errors should be used, or the heteroscedasticity should be modelled using weighted least squares to obtain more efficient estimates. Two types of heteroscedasticity tests including the Harvey test for heteroscedasticity and the autoregressive conditional heteroscedasticity (ARCH) test are applied in this study.

3.5.3.1 The Harvey Test for Heteroscedasticity

The Harvey test for heteroscedasticity was presented by Harvey (1976). It has the null hypothesis of no heteroscedasticity against the alternative hypothesis of heteroscedasticity. The heteroscedasticity is assumed to take the following form: $\sigma_t^2 = exp(X_t'\gamma)$, where$X'$is a vector of independent variables of the model under consideration. The test is based on the following auxiliary regression equation:

$$ln\hat{\varepsilon}_t^2 = \gamma_0 + \gamma_1 x_{1t} + \gamma_2 x_{2t} + \cdots + \gamma_k x_{kt} + u_t \qquad 3.13$$

where x_{it} are the explanatory variables of the model of number k, $\hat{\varepsilon}_t$ is the estimated residual of the model, γ_i are the parameters.

Under the null hypothesis of homoscedasticity, i.e., $H_0: \gamma_1 = \gamma_2 = \cdots = \gamma_k = 0$, the test statistic is nR^2, where n is the sample size and R^2 is the coefficient of determination of the auxiliary regression equation (equation 3.13). The test statistic has an asymptotical χ^2 distribution with k degrees of freedom (denoted as $\chi^2(k)$). When the value of nR^2 is greater than the critical value of $\chi^2(k)$, the null hypothesis of homoscedasticity is rejected, and accepted otherwise.

The Harvey test is similar to the Breusch-Pagan Lagrange multiplier (B-P LM) test, which was presented by Breusch and Pagan (1979) and extended by Koenker (1981). The B-P LM test has the null hypothesis of no heteroscedasticity against the alternative hypothesis of heteroscedasticity, and the heteroscedasticity is assumed to take the following form: $\sigma_t^2 = h(X_t'\gamma)$, where $X'$is a vector of independent variables of the model under consideration. The test is based on the following auxiliary regression equation:

$$\hat{\varepsilon}_t^2 = \gamma_0 + \gamma_1 x_{1t} + \gamma_2 x_{2t} + \cdots + \gamma_k x_{kt} + u_t \qquad 3.14$$

where x_{it} are the explanatory variables of the model of number k, $\hat{\varepsilon}_t$ is the estimated residual of the model, γ_i are the parameters.

Under the null hypothesis of homoscedasticity, i.e., $H_0: \gamma_1 = \gamma_2 = \cdots = \gamma_k = 0$ the test statistic is nR^2, where n is the sample size and R^2 is the coefficient of determination of the auxiliary regression equation (equation 3.14). The test statistic has an asymptotical χ^2 distribution with k degrees of freedom (denoted as $\chi^2(k)$). When the value of nR^2 is greater than the critical value of $\chi^2(k)$, the null hypothesis of homoscedasticity is rejected, and if the value of nR^2 is smaller than the critical value of $\chi^2(k)$, the null hypothesis of homoscedasticity is accepted.

3.5.3.2 The ARCH Test

The ARCH test is a Lagrange multiplier test for ARCH in the residuals presented by Engle (1982). This particular heteroscedasticity specification was motivated by the observation that in many financial time series, the magnitude of residuals appeared to be related to the magnitude of recent residuals. The presence of ARCH may result in loss of efficiency in OLS estimates. The test is based on the following auxiliary regression:

$$\hat{\varepsilon}_t^2 = \gamma_0 + \gamma_1 \hat{\varepsilon}_{t-1}^2 + \gamma_2 \hat{\varepsilon}_{t-2}^2 + \cdots + \gamma_p \hat{\varepsilon}_{t-p}^2 + u_t \qquad 3.15$$

where $\hat{\varepsilon}_t$ is the estimated residual of the model, p is the pre-selected lag length, and γ_i are the parameters.

The null hypothesis is that there is no ARCH up to order p in the residuals. Under the null hypothesis, the test statistic is Engle's LM test statistic, which is computed as nR^2, where n is the sample size and R^2 is the coefficient of determination of the auxiliary regression equation (equation 3.15). The LM test statistic is asymptotically distributed as χ^2 distribution with p degrees of freedom (denoted as $\chi^2(p)$). When the value of nR^2 is greater than the critical value of $\chi^2(k)$, the null hypothesis of no ARCH is rejected, and it is accepted if the statistic is smaller than the critical value.

3.5.4 The Ramsey Regression Equation Specification Error Test

The Ramsey RESET test was developed by Ramsey (1969) for detecting model misspecification from the following sources: omitting important explanatory variables, choosing incorrect functional form, and existence of correlation between explanatory variables and errors. Such misspecifications will result in biased and inconsistent least squares estimators, as well as invalid conventional inference procedures. The RESET test is conducted through three steps:

The first step is to estimate the proposed model (equation 3.16) using OLS and to retain the estimated values of the dependent variable y_t:

$$y_t = x_t \beta + \varepsilon_t \qquad 3.16$$

where y_t is the dependent variable, x_t is the vector of independent variables and ε_t is the error term.

The second step is based on the augmented regression (equation 3.17), which is to estimate the original dependent variable y_t against the original explanatory variables x_t and different powers of the fitted values of y_t from equation 3.16:

$$y_t = x_t \delta + Z_t \gamma + u_t \qquad 3.17$$

where $Z_t = (\hat{y}_t^2, \hat{y}_t^3, \hat{y}_t^4, \ldots)$, $\hat{y}_t$ is fitted value from equation 3.16 of y against x_t, and δ and γ are parameters. The estimated dependent variables with different powers are introduce one by one into the augmented equation (equation 3.17), with the lowest powered variable first. If only

$\hat{y}_t^2$ is included, the significance of its parameter can be tested using the t statistic.

The third step is to test the significance of the γ parameters using standard restriction test such as the F or Wald test. The null hypothesis of this test is that there is no misspecification, and when the γ parameters are indicated to be zero, the null hypothesis of no misspecification are accepted. However, when the null hypothesis is rejected if the γ parameters are indicated to be non-zero, which suggests the existence of misspecification, the test cannot detect the specific type of misspecification. Because there is no specific alternative hypothesis indicating how the model is mis-specified.

3.6 Forecasting Methods

3.6.1 Individual Forecasting Models

The selected individual models that are considered as candidate constituents in combinations serve not only as forecasting tools but also as sources of diversification in combination forecasts. If we see forecasts as information, combining forecasts are aggregation of information. According to Bates and Granger (1969), the combination of models that contain independent information is most likely to improve the forecasting performance. To ensure that constituent models contain as much independent information as possible, a variety of individual forecasting models which are different in modelling techniques, assumptions and explanatory variables are selected.

The individual forecasting models applied in this study are comprised of causal econometric and non-causal time series techniques with seasonal naive no-change forecasts being as benchmarks. As reviewed in chapter 2, in the current literature, popular econometric models based on time series data include the cointegration technique, ADLM, LI, VAR, TVP and SD models, all of which are utilized to generate individual forecasts. The chosen cointegration technique is the bounds test cointegration approach, which is selected based on the integration orders of model variables. All econometric techniques are applied with two model specifications which are different in identified explanatory variables. The ones that follow the mainstream econometric specification in the current tourism demand 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 time series techniques employed consist of the 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 constituent models, all of which are summarized in table 3-3. The detailed discussion of the bounds test cointegration approach can be found in section 3.4, that of the other causal econometric models is presented in section 2.3.2, and that of the non-causal time series techniques is provided in section 2.2.

To reduce overlapping information, some popular forecasting techniques are excluded. For example, the SARIMAX model are not included as it is in essence an integration of non-causal time series and causal econometric modelling techniques, which contains the same information as that in the SARIMA model and econometric models. Only one type of cointegration technique is selected not only due to the requirement of model variables but also owing to the fact that similar information is contained in different types of cointegration techniques. The STSM is excluded since there is already a type of state space model, i.e., the state space ETS model included. The AI-based forecasting techniques are not considered as they are beyond the researcher's expertise.

Table 3-3 Summary of Individual Forecasting Models

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

The TVP Model	$y_t = x_t'\beta_t + \varepsilon_t, \qquad \varepsilon_t \sim i.i.d.(0, H_t)$ $\beta_{t+1} = T_t\beta_t + \eta_t, \qquad \beta_1 \sim N(b_1, P_1), \eta_t \sim i.i.d.(0, Q_{t)}$	
The Simple Dynamic Model	$y_t = \sum_{j=1}^{n} \beta_{j0}\, x_{jt} + \phi_1 y_{t-s} + \varepsilon_t$	
Time Series Techniques		
The SARIMA Model	$\Phi_P(B^s)\phi_p(B)(1-B^s)^D(1-B)^d y_t = \mu + \Theta_Q(B^s)\theta_q(B)\varepsilon_t$	
The ETS Model	$\hat{y}_{t+k} = (L_t + kT_t)S_{t+k-p}$ $L_t = \alpha\frac{y_t}{S_{t-p}} + (1-\alpha)(L_{t-1} + T_{t-1})$ $T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$ $S_t = \delta\frac{y_t}{L_t} + (1-\delta)S_{t-p}$ or, $\hat{y}_{t+k} = L_t + kT_t + S_{t+k-p}$ $L_t = \alpha(y_t - S_{t-p}) + (1-\alpha)(L_{t-1} + T_{t-1})$ $T_t = \gamma(L_t - L_{t-1}) + (1-\gamma)T_{t-1}$ $S_t = \delta(y_t - L_t) + (1-\delta)S_{t-p}$	Non-causal models
The State Space ETS Model	$y_t = w'X_{t-1} + \varepsilon_t$ $X_t = FX_{t-1} + g\varepsilon_t$	

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

3.6.2 The Combination Forecasting Approach

3.6.2.1 Combination Forecasts

Combining individual forecasts can take advantage of more information embedded in different single models and therefore lead to more accurate predictions. The combination forecasts considered in this study are the weighted averages of the individual ones. Let $\hat{f}^{i}_{t+h,t}$ denote the ith individual out-of-sample h-step-ahead forecast of f_{t+h}, computed at time t, which is, the ith forecast in a group of single forecasts for a given country. The h-step-ahead combination forecasts have the following form:

$$f^{c}_{t+h,t} = \sum_{i=1}^{n} w_{it}\hat{f}^{i}_{t+h,t} \qquad 3.18$$

where $f^{c}_{t+h,t}$ denotes the combination forecast, w_{it} is the weight for the ith single forecast $\hat{f}^{i}_{t+h,t}$ at time t, and n is the number of the constituent forecasts, and h represents the forecasting horizon. In this study, one- to four-step-ahead forecasts are considered, so h takes values of 1, 2, 3, 4.

3.6.2.2 Weighting Schemes

This study evaluates the most popular statistical combination approaches in the current tourism demand literature consisting of the SA, VACO and DMSFE methods. Besides, two weighting schemes including the inverse-MAE method and the two-stage combination method have been introduced and tested. 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. Except SA, all other combination approaches assign unequal weights and the focus is on how to get the optimal weights.

The Simple Average Method

The SA method assigns equal weights to all single forecasts. Because it is easy to implement and has good track record in economic and business forecasting, the SA method is a common choice and always serves as a benchmark in the tourism demand literature. The composite forecast can be calculated as:

$$f^{c}_{t+h,t} = \sum_{i=1}^{n} \frac{1}{n}\hat{f}^{i}_{t+h,t} \qquad 3.19$$

where $f^{c}_{t+h,t}$ is the h-step-ahead combination forecast, and $\hat{f}^{i}_{t+h,t}$ is the ith individual out-of-sample h-step-ahead forecast, and n is the number of single constituents in the combination panel.

The Variance-Covariance Method

Weights in the VACO method are calculated to minimise the error variance of the combination forecasts (assuming unbiasedness of each single forecast). The principle can be illustrated using the case where there are two single forecasts in the combination panel. The combination forecasts from the two unbiased forecasting methods are given as:

$$f^{c}_{t+h,t} = w\hat{f}^{1}_{t+h,t} + (1-w)\hat{f}^{2}_{t+h,t} \quad 3.20$$

where $f^{c}_{t+h,t}$ is the h-step-ahead combination forecast derived from the h-step-ahead single forecasts of $\hat{f}^{1}_{t+h,t}$ and $\hat{f}^{2}_{t+h,t}$, and w is the combining weight for the single forecasts $\hat{f}^{1}_{t+h,t}$. The error of the combination forecast $e^{c}_{t+h,t}$ can be expressed as:

$$e^{c}_{t+h,t} = we^{1}_{t+h,t} + (1-w)e^{2}_{t+h,t} \quad 3.21$$

with the variance as

$$\sigma_c^2 = w^2\sigma_{11}^2 + (1-w^2)\sigma_{22}^2 + 2w(1-w)\sigma_{12} \quad 3.22$$

where $e^{1}_{t+h,t}$ and $e^{2}_{t+h,t}$ are the errors of the single forecasts $\hat{f}^{1}_{t+h,t}$ and $\hat{f}^{2}_{t+h,t}$ respectively, σ_{11}^2 and σ_{22}^2 are the unconditional forecast error variances and σ_{12} is the unconditional forecast error covariance. The weight that minimises the combination forecast variance is constructed as:

$$w^* = \frac{\sigma_{22}^2 - \sigma_{12}}{\sigma_{22}^2 + \sigma_{11}^2 - 2\sigma_{12}} \quad 3.23$$

It is obvious that the forecast error variance from the optimal combination is lower than the individual variance of σ_{11}^2 or σ_{22}^2, so combining is beneficial. In practice, unknown σ_{11}^2, σ_{22}^2 and σ_{12} can be estimated from data, and a possible estimator of the combination weight is:

$$W^* = \frac{\sum_{t=1}^{T} e_{1t}^2 - \sum_{t=1}^{T} e_{1t}e_{2t}}{\sum_{t=1}^{T} e_{1t}^2 + \sum_{t=1}^{T} e_{2t}^2 - 2\sum_{t=1}^{T} e_{1t}e_{2t}} \quad 3.24$$

where e_{1t} and e_{2t} are estimated individual forecast errors and T is the sample size. Much simpler formula is used in practical which neglects the sample covariance term:

$$w^* = \frac{\sum_{t=1}^{T} e_{1t}^2}{\sum_{t=1}^{T} e_{1t}^2 + \sum_{t=1}^{T} e_{2t}^2} \quad 3.25$$

This formula can be extended to multiple-forecasting models:

$$w_i = \frac{\left[\sum_{t=1}^{T} e_{it}^2\right]^{-1}}{\mathrm{sum}_{j=1}^{n}\left[\sum_{t=1}^{T} e_{jt}^2\right]^{-1}} \quad 3.26$$

where w_i 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. With training sample being from 2013Q1 to 2015Q4, $T = 12$ in this study.

The Discounted Mean Square Forecast Error Method

The DMSFE method calculates the optimal weight based on mean-square error and uses a discounting factor to give more weight to the more recent forecasts. The weights scheme can be written as:

$$w_{it} = \frac{\phi_{it}^{-1}}{\sum_{j=1}^{n}\phi_{jt}^{-1}}, \ \phi_{it} = \sum_{s=1}^{T-1}\alpha^{t-1-s}\left(f_{s+h} - \hat{f}^{i}_{s+h,s}\right)^2 \quad 3.27$$

where w_{it} is the weight of the ith individual forecast, α is a selected discounting factor with $0 < \alpha \ll 1$, $\hat{f}^{i}_{s+h,s}$ is the individual h-step-ahead forecast generated at time s from single model i for the actual value f_{s+h}, T is the sample size, and n is the number of single constituents in the combination panel. 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 which are close to 1: $\alpha = 0.95,0.90,0.85$ (e.g. Shen et al. 2008; 2011; Song et al. 2009). With training sample being from 2013Q1 to 2015Q4, $T = 12$ in this study.

The differences of VACO and DMSFE lie in two aspects. Firstly, DMSFE makes use of a discounting factor to weigh the more recent forecasts more heavily. Secondly, DMSFE ignores the covariance among the errors, which is justified by Clemen and Winkler (1986). They pointed out that when the correlations among the forecast errors are high, the combination weights tend to be more sensitive to the changes in the correlations, and the covariance should be ignored to avoid the instability caused by the interdependence between combination weights and the correlations among the forecasting errors.

The Inverse-MAE Method

The inverse-MAE combination method computes the combining weights based on the historical performance of the individual forecasts which is measured by MAE. The formula of the combining weights for ith individual forecasting model can be expressed as:

$$w_i = \frac{[MAE_i]^{-1}}{sum_{j=1}^{n}\left[MAE_j\right]^{-1}} \qquad 3.28$$

where $MAE_i = \frac{1}{T}\left|\sum_{t=1}^{T} f_{t+h} - \hat{f}^{i}_{t+h}\right|$ signifies the mean absolute error of the ith individual forecast $\hat{f}^{i}_{t+h}$, f_{t+h} is the actual value, T is the sample size and n is the number of single constituents in the combination panel. With training sample being from 2013Q1 to 2015Q4, $T = 12$ in this study. This method chooses MAE as the measure of the individual forecasting performance rather than the mean squared error (MSE), which is used in VACO and DMSFE. With the same individual forecasts, the better performing constituent forecasts are weighted lighter since the forecasting error is not amplified by the square operator.

The Two-Stage Combination Method

When combination is considered, it is natural to ask how wide a set of individual models should be included as potential constituents of combination forecasts. All the above-descried combination methods include all the single models in an identified combination panel to compute the composite forecasts no matter how bad some single models may perform. Some studies showed that the divergence of the performance of the constituent forecasts is an important reason of the inferiority of the composite forecasts (Teräsvirta et al. 2005; Shen et al. 2011).

The trimmed average weighting scheme, which combines individual forecasts by a simple arithmetic mean, excluding the worst performing $k\%$ of the models, is a method to test whether excluding the worst performing single models is beneficial. A trimming of 10%-30% is usually recommended (Jose and Winkler 2008; Lemke and Gabrys 2010). However, the trimming is only applied to the simple average method. To thoroughly evaluate whether forecasting accuracy can be improved by preselecting the included individual constituents based on their forecasting performance, the two-stage combination method is presented. In the first stage, the individual candidates are ranked based on their forecasting performance, which is measured by MSE based on data from 2013Q1 to 2015Q4. The worst 20% individual models are excluded from the combination panel. In the second step, composite forecasts are computed considering the left single models as potential constituents. All the above-mentioned weighting schemes are applied with the pre-selecting procedure, resulting in six new combination methods: the two-stage SA, the two-stage VACO, the two-stage DMSFE (with three discounting factors of $\alpha = 0.95, 0.90, 0.85$) and the two-stage inverse-MAE methods.

3.6.2.3 The Number of All Possible Combinations

There are 15 individual models available to combine for four forecasting horizons and seven destination-origin pairs. The 15 individual models are categorized into three groups: the first one consists of all the 15 models; the second one is comprised of 9 models including 6 traditional econometric models and 3 time series models; and the third one contains 9 models including 6 climate econometric models and 3 time series models. For each origin country, we need to consider all possible combinations for each group. The total number of all possible combinations N can be specified as:

$$N = \sum_{r=2}^{n} C_n^r \tag{3.29}$$

$$C_n^r = \frac{n!}{r!\,(n-r)!} \tag{3.30}$$

where n is the number of the individual models in each group.

In this study, the number of all possible combinations is 32752 when $n = 15$, 5007 when $n = 12$ for the two-stage combination method, and 502 when $n = 9$. To the best of my knowledge, this study represents the first attempt to examine more than 10 single forecasting models in the combination in the current tourism demand literature.

3.7 Forecasting Procedures and Accuracy Measures

3.7.1 The Recursive Individual Forecasting Procedure

This study follows the recursive forecasting procedure which is popular in the tourism demand forecasting literature (Song and Li 2008). The procedure is illustrated in figure 3-4. The individual models are firstly estimated over the period of 1994Q1 to 2012Q4, and the estimated models are used to forecast over the period of 2013Q1 to 2017Q4, resulting in 20 forecasts, which contains one one-step-ahead forecast, one two-step-ahead forecast, one three-step-ahead forecast, etc., and one 20-step-ahead forecast. Subsequently, the models are re-estimated based on data from 1994Q1 to 2013Q1, and forecasts are generated for 2013Q2 to 2017Q4, producing 19 forecasts, which consists of one one-step-ahead forecast, one two-step-ahead forecast, one three-step-ahead forecast, etc., and one 19-step-ahead forecast. Such a procedure is repeated until all observations are exhausted with one more observation added to the estimation period each time. Finally, 20 one-step-ahead, 19 two-step-ahead, 18 three-step-ahead and 17 four-step-ahead single forecasts are generated from each individual model for each origin country.

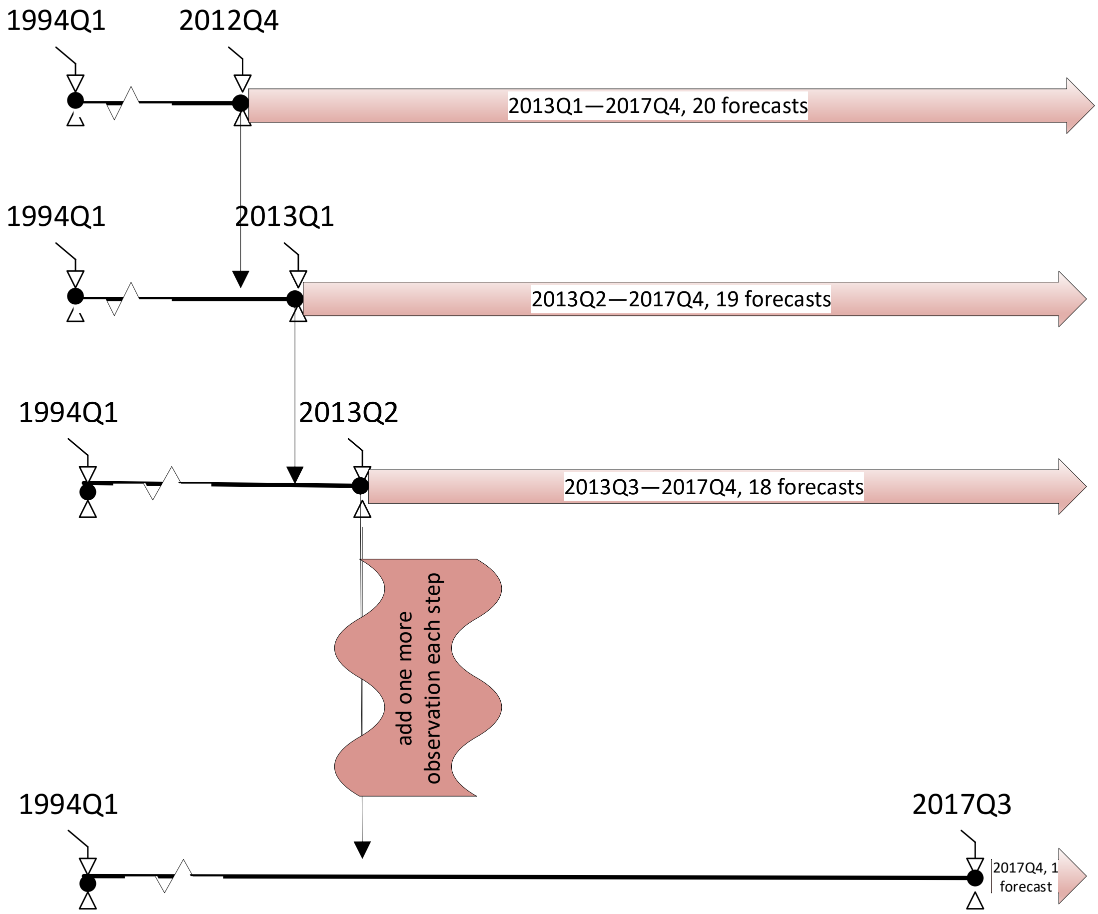

Figure 3-4 The Recursive Forecasting Procedure

Source: the author

3.7.2 The Recursive Weighting Procedure

The combining weights determined by different combination methods (except the SA scheme) depend on the historical performance of individual forecasts. And the weights are time-varying by applying the recursive weighting procedure (see figure 3-5[4]). For example, for the composite forecasts in 2016Q1, the historical performance of single forecasts from 2013Q1 to 2015Q4 are considered to construct the combining weights. For the combination forecasts in 2016Q2, the historical performance of the individual forecasts from 2013Q1 to 2016Q1 are taken into account to decide the combining weights. Such a procedure is repeated with one more forecast added to the training sample each time, updating the weights each period according to the historical individual forecasting performance.

4 Figure 3-5 illustrates the computation of the time-varying weights for one-step-ahead combination forecasts. The weights for two- to four-step-ahead combination forecasts are computed iteratively in the same way.

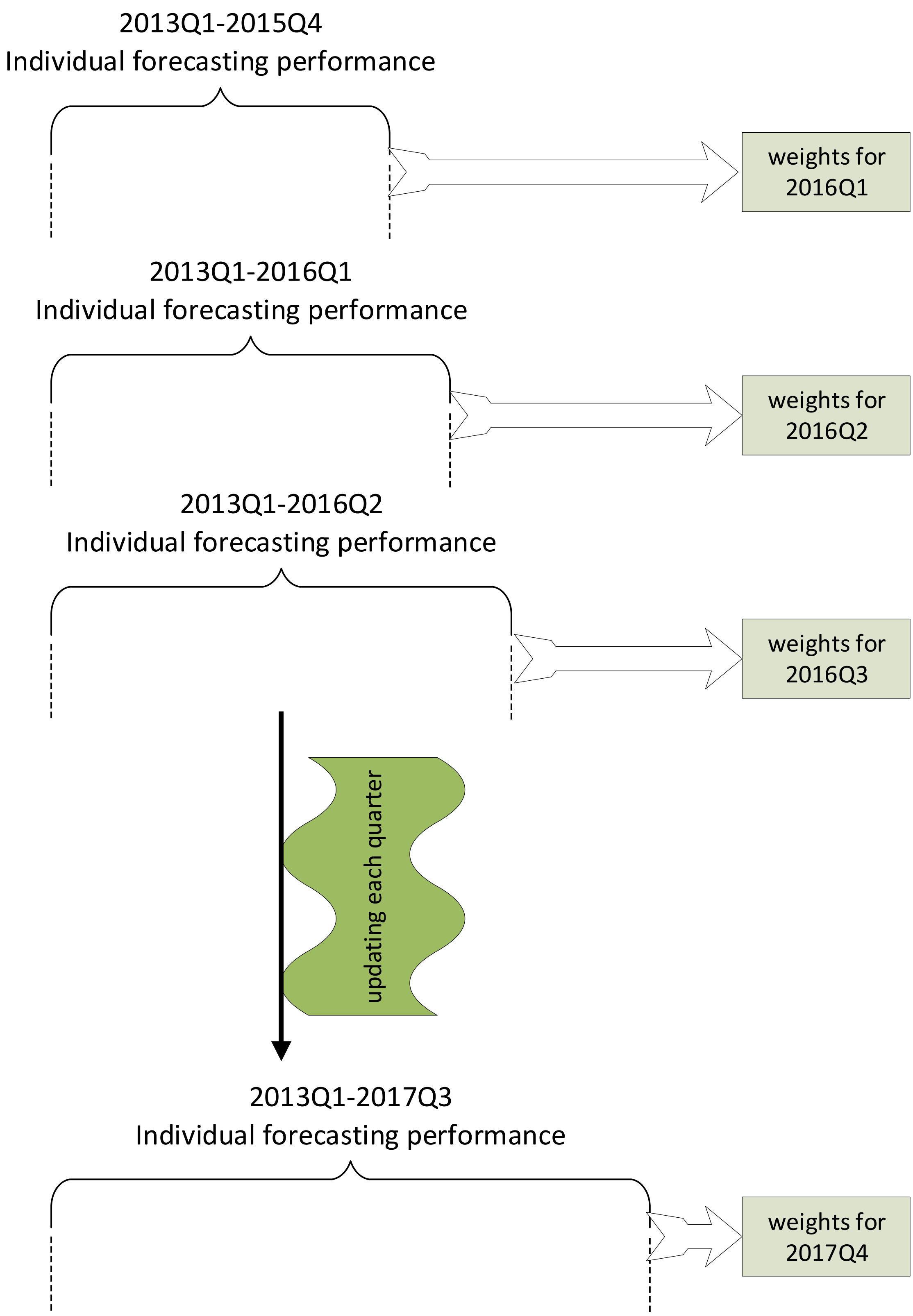

Figure 3-5 The Recursive Weighting Procedure for One-Step-Ahead Combination Forecasts

Source: the author

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

3.7.3 Forecasting Accuracy Measures

The forecasting accuracy depends on how close the h-step-ahead forecast values ($\hat{f}_{t+h,t}$) are to the actual values (f_{t+h}), which can be measured by the forecasting error, i.e., the difference between the forecast and the actual values. The forecasting error e_t can be specified as:

$$e^{i}_{t+h} = f_{t+h} - \hat{f}^{i}_{t+h,t} \qquad 3.31$$

where e^{i}_{t+h} is the h-step-ahead forecasting error of the ith model at time t, $\hat{f}^{i}_{t+h,t}$ is the h-step-ahead forecast value from the ith model at time t, and f_{t+h} is the actual value.

In theory, the forecasting errors are expected to be random series with zero mean if the forecasting model is correctly specified, and the sum of the forecasting errors is likely to be zero (Song, Witt and Li 2009). However, poorly-performed models may give very small forecasting errors because of the offset of the positive and negative errors.

In the current tourism demand literature, the forecasting performance of alternative competitors is evaluated based on loss functions, which either make use of the absolute values or the squared values of the forecasting errors. The three most popular loss functions in tourism studies including MAE, MAPE and RMSE are considered here. MAE, MAPE and RMSE can be specified as:

$$MAE_{i,h} = \frac{1}{T}\left|\sum_{t=1}^{T} f_{t+h} - \hat{f}^{i}_{t+h,t}\right| \qquad 3.32$$

$$MAPE_{i,h} = \frac{1}{T}\left|\sum_{t=1}^{T} \frac{f_{t+h} - \hat{f}^{i}_{t+h,t}}{f_{t+h}}\right| \qquad 3.33$$

$$RMSE_{i,h} = \sqrt{\frac{\sum_{t=1}^{T}\left(f_{t+h} - \hat{f}^{i}_{t+h,t}\right)^2}{T}} \qquad 3.34$$

where $MAE_{i,h}$, $MAPE_{i,h}$ and $RMSE_{i,h}$ are the h-step-ahead MAE, MAPE and RMSE for the ith model, f_{t+h} is the actual value, $\hat{f}^{i}_{t+h,t}$ is the h-step-ahead forecast from model i, and the comparison sample is from 2016Q1 to 2017Q4 with sample size $T = 8$.

The differences among the three measures mainly lie in two aspects. Firstly, MAE and RMSE are absolute measures, whose values are affected by the magnitudes of the demand variables. On the other hand, as a relative measure, the value of MAPE is independent on the unit or magnitudes of the demand variables, and hence is more favorable when making comparisons across different destination-origin pairs compared to MAE and RMSE. Secondly, MAE and MAPE are based on the linear loss function, which gives equal weights to all errors irrespective of the magnitude of the error, and RMSE is based on the quadratic loss function, which gives more weights to larger errors. So poorer forecasting performance is assessed by RMSE than by MAE if there are large forecasting errors.

There are different opinions regarding which type of loss function is favorable. The ones supporting the quadratic loss function think that larger errors affect forecasting performance more than smaller errors, and hence should be weighted more heavily. For example, Huss (1985) comments that: "*the RMSE places particularly strong emphasis on outliers. It is consistent with the assumption that the cost of forecasting errors increases exponentially with the size of the error.*" Similarly, Kling and Bessler (1985) states that: "*if one believes that large errors have a greater than proportional cost to decision makers than do small errors, then the RMS error is the more appropriate of the two evaluation criteria*." Others who disagree with the quadratic loss function argue that one very large forecasting error may be caused by an unforeseen exogenous shock, and therefore giving more weights to larger errors undermines the fairness of the overall evaluation. To provide different criteria regarding assessing forecasting performance, this study applies MAE, MAPE and RMSE as accuracy measures, and the smaller these measures are, the better the forecasts are.

3.8 Programming for Combination Forecasting

The codes for computing the combination forecasts as well as conducting forecasting comparison and analysis are written in Matlab 2018a, and the seven origins are studied one by one with the same program and different data. For each origin, individual forecasts are categorized into three groups, with the first one including all single forecasts, the second one consisting of forecasts from traditional econometric and time series models, and the last on being comprised of forecasts from climate econometric and time series models. The codes for each group are similar with different individual inputs, and for each group, the following steps are carried out, which is illustrated in figure 3-6.

Firstly, data for one- to four-step-ahead individual forecasts from each single model are imported and used for calculating the forecasting accuracy measures for single projections. The three measures of MAE, MAPE and RMSE are worked out individually, and the measures of the best and worst single forecasts are saved. Secondly, all possible combinations are identified considering all component forecasts, and different combining weights for each component are computed accounting for individual forecasts and the actual values of tourism demand. The different weighting schemes under study are treated respectively, and the recursive weighting procedure is programmed. Next, based on the results of all possible combinations and the corresponding weights, all combination forecasts are worked out and retained in a matrix. Besides, relevant results for comparison and analysis, which include the accuracy measures, the forecasting horizon, the number of component and the weighting scheme of each combination forecast are also computed and summarized into a cell array. From these results, the superior combination forecasts compared to the best single ones, and the inferior combination forecasts compared to the worst single ones are recognized, and the superior and inferior percentages are calculated. In addition, the best forecasts from each comparison are identified and the number of components comprising it are worked out. Finally, the results of combination forecasts and the relevant results are integrated to get the frequencies of each component forecasts in the superior composite ones. The iteration loops are used when needed throughout the program, and the detailed codes are provided in Appendix 2.

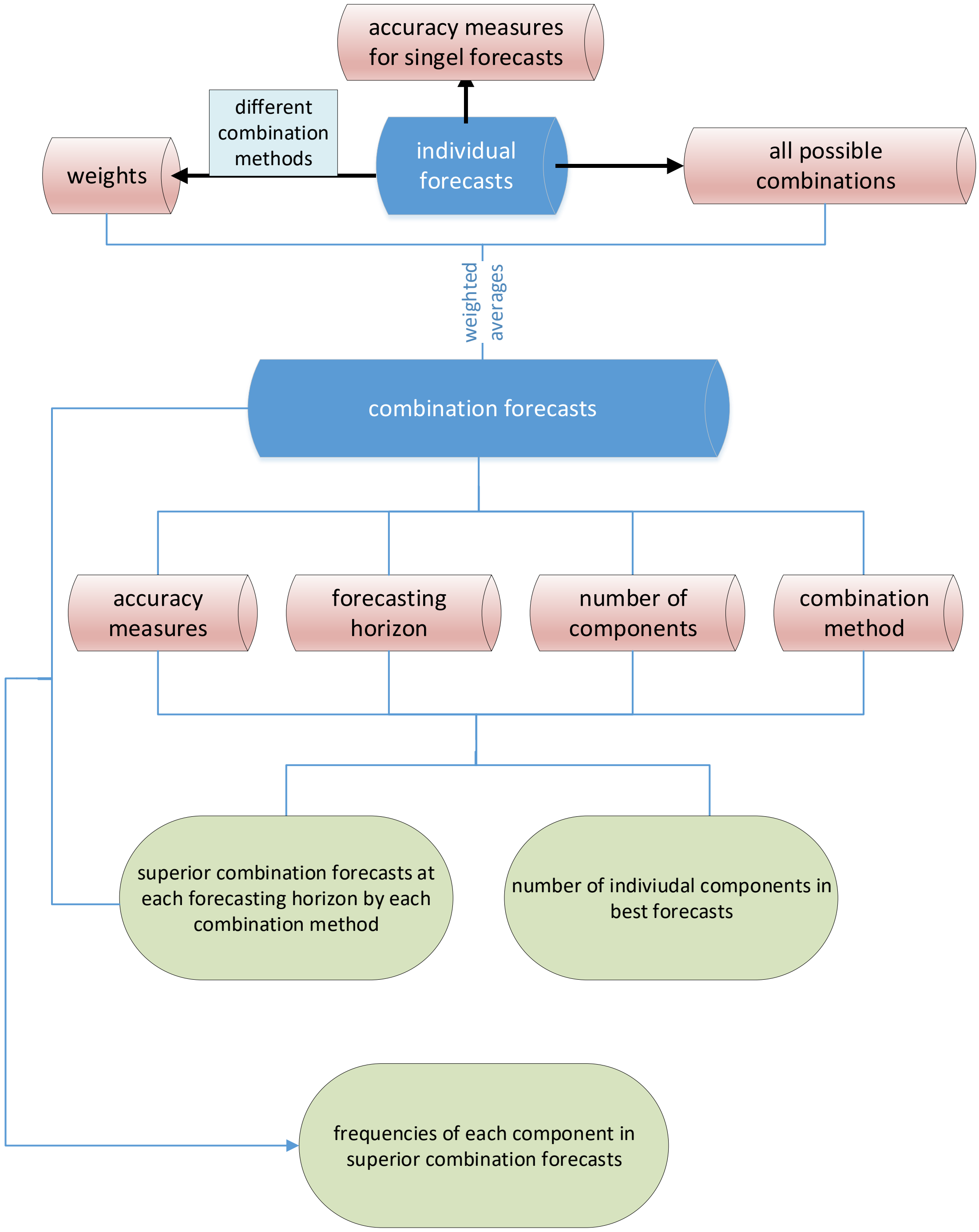

Figure 3-6 Flowchart for The Program for Combination Forecasts

Source: the author

Chapter 4 Results and Discussion: Impact Analysis

4.1 Introduction

This chapter employs the cointegration test to evaluate whether there exist long-run equilibrium relationships among tourism demand and its influencing factors. To evaluate the impact of the climate condition on tourism demand and to assess whether there are cointegration relationships existing for the seven origins when the climate factor is included, the climate econometric models are considered. The identified influencing factors include income, own price, substitute price, the climate factor, which is represented by the UK's TCI, and several one-off events.

Firstly, unit root tests are run to check the integration orders of the variables under study, and the results are presented in section 4.2, which show that the variables are integrated of order zero or one. Since the variables are not integrated of the same order, the ADLM bounds test approach is chosen to test cointegration, and the bounds test results are summarized in section 4.3. When a long-run relationship is supported by the bounds test, the parameters are estimated afterwards. The impacts of different factors on tourism demand are discussed and a cross-country comparison is conducted, which are reported in section 4.4. Section 4.4 also presents the results of the diagnostic tests. Section 4.5 provides a summary of this chapter. The empirical analysis in this chapter is based on data from 1994Q1 to 2017Q4.

4.2 Results of Unit Root Tests

Before estimation, unit root tests are run to test the stationarity of model variables in this study. The objective is to check the integration orders of the variables, and to choose the appropriate cointegration test method. The ADF unit root test (Dickey and Fuller 1981) and the PP unit root test (Phillips and Perron 1988) are commonly used in the tourism demand literature, and the latter one is chosen in this study because of its advantages compared to the ADF test. The PP test, which takes the same estimation procedure as the ADF one, corrects the statistic to allow for autocorrelations and heteroscedasticity, and does not require to select the level of serial correlation as in the ADF test.

The variables under consideration include the dependent and independent variables for the seven origin-destination pairs, i.e., tourist arrivals (lny_t), income ($lnGDP_t$), own price ($lnrp_t$) and substitute price ($lnsp_t$) for the seven markets respectively, as well as the climate condition in the destination ($lnTCI_{UK,t}$). All the variables are tested in both the level and first-differenced forms. The unit root test results are presented in Table 4-1.

Table 4-1 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: The superscript * means significant at the 10% level, ** means significant at the 5% level, and *** means significant at the 1% level.

The unit root test results indicate that model variables are integrated of different orders. The dependent variables (lny_t) and the climate determinant ($lnTCI_{UK,t}$) are stationary at the 1% significance level. Except the income ($lnGDP_t$) for the Italian market, which is stationary at the 1% significance level, all economic determinants under consideration ($lnGDP_t$, $lnrp_t$, $lnsp_t$) are non-stationary and integrated of order one (denoted by $I(1)$) at the 1% significant level.

The EG-ECM, WB-ECM and JML-ECM are inappropriate cointegration test methods, as they all require that all variables under study are integrated of order one. Since no variable is integrated of order two or above at the 5% significant level, which satisfies the assumption of the bound test (Psarian et al. 2001), the ADLM bound test approach is appropriate and valid to generate long-run estimates for all the origin countries under study.

4.3 Results of Bounds Test

An important step in the ADLM bounds test procedure is the choice of the lag order on the first-differenced variables in equation 3.7. This study applies Eviews 10.0 to auto-select the appropriate lag structure for the bounds test models based on AIC, and the maximum lag length is set to be four considering that quarterly data are used in the estimation. The ordering of the variables in the bounds test models is: $lny_t, lnGDP_t, lnrp_{t,} lnsp_t, lnTCI_{UK,t}$. Bounds tests are run for all the models. The lag structure and computed F-statistics are summarized in table 4-2.

Table 4-2 The Lag Structure and The Bounds Test Statistics

	France	Germany	Irish Republic	Italy	The Netherlands	Spain	The US
Lag structure	(4,3,2,1,1)	(4,4,4,34)	(4,2,2,0,0)	(4,0,0,0,4)	(4,3,2,1,1)	(4,3,3,4,4)	(4,2,2,0,1)
F-statistic	5.14***	4.02**	10.40***	5.05***	5.12***	4.83***	4.34**

Notes: The bounds critical values are reported based on Eviews 10.0: at the 10% significance level, the lower bound value is 2.2, and the upper bound value is 3.09; at the 5% significance level, the lower bound value is 2.56, and the upper bound value is 3.49; and at the 1% significance level, the lower bound value is 3.29, and the upper bound value is 4.37.

Table 4-2 shows the divergence of the lag structure across the seven markets, which implies that the dynamic relationships in tourism demand and its influencing factors are different in the seven cases. The statistics of the cointegration tests are higher than the upper critical bound value at the 5% significance level for all origins, which means that we can reject the null hypothesis of 'no cointegration' at the 5% significance level for all markets. In other words, for all seven origins, the test results demonstrate that there exist long-run relationships among tourist arrivals, income, own price, substitute price and the UK's climate condition in the data sample considered.

4.4 Estimation Results and The Impact of Influencing Factors

Since the results support the existence of cointegration relationships among tourism demand and the identified determinants, the long-run coefficients are estimated afterwards. Table 4-3 reports the results of the estimation with tourist arrivals as the dependent variable.

Table 4-3 Estimation Results (Dependent Variable: lny_t)

	France	Germany	Irish Republic	Italy	The Netherlands	Spain	The US
Long-run coefficient estimates							
$lnGDP_t$	1.30*** (8.08)	2.13*** (2.74)	0.07 (0.47)	1.39*** (2.05)	0.18 (0.48)	0.21 (0.11)	1.75* (1.95)
$lnrp_t$	-1.76*** (-3.88)	-1.35** (-2.11)	-0.66** (-2.18)	-0.95 (-1.23)	-1.69*** (-2.70)	-0.42 (-0.43)	-0.32 (-0.61)
$lnsp_t$	7.54** (2.05)	-6.67 (-1.22)	0.12 (0.16)	-5.64*** (-2.90)	-0.46 (-0.17)	-9.45* (-1.70)	1.83*** (3.23)
$lnTCI_{UK,t}$	0.96* (1.83)	1.37 (0.87)	1.34*** (3.06)	1.39*** (2.63)	2.48** (2.14)	7.73** (2.48)	2.15** (2.10)
D_{911}	-0.42*** (-3.30)			-0.63*** (-3.73)	-0.28* (-1.70)		-0.70** (-2.46)
D_{DS}		-0.30*** (-2.09)	-0.22* (-1.81)	-0.38*** (-2.09)			
D_{2008}							-0.45*** (-3.42)
C	0.69 (0.08)	-21.19 (-1.43)	5.86*** (2.20)	-30.22 (-1.51)		-21.63 (-0.99)	-22.27*** (-1.46)
ECM_{t-1}	-0.42*** (-5.20)	-0.32*** (-3.95)	-0.58*** (-8.15)	-0.31*** (-4.21)	-0.42*** (-5.19)	-0.33*** (-5.58)	-0.35*** (-5.37)
Diagnostic tests							
B-G LM	12.80 [0.12]	1.07 [0.89]	6.68 [0.15]	8.48 [0.08]	8.93* [0.07]	3.14 [0.20]	0.76 [0.68]
Harvey	19.89 [0.22]	23.62 [0.48]	11.84 [0.54]	23.70* [0.06]	20.89 [0.40]	28.77 [0.15]	19.04 [0.16]
ARCH	0.31 [0.57]	4.17 [0.99]	1.69 [0.19]	0.03 [0.85]	2.16 [0.14]	0.83 [0.35]	0.03 [0.85]
JB Normality	8.20 [0.11]	6.89* [0.06]	1.57 [0.45]	5.19* [0.07]	6.96 [0.30]	1.26 [0.53]	2.14 [0.14]
RESET	0.07 [0.78]	1.72 [0.19]	0.96 [0.32]	2.27 [0.13]	0.45 [0.50]	4.01* [0.07]	0.29 [0.59]

Notes: Two decimal places are retained for all values. The superscripts (*, ** and ***) denote that the statistical tests are significant at the 10%, 5% and 1% significance levels respectively. The numbers in parentheses are the t-statistics, and the numbers in square brackets are p-values of the tests.

4.4.1 Income

The estimated coefficients of income ($lnGDP_t$) are of expected signs for all markets, which is consistent with demand theory. A comparison across the seven origins show that income affects tourism demand to different extents.

For the French, Germany, Italian and American origins, income is proved to be significant at the 10% significance level, with estimated values all above one, demonstrating that travel to the UK are considered as luxury goods by consumers from these countries. The income elasticities are estimated to be 1.30 for the French market, 2.13 for the German origin, 1.39 for the Italian case and 1.75 for the American states, suggesting that 10% increase in the origin's income will raise tourist arrivals from France by 13%, tourism demand from Germany by 21.3%, traveler flows from Italy by 13.9% and tourist flows from the US by 17.5%. It implies that Germans are the most responsive to income fluctuations among travelers from these four countries.

With regard to tourism demand from Irish Republic, the Netherlands and Spain, income elasticities are found to be statistically insignificant, and the estimated values are between zero and one. The insignificant estimates indicate that residents in these countries do not regard income as an important influencing factor when they are making travel decisions to the UK. The estimated values show that one percent change in income in the home country will lead to less than one percent change in inbound tourism demand to the UK, demonstrating that Irish, Dutch and Spanish holiday makers consider travel to the UK as necessary goods.

The income elasticity estimates show that for four out of seven cases, tourism to UK is considered as a luxury good. However, its influence is not always significant. There are other studies reporting insignificant estimates of income elasticity (Kulendran and King 1997; Kim and Song 1998; Song et al. 2003b; Seetaram et al. 2014). All estimates that are between zero and one are related to European origins, suggesting that travel distance has a role to play in the effect that income has on tourism demand. The empirical results of this study show that some short-haul travellers are income-indifferent and income-inelastic.

4.4.2 Own Price

Own price ($lnrp_t$) is found to have the expected signs for all origins with diverse effects across different cases.

Estimates of own price elasticities are significant at the 5% significance level in four cases: the French, the Germany, the Irish and the Dutch ones, with estimated values to be -1.76, -1.35, -0.66 and -1.69 respectively. The results mean that 10% increase in the own price level will lead to 17.6%, 13.5%, 6.6% and 16.9% reductions in tourist arrivals from France, Germany, Irish Republic and the Netherlands respectively. It suggests that leisure travel to the UK is price-elastic for the French, German and Dutch markets, with French visitors being the most sensitive; and is price-inelastic for the Irish origin.

With respect to tourist arrivals from Italy, Spain and the US, own price is estimated to be statistically insignificant, and the absolute values of the estimates are between zero and one. It shows that own price has insignificant effects on tourism demand from these countries. The estimated values show that one percent change in own price will lead to less than one percent change in tourism demand to the UK from Italy, Spain and the US, demonstrating that tourism demand from these countries are price-inelastic.

The estimate for the American market (-0.32) has the smallest absolute value among the seven cases. Being the only visitors from another continent, the Americans do not consider own price as an important factor when making travel decisions to the UK. One possible explanation is that airfare, which accounts for a large portion in long-haul travellers' expenditure, is excluded from the computation of own price in this study. And long-haul holiday makers are usually more sensitive to the fluctuations in airfare than to that in the relative CPI between the destination and the home country.

4.4.3 Substitute Price

The estimated coefficients of substitute price ($lnsp_t$) present great divergence across different origins. The estimates are statistically significant at the 10% significance level for four cases: the French, Italian, Spanish and American ones; and are insignificant in the other three markets: the German, Irish and Dutch ones.

As with tourism demand from France and the US, substitute price is proved to be significant at the 5% significance level, and the estimates are 7.54 and 1.83 respectively, meaning that 10% increase in the price level in the competing destinations will lead to 75.4% and 18.3% raise in tourist arrivals to the UK from France and the US respectively, which implies that for the French and Americans, leisure travel to the alternative countries are substitute goods to holiday trips to the UK, and French tourists are more responsive to substitute prices changes than American visitors.

For the Italian and Spanish markets, substitute price is estimated to be -5.64 and -9.45 at the 10% significance level respectively, showing that 10% increase in the price level in the alternative destinations will reduce UK inbound tourism demand from Italy by 56.4%, and that from Spain by 94.5%, suggesting that Italians and the Spanish regard the competing countries as complimentary destinations to the UK, and visitors from Spain are extremely sensitive to fluctuations in the price level in the alternative destinations.

With respect to tourism demand from Germany, Irish Republic and the Netherlands, substitute price is estimated to be insignificant with the estimates being -6.67, 0.12 and -0.46 respectively. The insignificant estimates indicate that the price level in the alternative destinations do not have influential effects on tourist arrivals form these countries. The estimate with the smallest absolute value among the seven cases is the one for the Irish origin. The insignificant effect of substitute price on tourism demand from Irish Republic shows the importance of the UK as a tourism destination for the Irish, which can be explained by the close bonds in history, culture and geography between these two countries.

4.4.4 One-off Events

Three one-off events show significant effects. The 9/11 terrorist attack (D_{911}) is found to have important impacts on tourism demand from France, Italy, the Netherlands and the US, with estimated coefficients being -0.42, -0.63, -0.28 and -0.70 respectively. The results show that the attack caused 42%, 63%, 28% and 70% loss in tourist arrivals from these four countries respectively, with the US being the most affected market.

The outbreak of foot-and-mouth disease (D_{DS}) demonstrate influential effects for three European countries: Germany, Irish Republic and Italy, and the induced decrease in tourism demand are 30%, 22% and 38% respectively.

For the American case, the coefficient of the 2008 financial crisis (D_{2008}) is estimated to be -0.45 at the 1% significance level, indicating that the crisis caused 45% reduction in tourist arrivals form the US. For other origins, which are all European countries, the 2008 financial crisis is proved to have no significant impact on inbound tourism demand to the UK, which makes the US is the only affected market.

The terrorist bombing in London on 7th of June 2005 ($\mathrm{D_{BM}}$) and the Olympic Games held in London in 2012 ($\mathrm{D_{OL}}$) are found to have no significant effects on all markets and are excluded from the table.

4.4.5 The Climate Factor

The demand elasticities with regard to the climate factor ($lnTCI_{UK,t}$) are estimated to have positive signs for all markets and are proved to be significant at the 10% significance level with only one exception (the German market). The estimates are 0.96, 1.34, 1.39, 2.48, 7.73 and 2.15 for the French, Irish, Italian, Dutch, Spanish and American cases respectively, demonstrating that better climate conditions in the UK contribute to more tourist arrivals from these countries. The results mean that in the long run, 10% rise in the UK's TCI will induce 9.6%, 13.4%, 13.9%, 24.8%, 77.3% and 21.5% increase in visitor flows from France, Irish Republic, Italy, the Netherlands, Spain and the US respectively, rendering that the Spanish is the most sensitive to the changes in the destination's climate condition. With respect to tourism demand from Germany, the coefficient is estimated to be statistically insignificant with a positive value, suggesting that travellers from Germany do not consider climate as an important tourism demand influencing factor.

The empirical results are in accordance with the assumption, which is that the more favourable the climate condition in the destination is, the more inbound tourism demand is induced. It is also shown that the effect of the climate factor on tourism demand is divergent across seven markets, with the effect being estimated to be insignificant for one origin. As a result, whether the climate factor is an important demand determinant should be evaluated case by case. Excluding the climate factor from the econometric model without empirical validation is not recommended.

4.4.6 The Error Correction Term

The coefficients on the error correction term (EC_{t-1}) represent the adjustment speeds to the equilibrium following shocks to the system. The adjustment speed shows the extent to which the variables have the tendency to converge towards the long-run equilibrium value following shocks. According to Banerjee et al. (1993), a negative and statistically significant error correction term confirms the existence of a long-run cointegration relationship among the variables. The coefficients on EC_{t-1} for the seven markets are all estimated to be significant at the 1% significance level with negative values, reinforcing the findings that long-run relations exist in the system for all cases.

For example, the coefficient estimates for the Irish and Spanish markets are -0.58, -0.33 respectively, showing that if all other factors remain unchanging, 58% of tourist arrivals disequilibrium of the Irish origin and 33% of tourism demand disequilibrium of the Spanish case are corrected in one quarter. Interpretatively, it requires about two quarters for the tourism demand from Irish Republic, and about three seasons of the tourism demand from Spain to restore equilibrium, suggesting that the Irish market regains the long-run equilibrium more quickly following shocks.

4.4.7 Diagnostic Tests

Several diagnostic tests including the B-G LM test for serial correlation, the Harvey test for heteroscedasticity, the ARCH test, the J-B test for non-normality and the Ramsey RESET for misspecification are run after estimation. Except the Ramsey RESET test, all the statistics are chi-square statistics. The Ramsey RESET statistic is an F statistic.

The results are summarized in Table 4-3, with the p values of the tests being reported in the table. The results show that a few tests are passed at the 10% significance level: the B-G LM test for the Dutch market, the Harvey test for the Italian origin, the J-B test for the Germany and Italian models, and the RESET test for the Spanish case. All the other tests are passed by all models at the 5% significance level, demonstrating that the models are well specified and are qualified for forecasting.

4.5 Summary

In this chapter, cointegration analysis is conducted for seven origin-destination pairs to reveal the long-run relationships between tourism demand and its influencing factors, which include income, own price, substitute price and the destination's climate condition. Based on the integration orders of the model variables, the bounds test approach is chosen to test cointegration, and the results confirm the existence of long-run relationships between tourism demand and the explanatory variables. The estimated coefficients for the economic factors and the climate factor are all of expected signs, which is in accordance with demand theory and supports the assumptions of this study. The effects of different factors are diverse across seven cases.

For the French market, all the explanatory variables are found to be statistically significant at the 10% significance level, with the UK's TCI being the least important factor. For the Italian and American cases, own price is the only factor that has insignificant effect, while income, substitute price and the climate factor all have crucial impacts on tourism demand. For the German origin, income and own price are the determinants that are proved to have significant effects on tourism demand, and the climate factor is estimated to be insignificant with a positive value.

Income is found to be insignificant in three European cases: the Irish, the Dutch and the Spanish. With respect to tourism demand from Irish Republic and the Netherlands, own price and the climate factor are the ones that have influential effects; while for the Spanish market, the climate factor is the most crucial influencing factor, being significant at the 5% significance level, and substitute price is the only economic factor that has important effect.

The ranges of demand elasticity estimates are different from the results reported by other studies. According to Cortes-Jimenez and Blake (2011), who evaluated the drivers of inbound tourism demand to UK from the same seven source markets, for leisure travel, the income elasticity estimates ranged from 1.37 to 2.10; the own price elasticities were estimated to be from -0.14 to -2.48; and the elasticities with regard to substitute price were from -7.01 to 4.15. Through a meta-regression analysis of 195 studies published from 1961-2011, Peng et al. (2015) showed that origin, destination, time period, modelling method, data frequency, the inclusion/omission of other explanatory variables and their measures, and sample size all significantly influence the demand elasticities generated by a model.

The significant one-off events include the 9/11 terrorist attack for the French, Italian, Dutch and American origins, the outbreak of foot-and-mouth disease for tourism demand from Germany, Irish Republic and Italy, and the 2008 financial crisis for tourist arrivals from the US. Spain is the only country that does not affected by any of the events under consideration.

The climate condition in the destination, which is proxied by TCI in the UK, is proved to have significant effects on tourism demand in six out of seven cases, with estimates ranging from 0.96 to 7.73. As with the Irish and Spanish origins, the climate factor is the most important influencing factor. Only tourist arrivals from Germany is found to be not influenced by the destination's climate condition. The empirical results demonstrate that the climate factor has an important role to play in tourism demand and should be considered in the modelling process.

Chapter 5 Results and Discussion: Forecasts Comparison

5.1 Introduction

This chapter presents the core results of this study, which are the performances of various forecasts and comparisons among them. In addition, in-depth analyses are provided to answer key research questions.

The rest of this chapter is organized into five sections. Section 5.2 is about the forecasting performance of different individual models including traditional econometric models, climate econometric models and time series models, which are measured for seven origins at one-to-four forecasting horizons by MAE, MAPE and RMSE respectively. The seasonal naïve no-change forecasts are used as forecasting benchmarks. The comparisons are conducted in three ways: the country-specific assessment for seven markets, the forecasting-horizon-specific evaluation for four forecasting lengths and the general comparison for the overall performance of each individual model, and the results are shown in section 5.2.1, 5.2.2 and 5.2.3 respectively. The MAE, MAPE and RMSE of each individual model are normalized relative to the seasonal naïve no-change forecast. The ratios which are below one indicate that there are predictive gains compared to the no-change forecast. An analysis on whether including the climate determinant can improve the forecasting performance of econometric models is provided in section 5.2.4.

Section 5.3 compares the forecasting ability between the combination forecasting approach and the individual forecasting approach. Except the seasonal naïve no-change model, all individual forecasting models serve as candidate components for combination. There are three groups of individual models to be combined which are studied respectively: all econometric and time series models; traditional econometric and time series models; and climate econometric and time series models. The results for these three combination groups are presented in section 5.3.1, 5.3.2 and 5.3.3 respectively. The combination methods under consideration include the SA, VACO, DMSFE methods, as well as the newly-introduced inverse-MAE and two-stage combination methods. For each combination group, the percentages of the superior combination forecasts compared to the best single ones are reported and compared in three ways: the country-specific assessment across seven origins, the forecasting-horizon-specific evaluation over four forecasting lengths, and the general comparison for the overall performance of each combination method. An evaluation on which combination group provides the most accurate composite forecasts is provided in section 5.3.4, where whether including econometric models with different explanatory variables in combination can improve the forecasting performance is assessed.

Section 5.4 and section 5.5 expand the study to answer the following questions: which models to combine and how many models to combine? In section 5.4, the frequencies of each individual component in the superior combination forecasts are summarized and compared for each origin. Firstly, the frequencies are reported for each combination method respectively, and then the general chances of each individual model for each market are provided by taking averages of across 12 combination methods. In addition, an assessment is conducted to reveal whether more accurate individual forecasts have higher chances to construct the superior combination forecasts.

In section 5.5, the most accurate forecasts among all individual and combination candidates are identified for every comparison and each origin, and the numbers of components, which can range from 1 to 15 are reported.

A conclusion of this chapter is provided in section 5.6.

5.2 Forecasting Performance of Individual Models

As illustrated in chapter 3, the seasonal naïve no-change model is selected as the forecasting benchmark. Other individual forecasting models, which are considered as constituent candidates for combination, fall into three categories: traditional econometric models, climate econometric models and time series models. Econometric modelling techniques considered include the bounds test cointegration approach, the ADLM, LI, VAR, TVP and SD models; and time series models consist of the SARIMA, ETS and state space ETS models, resulting in 15 individual forecasting models. The difference between traditional and climate econometric models lies in whether the climate factor is considered as a demand determinant and included as an explanatory variable.

All individual forecasts are generated by Eviews 10.0. In view of the utilization of quarterly data, the maximum lag length is set to be four for the bounds test cointegration approach, the ADLM, LI and VAR models. The general-to-specific modeling approach recommended by Song, Witt and Li (2009) is adopted to achieve the final ADLM. The insignificant variables judged by the t values are eliminated one by one starting with the most insignificant ones until all remaining coefficients are statistically significant. The Block Exogeneity Wald Tests are run for the VAR models, and the variables that are found to be exogenous are excluded from the system before forecasting. The state variables (parameters) in the TVP models are firstly assumed to follow a random walk process, and if the estimated residuals are insignificant, the parameters are then reduced to be time-invariant coefficients. The initial values are specified based on the estimated results of OLS, and all state space models achieve convergence.

The automatic ARIMA forecasting function provided by Eviews 10.0 is used to generate forecasts from the SARIMA models. The maximum orders for the AR and MA parts are set to be four and the maximum orders for the SAR and SMA parts are set to be one. And the specific orders are selected based on AIC. The types of the error, trend and seasonal components in the state space ETS models are automatically selected based on AIC, according to which the smoothing methods of the ETS models are chosen.

The ratios of the accuracy measures including MAE, MAPE and RMSE of the 15 individual models compared to those of the seasonal naïve no-change model are provided for each case and the comparisons are conducted in three ways. Firstly, the forecasting performance is evaluated for different origins, which is a market-specific assessment and reveals which individual models perform better for forecasting tourism demand from a particular market. Secondly, the forecasting ability is assessed for different forecasting horizons, which is a forecasting-horizon-specific evaluation and demonstrates which individual models are superior for forecasting tourism demand from seven leading markets to the UK at a specific forecasting length. And lastly, a general comparison is made by taking averages across all origins and over all forecasting horizons, which shows the performance of different individual forecasting models when general tourism demand to the UK is concerned.

5.2.1 Comparison for Different Origins

Firstly, performance of different forecasting models is examined at a disaggregated level at each forecasting horizon for each origin (see table 5-1), and then at an aggregated level by taking averages over all time horizons for each origin (see table 5-2). As shown in table 5-1 and table 5-2 which highlight the ranks of the top-three performing models for each comparison, no single model is superior to others across all markets or over all forecasting horizons irrespective of which accuracy measure is used. Table 5-2 shows that for the French, Italian and Dutch cases, most ratios are below one, which means that most selected individual models are superior in forecasting performance compared to the benchmark model. And for other origins, all ratios are below one indicating that all selected single models can forecast better than the seasonal naïve no-change model.

For each origin, individual models' performance shows a degree of consistency across different forecasting horizons (see table 5-1). For instance, for forecasting tourism demand from France, the climate bounds test model is superior to others for most forecasting horizons irrespective of accuracy measure. Similarly, the climate LI model outperforms the others for one- to four-step-forecasts as with projecting tourist arrivals from Irish Republic. For the Spanish case, the traditional TVP model ranks number one for 11 out of 12 cases based on three accuracy measures.

On the other hand, individual models' forecasting performance varies greatly according to the origin market under consideration. Some models may perform very well for one origin, but they cannot generate accurate forecasts for others. As shown in table 5-2, when forecasting accuracy is judged by MAE, the finest three models for the case of France are the climate bounds test, SARIMA and state ETS models, and the worst three models are the climate LI, climate SD and traditional LI models. For tourism demand from Germany, the ETS, state ETS and traditional VAR models generate the most accurate forecasts, and the climate LI, climate VAR models and climate ADLM are at the bottom of the list. The climate LI model, however, shows superior forecasting ability for the Irish origin, ranking number one, followed by the climate SD and traditional TVP models. The poorest performance is seen from the traditional ADLM.

With respect to tourism demand from Italy and the Netherlands, the ETS and state space ETS models are hard to beat. For the Italian case, the climate ADLM model remains at the third place, followed by the SARIMA model; and the least accurate one is the traditional LI model, which demonstrates satisfactory forecasting power for the Dutch market, ranking number four after the SARIMA model. However, its climate counterpart, i.e., the climate LI model is at the bottom of the list. When forecasting tourist arrivals from Spain, the traditional TVP model generates the most accurate forecasts, followed by the state space ETS and SARIMA model. The least accurate forecasts are generated from the climate SD and traditional LI model. For the case of the US, the top performing models are the climate ADLM, SARIMA and ETS models, while the climate LI model is the worst.

As far as different accuracy measures are concerned, it is demonstrated that the forecasting performance of individual models shows a degree of consistency across three accuracy measures for all origins (see table 5-2). Take the rankings judged by MAE and MAPE for example, for the French, German, Irish, Italian and Spanish cases, the top-four and the least accurate models in terms of MAPE are the same as with those measured by MAE. For the German case, the traditional LI model, which ranks number four judged by MAE, replaced the traditional VAR model to gain the third place measured by MAPE. When forecasting tourism demand from Spain, the third place is taken by the climate ADLM judged by MAPE, which is the fourth in the list measured by MAE. With respect to the Dutch and American markets, the bottom-three performing models remain the same, while the top-three performing ones vary. For tourist arrivals from the Netherlands, the best three models are the ETS, state space ETS and SARIMA models measured by MAE, and the SARIMA model is replaced by the traditional LI model judged by MAPE. For the American market, the top three places are gained by the climate ADLM, SARIMA and ETS models according to MAE, and by the SARIMA, traditional ADLM and traditional SD models based on MAPE.

Generally, although the individual rankings of different models do change across three accuracy measures in some cases, the models that remain in the top half list do not change with only a few exceptions. For instance, for the Italian market, the traditional TVP model gains the sixth place measured by RMSE, and it drops to the ninth place judged by MAPE; and with forecasting tourism demand from the Netherlands, the same model is the seventh in the list in terms of MAE and ranks number twelve considering MAPE. With respect to tourism demand from Irish Republic, the climate VAR model is the fifth best in terms of RMSE, which drops to the twelfth place judged by MAPE. For the American origin, the climate SD model is the eleventh in position judged by MAE, and it becomes the seventh best when RMSE is applied.

The market-specific evaluation shows that the forecasting ability of different individual models mainly depend on the origin-destination pair under study. For one market, one individual model can provide consistent performance for one- to four-step-ahead forecasts, and the three accuracy criteria of MAE, MAPE and RMSE yield similar assessments. It implies that the nature of data, i.e., the DGP decides the performance of different modelling techniques. And the DGP varies greatly among the seven markets under consideration.

For most cases, the time series techniques forecast better than the econometric models. For example, based on MAE, at least two time series models gain positions among the top-three performing forecasters for all origins except the Irish one, which demonstrates that time series models have strong abilities when forecasting tourism demand from France, Germany, Italy, the Netherlands, Spain and the US to the UK. It also implies that, for these six markets, the historical pattern of the tourism demand variable captures its intrinsic evolution well enough to generate accurate forecasts. For tourist arrivals from Irish Republic, time series techniques are outperformed by four climate econometric and two traditional econometric models. The leading indicator specification including the climate determinant generates the most accurate forecasts, showing that the historical values of economic and climate explanatory variables explain tourism demand from Irish Republic well.

Based on three accuracy measures, at least one of the climate econometric models gains a place in the top five list for all markets except for the German one: the climate bounds test model for the French case, the climate LI model for the Irish market, and the climate ADLM for the Italian, Dutch, Spanish and American origins. The poor performance of the climate econometric models for the German case is mainly due to the fact that the climate factor is proved to have insignificant effect on tourism demand from Germany, which is shown in table 4-3. The climate specification, which includes the climate factor as an explanatory variable is unlikely to forecast well, and the traditional specification is appropriate to reflect tourism demand from Germany. On the other hand, for other origins, it shows that the climate variable can facilitate forecasting tourism demand, and the DGP decides which type of econometric modelling technique is appropriate. It is recommended that the climate factor should be included when forecasting tourism demand on the condition that it is a significant influencing factor.

5.2.2 Comparison for Different Forecasting Horizons

Table 5-3 presents the forecasting-horizon-specific accuracy assessment for four forecasting horizons measured by MAE, MAPE and RMSE respectively, which are obtained by taking averages across seven origin countries, and the ratios of the top-three performing models for each comparison are highlighted. It is demonstrated that all ratios are below one based on MAE and MAPE, and most ratios are below one according to RMSE. It means that most selected individual models can generate more accurate forecasts than the benchmark model as far as these four forecasting horizons are concerned.

As with the relative forecasting performance of different selected individual models, it is clear that time series models are hard to beat for the one- to four-quarter-ahead forecasts regardless of which criterion is used to judge the forecasting performance. The best-performing and most consistent time series technique is the state space ETS model, which ranks number one for eight out of twelve comparisons. The SARIMA technique outperforms the ETS model when the forecasting length is one quarter or two quarters, while it is beaten by the ETS model as far as the three- and four-step-ahead forecasts are concerned.

The only chance of the econometric models to gain a place among the top-three performing forecasters is seen from the four-step-ahead forecasts: the traditional TVP model is the third in place when the accuracy is measured by MAPE.

The least accurate forecasts are generated by the climate LI model irrespective of forecasting length or accuracy measure. Its traditional counterpart, i.e., the traditional LI model does not perform well either, remaining in the bottom half of the list for all comparisons. It indicates that the LI model, which ignores the effect of lagged dependent variable is not a good choice for one- to four-step-ahead forecasts no matter whether the climate factor is considered as a determinant.

For one- and two-step-ahead forecasts, the best performing econometric model is the climate ADLM, which is the fourth in position regardless of accuracy measures; and for three- and four-step-ahead forecasts, the traditional VAR and traditional TVP models are superior compared to other econometric methods. The forecasting performance of the traditional TVP models is satisfactory except for the one-step-ahead projections. On the other hand, the climate VAR and climate TVP models always perform poorly.

In general, the three accuracy measures provide consistent results regarding individual models' rankings in terms of their forecasting performance with only a few exceptions. For example, the traditional ADLM model is ranked much lower according to MAPE than based on MAE and RMSE when one-step-ahead forecasts are concerned. And the climate TVP model achieves much higher position judged by RMSE than measured by MAE and MAPE as with four-step-ahead forecasts.

The forecasting-horizon-specific assessment shows that for projecting quarterly international tourism demand to the UK, the state space ETS model is the best choice for one- to four-step-ahead forecasts. As with the econometric models, the climate ADLM technique generates good forecasts if the forecasting length is one quarter or two quarters, and the traditional VAR and traditional TVP models are the most accurate ones when the forecasting horizon is three or four seasons.

5.2.3 General Comparison among Various Individual Models

The overall forecasting performance of various individual models, which is obtained by averaging across seven markets and over four forecasting horizons is summarized in Table 5-4, and the ratios of the top-three performing models of each comparison are highlighted. It is shown that according to MAE and MAPE, all ratios are below one indicating that generally all selected models are superior than the benchmark model. And based on RMSE, only the ratio of the climate LI model is above one, meaning that except the climate LI model, all selected models forecast better than the benchmark in general.

All accuracy measures consistently show that the state space ETS model behaves the best, followed by the SARIMA and the ETS models; while the climate LI model provides the poorest forecasts. The climate ADLM outperforms other econometric models, and the traditional VAR and traditional TVP models are the next best econometric techniques. Except the bounds test model and the ADLM, the traditional specification beats the climate one in terms of generating more accurate forecasts.

The forecasting superiority of the non-causal time series models compared to the causal econometric techniques is in line with the empirical findings of some existing studies (e.g. Witt and Witt 1995; Kulendran and King 1997; Athanasopoulos et al. 2011), but it contrasts with the empirical results of others (e.g. Fildes et al. 2011; Song et al. 2011a; Gunter and Önder 2015).

5.2.4 Whether Including Climate Variables Can Improve The Forecasting Ability of Econometric Models?

In this study, different econometric techniques are applied with two different model specifications: with and without the climate determinant. To evaluate whether introducing the climate factor is beneficial, the forecasting accuracy measures of traditional econometric models are compared to those of their climate counterparts for each modelling technique and each origin, which are shown in table 5-5 with the better-performing specifications being highlighted.

There is no clear-cut evidence showing which model specification is superior in terms of generating more accurate forecasts, and the relative forecasting performance of the climate and traditional econometric models depends on the origin country under consideration, the accuracy measure used as well as the econometric technique applied.

For the French, German, Dutch, Spanish and American markets, the traditional econometric models generally perform better than the climate ones. As far as forecasting tourism demand from France is concerned, five out of six traditional econometric models perform better, and only the climate bounds test cointegration technique outperforms its traditional counterpart. For the German case, all techniques with the traditional specification generate more accurate forecasts. For the Dutch origin, the climate ADLM, climate VAR and climate SD models forecast more accurately based on MAE and MAPE. But judged by RMSE, the traditional VAR model without the climate determinant beats its climate rival. As with the Spanish market, better climate econometric models include the climate ADLM and the climate LI model when MAE and MAPE are used as accuracy measures; and judged by RMSE, the climate bounds test technique is added to the list. Regarding tourist arrivals from the US, the results are less consistent when different measures are used. The climate bounds test, climate ADLM and climate VAR models are superior judged by MAE, only the climate bounds test technique beats its traditional counterpart in terms of MAPE, and the climate bounds test and climate ADLM generate better forecasts evaluated by RMSE.

As with forecasting tourism demand from Irish Republic and Italy, the climate econometric models generally perform better than the traditional ones. For the Irish case, four out of six climate models including the climate bounds test, climate ADLM, climate LI and climate SD models are superior no matter which accuracy measure is used. For the Italian market, the climate ADLM, climate LI and climate VAR models outperform the traditional ones based on MAE; when judged by MAPE, the above-mentioned three climate models and the climate TVP model are superior; and based on RMSE, the climate ADLM and climate LI model outperform their traditional counterparts.

To conduct a country-specific evaluation at the aggregate level, the forecasting performance of climate and traditional econometric models are compared for each origin by taking averages over six econometric modelling techniques, and the results are presented in table 5-6 with the superior specifications highlighted. It shows that the results are consistent across three accuracy measures for all markets. To forecast tourism demand from Irish Republic and Italy, the climate specification is superior compared to the traditional one, while for projecting tourist arrivals from France, Germany, the Netherlands, Spain and the US, traditional econometric models are better.

A general comparison is also made by taking averages over seven origins, and the results are summarized in table 5-7 with the better specifications highlighted. It demonstrates that econometric models with traditional specification generally perform better.

The relatively sound performance of the climate specification for forecasting tourism demand from Irish Republic and Italy is supported by the estimation results that the climate determinant has significant effect on tourism demand, which is presented in section 4.4. As shown in table 4-3, the climate variable is found to have significant effect on tourism demand at the 1% significance level for the Irish and Italian cases. On the other hand, for the German market, it is demonstrated that there exists no significant long-run effect of the climate determinant on tourism demand, which explains the poor performance of the climate econometric models when forecasting tourism demand from Germany. It is recommended that for country-specific studies, the long-run relationship between the climate factor and tourism demand should be evaluated firstly. If the impact of the climate factor on tourism demand is found to be significant, the climate determinant should be included when forecasting tourism demand, which can contribute to more accurate forecasts.

In addition, it is important to consider the market differences among origin countries when forecasting tourism demand. Whether the destination's climate condition is an important influencing factor of tourism demand is fundamentally determined by tourists' attitudes towards climate attributes in their decision-making process. It is reasonable to assume that tourists from different origins have different opinions, which is confirmed by the empirical results presented in chapter 4 of this study. Therefore, there is no general answer as to whether including the climate factor can improve the forecasting ability of econometric models, which depends on the market under study

Table 5-1 Forecasting Performance of Individual Models at Different Forecasting Horizons across Seven Origins

origin	step	bounds©	ETS	SARIMA	ADLM©	State ETS	LI©	VAR©	TVP©	SD©	bounds	ADLM	LI	VAR	TVP	SD
MAE ratios																
France	1	0.633 1	0.891 12	0.719 2	0.840 9	0.730 3	1.514 15	0.822 8	0.906 13	0.993 14	0.780 6	0.754 4	0.852 11	0.773 5	0.809 7	0.851 10
	2	0.697 2	0.917 10	0.690 1	0.866 8	0.735 3	1.410 15	0.974 13	0.946 12	1.014 14	0.797 7	0.784 5	0.918 11	0.778 4	0.786 6	0.878 9
	3	0.643 1	0.830 9	0.691 3	0.822 8	0.707 4	1.284 15	0.929 12	0.862 10	0.989 14	0.772 6	0.783 7	0.959 13	0.659 2	0.763 5	0.876 11
	4	0.617 1	0.884 10	0.688 3	0.863 8	0.684 2	1.313 15	0.939 12	0.864 9	1.039 14	0.787 6	0.817 7	1.015 13	0.759 4	0.777 5	0.900 11
Germany	1	0.700 12	0.478 2	0.682 10	0.775 13	0.461 1	0.940 15	0.847 14	0.656 9	0.687 11	0.563 5	0.616 7	0.503 4	0.503 3	0.637 8	0.579 6
	2	0.649 7	0.487 1	0.697 11	0.823 13	0.502 2	0.976 15	0.829 14	0.706 12	0.686 10	0.574 4	0.653 8	0.525 3	0.584 5	0.658 9	0.601 6
	3	0.599 7	0.461 2	0.712 11	0.823 13	0.455 1	0.971 15	0.888 14	0.753 12	0.659 9	0.563 5	0.679 10	0.555 4	0.524 3	0.638 8	0.595 6
	4	0.636 7	0.495 1	0.788 11	0.846 13	0.539 3	0.950 14	0.992 15	0.794 12	0.685 9	0.580 5	0.704 10	0.576 4	0.515 2	0.664 8	0.601 6
Irish Republic	1	0.553 4	0.604 7	0.638 9	0.659 11	0.600 6	0.406 1	0.620 8	0.708 13	0.476 2	0.742 14	0.816 15	0.653 10	0.557 5	0.547 3	0.669 12
	2	0.592 6	0.621 9	0.579 5	0.722 13	0.606 7	0.408 1	0.617 8	0.707 12	0.490 2	0.764 14	0.883 15	0.623 10	0.575 4	0.575 3	0.667 11
	3	0.552 3	0.668 9	0.643 7	0.785 14	0.661 8	0.436 1	0.627 6	0.720 12	0.493 2	0.781 13	0.954 15	0.685 11	0.564 5	0.560 4	0.681 10

	4	0.572 5	0.620 7	0.722 13	0.817 14	0.649 10	0.407 1	0.604 6	0.709 11	0.484 2	0.717 12	0.969 15	0.638 9	0.562 3	0.571 4	0.638 8
Italy	1	0.838 10	0.648 3	0.672 4	0.620 1	0.626 2	0.751 5	0.780 8	0.852 11	0.886 13	0.762 7	0.760 6	1.196 15	0.912 14	0.877 12	0.819 9
	2	0.871 11	0.617 1	0.671 4	0.655 3	0.627 2	0.784 6	0.796 7	0.931 13	1.044 14	0.824 8	0.784 5	1.262 15	0.916 12	0.865 9	0.868 10
	3	0.873 10	0.607 2	0.708 4	0.701 3	0.603 1	0.805 6	0.804 5	0.867 9	1.000 14	0.843 8	0.805 7	1.292 15	0.929 13	0.876 11	0.895 12
	4	0.900 13	0.605 2	0.657 3	0.698 4	0.566 1	0.844 10	0.754 5	0.802 6	1.022 14	0.816 7	0.828 9	1.330 15	0.861 11	0.821 8	0.877 12
the Netherlands	1	0.944 11	0.747 4	0.715 1	0.771 5	0.725 3	1.133 14	0.907 9	1.358 15	0.954 12	0.900 8	0.862 6	0.717 2	0.908 10	0.870 7	1.000 13
	2	1.009 13	0.696 1	0.776 3	0.787 5	0.714 2	1.133 15	0.953 11	0.933 10	0.997 12	0.926 9	0.876 6	0.776 4	0.920 8	0.885 7	1.046 14
	3	0.998 12	0.768 3	0.739 1	0.804 5	0.746 2	1.130 15	0.932 8	0.923 7	1.002 13	0.970 10	0.892 6	0.780 4	0.985 11	0.955 9	1.062 14
	4	0.993 11	0.660 1	0.663 2	0.838 5	0.689 3	1.116 15	0.965 9	0.946 8	1.039 13	0.929 6	0.935 7	0.709 4	1.020 12	0.980 10	1.045 14
Spain	1	0.831 13	0.714 7	0.624 3	0.655 4	0.583 2	0.728 10	0.661 5	0.701 6	0.837 14	0.780 11	0.780 12	0.868 15	0.719 8	0.583 1	0.724 9
	2	0.856 14	0.782 11	0.646 4	0.640 2	0.643 3	0.721 6	0.764 9	0.772 10	0.925 15	0.735 8	0.804 12	0.853 13	0.724 7	0.564 1	0.712 5
	3	0.877 14	0.783 10	0.629 2	0.702 5	0.646 3	0.749 7	0.777 8	0.779 9	0.926 15	0.809 11	0.829 12	0.874 13	0.717 6	0.551 1	0.693 4
	4	0.874 14	0.778 10	0.664 3	0.678 4	0.601 2	0.776 8	0.777 9	0.771 7	0.871 13	0.862 11	0.867 12	0.904 15	0.695 5	0.545 1	0.717 6

US	1	0.752 11	0.644 4	0.395 1	0.561 2	0.745 9	0.907 15	0.692 6	0.741 7	0.876 14	0.753 12	0.620 3	0.743 8	0.749 10	0.849 13	0.662 5
	2	0.971 14	0.661 4	0.672 5	0.607 1	0.618 2	0.969 13	0.998 15	0.958 12	0.846 8	0.890 11	0.650 3	0.782 7	0.874 9	0.882 10	0.685 6
	3	0.898 13	0.619 2	0.685 4	0.614 1	0.829 8	0.990 15	0.816 7	0.917 14	0.847 11	0.852 12	0.680 3	0.830 9	0.837 10	0.809 6	0.728 5
	4	0.797 10	0.667 2	0.683 4	0.635 1	0.670 3	1.038 15	0.733 6	0.838 12	0.774 8	0.987 14	0.692 5	0.850 13	0.799 11	0.781 9	0.769 7

MAPE ratios

France	1	0.662 1	0.966 12	0.801 3	0.829 6	0.815 4	1.568 15	0.836 7	0.967 13	0.993 14	0.840 9	0.768 2	0.966 11	0.824 5	0.920 10	0.840 8
	2	0.717 1	0.961 10	0.733 2	0.846 6	0.780 3	1.429 15	1.002 12	0.998 11	1.025 14	0.851 7	0.793 4	1.006 13	0.816 5	0.881 9	0.858 8
	3	0.690 1	0.933 11	0.749 3	0.821 6	0.771 4	1.317 15	0.984 12	0.929 10	0.994 13	0.857 7	0.803 5	1.037 14	0.697 2	0.882 9	0.866 8
	4	0.666 1	0.956 11	0.728 2	0.862 6	0.744 3	1.368 15	0.997 12	0.932 10	1.017 13	0.880 7	0.837 5	1.106 14	0.818 4	0.903 9	0.890 8
Germany	1	0.645 11	0.441 2	0.636 10	0.713 13	0.437 1	0.996 15	0.754 14	0.567 9	0.650 12	0.510 6	0.545 8	0.447 3	0.466 4	0.537 7	0.504 5
	2	0.598 9	0.433 1	0.604 12	0.752 14	0.443 2	1.018 15	0.724 13	0.601 10	0.604 11	0.516 4	0.577 8	0.459 3	0.541 6	0.544 7	0.518 5
	3	0.558 8	0.421 2	0.620 11	0.757 13	0.414 1	1.026 15	0.758 14	0.638 12	0.587 9	0.499 4	0.601 10	0.486 3	0.507 5	0.531 7	0.518 6
	4	0.588 8	0.459 1	0.733 12	0.785 13	0.502 3	1.032 15	0.845 14	0.674 11	0.617 9	0.520 5	0.627 10	0.506 4	0.488 2	0.554 7	0.529 6

Irish Republic	1	0.567 6	0.561 5	0.640 9	0.806 13	0.568 7	0.341 1	0.743 12	0.929 14	0.432 2	0.683 11	1.074 15	0.668 10	0.524 4	0.524 3	0.624 8
	2	0.635 6	0.753 12	0.711 9	0.815 13	0.727 10	0.338 1	0.749 11	0.924 14	0.438 2	0.690 8	1.099 15	0.663 7	0.537 4	0.537 3	0.623 5
	3	0.485 3	0.707 9	0.738 11	0.823 13	0.701 8	0.339 1	0.772 12	0.965 14	0.471 2	0.722 10	1.133 15	0.663 7	0.536 4	0.544 5	0.645 6
	4	0.623 5	0.738 9	0.911 13	0.839 12	0.796 11	0.336 1	0.752 10	0.938 14	0.474 2	0.675 8	1.129 15	0.657 7	0.540 3	0.548 4	0.633 6
Italy	1	0.930 11	0.708 3	0.725 4	0.674 1	0.677 2	0.846 8	0.833 7	0.911 10	0.963 12	0.819 6	0.815 5	1.275 15	1.011 14	0.973 13	0.896 9
	2	0.945 10	0.678 2	0.715 4	0.709 3	0.676 1	0.875 7	0.844 6	0.979 12	1.133 14	0.883 8	0.831 5	1.348 15	0.997 13	0.940 9	0.951 11
	3	0.942 10	0.655 2	0.753 3	0.758 4	0.630 1	0.898 8	0.832 5	0.899 9	1.070 14	0.896 7	0.850 6	1.375 15	1.006 13	0.946 11	0.976 12
	4	1.002 13	0.661 2	0.715 3	0.768 4	0.600 1	0.946 10	0.801 5	0.856 6	1.136 14	0.890 8	0.884 7	1.426 15	0.960 11	0.912 9	0.978 12
the Netherlands	1	0.967 9	0.872 6	0.798 2	0.825 3	0.829 4	1.352 14	0.948 8	1.456 15	0.979 10	0.940 7	0.862 5	0.787 1	1.002 12	0.997 11	1.036 13
	2	1.006 12	0.812 2	0.915 6	0.819 3	0.799 1	1.322 15	0.982 8	1.005 11	1.012 13	0.934 7	0.844 5	0.836 4	0.987 9	0.987 10	1.063 14
	3	0.984 9	0.887 5	0.907 6	0.840 2	0.833 1	1.344 15	0.965 7	1.008 10	1.015 11	0.981 8	0.859 3	0.860 4	1.053 12	1.064 13	1.074 14
	4	0.985 8	0.760 3	0.740 1	0.876 5	0.751 2	1.314 15	0.999 9	1.033 10	1.052 11	0.953 7	0.901 6	0.772 4	1.095 14	1.093 13	1.065 12
Spain	1	0.966 13	0.859 8	0.783 5	0.779 4	0.712 1	0.806 6	0.772 2	0.814 7	0.970 14	0.944 12	0.922 11	0.970 15	0.898 10	0.776 3	0.893 9

	2	0.979 14	0.883 11	0.758 3	0.705 2	0.763 4	0.788 5	0.881 10	0.880 9	1.075 15	0.866 8	0.906 12	0.926 13	0.835 6	0.687 1	0.849 7
	3	0.976 14	0.888 10	0.713 2	0.758 3	0.791 4	0.825 7	0.876 9	0.865 8	1.056 15	0.916 11	0.927 12	0.958 13	0.815 6	0.661 1	0.809 5
	4	0.979 12	0.862 8	0.744 4	0.740 3	0.679 2	0.855 7	0.895 10	0.873 9	1.017 15	1.010 14	0.963 11	0.991 13	0.774 5	0.648 1	0.837 6
US	1	0.742 10	0.699 6	0.452 1	0.629 3	0.746 11	0.842 14	0.698 5	0.719 9	0.890 15	0.761 12	0.628 2	0.716 8	0.711 7	0.772 13	0.658 4
	2	0.935 13	0.749 7	0.673 4	0.670 3	0.683 5	0.863 11	0.975 15	0.941 14	0.806 8	0.876 12	0.637 2	0.711 6	0.816 9	0.854 10	0.636 1
	3	0.889 14	0.667 3	0.693 5	0.679 4	0.804 10	0.873 13	0.811 11	0.920 15	0.798 8	0.852 12	0.661 1	0.749 6	0.799 9	0.750 7	0.662 2
	4	0.786 12	0.637 1	0.688 5	0.693 6	0.641 2	0.916 14	0.721 8	0.854 13	0.727 9	1.145 15	0.672 3	0.768 11	0.739 10	0.719 7	0.682 4

RMSE ratios

France	1	0.675 1	0.945 9	0.744 2	0.957 12	0.775 3	1.579 15	0.948 10	0.960 13	1.112 14	0.821 6	0.873 8	0.851 7	0.785 4	0.805 5	0.956 11
	2	0.716 1	0.932 9	0.738 2	0.986 12	0.752 3	1.476 15	1.053 13	0.982 11	1.155 14	0.832 6	0.902 7	0.916 8	0.792 4	0.798 5	0.977 10
	3	0.642 1	0.846 7	0.749 3	0.955 10	0.762 4	1.408 15	0.999 13	0.896 8	1.150 14	0.795 6	0.916 9	0.987 11	0.686 2	0.793 5	0.989 12
	4	0.625 1	0.917 8	0.740 2	0.987 10	0.752 3	1.389 15	1.019 12	0.916 7	1.233 14	0.808 6	0.945 9	1.023 13	0.773 4	0.794 5	1.011 11
Germany	1	0.737 9	0.652 2	0.766 12	0.793 13	0.654 3	0.940 15	0.880 14	0.725 8	0.750 10	0.675 4	0.707 7	0.703 6	0.623 1	0.763 11	0.700 5

	2	0.682 4	0.619 1	0.778 12	0.819 13	0.629 2	0.965 15	0.852 14	0.765 10	0.755 9	0.677 3	0.729 8	0.723 7	0.683 5	0.778 11	0.709 6
	3	0.653 4	0.569 1	0.800 11	0.824 13	0.569 2	0.968 15	0.939 14	0.800 12	0.739 7	0.680 5	0.748 9	0.744 8	0.648 3	0.777 10	0.710 6
	4	0.691 4	0.609 1	0.877 13	0.846 12	0.633 2	0.959 14	1.039 15	0.842 11	0.759 7	0.698 5	0.768 9	0.764 8	0.653 3	0.804 10	0.725 6
Irish Republic	1	0.561 6	0.621 7	0.659 10	0.710 13	0.624 8	0.456 1	0.552 5	0.654 9	0.504 2	0.783 14	0.920 15	0.674 11	0.528 3	0.529 4	0.701 12
	2	0.590 6	0.622 8	0.627 9	0.764 13	0.618 7	0.467 1	0.554 4	0.655 11	0.514 2	0.783 14	0.964 15	0.654 10	0.536 3	0.555 5	0.695 12
	3	0.586 6	0.685 10	0.666 7	0.854 13	0.681 8	0.509 1	0.577 5	0.681 9	0.527 2	0.857 14	1.061 15	0.742 12	0.528 3	0.535 4	0.719 11
	4	0.612 6	0.688 8	0.743 12	0.885 14	0.700 9	0.484 1	0.565 5	0.671 7	0.530 2	0.834 13	1.097 15	0.725 11	0.537 3	0.546 4	0.706 10
Italy	1	1.226 9	0.935 2	0.991 3	1.004 4	0.914 1	1.257 11	1.228 10	1.279 13	1.286 14	1.111 5	1.190 6	1.707 15	1.260 12	1.221 8	1.204 7
	2	1.251 9	0.932 2	0.987 3	1.041 4	0.920 1	1.282 12	1.245 8	1.377 13	1.466 14	1.165 5	1.230 7	1.747 15	1.271 11	1.219 6	1.258 10
	3	1.271 7	0.885 1	1.021 3	1.079 4	0.920 2	1.314 12	1.352 13	1.276 8	1.486 14	1.182 5	1.284 9	1.779 15	1.302 10	1.240 6	1.310 11
	4	1.327 12	0.943 2	0.970 3	1.087 4	0.888 1	1.365 13	1.224 9	1.204 8	1.570 14	1.138 5	1.323 11	1.819 15	1.196 7	1.151 6	1.320 10
the Netherlands	1	0.892 11	0.697 3	0.670 1	0.745 4	0.673 2	1.019 14	0.875 10	1.452 15	0.941 12	0.871 9	0.848 7	0.753 5	0.854 8	0.787 6	0.986 13
	2	0.978 12	0.640 1	0.707 3	0.761 5	0.648 2	0.999 14	0.914 11	0.828 7	0.982 13	0.904 10	0.868 8	0.759 4	0.868 9	0.799 6	1.015 15

	3	0.973 12	0.706 3	0.670 1	0.784 5	0.680 2	1.018 14	0.911 9	0.833 6	1.017 13	0.951 11	0.893 8	0.768 4	0.922 10	0.854 7	1.045 15
	4	0.983 12	0.640 2	0.637 1	0.805 5	0.670 3	1.011 13	0.932 9	0.850 6	1.048 15	0.939 10	0.921 8	0.753 4	0.947 11	0.873 7	1.044 14
Spain	1	0.869 13	0.790 6	0.729 4	0.721 3	0.666 2	0.850 11	0.812 8	0.808 7	0.915 14	0.866 12	0.837 10	0.939 15	0.751 5	0.621 1	0.827 9
	2	0.883 10	0.889 12	0.769 5	0.713 2	0.741 3	0.843 7	0.894 13	0.886 11	0.979 15	0.846 8	0.858 9	0.920 14	0.767 4	0.598 1	0.820 6
	3	0.894 11	0.868 8	0.722 2	0.778 5	0.737 3	0.841 7	0.893 10	0.911 13	0.969 15	0.901 12	0.880 9	0.929 14	0.759 4	0.592 1	0.808 6
	4	0.899 10	0.851 7	0.706 3	0.749 5	0.655 2	0.855 8	0.874 9	0.901 11	0.928 13	0.942 15	0.910 12	0.930 14	0.749 4	0.593 1	0.822 6
US	1	1.081 11	0.731 4	0.400 1	0.626 2	0.878 7	1.018 10	1.122 12	1.198 14	0.901 8	1.160 13	0.719 3	0.866 6	0.992 9	1.224 15	0.745 5
	2	1.237 13	0.764 4	0.740 3	0.654 2	0.630 1	1.076 11	1.295 14	1.380 15	0.870 7	1.227 12	0.765 5	0.913 8	1.035 10	1.006 9	0.810 6
	3	1.020 13	0.671 2	0.736 3	0.668 1	0.959 10	1.110 14	0.943 9	1.165 15	0.867 5	1.007 12	0.802 4	0.959 11	0.897 7	0.910 8	0.892 6
	4	0.953 10	0.776 4	0.751 3	0.693 1	0.740 2	1.166 14	0.852 7	0.993 12	0.840 5	1.221 15	0.840 6	0.999 13	0.907 8	0.916 9	0.960 11

Notes: There are two figures presented in each cell with the ones in the first line being the accuracy measure ratios and the ones below them being the ranks. All MAE, MAPE and RMSE ratios have been normalized relative to the seasonal naive no-change forecast, and three decimal places are retained for all ratios. Ratios below one indicate predictive gains relative to the seasonal naive no-change forecast. The ratios and ranks of the top three individual models are highlighted for each comparison.

Table 5-2 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 1	0.881 10	0.697 2	0.848 8	0.714 3	1.380 15	0.916 12	0.895 11	1.009 14	0.784 6	0.785 7	0.936 13	0.742 4	0.784 5	0.876 9
Germany	0.646 7	0.480 1	0.720 11	0.817 13	0.489 2	0.959 15	0.889 14	0.727 12	0.679 10	0.570 5	0.663 9	0.540 4	0.531 3	0.649 8	0.594 6
Irish Republic	0.567 5	0.628 7	0.646 9	0.746 13	0.629 8	0.414 1	0.617 6	0.711 12	0.486 2	0.751 14	0.905 15	0.650 10	0.564 4	0.563 3	0.663 11
Italy	0.870 12	0.619 2	0.677 4	0.669 3	0.606 1	0.796 7	0.784 5	0.863 10	0.988 14	0.811 8	0.794 6	1.270 15	0.904 13	0.860 9	0.865 11
the Netherlands	0.986 11	0.718 1	0.723 3	0.800 5	0.719 2	1.128 15	0.939 9	1.040 14	0.998 12	0.931 8	0.891 6	0.745 4	0.958 10	0.922 7	1.039 13
Spain	0.859 13	0.764 10	0.641 3	0.669 4	0.618 2	0.744 7	0.745 8	0.756 9	0.890 15	0.796 11	0.820 12	0.875 14	0.714 6	0.561 1	0.712 5
US	0.855 12	0.648 3	0.609 2	0.604 1	0.716 6	0.976 15	0.810 8	0.863 13	0.836 11	0.870 14	0.661 4	0.801 7	0.815 9	0.830 10	0.711 5
MAPE ratios															
France	0.684 1	0.954 10	0.753 2	0.840 6	0.778 3	1.421 15	0.955 11	0.956 12	1.007 13	0.857 7	0.800 5	1.029 14	0.789 4	0.897 9	0.864 8
Germany	0.598 9	0.439 1	0.648 12	0.752 13	0.449 2	1.018 15	0.770 14	0.620 11	0.614 10	0.511 5	0.587 8	0.475 3	0.501 4	0.542 7	0.517 6
Irish Republic	0.577 5	0.690 8	0.750 11	0.821 13	0.698 10	0.339 1	0.754 12	0.939 14	0.454 2	0.692 9	1.109 15	0.663 7	0.534 3	0.538 4	0.631 6
Italy	0.955 12	0.675 2	0.727 3	0.727 4	0.646 1	0.891 8	0.827 5	0.911 9	1.075 14	0.872 7	0.845 6	1.356 15	0.994 13	0.943 10	0.950 11
the Netherlands	0.985 9	0.833 3	0.840 5	0.840 4	0.803 1	1.333 15	0.973 8	1.126 14	1.014 10	0.952 7	0.867 6	0.814 2	1.034 11	1.035 12	1.059 13

Spain	0.975 14	0.873 10	0.750 4	0.746 3	0.736 2	0.818 5	0.856 8	0.858 9	1.029 15	0.934 12	0.930 11	0.961 13	0.831 6	0.693 1	0.847 7
US	0.838 12	0.688 5	0.626 1	0.668 4	0.719 6	0.873 14	0.801 10	0.859 13	0.805 11	0.909 15	0.649 2	0.736 7	0.766 8	0.773 9	0.659 3
RMSE ratios															
France	0.665 1	0.910 8	0.743 2	0.971 11	0.760 4	1.463 15	1.005 13	0.938 9	1.163 14	0.814 6	0.909 7	0.944 10	0.759 3	0.798 5	0.983 12
Germany	0.691 5	0.612 1	0.805 12	0.821 13	0.621 2	0.958 15	0.928 14	0.783 11	0.751 9	0.682 4	0.738 8	0.733 7	0.651 3	0.780 10	0.711 6
Irish Republic	0.587 6	0.654 7	0.674 10	0.803 13	0.655 8	0.479 1	0.562 5	0.665 9	0.519 2	0.814 14	1.010 15	0.699 11	0.532 3	0.541 4	0.705 12
Italy	1.268 10	0.924 2	0.992 3	1.053 4	0.911 1	1.305 13	1.262 9	1.284 12	1.452 14	1.149 5	1.257 7	1.763 15	1.257 8	1.208 6	1.273 11
the Netherlands	0.957 11	0.671 2	0.671 3	0.774 5	0.668 1	1.012 14	0.908 9	0.991 12	0.997 13	0.916 10	0.883 7	0.758 4	0.898 8	0.828 6	1.023 15
Spain	0.886 12	0.850 8	0.731 3	0.740 4	0.700 2	0.848 7	0.868 9	0.877 11	0.948 15	0.889 13	0.871 10	0.930 14	0.756 5	0.601 1	0.819 6
US	1.073 12	0.736 3	0.657 1	0.660 2	0.802 5	1.092 13	1.053 11	1.184 15	0.869 7	1.154 14	0.782 4	0.935 8	0.957 9	1.014 10	0.852 6

Notes: There are two figures presented in each cell with the ones in the first line being the accuracy measure ratios and the ones below them being the ranks. All MAE, MAPE and RMSE ratios have been normalized relative to the seasonal naive no-change forecast, and three decimal places are retained for all ratios. Ratios below one indicate predictive gains relative to the seasonal naive no-change forecast. The ratios and ranks of the top three individual models are highlighted for each comparison.

Table 5-3 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 7	0.729 3	0.691 2	0.751 4	0.683 1	0.964 15	0.798 10	0.909 14	0.839 13	0.785 8	0.802 11	0.833 12	0.777 5	0.785 9	0.779 6
2	0.831 10	0.753 3	0.730 2	0.759 4	0.696 1	0.948 15	0.880 13	0.904 14	0.870 12	0.802 8	0.813 9	0.850 11	0.790 7	0.776 5	0.785 6
3	0.789 7	0.737 2	0.739 3	0.776 6	0.706 1	0.946 15	0.857 12	0.889 14	0.856 11	0.818 9	0.833 10	0.875 13	0.773 5	0.768 4	0.793 8
4	0.804 8	0.725 2	0.768 4	0.795 6	0.673 1	0.967 15	0.859 9	0.880 13	0.863 11	0.868 12	0.859 10	0.889 14	0.774 5	0.751 3	0.802 7
	RMSE ratios														
1	0.863 7	0.767 3	0.708 1	0.794 4	0.741 2	1.017 15	0.917 12	1.011 14	0.916 11	0.898 10	0.870 8	0.927 13	0.827 5	0.850 6	0.874 9
2	0.905 9	0.771 3	0.764 2	0.820 4	0.705 1	1.015 15	0.972 13	0.982 14	0.960 12	0.919 10	0.902 8	0.948 11	0.850 6	0.822 5	0.898 7
3	0.863 7	0.747 1	0.766 3	0.849 6	0.758 2	1.024 15	0.945 12	0.937 10	0.965 13	0.910 8	0.941 11	0.987 14	0.820 5	0.814 4	0.925 9
4	0.870 7	0.775 3	0.775 2	0.865 6	0.720 1	1.033 15	0.929 9	0.911 8	0.987 13	0.940 10	0.972 12	1.002 14	0.823 5	0.811 4	0.941 11

Notes: There are two figures presented in each cell with the ones in the first line being the accuracy measure ratios and the ones below them being the ranks. All MAE, MAPE and RMSE ratios have been normalized relative to the seasonal naive no-change forecast, and three decimal places are retained for all ratios. Ratios below one indicate predictive gains relative to the seasonal naive no-change forecast. The ratios and ranks of the top three individual models are highlighted for each comparison.

Table 5-4 General Forecasting Performance of Individual Models

measure	bounds ©	ETS	SARIMA	ADLM ©	State ETS	LI©	VAR©	TVP©	SD©	bounds	ADLM	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 figures presented in each cell with the ones in the first line being the accuracy measure ratios and the ones below them being the ranks. All MAE, MAPE and RMSE ratios have been normalized relative to the seasonal naive no-change forecast, and three decimal places are retained for all ratios. Ratios below one indicate predictive gains relative to the seasonal naive no-change forecast. The ratios and ranks of the top three individual models are highlighted for each comparison.

Table 5-5 Comparison Between Traditional and Climate Econometric Models for Each Modelling Technique

measure	model	France		Germany		Irish Republic		Italy		the Netherlands		Spain		US	
		climate	traditional	climate	traditional	climate	traditional	climate	traditional	climate	traditional	climate	traditional	climate	traditional
MAE	bounds	45935.65	55613.09	46948.36	41434.04	22964.69	30408.67	34967.91	32591.10	40055.09	37835.52	42969.63	39830.29	76701.00	78097.94
	ADLM	60129.65	55652.81	59339.90	48195.05	30191.85	36664.17	26865.89	31912.19	32516.95	36223.07	33444.39	40998.01	54220.10	59298.09
	LI	97914.97	66374.09	69709.38	39216.79	16775.49	26311.20	31974.04	51028.76	45837.14	30291.01	37182.51	43734.32	87580.37	71917.33
	VAR	64979.28	52660.94	64615.63	38621.81	24976.53	22855.01	31480.67	36331.97	38169.45	38932.54	37250.22	35696.25	72679.52	73109.46
	TVP	63453.79	55596.78	52865.56	47185.17	28781.93	22809.35	34675.14	34531.77	42260.14	37486.54	37794.29	28033.93	77474.72	74512.77
	SD	71573.14	62157.03	49376.79	43163.45	19675.74	26860.38	39686.85	34729.66	40550.63	42204.61	44501.58	35582.53	75015.90	63809.85
MAPE	bounds	10.08%	12.64%	12.63%	10.81%	12.65%	15.17%	15.87%	14.49%	19.17%	18.52%	20.14%	19.29%	20.46%	22.19%
	ADLM	12.38%	11.79%	15.89%	12.41%	17.98%	24.30%	12.08%	14.04%	16.35%	16.86%	15.40%	19.20%	16.30%	15.85%
	LI	20.95%	15.17%	21.52%	10.03%	7.42%	14.53%	14.81%	22.53%	25.94%	15.83%	16.90%	19.85%	21.32%	17.97%
	VAR	14.08%	11.63%	16.28%	10.58%	16.53%	11.71%	13.75%	16.51%	18.94%	20.12%	17.68%	17.16%	19.57%	18.71%
	TVP	14.10%	13.22%	13.10%	11.45%	20.57%	11.79%	15.14%	15.66%	21.90%	20.14%	17.72%	14.31%	20.97%	18.89%
	SD	14.85%	12.73%	12.99%	10.93%	9.94%	13.83%	17.87%	15.79%	19.73%	20.61%	21.26%	17.50%	19.66%	16.10%
RMSE	bounds	54798.65	67095.74	58619.84	57894.22	28285.84	39211.36	41737.34	37817.13	49766.80	47683.60	47925.31	48060.51	119298.05	128304.11
	ADLM	80057.96	74936.15	69660.46	62631.89	38670.88	48656.86	34634.80	41351.60	40255.25	45920.36	40027.24	47126.26	73391.66	86899.49
	LI	120618.46	77839.09	81313.55	62232.73	23066.25	33657.03	42926.49	58005.39	52643.01	39452.77	45838.00	50274.32	121458.09	103907.69
	VAR	82845.75	62570.18	78726.05	55287.37	27058.01	25626.40	41526.95	41372.36	47247.39	46717.37	46950.99	40910.09	117056.00	106453.05
	TVP	77358.71	65756.39	66449.21	66221.78	32040.49	26066.91	42248.97	39737.62	51542.53	43100.73	47409.37	32512.46	131639.18	112761.23
	SD	95847.65	81072.53	63714.84	60332.17	24981.80	33959.58	47782.94	41881.08	51873.06	53202.06	51260.90	44308.65	96659.65	94713.18

Notes: Two decimal places are retained for all MAE, MAPE and RMSE values and the measures of the superior model specification for each modelling technique and each origin are highlighted.

Table 5-6 Comparison Between Traditional and Climate Econometric Models for Each Origin

measure	France		Germany		Irish Republic		Italy		the Netherlands		Spain		US	
	climate	traditional	climate	traditional	climate	traditional	climate	traditional	climate	traditional	climate	traditional	climate	traditional
MAE	67331.08	58009.12	57142.6	42969.38	23894.37	27651.46	33275.08	36854.24	39898.23	37162.21	38857.1	37312.56	73945.27	70124.24
MAPE	14.41%	12.86%	15.40%	11.04%	14.18%	15.22%	14.92%	16.51%	20.34%	18.68%	18.19%	17.88%	19.71%	18.28%
RMSE	85254.53	71545.01	69747.32	60766.69	29017.21	34529.69	41809.58	43360.86	48888.01	46012.82	46568.64	43865.38	109917.1	105506.5

Notes: Two decimal places are retained for all MAE, MAPE and RMSE values and the measures of the superior model specification for each modelling technique and each origin are highlighted.

Table 5-7 General Comparison Between Traditional and Climate Econometric Models

	climate	traditional
MAE	47763.39	44297.6
MAPE	16.74%	15.78%
RMSE	61600.34	57940.99

Notes: Two decimal places are retained for all MAE, MAPE and RMSE values and the measures of the superior model specification for each modelling technique and each origin are highlighted.

5.3 Forecasting Performance of Combination Methods

The selected individual models are categorized into three groups: the first group which includes all models; the second one which contains traditional econometric and time series models; and the third set which consists of climate econometric and time series models. Combination is conducted for each group respectively using different weighting schemes.

To assess whether combining individual models can contribute to more accurate forecasts, the forecasting accuracy of combination forecasts is compared to that of the best single ones. The percentages of superior combination forecasts compared to the best single ones are worked out, and if the superior percentage is above zero, it is concluded that the most accurate forecast is given by the combination forecasting approach. In addition, the performance of combination forecasts is compared to that of the worst individual ones, and the percentages of inferior combination forecasts to the worst single ones are computed. If the inferior percentage is zero, it shows that all combination forecasts are better than the worst single ones, suggesting that the least accurate forecast is not generated from the combination approach. When the superior percentage is above zero, and the inferior percentage is zero, it is concluded that combining individual forecasts is beneficial.

5.3.1 Combining All Models

Firstly, all models are combined using the SA, VACO, DMSFE, inverse-MAE, and the two-stage combination methods. The comparisons between the combination and the individual forecasting approach are conducted in three ways measured by MAE, MAPE and RMSE respectively. At first, forecasting performance is evaluated for different origins, which is a market-specific assessment and reveals whether combination forecasts are better than individual ones and the performance rankings of different weighting schemes for forecasting tourism demand from a particular market. Afterwards, forecasting accuracy is assessed for different forecasting horizons, which is a forecasting-horizon-specific evaluation and demonstrates whether combination is beneficial and which combination methods are superior for forecasting tourism demand to the UK at a specific horizon length. At last, a general comparison by taking averages across all origins and over all forecasting horizons are made, which shows whether combining individual forecasts results in more accurate projections and the performance rankings of various weighting methods as far as general tourism demand to the UK is concerned.

5.3.1.1 Comparison across Different origins

At first, the performance of different combination forecasts is examined at a disaggregated level for all forecasting lengths for seven origins. Table 5-8 to table 5-10 demonstrate the percentages of the superior combination forecasts compared to the best single ones and the performance ranking of every combination method for each forecasting horizon and each origin measured by MAE, MAPE and RMSE respectively. The highest three superior percentages for each comparison are highlighted. It shows that all superior percentages are above zero, implying that the most accurate forecasts are generated through the combination forecasting approach for one- to four-step-ahead forecasts for seven markets.

In general, the two-stage weighting schemes are outperformed by the one-stage ones, which take up the top-three performing weighting schemes for most cases with only three exceptions: the TS-DMSFE ($\alpha = 0.85$) method ranks number three for the one-step-ahead forecasts for the Italian market when MAPE is used as the accuracy criterion (table 5-9), and the TS-SA method is the third in position for the one-step-ahead forecasts for the American market based on MAPE and RMSE (table 5-9 and table 5-10). The Inverse-MAE method performs very well, especially for the German, Irish and American cases.

The forecasting ability of different combination methods changes according to the origin market under consideration and the accuracy measure used.

When forecasting accuracy is measured by MAE (see table 5-8), the lowest percentage of the superior combination forecasts is 4.77%, which is gained by the TS-SA method for the one-step-ahead forecasts for the Dutch market. It suggests that 4.77% of one-step-ahead combination forecasts obtained through the TS-SA method are better than the best single one-step-ahead projection as far as tourism demand from the Netherlands is concerned. The highest superior percentage of 29.93% is obtained by the DMSFE ($\alpha = 0.85$) method for the three-step-ahead projections for the Irish market, which means that 29.93% of combination forecasts generated through the DMSFE ($\alpha = 0.85$) method are more accurate than the best single one when it comes to three-quarter-ahead forecasts for the Irish Republic. Judged by MAPE and RMSE (see table 5-9 and table 5-10), the lowest superior percentages are 3.04% (by the TS-VACO method for the two-step-ahead American forecasting) and 8.91% (by the TS-SA method for the one-step-ahead Dutch forecasting) respectively, and the highest superior percentages are 29.92% (by the DMSFE ($\alpha = 0.85$) method for the three-step-ahead Irish forecasting) and 29.95% (by the DMSFE method ($\alpha = 0.85$) for the one-step-ahead Spanish forecasting) respectively.

For each origin, there exists no great difference in the superior percentages generated by the same weighting scheme across different forecasting lengths (see each column of table 5-8 to table 5-10).

For example, for the German market measured by MAE (see table 5-8), the superior percentages gained by all one-stage combination methods for one- to four-step-ahead forecasts are all between 29% and 30%. Likewise, for the Spanish origin judged by RMSE (see table 5-10), the superior percentages obtained by all two-stage weighting schemes for four forecasting lengths all ranges from 23% to 25%. A few exceptions are seen from the Dutch and American cases. As with the evaluation based on MAE for the Dutch market (see table 5-8), the superior percentages generated by all weighting methods for the one-step-ahead forecasts are much lower compared to those for other forecasting lengths, and the biggest gap is above 20%. Similarly, for the American case judged by MAPE (see table 5-9), the superior percentages for the second- and third-step-ahead forecasts are much lower than those for the first- and fourth-step-ahead ones, and the biggest difference is nearly 20%.

In addition, for projecting tourism demand from a particular origin at a specific forecasting length, the difference in the superior percentages gained by various weighting schemes is not great (see each row of table 5-8 to table 5-10). For instance, for one-step-ahead forecasts for the Irish origin based on MAPE and RMSE, the superior percentages generated by all one-stage combination methods are around 29%, and those produced by all two-stage weighting schemes are about 24%.

Next, the performance of different combination forecasts is examined at an aggregated level for seven origins. Table 5-11 to table 5-13 show the average percentage of the superior combination forecasts compared to the best single ones and the performance ranking of each combination method over the four forecasting horizons under study for each origin measured by MAE, MAPE and RMSE respectively. The top-three combination methods are highlighted for each comparison.

The superior percentages range from 10.35%, which is achieved by the TS-VACO method for forecasting tourism demand from the US judged by MAPE, to 29.86%, which is provided by the SA method for the Spanish case measured by RMSE. It means that the best forecasts are always produced through the combination forecasting approach for seven origins regardless of combination method, with at least 10.35% composite forecasts being more accurate than the best constituent one as far as one combination method for one market is concerned. The two-stage combination methods, which always remain at the bottom half of the list for every market irrespective of accuracy measure, provide unsatisfactory combination forecasts. On the other hand, the one-stage weighting schemes perform very well, all of which have some opportunities to rank top two.

For most cases, the performance ranking of one particular combination method shows a degree of consistency across different markets. For example, according to table 5-11 which is judged by MAE, for the French, Italian and Spanish markets, the highest three percentages are provided by the DMSFE methods, and the worst performance is seen from the TS-SA, TS-VACO and TS-INVERSE-MAE schemes. The DMSFE ($\alpha = 0.85/0.90$) methods always behave well, except for the German and American cases, when they stay at the bottom among all one-stage weighting schemes. The Inverse-MAE method shows good forecasting ability, gaining the first place for forecasting tourism demand from Germany and the Irish Republic, and ranking number two for the American origin. The SA method, which is the worst one among all one-stage methods for five out of seven cases, for the first time shows its superior performance for forecasting tourism demand from the US, ranking number one, followed by the inverse-MAE and VACO methods.

Table 5-11 and table 5-12 show that MAE and MAPE provide consistent evaluation regarding the performance ranking of different combination forecasts for the French, German and Spanish origins with the best-three and the worst-three methods staying unchanged. But for other markets, the forecasting performance evaluations based on MAE and MAPE are inconsistent. As with tourism demand from the Irish Republic, the TS-SA, which ranks number eight based on MAE, takes up the last position measured by MAPE. The SA method, which performs unsatisfactorily according to MAE for the Italian and Dutch cases, achieves the second and the third places respectively judged by MAPE. According to MAPE, the inverse-MAE method acts the best for projecting tourist arrivals from the Netherlands, which ranks number five based on MAE. For the American market, the VACO and the TS-VACO methods are evaluated worse by MAPE than by MAE.

According to table 5-11 and table 5-13, MAE and RMSE yield consistent assessment regarding the performance ranking of different combination forecasts for the French origin; but for other cases, inconsistency is shown. For example, for the German and Spanish markets, the SA method gains the first place in the list based on RMSE, which ranks number four and number six respectively according to MAE. When RMSE is used to measure forecasting accuracy, better results are achieved by the VACO method for forecasting tourism demand from the Irish Republic and Italy; and for the Irish, Italian and Spanish cases, the DMSFE ($\alpha = 0.85$) method generate less accurate forecasts.

It is also shown that for the same origin market, there exists no great difference among the superior percentages achieved by different weighting schemes, especially within the one-stage or the two-stage group (each row of table 5-11 to table 5-13). The greatest difference in the superior percentages generated by the best and the worse schemes is always below 10% with only one exception, which is 10.11% seen from the American origin measured by RMSE (table 5-13). It is between the highest percentage of 28.69% achieved by the SA method and the lowest percentage of 18.58% produced by the TS-DMSFE ($\alpha = 0.85$) scheme. The DMSFE ($\alpha = 0.85$) method, which is the poorest-performing one among the one-stage schemes for this case, provides a superior percentage of 24.79%, resulting in the greatest difference among the one-stage group being only 3.90%. When MAE and MAPE is used to measure accuracy, the greatest differences are also seen from the American case, which are both below 10%.

On the contrary, there exists divergence in the superior percentages generated by the same weighting scheme for different markets, which is sharper if the combination method belongs to the two-stage combination approach (each column of table 5-11 to table 5-13). For example, judged by MAE, the highest superior percentage of the SA method is 29.67% for the German market, and the lowest one is 18.40% for the Italian origin, resulting in a gap of 11.27%. Measured by MAPE, the superior percentages achieved by the TS-VACO method for seven markets ranges from 10.35% for the American case to 24.01% for the Irish origin, which means that the biggest gap is 13.66%. Similarly, the sharpest divergence provided by the TS-SA method according to RMSE is 12.50%, which is between 24.08% for the Spanish case and 11.58% for the Dutch market.

The percentages of the inferior combination forecasts compared to the worst single ones are zero for all cases, which means that all composite forecasts are better than the worst constituent ones regardless of weighting scheme or accuracy measure for seven origins. It shows that the least accurate forecasts are not generated through the combination forecasting approach. Therefore, it is concluded that combination contributes to better forecasts with higher degree of accuracy and lower risk of forecasting failure for seven markets as far as the first combination group is concerned.

5.3.1.2 Comparison over Different Forecasting Horizons

For a forecasting-horizon-specific evaluation, the average percentages of the superior combination forecasts compared to the best single ones for four forecasting horizons measured by MAE, MAPE and RMSE respectively are obtained by averaging across seven origin countries, which are presented in table 5-14. Table 5-14 also provides the performance ranking of every combination method for each forecasting horizon and the percentages achieved by the top-three performing weighting schemes for each comparison are highlighted.

It shows that for four forecasting lengths, all superior percentages achieved by different combination methods are above 19%, which means that at least 19% combination forecasts obtained through any weighting scheme are better than the best single projection, suggesting that combing individual forecasts can improve forecasting accuracy irrespective of forecasting horizon, weighting scheme or accuracy measure.

The highest superior percentage of 29.26% is achieved by the VACO method for the four-step-ahead forecasts measured by RMSE, and the lowest superior percentage of 19.53% is produced by the TS-SA scheme for the one-step-ahead forecasts judged by MAE. The two-stage combination methods are worse than the one-stage ones for all cases, with the poorest performing method being the TS-SA, which stays at the bottom for ten out of twelve comparisons. The inverse-MAE method performs well, especially for the one-step-ahead forecasts, ranking number one no matter which accuracy measure is used.

The performance ranking of various weighting schemes varies according to the forecasting length under consideration and the accuracy measure used. For the one-step-ahead forecasts, the inverse-MAE method produces the best results regardless of accuracy measure, followed by the VACO and the DMSFE ($\alpha = 0.95$) methods judged by MAE and RMSE, and the SA and the the DMSFE ($\alpha = 0.90$) schemes based on MAPE. The TS-SA, TS-DMSFE ($\alpha = 0.85$) and TS-DMSFE ($\alpha = 0.90$) methods always remain at the bottom irrespective of accuracy measure.

As far as the second-step-ahead forecasts are concerned, the VACO method ranks number one according to MAE and RMSE, and the SA method is the first in position measured by MAPE. The TS-SA and TS-inverse-MAE methods rank number twelve and eleven respectively according to three measures. And the tenth position, which is taken by the TS-DMSFE ($\alpha = 0.85$) method judged by MAE and RMSE, is obtained by the TS-VACO scheme measured by MAPE.

Three accuracy measures provide inconsistent assessments regarding the performance ranking of different combination methods for the three-step-ahead forecasts. Based on MAE, the top-three weighting schemes are the DMSFE methods, and the bottom-three ones are the TS-SA, TS-inverse-MAE and the TS-VACO methods. Judged by MAPE, the DMSFE ($\alpha = 0.85$) method ranks number one, followed by the SA and DMSFE ($\alpha = 0.90$) schemes, and the TS-VACO, TS-inverse-MAE and TS-DMSFE ($\alpha = 0.95$) methods take the last three positions. Measured by RMSE, the VACO method shows superiority compared to other schemes, and the DMSFE ($\alpha = 0.95$) and DMSFE

($\alpha = 0.90$) methods follow it. And the TS-SA, TS-inverse-MAE and TS-VACO schemes stay at the bottom of the list.

When it comes to the four-step-ahead forecasts, assessed by MAE and MAPE, the DMSFE ($\alpha = 0.85$) method displays the finest forecasting performance, followed by the inverse-MAE and DMSFE ($\alpha = 0.90$) schemes; and the TS-SA, TS-VACO and TS-DMSFE ($\alpha = 0.95$) methods are the inferior ones. Evaluated by RMSE, the VACO method behaves the best, and the DMSFE ($\alpha = 0.95$) and DMSFE ($\alpha = 0.90$) schemes come after it. And the last three places are taken by the TS-SA, TS-inverse-MAE and TS-DMSFE ($\alpha = 0.85$) methods.

For the same forecasting length, various weighting schemes produce similar superior percentages (see each row of table 5-14). For example, the difference in the superior percentages of the best and the worst weighting schemes for the one-step-ahead forecasts are 4.79% (between 24.32% yielded by the inverse-MAE method and 19.53% produced by the TS-SA method), 5.42% (between 26.34% yielded by the inverse-MAE method and 20.92% produced by the TS-DMSFE ($\alpha = 0.85$) method) and 5.80% (between 26.07% yielded by the inverse-MAE method and 20.27% produced by the TS-SA method) measured by MAE, MAPE and RMSE respectively.

Similarly, there is not great difference in the superior percentages generated by the same weighting scheme for different forecasting lengths (see each column of table 5-14). For instance, measured by MAPE, the highest superior percentage generated by the inverse-MAE method is 27.12% for the four-step-ahead forecasts, and the lowest percentage is 25.48% for the third-step-ahead forecasts, which means that the gap is only 1.64%. For the TS-inverse-MAE scheme, the biggest gap judged by MAPE is 2.06% between 21.79% for the four-step-ahead forecasts and 19.73% for the third-step-ahead projections.

The percentages of the inferior combination forecasts compared to the worst single ones are zero for all cases, which means that all composite forecasts are better than the worst constituent ones regardless of weighting scheme or accuracy measure for four forecasting lengths. It shows that the least accurate forecasts are not generated through the combination forecasting approach. Therefore, it is concluded that the combination forecasting approach is superior with higher degree of accuracy and lower risk of forecasting failure for one- to four-step-ahead forecasts as far as the first combination group is concerned.

5.3.1.3 General Comparison among Various Weighting Schemes

Table 5-15 provides the average percentages of the superior forecasts compared to the best single ones and the performance ranking of each weighting scheme over four forecasting horizons and seven origins under study measured by MAE, MAPE and RMSE respectively. The top-three superior percentages for each comparison are highlighted. It is shown that all superior percentages are above 20%, which means that at least 20% of combination forecasts produced by any weighting scheme are better than the best single projection, suggesting that combing individual forecasts can improve forecasting accuracy irrespective of weighting scheme or accuracy measure.

The highest superior percentage of 27.84% is achieved by the VACO method measured by RMSE, and the lowest superior percentage of 20.43% is produced by the TS-SA scheme judged by MAE. The two-stage combination methods are inferior compared to the one-stage ones no matter which accuracy measure is used, with the poorest performing method being the TS-SA, which stays at the bottom according to MAE and RMSE and ranks number eleven based on MAPE. The inverse-MAE method performs very well, ranking number one based on MAE and MAPE.

The performance ranking of different weighting schemes is affected by accuracy measure. Based on MAE, the top-three weighting schemes are the inverse-MAE, the VACO and the DMSFE ($\alpha = 0.95$) methods. According to MAPE, the inverse-MAE, the DMSFE ($\alpha = 0.85$) and the SA schemes are the best three ones. And judged by RMSE, the VACO, the DMSFE ($\alpha = 0.95$) and the DMSFE ($\alpha = 0.90$) methods perform the best. The SA method only shows good performance when accuracy is measured by MAPE, and it remains as the worst one-stage method assessed by MAE and RMSE. The DMSFE methods perform generally well, and the best discounting factor in terms of generating more accurate forecasts changes according to accuracy measure.

The inverse-MAE method, which performs the best according to MAE and MAPE, is only the fifth in position according to RMSE, based on which, the VACO scheme ranks number one. It can be explained by the way how combining weights are computed by these two weighting schemes, or more specifically, how historical individual forecasting errors affect combining weights. As shown in equation 3.28, for the inverse-MAE method, the weights of individual forecasts are determined by their MAEs: the smaller a constituent's MAE is, the heavier it is weighted. As with the VACO method, it is the MSE that affects the combining weights: the smaller an individual forecast's MSE is, the larger weight it is assigned to. The individual forecasting error is amplified by the squaring operator in the VACO method. As demonstrated in equation 3.34, the forecasting error is also amplified by the squaring operator in RMSE, which is the only accuracy measure considered in this study that uses the squared values of the forecasting errors in the loss function.

The differences between the superior percentages of the best and the worst combination methods are 5.99% based on MAE (between 20.43% by the TS-SA method and 26.42% by the inverse-MAE method), 5.56% according to MAPE (between 20.56% by the TS-VACO method and 26.12% by the inverse-MAE scheme) and 6.77% measured by RMSE (between 21.07% by the TS-SA method and 27.84% by the VACO method), which is not great in percent. But considering the great number of all possible combination forecasts, the number of superior combination forecasts achieved by different weighting schemes are not small.

The percentages of the inferior combination forecasts compared to the worst single ones are zero for all cases, which means that all composite forecasts are better than the worst constituent ones regardless of weighting scheme or accuracy measure. It shows that the least accurate forecasts are not generated through the combination forecasting approach. Therefore, it is concluded that combining individual forecasts are beneficial with higher degree of accuracy and lower risk of forecasting failure as far as the first combination group is concerned.

Table 5-8 Superior Percentage and Rank of Each Combination Method at Different Forecasting Horizons for Seven Origins (the First Combination Group, MAE)

	step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE	TS-SA	TS-VACO	TS-DMSFE (.85)	TS-DMSFE (.90)	TS-DMSFE (.95)	TS-INVERSE -MAE
France	1	24.67% 6	28.14% 4	28.62% 1	28.52% 2	28.35% 3	27.70% 5	20.43% 12	23.81% 10	23.99% 7	23.93% 8	23.89% 9	23.21% 11
	2	28.82% 6	29.28% 5	29.47% 1	29.42% 2	29.35% 3	29.32% 4	22.45% 12	23.87% 7	23.83% 10	23.85% 9	23.87% 7	23.39% 11
	3	29.52% 6	29.85% 4	29.85% 2	29.85% 1	29.85% 2	29.70% 5	23.25% 12	24.17% 7	24.13% 10	24.15% 9	24.17% 7	23.73% 11
	4	28.71% 6	28.91% 5	29.01% 1	28.97% 2	28.94% 4	28.95% 3	23.25% 12	24.13% 7	24.07% 10	24.09% 9	24.13% 7	23.81% 11
Germany	1	29.87% 2	29.85% 3	29.80% 6	29.81% 5	29.83% 4	29.88% 1	24.33% 7	24.31% 8	24.31% 8	24.31% 8	24.31% 8	24.31% 8
	2	29.72% 4	29.74% 2	29.67% 6	29.70% 5	29.72% 3	29.80% 1	24.01% 8	23.97% 9	23.93% 12	23.95% 11	23.97% 9	24.05% 7
	3	29.71% 3	29.72% 2	29.65% 6	29.67% 5	29.69% 4	29.78% 1	24.45% 7	24.39% 9	24.37% 11	24.37% 11	24.39% 9	24.41% 8
	4	29.38% 6	29.52% 2	29.45% 5	29.47% 4	29.49% 3	29.59% 1	22.31% 12	22.93% 8	22.71% 11	22.75% 10	22.83% 9	22.97% 7
Irish Republic	1	28.69% 6	29.11% 2	29.12% 1	29.09% 3	29.09% 5	29.09% 3	23.77% 8	23.33% 9	23.33% 9	23.27% 11	23.25% 12	23.81% 7
	2	29.85% 2	29.85% 3	29.81% 6	29.82% 5	29.83% 4	29.89% 1	24.07% 12	24.21% 10	24.23% 7	24.23% 7	24.21% 10	24.23% 7
	3	29.89% 6	29.92% 5	29.93% 1	29.92% 3	29.92% 4	29.93% 1	24.19% 10	24.19% 10	24.25% 8	24.21% 9	24.19% 10	24.27% 7
	4	29.58% 2	29.45% 6	29.55% 3	29.51% 4	29.48% 5	29.70% 1	23.61% 12	23.63% 11	23.77% 8	23.73% 9	23.71% 10	23.95% 7
Italy	1	16.08% 10	16.95% 6	18.42% 1	18.06% 2	17.57% 3	16.26% 9	15.48% 11	16.30% 8	17.36% 4	17.22% 5	16.88% 7	15.48% 11

	2	22.91% 6	23.80% 4	24.84% 1	24.64% 2	24.29% 3	23.31% 5	18.87% 12	19.69% 10	20.37% 7	20.33% 8	20.05% 9	19.29% 11
	3	20.10% 3	19.08% 6	20.54% 1	19.95% 4	19.53% 5	20.28% 2	15.46% 10	14.70% 12	16.60% 7	15.94% 8	15.22% 11	15.68% 9
	4	14.52% 11	16.92% 5	19.57% 1	18.51% 2	17.63% 4	17.89% 3	13.26% 12	15.14% 10	16.30% 6	15.78% 7	15.44% 9	15.66% 8
The Netherlands	1	7.21% 7	9.59% 1	8.48% 5	8.87% 3	9.23% 2	8.55% 4	4.77% 12	7.21% 6	6.37% 10	6.69% 9	6.95% 8	5.91% 11
	2	19.51% 6	23.96% 1	23.23% 4	23.47% 3	23.72% 2	20.97% 5	10.49% 12	16.24% 7	15.84% 10	16.00% 9	16.12% 8	12.18% 11
	3	22.24% 6	25.81% 1	25.09% 4	25.33% 3	25.58% 2	22.57% 5	13.86% 12	18.81% 7	18.19% 10	18.37% 9	18.59% 8	13.92% 11
	4	27.98% 6	28.98% 1	28.93% 4	28.96% 3	28.96% 2	28.46% 5	22.31% 12	23.33% 7	23.29% 10	23.31% 8	23.31% 8	22.81% 11
Spain	1	29.89% 6	29.91% 4	29.92% 1	29.92% 1	29.91% 3	29.90% 5	24.01% 12	24.07% 7	24.05% 11	24.07% 7	24.07% 7	24.07% 7
	2	29.61% 5	29.61% 4	29.75% 1	29.71% 2	29.67% 3	29.58% 6	23.89% 11	23.93% 10	24.13% 7	24.05% 8	23.99% 9	23.89% 11
	3	29.33% 6	29.48% 4	29.61% 1	29.57% 2	29.52% 3	29.38% 5	23.59% 12	23.69% 10	23.79% 7	23.77% 8	23.73% 9	23.65% 11
	4	29.78% 6	29.80% 5	29.85% 1	29.83% 2	29.82% 3	29.82% 4	23.79% 11	23.77% 12	23.91% 7	23.85% 9	23.83% 10	23.87% 8
The US	1	29.35% 1	24.21% 3	20.22% 9	21.46% 7	22.85% 6	28.85% 2	23.95% 4	20.75% 8	18.13% 12	18.97% 11	19.91% 10	23.57% 5
	2	23.22% 1	20.41% 3	18.50% 6	19.09% 5	19.72% 4	22.19% 2	13.28% 7	11.14% 9	10.11% 12	10.37% 11	10.76% 10	12.06% 8
	3	29.32% 1	28.53% 6	28.84% 3	28.76% 4	28.63% 5	29.01% 2	22.11% 7	20.37% 12	21.23% 9	20.99% 10	20.71% 11	21.37% 8
	4	29.32% 2	28.76% 6	28.82% 3	28.80% 4	28.78% 5	29.34% 1	22.89% 8	21.77% 12	21.93% 9	21.89% 10	21.85% 11	22.93% 7

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the top three combination methods are highlighted.

Table 5-9 Superior Percentage and Rank of Each Combination Method at Different Forecasting Horizons for Seven Origins (the First Combination Group, MAPE)

origin	step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE	TS-SA	TS-VACO	TS-DMSFE (.85)	TS-DMSFE (.90)	TS-DMSFE (.95)	TS-INVERSE -MAE
France	1	19.96% 11	24.59% 4	25.77% 1	25.49% 2	25.07% 3	24.57% 5	17.32% 12	22.03% 9	22.73% 6	22.55% 7	22.33% 8	21.47% 10
	2	28.60% 6	29.04% 5	29.20% 1	29.16% 2	29.12% 4	29.14% 3	22.13% 12	23.33% 10	23.37% 7	23.37% 7	23.37% 7	23.15% 11
	3	28.41% 6	29.12% 4	29.13% 3	29.15% 1	29.15% 2	28.77% 5	22.45% 12	23.63% 7	23.53% 10	23.59% 9	23.63% 7	23.03% 11
	4	28.57% 6	28.72% 5	28.78% 1	28.75% 3	28.73% 4	28.78% 1	23.55% 12	24.13% 7	24.05% 10	24.09% 9	24.11% 8	23.95% 11
Germany	1	29.65% 4	29.70% 2	29.63% 6	29.65% 5	29.67% 3	29.78% 1	24.31% 10	24.33% 8	24.31% 10	24.31% 10	24.33% 8	24.35% 7
	2	29.45% 6	29.62% 2	29.51% 5	29.53% 4	29.59% 3	29.69% 1	23.89% 12	23.97% 8	23.97% 8	23.97% 8	23.97% 8	24.01% 7
	3	29.37% 6	29.57% 2	29.45% 5	29.49% 4	29.52% 3	29.63% 1	24.47% 7	24.45% 12	24.47% 7	24.47% 7	24.47% 7	24.47% 7
	4	28.97% 6	29.30% 2	29.17% 5	29.20% 4	29.25% 3	29.38% 1	21.49% 12	22.55% 7	22.17% 11	22.27% 10	22.39% 9	22.51% 8
Irish Republic	1	29.46% 6	29.64% 2	29.64% 3	29.63% 5	29.63% 4	29.65% 1	24.15% 8	23.87% 10	23.89% 9	23.85% 11	23.85% 11	24.19% 7
	2	29.83% 3	29.84% 2	29.82% 5	29.82% 5	29.83% 4	29.88% 1	23.99% 12	24.15% 11	24.17% 7	24.17% 7	24.17% 7	24.17% 7
	3	29.80% 6	29.90% 4	29.92% 1	29.92% 2	29.91% 3	29.90% 4	24.03% 12	24.13% 11	24.19% 7	24.19% 7	24.15% 10	24.19% 7
	4	29.62% 6	29.67% 5	29.74% 2	29.73% 3	29.70% 4	29.79% 1	23.61% 12	23.91% 11	23.97% 9	23.99% 8	23.95% 10	24.03% 7
Italy	1	18.42% 2	17.26% 11	18.62% 1	18.28% 4	17.81% 7	17.52% 9	17.78% 8	17.40% 10	18.41% 3	18.25% 5	17.93% 6	17.22% 12

	2	23.57% 4	23.35% 5	24.58% 1	24.33% 2	23.88% 3	23.25% 6	19.97% 11	20.11% 10	20.85% 7	20.77% 8	20.47% 9	19.97% 11
	3	19.71% 1	16.25% 6	17.71% 3	17.15% 4	16.66% 5	18.62% 2	15.68% 7	12.92% 12	14.50% 9	13.94% 10	13.42% 11	14.76% 8
	4	17.44% 7	17.55% 5	20.00% 1	19.03% 3	18.18% 4	19.29% 2	16.40% 12	16.58% 11	17.54% 6	17.12% 9	16.82% 10	17.34% 8
the Netherlands	1	25.13% 3	25.42% 2	24.33% 6	24.63% 5	25.00% 4	26.27% 1	15.74% 12	18.07% 7	17.02% 11	17.36% 10	17.70% 9	17.74% 8
	2	29.42% 1	29.23% 3	29.11% 6	29.16% 5	29.20% 4	29.30% 2	23.03% 7	22.83% 9	22.59% 12	22.69% 11	22.77% 10	22.87% 8
	3	29.13% 2	29.18% 1	28.94% 5	29.03% 4	29.10% 3	28.89% 6	22.55% 7	22.55% 7	22.23% 11	22.37% 10	22.43% 9	21.99% 12
	4	29.17% 6	29.46% 1	29.41% 4	29.42% 3	29.45% 2	29.31% 5	23.67% 12	23.89% 7	23.87% 8	23.87% 8	23.87% 8	23.87% 8
Spain	1	29.86% 6	29.88% 4	29.89% 1	29.89% 2	29.89% 2	29.88% 4	23.97% 8	23.97% 8	23.97% 8	23.95% 12	23.97% 8	24.01% 7
	2	29.26% 4	29.23% 5	29.52% 1	29.44% 2	29.34% 3	29.20% 6	23.61% 11	23.63% 10	23.89% 7	23.79% 8	23.69% 9	23.61% 11
	3	29.81% 6	29.83% 4	29.89% 1	29.87% 2	29.85% 3	29.82% 5	24.07% 12	24.11% 10	24.17% 7	24.15% 8	24.13% 9	24.09% 11
	4	29.78% 6	29.83% 4	29.87% 1	29.85% 2	29.84% 3	29.82% 5	23.91% 11	23.91% 11	24.03% 7	23.97% 9	23.93% 10	23.99% 8
US	1	28.29% 1	20.61% 5	19.25% 8	19.68% 7	20.12% 6	26.74% 2	23.29% 3	17.12% 9	16.12% 12	16.36% 11	16.76% 10	22.27% 4
	2	9.71% 1	8.09% 3	7.94% 6	7.97% 5	8.02% 4	8.43% 2	3.97% 7	3.04% 12	3.18% 8	3.14% 10	3.10% 11	3.18% 8
	3	15.05% 4	14.27% 5	17.41% 1	16.47% 2	15.43% 3	12.72% 6	7.05% 9	5.91% 11	8.25% 7	7.43% 8	6.73% 10	5.57% 12
	4	23.91% 2	22.67% 6	24.00% 1	23.66% 3	23.20% 5	23.45% 4	17.28% 7	15.34% 12	16.42% 9	16.18% 10	15.82% 11	16.84% 8

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the top three combination methods are highlighted.

Table 5-10 Superior Percentage and Rank of Each Combination Method at Different Forecasting Horizons for Seven Origins (the First Combination Group, RMSE)

origin	step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE	TS-SA	TS-VACO	TS-DMSFE (.85)	TS-DMSFE (.90)	TS-DMSFE (.95)	TS-INVERSE -MAE
France	1	20.88% 11	24.61% 5	25.92% 1	25.55% 2	25.11% 3	24.65% 4	18.21% 12	21.99% 9	22.73% 6	22.55% 7	22.27% 8	21.59% 10
	2	27.86% 6	28.95% 4	29.23% 1	29.16% 2	29.06% 3	28.89% 5	21.69% 12	23.43% 10	23.57% 7	23.55% 8	23.47% 9	23.11% 11
	3	26.87% 6	28.20% 4	28.39% 1	28.35% 2	28.29% 3	27.78% 5	21.33% 12	23.45% 7	23.45% 7	23.45% 7	23.45% 7	22.59% 11
	4	28.47% 6	29.07% 2	29.09% 1	29.07% 3	29.07% 3	28.99% 5	23.07% 12	24.05% 7	24.01% 10	24.03% 8	24.03% 8	23.79% 11
Germany	1	29.88% 1	29.83% 3	29.82% 6	29.83% 5	29.83% 4	29.85% 2	24.19% 7	24.17% 9	24.15% 12	24.17% 9	24.17% 9	24.19% 7
	2	29.88% 1	29.83% 3	29.82% 6	29.82% 5	29.83% 4	29.85% 2	23.91% 7	23.83% 9	23.79% 12	23.81% 11	23.83% 9	23.87% 8
	3	29.85% 1	29.79% 3	29.78% 6	29.78% 4	29.78% 4	29.83% 2	24.25% 7	24.17% 10	24.19% 9	24.17% 10	24.17% 10	24.21% 8
	4	29.67% 3	29.68% 2	29.65% 6	29.65% 5	29.67% 3	29.72% 1	23.39% 12	23.53% 7	23.43% 10	23.43% 10	23.49% 9	23.53% 7
Irish Republic	1	29.76% 6	29.87% 2	29.86% 4	29.87% 2	29.87% 1	29.83% 5	24.43% 8	24.43% 8	24.43% 8	24.43% 8	24.43% 8	24.45% 7
	2	29.76% 5	29.78% 2	29.75% 6	29.76% 4	29.77% 3	29.83% 1	23.89% 12	24.19% 9	24.19% 9	24.21% 7	24.21% 7	24.13% 11
	3	29.78% 6	29.85% 3	29.83% 5	29.85% 3	29.85% 2	29.87% 1	24.01% 12	24.11% 10	24.11% 10	24.13% 8	24.13% 8	24.21% 7
	4	29.76% 6	29.85% 5	29.86% 2	29.86% 2	29.86% 2	29.87% 1	23.87% 12	24.19% 9	24.23% 7	24.21% 8	24.19% 9	24.15% 11
Italy	1	26.06% 6	27.12% 4	27.42% 1	27.41% 2	27.32% 3	26.88% 5	18.97% 12	20.79% 10	21.13% 7	21.11% 8	21.03% 9	20.19% 11

	2	28.25% 6	28.55% 4	28.51% 5	28.59% 2	28.59% 1	28.57% 3	21.07% 12	21.97% 10	22.05% 8	22.13% 7	22.03% 9	21.85% 11
	3	28.67% 6	28.83% 2	28.73% 5	28.78% 4	28.83% 3	28.92% 1	22.03% 12	22.59% 11	22.63% 8	22.63% 8	22.63% 8	22.69% 7
	4	28.81% 6	29.14% 1	28.86% 5	28.96% 4	29.06% 3	29.11% 2	21.85% 12	22.69% 7	22.29% 11	22.43% 10	22.55% 8	22.53% 9
the Netherlands	1	12.18% 8	15.91% 4	16.55% 1	16.28% 2	16.07% 3	13.96% 5	8.91% 12	12.00% 10	12.40% 6	12.22% 7	12.10% 9	10.21% 11
	2	17.84% 6	24.55% 1	24.18% 4	24.29% 3	24.41% 2	20.36% 5	9.05% 12	16.42% 7	16.24% 10	16.28% 9	16.38% 8	11.40% 11
	3	17.81% 6	24.40% 1	23.82% 4	24.02% 3	24.20% 2	19.19% 5	10.65% 12	17.72% 7	17.26% 10	17.40% 9	17.60% 8	11.56% 11
	4	23.13% 6	27.48% 1	27.34% 4	27.38% 3	27.43% 2	25.37% 5	17.74% 12	21.83% 7	21.69% 10	21.73% 9	21.77% 8	19.93% 11
Spain	1	29.94% 4	29.93% 6	29.95% 1	29.94% 3	29.94% 4	29.95% 2	24.13% 10	24.11% 12	24.19% 7	24.17% 8	24.13% 10	24.15% 9
	2	29.74% 1	29.60% 6	29.64% 3	29.62% 4	29.60% 5	29.68% 2	24.17% 7	24.07% 11	24.11% 9	24.09% 10	24.07% 11	24.13% 8
	3	29.86% 1	29.84% 4	29.84% 3	29.84% 4	29.84% 4	29.85% 2	24.09% 10	24.11% 8	24.09% 10	24.09% 10	24.11% 8	24.13% 7
	4	29.89% 1	29.86% 6	29.88% 2	29.87% 4	29.87% 5	29.88% 2	23.95% 8	23.87% 12	23.93% 9	23.91% 10	23.91% 10	23.97% 7
US	1	28.28% 1	21.66% 5	18.07% 9	19.12% 7	20.28% 6	27.41% 2	23.05% 3	18.51% 8	15.92% 12	16.58% 11	17.46% 10	22.45% 4
	2	27.46% 1	24.83% 3	22.60% 6	23.30% 5	24.03% 4	26.96% 2	18.15% 7	14.82% 9	13.42% 12	13.88% 11	14.30% 10	17.00% 8
	3	29.50% 1	28.66% 6	28.76% 3	28.73% 4	28.69% 5	29.28% 2	22.65% 7	20.57% 12	21.07% 9	20.95% 10	20.77% 11	21.79% 8
	4	29.54% 6	29.70% 4	29.74% 1	29.73% 2	29.72% 3	29.67% 5	23.43% 12	23.77% 10	23.91% 7	23.89% 8	23.83% 9	23.71% 11

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the top three combination methods are highlighted.

Table 5-11 Superior Percentage and Rank of Each Combination Method for Each Origin (the First Combination Group, MAE)

origin	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE	TS-SA	TS-VACO	TS-DMSFE (.85)	TS-DMSFE (.90)	TS-DMSFE (.95)	TS-INVERSE-MAE
France	27.93% 6	29.05% 4	29.24% 1	29.19% 2	29.12% 3	28.92% 5	22.34% 12	23.99% 10	24.00% 8	24.00% 8	24.01% 7	23.53% 11
Germany	29.67% 4	29.71% 2	29.64% 6	29.66% 5	29.69% 3	29.76% 1	23.77% 12	23.90% 8	23.83% 11	23.84% 10	23.87% 9	23.93% 7
Irish Republic	29.50% 6	29.58% 4	29.60% 2	29.59% 3	29.58% 5	29.65% 1	23.91% 8	23.84% 11	23.89% 9	23.86% 10	23.84% 11	24.06% 7
Italy	18.40% 6	19.19% 5	20.84% 1	20.29% 2	19.75% 3	19.44% 4	15.77% 12	16.46% 11	17.66% 7	17.32% 8	16.90% 9	16.53% 10
the Netherlands	19.23% 6	22.08% 1	21.43% 4	21.66% 3	21.87% 2	20.14% 5	12.86% 12	16.40% 7	15.92% 10	16.09% 9	16.24% 8	13.71% 11
Spain	29.65% 6	29.70% 4	29.78% 1	29.76% 2	29.73% 3	29.67% 5	23.82% 12	23.86% 11	23.97% 7	23.93% 8	23.90% 9	23.87% 10
The US	27.80% 1	25.48% 3	24.10% 6	24.53% 5	25.00% 4	27.35% 2	20.56% 7	18.51% 9	17.85% 12	18.05% 11	18.31% 10	19.98% 8

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the top three combination methods are highlighted.

Table 5-12 Superior Percentage and Rank of Each Combination Method for Each Origin (the First Combination Group, MAPE)

origin	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE	TS-SA	TS-VACO	TS-DMSFE (.85)	TS-DMSFE (.90)	TS-DMSFE (.95)	TS-INVERSE-MAE
France	26.39% 6	27.87% 4	28.22% 1	28.14% 2	28.02% 3	27.81% 5	21.36% 12	23.28% 10	23.42% 7	23.40% 8	23.36% 9	22.90% 11
Germany	29.36% 6	29.55% 2	29.44% 5	29.47% 4	29.51% 3	29.62% 1	23.54% 12	23.82% 8	23.73% 11	23.75% 10	23.79% 9	23.83% 7
Irish Republic	29.68% 6	29.76% 5	29.78% 2	29.78% 3	29.77% 4	29.81% 1	23.94% 12	24.01% 11	24.05% 8	24.05% 9	24.03% 10	24.14% 7
Italy	19.79%	18.61%	20.22%	19.69%	19.13%	19.67%	17.46%	16.75%	17.83%	17.52%	17.16%	17.32%

	2	6	1	3	5	4	9	12	7	8	11	10
the Netherlands	28.21% 3	28.32% 2	27.95% 6	28.06% 5	28.19% 4	28.44% 1	21.25% 12	21.83% 7	21.43% 11	21.57% 10	21.69% 8	21.61% 9
Spain	29.68% 6	29.69% 4	29.79% 1	29.76% 2	29.73% 3	29.68% 5	23.89% 12	23.90% 11	24.01% 7	23.96% 8	23.93% 9	23.92% 10
the US	19.24% 1	16.41% 6	17.15% 3	16.94% 4	16.69% 5	17.84% 2	12.90% 7	10.35% 12	10.99% 9	10.77% 10	10.60% 11	11.96% 8

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the top three combination methods are highlighted.

Table 5-13 Superior Percentage and Rank of Each Combination Method for Each Origin (the First Combination Group, RMSE)

origin	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE	TS-SA	TS-VACO	TS-DMSFE (.85)	TS-DMSFE (.90)	TS-DMSFE (.95)	TS-INVERSE -MAE
France	26.02% 6	27.71% 4	28.16% 1	28.03% 2	27.88% 3	27.58% 5	21.08% 12	23.23% 10	23.44% 7	23.39% 8	23.30% 9	22.77% 11
Germany	29.82% 1	29.78% 3	29.77% 6	29.77% 5	29.78% 4	29.81% 2	23.93% 8	23.92% 9	23.89% 12	23.89% 11	23.91% 10	23.95% 7
Irish Republic	29.76% 6	29.84% 3	29.83% 5	29.83% 4	29.84% 2	29.85% 1	24.05% 12	24.23% 11	24.24% 8	24.24% 7	24.24% 8	24.23% 10
Italy	27.95% 6	28.41% 3	28.38% 4	28.43% 2	28.45% 1	28.37% 5	20.98% 12	22.01% 10	22.02% 9	22.07% 7	22.06% 8	21.81% 11
the Netherlands	17.74% 6	23.09% 1	22.97% 4	22.99% 3	23.03% 2	19.72% 5	11.58% 12	16.99% 7	16.90% 10	16.91% 9	16.96% 8	13.28% 11
Spain	29.86% 1	29.81% 6	29.83% 3	29.82% 4	29.81% 5	29.84% 2	24.08% 8	24.04% 12	24.08% 9	24.06% 10	24.05% 11	24.09% 7
the US	28.69% 1	26.21% 3	24.79% 6	25.22% 5	25.68% 4	28.33% 2	21.82% 7	19.42% 9	18.58% 12	18.82% 11	19.09% 10	21.24% 8

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the top three combination methods are highlighted.

Table 5-14 Superior Percentage and Rank of Each Combination Method for Each Forecasting Horizon (the First Combination Group)

step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE	TS-SA	TS-VACO	TS-DMSFE (.85)	TS-DMSFE (.90)	TS-DMSFE (.95)	TS-INVERSE-MAE
MAE												
1	23.68% 4	23.97% 2	23.51% 6	23.68% 5	23.83% 3	24.32% 1	19.53% 12	19.97% 8	19.65% 11	19.78% 10	19.89% 9	20.05% 7
2	26.23% 6	26.66% 1	26.47% 4	26.55% 3	26.62% 2	26.44% 5	19.58% 12	20.43% 7	20.35% 10	20.39% 9	20.42% 8	19.87% 11
3	27.16% 6	27.48% 4	27.64% 1	27.58% 2	27.53% 3	27.23% 5	20.98% 12	21.47% 10	21.79% 7	21.68% 8	21.57% 9	21.00% 11
4	27.04% 6	27.48% 5	27.88% 1	27.72% 2	27.59% 4	27.68% 3	21.63% 12	22.10% 11	22.28% 8	22.20% 9	22.15% 10	22.28% 7
MAPE												
1	25.83% 2	25.30% 6	25.30% 5	25.32% 3	25.31% 4	26.34% 1	20.93% 11	20.97% 9	20.92% 12	20.94% 10	20.98% 8	21.60% 7
2	25.69% 1	25.49% 6	25.67% 2	25.63% 3	25.57% 4	25.56% 5	20.08% 12	20.15% 10	20.29% 7	20.27% 8	20.22% 9	20.13% 11
3	25.90% 2	25.45% 6	26.06% 1	25.87% 3	25.66% 4	25.48% 5	20.04% 8	19.67% 12	20.19% 7	20.02% 9	19.85% 10	19.73% 11
4	26.78% 5	26.74% 6	27.28% 1	27.09% 3	26.91% 4	27.12% 2	21.41% 12	21.47% 11	21.72% 8	21.64% 9	21.55% 10	21.79% 7
RMSE												
1	25.28% 6	25.56% 2	25.37% 5	25.43% 4	25.49% 3	26.07% 1	20.27% 12	20.86% 8	20.71% 11	20.75% 10	20.80% 9	21.03% 7
2	27.26% 6	28.01% 1	27.68% 5	27.79% 3	27.90% 2	27.73% 4	20.27% 12	21.24% 7	21.05% 10	21.13% 9	21.18% 8	20.78% 11
3	27.48% 6	28.51% 1	28.45% 4	28.48% 3	28.50% 2	27.82% 5	21.28% 12	22.39% 10	22.40% 9	22.40% 8	22.41% 7	21.60% 11
4	28.47% 6	29.26% 1	29.20% 4	29.22% 3	29.24% 2	28.94% 5	22.47% 12	23.42% 7	23.35% 10	23.37% 9	23.39% 8	23.08% 11

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the top three combination methods are highlighted.

Table 5-15 General Comparison among Various Combination Methods (the First Combination Group, Superior Percentage and Rank)

measure	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE	TS-SA	TS-VACO	TS-DMSFE (.85)	TS-DMSFE (.90)	TS-DMSFE (.95)	TS-INVERSE -MAE
MAE	26.03% 6	26.40% 2	26.38% 5	26.38% 4	26.39% 3	26.42% 1	20.43% 12	20.99% 10	21.02% 7	21.01% 8	21.01% 9	20.80% 11
MAPE	26.05% 3	25.74% 6	26.08% 2	25.98% 4	25.86% 5	26.12% 1	20.62% 11	20.56% 12	20.78% 8	20.72% 9	20.65% 10	20.81% 7
RMSE	27.12% 6	27.84% 1	27.67% 4	27.73% 3	27.78% 2	27.64% 5	21.07% 12	21.98% 7	21.88% 10	21.91% 9	21.94% 8	21.62% 11

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the top three combination methods are highlighted.

5.3.2 Combining Traditional Econometric and Time Series Models

According to the empirical results presented in section 5.3.1, the two-stage combination methods are always inferior to the one-stage ones for all comparisons. As a result, they are excluded from the list of applied weighting schemes for combining the second and the third groups.

For the second group, the six weighting schemes for computing combination forecasts are the SA, the VACO, the DMSFE ($\alpha = 0.85, 0.90, 0.95$) and the inverse-MAE methods, and the nine individual models to be combined are the three time series techniques (ETS, state ETS and SARIMA) as well as the six traditional econometric models (traditional bounds test, traditional ADLM, traditional LI, traditional VAR, traditional TVP and traditional SD). The comparisons are conducted in three ways in accordance with what are used in combining the first group: the country-specific, the forecasting-horizon-specific and the general assessments.

5.3.2.1 Comparison across Different Origins

Firstly, the comparison is conducted at the disaggregated level, and the percentages of the superior combination forecasts compared to the best single ones and the forecasting ranking of each combination method for each forecasting horizon and each origin assessed by three measures are presented in table 5-16 to table 5-18 respectively. The highest superior percentages for each comparison are highlighted.

It is demonstrated that all superior percentages are above zero, which means that the most accurate forecasts are gained by the combination forecasting approach for all cases. Every weighting scheme has a chance to be the best performing one. The superior percentages range from 2.99% generated from the DMSFE ($\alpha = 0.85$) method for the second-step-ahead forecasts for the Dutch market judged by MAPE, to 30.08% achieved for the German case by all weighting schemes for the two-step-ahead projections measured by MAE, and by the SA and inverse-MAE methods for the one-step-ahead forecasts according to MAPE.

The forecasting length, the origin market and the accuracy measure all affect the performance ranking of different weighting schemes. For example, as with projecting tourist arrivals from the Netherlands measured by MAE, the inverse-MAE method, which is the best when the forecasting length is one quarter, ranks number five for the second- and four-step-ahead forecasts and is the third in position for the third-step-ahead projections. The SA method performs very well for the French market, ranking number one or two for eleven out of twelve comparisons, but it is the lease accurate method for the Italian origin, always staying at the bottom of the list. When the first-step-ahead forecasts for the Spanish case is concerned, the VACO method provides the best results according to MAE and MAPE, but judged by RMSE, it is only the fourth in place.

For each origin, the difference in the superior percentages produced by the same weighting scheme across four forecasting lengths is not great. A few exceptions are in seen from the American and Dutch cases. For instance, as far as tourism demand from the US is concerned, judged by MAE, the superior percentages generated from any scheme are all above 29% for the four-step-ahead forecasts, and they are all around 7% for the second-step-forecasts, resulting in the gap being more than 20%. As with the performance of forecasting tourism demand from the Netherlands judged by MAE, the difference in the superior percentages produced through the VACO and DMSFE methods for different horizons are more than 10%.

In addition, for one particular market, the difference in the superior percentages obtained by different combination methods for the same forecasting length is not great for most cases (see each row of table 5-16 to table 5-18). The biggest gap is seen from the four-step-ahead Italian market forecasting measured by MAE. The superior percentage of the best-performing DMSFE ($\alpha = 0.85$) method is 24.70%, and that of the bottom-ranked SA method is 13.55%, resulting in a gap of 11.15%.

Next, the assessment is conducted at the aggregated level, and is presented in table 5-19 to table 5-21, which shows the average percentages of the superior combination forecasts compared to the best single ones and the forecasting ranking of each combination method over the four forecasting horizons for seven origins measured by MAE, MAPE and RMSE respectively. The highest superior percentages of each comparison are highlighted.

The superior percentages range from 6.62%, which is achieved by the DMSFE ($\alpha = 0.85$) method for forecasting tourism demand from the Netherlands judged by MAPE to 29.33%, which is provided by the VACO and inverse-MAE schemes for the Irish case measured by MAPE. It means that the most accurate forecasts are always produced by the combination forecasting approach for seven origins regardless of combination method or accuracy measure, with at least 6.62% of composite forecasts being superior compared to the best constituent ones as far as one combination method for one market is concerned.

The performance ranking of different combination methods changes when the concerned origin country varies. For instance, the SA method behaves very well for the French and American markets, always ranking number one or two, but it stays at the bottom for all comparisons when it comes to the Italian and Spanish cases. Similarly, the VACO method, which provides good results for forecasting tourism demand from the Netherlands, remains at the bottom of the French list. The inverse-MAE method is the most accurate weighting scheme for the French, German and Irish origins, but it fails to behave well for the Italian and Spanish markets.

For most cases, MAE, MAPE and RMSE provide consistent assessments regarding the performance ranking of different weighting schemes. For example, for forecasting tourist arrivals from France, Germany and Irish Republic, the inverse-MAE method behaves the best regardless of accuracy measure. The DMSFE ($\alpha = 0.85$) method ranks number one as with forecasting tourism demand from Italy and Spain no matter which criterion is used. And the VACO method for the Dutch case and the SA method for the American origin are always in the top two list.

However, there are a few exceptions. For instance, the SA method for the Irish market ranks number two according to MAE and RMSE, but based on MAPE, it drops to the last place. Besides, it is the sixth in position for the German origin judged by MAE and MAPE, but it ranks number three when RMSE is used. The inverse-MAE method is placed at number five for forecasting tourism demand from the Netherlands according to MAE and RMSE, but it Is the best-performing one based on MAPE.

For the same origin, the difference among the forecasting abilities of different combination methods is small, and for some cases more than one method produce the same superior percentage (see each row of table 5-19 to table 5-21). For example, for the French market, the highest superior percentage of 25.10% judged by MAE is achieved by the SA and inverse-MAE methods, and measured by MAPE, the best result of 27.04% is provided by the SA, DMSFE ($\alpha = 0.85$) and inverse-MAE methods.

On the other hand, for the same weighting scheme, the divergence of the superior percentages generated for different markets can be considerable (see each row of table 5-19 to table 5-21). For instance, judged by MAPE, the superior percentages provided by all weighting schemes for the Dutch market are less than 9%, and those for the Irish case are greater than 28%, which means that the gap can be as large as more than 20%.

The percentages of the inferior combination forecasts compared to the worst single ones are zero for seven origins, which means that all composite forecasts are better than the worst constituent ones no matter which weighting scheme is applied, or which accuracy measure is used. Therefore, it is concluded that combination contributes to better forecasts with higher degree of accuracy and lower risk of forecasting failure for seven markets as far as the second combination group is concerned.

Table 5-16 Superior Percentage and Rank of Each Combination Method at Different Forecasting Horizons for Seven Origins (the Second Combination Group, MAE)

origin	step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse-MAE
France	1	22.51% 1	20.92% 6	22.11% 3	21.71% 4	21.31% 5	22.31% 2
	2	26.10% 2	25.10% 6	25.50% 4	25.70% 3	25.50% 4	26.29% 1
	3	26.69% 1	26.29% 5	26.49% 3	26.49% 3	26.29% 5	26.69% 1
	4	25.10% 1	24.70% 3	24.10% 6	24.30% 5	24.50% 4	25.10% 1
Germany	1	29.68% 3	29.88% 1	29.68% 3	29.68% 3	29.68% 3	29.88% 1
	2	30.08% 1	30.08% 1	30.08% 1	30.08% 1	30.08% 1	30.08% 1
	3	29.08% 1	29.08% 1	29.08% 1	29.08% 1	29.08% 1	29.08% 1
	4	25.30% 6	26.29% 2	26.10% 3	26.10% 3	26.10% 3	26.69% 1
Irish Republic	1	26.29% 2	26.10% 3	25.50% 6	25.70% 5	25.90% 4	26.49% 1
	2	28.88% 1	28.09% 3	27.69% 5	27.69% 5	27.89% 4	28.88% 1
	3	29.48% 6	29.88% 1	29.68% 4	29.88% 1	29.88% 1	29.68% 4
	4	28.69% 5	28.88% 2	28.49% 6	28.88% 2	28.88% 2	29.28% 1
Italy	1	15.34% 6	21.71% 4	22.91% 1	22.31% 2	21.91% 3	19.92% 5
	2	18.33% 6	25.10% 4	26.10% 1	25.70% 2	25.50% 3	22.71% 5
	3	21.51% 6	25.50% 4	26.10% 1	25.70% 2	25.70% 2	24.90% 5
	4	13.55% 6	22.11% 4	24.70% 1	23.90% 2	23.11% 3	21.12% 5
the Netherlands	1	10.96% 4	11.75% 2	9.76% 6	10.56% 5	11.16% 3	12.55% 1
	2	11.75% 6	15.94% 1	15.34% 4	15.54% 3	15.94% 1	13.15% 5
	3	7.97% 6	10.36% 1	8.17% 5	8.76% 3	9.76% 2	8.76% 3
	4	13.75% 6	23.11% 1	22.31% 4	22.91% 2	22.91% 2	18.33% 5
Spain	1	29.08% 6	29.28% 1	29.28% 1	29.28% 1	29.28% 1	29.28% 1
	2	24.90% 5	25.10% 3	25.30% 1	25.30% 1	25.10% 3	24.90% 5

	3	25.50% 6	26.89% 4	27.29% 1	27.29% 1	27.09% 3	25.90% 5
	4	25.50% 6	26.89% 4	27.29% 1	27.29% 1	27.09% 3	26.49% 5
US	1	26.89% 1	21.51% 3	19.52% 6	20.12% 5	20.92% 4	26.10% 2
	2	6.57% 5	7.17% 1	6.97% 2	6.97% 2	6.97% 2	6.37% 6
	3	15.34% 1	14.14% 5	15.34% 1	14.74% 3	14.74% 3	13.75% 6
	4	29.48% 1	29.08% 3	29.08% 3	29.08% 3	29.08% 3	29.28% 2

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

Table 5-17 Superior Percentage and Rank of Each Combination Method at Different Forecasting Horizons for Seven Origins (the Second Combination Group, MAPE)

origin	step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse-MAE
France	1	22.91% 2	21.71% 6	23.31% 1	22.91% 2	22.31% 5	22.91% 2
	2	27.89% 2	27.29% 5	27.89% 2	27.69% 4	27.29% 5	28.09% 1
	3	29.68% 1	29.68% 1	29.68% 1	29.68% 1	29.68% 1	29.68% 1
	4	27.69% 1	27.29% 3	27.29% 3	27.29% 3	27.29% 3	27.49% 2
Germany	1	30.08% 1	29.88% 3	29.88% 3	29.88% 3	29.88% 3	30.08% 1
	2	29.48% 1	29.48% 1	29.28% 4	29.28% 4	29.28% 4	29.48% 1
	3	26.89% 6	27.49% 1	27.49% 1	27.49% 1	27.49% 1	27.29% 5
	4	22.11% 6	23.71% 2	23.71% 2	23.71% 2	23.71% 2	23.90% 1
Irish Republic	1	28.49% 3	28.69% 2	28.09% 6	28.29% 5	28.49% 3	28.88% 1
	2	29.08% 2	29.08% 2	28.49% 6	28.69% 5	28.88% 4	29.48% 1
	3	28.88% 6	29.88% 1	29.68% 2	29.68% 2	29.68% 2	29.68% 2
	4	28.88% 6	29.68% 1	29.68% 1	29.68% 1	29.68% 1	29.28% 5
Italy	1	16.93% 6	22.11% 4	22.71% 1	22.51% 2	22.31% 3	20.72% 5
	2	17.73% 6	24.30% 4	25.50% 1	25.30% 2	24.90% 3	21.51% 5
	3	19.52% 6	23.11% 5	24.10% 1	23.31% 2	23.31% 2	23.31% 2
	4	14.14% 6	21.31% 4	24.30% 1	23.11% 2	22.11% 3	21.31% 4

the Netherlands	1	9.96% 2	8.96% 3	6.77% 6	7.37% 5	8.17% 4	11.16% 1
	2	4.18% 2	3.78% 3	2.99% 6	3.19% 5	3.39% 4	4.58% 1
	3	6.57% 2	5.18% 3	3.59% 6	3.78% 5	4.38% 4	7.17% 1
	4	9.16% 6	14.94% 1	13.15% 4	13.75% 3	14.54% 2	11.95% 5
Spain	1	28.29% 6	28.88% 1	28.88% 1	28.88% 1	28.88% 1	28.88% 1
	2	24.50% 5	24.70% 4	25.10% 1	24.90% 2	24.90% 2	24.30% 6
	3	26.10% 6	27.29% 4	27.89% 1	27.89% 1	27.69% 3	26.49% 5
	4	25.50% 6	26.89% 2	27.09% 1	26.89% 2	26.89% 2	26.29% 5
US	1	27.09% 2	26.69% 3	26.10% 6	26.29% 5	26.49% 4	27.49% 1
	2	27.29% 2	26.49% 3	26.29% 5	26.29% 5	26.49% 3	27.49% 1
	3	29.88% 1	29.68% 2	29.68% 2	29.68% 2	29.68% 2	29.68% 2
	4	28.09% 6	28.49% 1	28.29% 5	28.49% 1	28.49% 1	28.49% 1

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

Table 5-18 Superior Percentage and Rank of Each Combination Method at Different Forecasting Horizons for Seven Origins (the Second Combination Group, RMSE)

origin	step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse-MAE
France	1	27.09% 1	26.10% 6	26.89% 3	26.69% 4	26.49% 5	27.09% 1
	2	27.29% 2	26.49% 6	27.09% 3	26.89% 4	26.69% 5	27.49% 1
	3	29.68% 5	29.88% 1	29.88% 1	29.88% 1	29.88% 1	29.68% 5
	4	27.69% 1	27.49% 3	27.29% 5	27.29% 5	27.49% 3	27.69% 1
Germany	1	29.28% 1	28.88% 3	28.88% 3	28.88% 3	28.88% 3	29.08% 2
	2	27.69% 1	27.49% 3	26.89% 6	27.29% 4	27.29% 4	27.69% 1
	3	27.89% 1	27.49% 3	27.29% 5	27.29% 5	27.49% 3	27.69% 2
	4	24.30% 6	25.50% 2	25.30% 5	25.50% 2	25.50% 2	25.70% 1
Irish Republic	1	28.49% 6	29.28% 1	28.88% 4	29.08% 2	29.08% 2	28.88% 4
	2	28.29% 1	26.89% 3	26.49% 5	26.49% 5	26.69% 4	28.09% 2

	3	28.88% 1	27.89% 3	27.29% 6	27.49% 4	27.49% 4	28.88% 1
	4	28.49% 2	28.29% 3	28.09% 4	28.09% 4	28.09% 4	28.88% 1
Italy	1	24.50% 6	28.09% 2	28.29% 1	28.09% 2	27.89% 4	27.49% 5
	2	26.49% 6	28.29% 4	28.69% 1	28.49% 2	28.49% 2	28.29% 4
	3	27.29% 6	28.69% 4	29.08% 1	28.88% 2	28.88% 2	28.49% 5
	4	27.69% 6	29.08% 1	29.08% 1	29.08% 1	29.08% 1	28.49% 5
the Netherlands	1	10.96% 6	12.55% 1	11.95% 4	12.15% 3	12.35% 2	11.35% 5
	2	18.92% 6	25.10% 2	24.90% 4	25.10% 2	25.30% 1	21.12% 5
	3	16.33% 6	23.90% 1	22.91% 4	23.11% 3	23.71% 2	17.73% 5
	4	22.51% 6	27.69% 1	27.29% 4	27.49% 2	27.49% 2	25.50% 5
Spain	1	29.28% 4	29.28% 4	29.48% 1	29.48% 1	29.28% 4	29.48% 1
	2	26.89% 4	27.09% 2	27.29% 1	27.09% 2	26.89% 4	26.89% 4
	3	28.09% 6	28.49% 4	28.88% 1	28.69% 2	28.69% 2	28.49% 4
	4	22.31% 6	24.50% 3	24.90% 1	24.90% 1	24.50% 3	23.90% 5
US	1	26.69% 1	22.11% 3	19.72% 6	20.92% 5	21.51% 4	25.70% 2
	2	18.73% 1	17.73% 6	17.93% 3	17.93% 3	17.93% 3	18.33% 2
	3	29.28% 3	29.08% 4	29.68% 1	29.48% 2	29.08% 4	29.08% 4
	4	27.49% 6	27.89% 1	27.69% 3	27.69% 3	27.89% 1	27.69% 3

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

Table 5-19 Superior Percentage and Rank of Each Combination Method for Each Origin (the Second Combination Group, MAE)

origin	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE
France	25.10% 1	24.25% 6	24.55% 3	24.55% 3	24.40% 5	25.10% 1
Germany	28.54% 6	28.83% 2	28.74% 3	28.74% 3	28.74% 3	28.93% 1
Irish Republic	28.34% 2	28.24% 3	27.84% 6	28.04% 5	28.14% 4	28.59% 1
Italy	17.18% 6	23.61% 4	24.95% 1	24.40% 2	24.05% 3	22.16% 5
the Netherlands	11.11% 6	15.29% 1	13.89% 4	14.44% 3	14.94% 2	13.20% 5
Spain	26.25% 6	27.04% 4	27.29% 1	27.29% 1	27.14% 3	26.64% 5
the US	19.57% 1	17.98% 3	17.73% 5	17.73% 5	17.93% 4	18.87% 2

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

Table 5-20 Superior Percentage and Rank of Each Combination Method for Each Origin (the Second Combination Group, MAPE)

origin	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE
France	27.04% 1	26.49% 6	27.04% 1	26.89% 4	26.64% 5	27.04% 1
Germany	27.14% 6	27.64% 2	27.59% 3	27.59% 3	27.59% 3	27.69% 1
Irish Republic	28.83% 6	29.33% 1	28.98% 5	29.08% 4	29.18% 3	29.33% 1
Italy	17.08% 6	22.71% 4	24.15% 1	23.56% 2	23.16% 3	21.71% 5
the Netherlands	7.47% 4	8.22% 2	6.62% 6	7.02% 5	7.62% 3	8.72% 1
Spain	26.10% 6	26.94% 4	27.24% 1	27.14% 2	27.09% 3	26.49% 5
the US	28.09% 2	27.84% 3	27.59% 6	27.69% 5	27.79% 4	28.29% 1

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

Table 5-21 Superior Percentage and Rank of Each Combination Method for Each Origin (the Second Combination Group, RMSE)

origin	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE
France	27.94% 2	27.49% 6	27.79% 3	27.69% 4	27.64% 5	27.99% 1
Germany	27.29% 3	27.34% 2	27.09% 6	27.24% 5	27.29% 3	27.54% 1
Irish Republic	28.54% 2	28.09% 3	27.69% 6	27.79% 5	27.84% 4	28.69% 1
Italy	26.49% 6	28.54% 4	28.78% 1	28.64% 2	28.59% 3	28.19% 5
the Netherlands	17.18% 6	22.31% 1	21.76% 4	21.96% 3	22.21% 2	18.92% 5
Spain	26.64% 6	27.34% 3	27.64% 1	27.54% 2	27.34% 3	27.19% 5
the US	25.55% 1	24.20% 3	23.75% 6	24.00% 5	24.10% 4	25.20% 2

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination method for each comparison are highlighted.

5.3.2.2 Comparison over Different Forecasting Horizons

To reveal the performance of different combination methods at different forecasting horizons, the average percentage of the superior combination forecasts compared to the best single ones and the forecasting ranking of each combination method over seven origins for four forecasting horizons measured by three measures are presented in table 5-22. The highest superior percentages of each comparison are highlighted It is shown that for each forecasting length, combining individual forecasts are beneficial with superior percentages being above 20% for all cases, which means that at least 20% of composite forecasts are better than the best constituent ones no matter which weighting scheme is used for four forecasting lengths.

The highest superior percentage is 27.92%, which is achieved by the VACO method for the third-step-ahead forecasts measured by RMSE, and the lowest superior percentage is 20.94%, which is obtained by the SA scheme for the two-step-ahead projections judged by MAE.

The performance ranking of different combination methods depends on the forecasting length under study and accuracy measure applied.

For the one-step-ahead forecasts, the inverse-MAE method behaves the best regardless of accuracy measure, and the least accurate scheme is the DMSFE ($\alpha = 0.85$) method based on MAE and RMSE, and the SA method according to MAPE. When the forecasting length is two quarters, the DMSFE methods are superior to others, while the SA method is always the last in position no matter which accuracy measure is used. Different accuracy measures provide controversial results as to which is the best performing weighting scheme as far as the three- and four-step-ahead forecasts are concerned: the best ones are the DMSFE ($\alpha = 0.95$) and DMSFE ($\alpha = 0.90$) methods based on MAE; the inverse-MAE and the DMSFE ($\alpha = 0.85$) according to MAPE; and the VACO method judged by RMSE. However, three accuracy measures yield consistent assessment regarding the least accurate combination methods for these two forecasting lengths: the SA method always performs the worst.

The difference among the performance of different weighting schemes for a specific forecasting length is quite small (see each row of table 5-22). For example, measured by RMSE, the biggest gaps in the superior percentages achieved by the most and least accurate weighting schemes are merely 0.71% (between 25.58% and 24.87%), 0.71% (between 25.61% and 24.90%), 1.14% (between 27.92% and 26.78%) and 1.43 (between 27.21% and 25.78%) for one- to four-step-ahead forecasts respectively.

Besides, the same weighting scheme generate similar superior percentages for different forecasting lengths (see each column of table 5-22). For instance, the superior percentages produced by the inverse-MAE method range from 21.77% for the second-step-ahead forecasts measured by MAE to 27.15% for the third-step-ahead forecasts judged by RMSE, resulting in a difference of 5.38%.

The percentages of the inferior combination forecasts compared to the worst single ones are zero for all cases, which means that all composite forecasts are better than the worst constituent ones regardless of weighting scheme or accuracy measure for four forecasting lengths. It shows that the least accurate forecasts are not generated through the combination forecasting approach. Therefore, it is concluded that the combination forecasting approach is superior with higher degree of accuracy and lower risk of forecasting failure for one- to four-step-ahead forecasts for the second combination group.

Table 5-22 Superior Percentage and Rank of Each Combination Method for Each Forecasting Horizon (the Second Combination Group)

step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse-MAE
MAE						
1	22.96% 3	23.02% 2	22.68% 6	22.77% 5	22.88% 4	23.79% 1
2	20.94% 6	22.37% 4	22.43% 1	22.43% 1	22.43% 1	21.77% 5
3	22.22% 6	23.16% 3	23.16% 2	23.13% 4	23.22% 1	22.68% 5
4	23.05% 6	25.87% 4	26.01% 2	26.07% 1	25.95% 3	25.18% 5
MAPE						
1	23.39% 6	23.85% 2	23.68% 5	23.73% 4	23.79% 3	24.30% 1
2	22.88% 6	23.59% 4	23.65% 1	23.62% 2	23.59% 3	23.56% 5
3	23.93% 6	24.62% 2	24.59% 3	24.50% 5	24.56% 4	24.76% 1
4	22.22% 6	24.62% 4	24.79% 1	24.70% 2	24.67% 3	24.10% 5
RMSE						
1	25.18% 3	25.18% 2	24.87% 6	25.04% 5	25.07% 4	25.58% 1
2	24.90% 6	25.58% 4	25.61% 1	25.61% 1	25.61% 1	25.41% 5
3	26.78% 6	27.92% 1	27.86% 3	27.83% 4	27.89% 2	27.15% 5
4	25.78% 6	27.21% 1	27.09% 4	27.15% 2	27.15% 3	26.84% 5

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

5.3.2.3 General Comparison among Various Weighting Schemes

The general comparison is conducted by taking averages over four forecasting horizons and seven origins for each weighting scheme, and the average percentages of the superior combination forecasts compared to the best single ones and the forecasting ranking of each combination method measured by MAE, MAPE and RMSE are presented in table 5-23. The highest superior percentages of each comparison are highlighted.

It is demonstrated that all superior percentages are above 22%, which means that at least 22% of combination forecasts produced by any weighting scheme are better than the best single projections, suggesting that combing individual forecasts can improve forecasting accuracy irrespective of weighting scheme or accuracy measure. The highest superior percentage of 26.47% is achieved by the VACO method measured by RMSE, and the lowest superior percentage of 22.30% is produced by the SA scheme judged by MAE. The inverse-MAE method performs satisfactorily, ranking number one based on MAPE.

The performance ranking of different weighting schemes changes according to accuracy measure. The DMSFE ($\alpha = 0.95$) method generates the most accurate combination forecasts when MAE is used to assess accuracy, followed by the VACO and DMSFE ($\alpha = 0.90$) methods. Judged by MAPE, the inverse-MAE, the DMSFE ($\alpha = 0.85$) and the VACO are the best three schemes. The top-three performing methods based on RMSE are the VACO, the DMSFE ($\alpha = 0.95$) and the DMSFE ($\alpha = 0.90$) schemes. The SA method remains at the bottom no matter which accuracy measure is used.

A comparison with the results generated by combining all models in the first group (see table 5-15) shows that, the first-placed methods based on MAPE and RMSE are the same (being the inverse-MAE and the VACO methods respectively), and the SA scheme is inferior to other one-stage combination methods according to MAE and RMSE.

The difference between the performances of different weighting schemes is small. The biggest gaps between the superior percentages generated by the best and the worst weighting schemes are 1.32% based on MAE (between 22.30% of the SA method and 23.62% of the DMSFE ($\alpha = 0.95$) method), 1.07% according to MAPE (between 23.11% of the SA method and 24.18% of the inverse-MAE method) and 0.81% measured by RMSE (between 25.66% of the SA method and 26.47% of the VACO method). It means that there is no great difference in the forecasting ability of different weighting schemes.

The percentages of the inferior combination forecasts compared to the worst single ones are zero for all cases, which means that all composite forecasts are better than the worst constituent ones regardless of weighting scheme or accuracy measure. It shows that the least accurate forecasts are not generated through the combination forecasting approach. Therefore, it is concluded that combining individual forecasts are beneficial with higher degree of accuracy and lower risk of forecasting failure as with the second combination group.

Table 5-23 General Comparison among Various Combination Methods (the Second Combination Group, Superior Percentage and Rank)

measure	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE
MAE	22.30% 6	23.61% 2	23.57% 4	23.60% 3	23.62% 1	23.36% 5
MAPE	23.11% 6	24.17% 3	24.17% 2	24.14% 5	24.15% 4	24.18% 1
RMSE	25.66% 6	26.47% 1	26.36% 4	26.41% 3	26.43% 2	26.25% 5

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best combination method are highlighted.

5.3.3 Combining Climate Econometric and Time Series Models

In accordance with combining the first and second groups, the comparisons are conducted in three ways: the country-specific assessment, the forecasting-horizon-specific evaluation and the general comparison.

In this group, the applied six one-stage weighting schemes include the SA, the VACO, the DMSFE ($\alpha = 0.85, 0.90, 0.95$) and the inverse-MAE methods, and the nine individual models to be combined consist of three time series techniques (ETS, state ETS and SARIMA) as well as the six climate econometric models (climate bounds test, climate ADLM, climate LI, climate VAR, climate TVP and climate SD).

5.3.3.1 Comparison across Different Origins

Firstly, to conduct the comparison at the disaggregated level, the percentages of the superior combination forecasts compared to the best single ones and the forecasting ranking of each combination method for each forecasting horizon and each origin are presented in table 5-24 to table 5-26. The highest superior percentages of each comparison are highlighted.

As shown in table 5-24 to table 5-26, all superior percentages are above 4%, which implies that the most accurate forecasts are produced by the combination forecasting approach for all cases. Each weighting scheme has a chance to generate the best results. The highest superior percentage of 30.28% is generated by the SA, VACO, DMSFE ($\alpha = 0.90$), DMSFE ($\alpha = 0.95$) and inverse-MAE methods for the third-step-ahead forecasts and by all schemes for the four-step-ahead projections for the Irish market judged by RMSE. The lowest superior percentage of 4.58% is seen from the American case when MAE is used as accuracy measure, which is obtained by the SA method for the four-step-ahead forecasts.

The performance ranking of different combination forecasts varies according to forecasting length, origin market as well as accuracy measure. For example, for forecasting performance for the American case, the SA method is superior to others when the forecasting length is one quarter, but it ranks number five of six when the three- and four-step-ahead forecasts are concerned measured by MAE and RMSE; and based on MAPE, it is the best choice for all forecasting horizons except the three one. According to MAE and RMSE, the inverse-MAE method is the best one for forecasting tourism demand from Irish Republic for all four forecasting horizons, but it performs unsatisfactorily for the French market. The SA method, which ranks number one according to RMSE for one- to four-step-ahead forecasts for the German origin, stays at the bottom when accuracy is measured by MAPE.

For each origin, the difference in the superior percentages gained through the same combination method across four forecasting lengths is not great for most cases. The biggest difference is seen from the American case. Measured by MAE, the superior percentages produced by all weighting methods for the one-step-ahead forecasts are much higher than those for other forecasting lengths, with the greatest gap being 25.30% provided by the SA method (between 4.58% for the four-step-ahead forecasts and 29.88% for the one-step-ahead forecasts).

In addition, there exists no great divergence in the superior percentages of different combination methods for the same forecasting length concerning one particular market (see each row of table 5-24 to table 5-26). Again, the American market provides the biggest gap, which is 7.97% between the highest percentage of 27.89% of the SA method and the lowest one of 19.92% of the DMSFE ($\alpha = 0.85$) method judged by RMSE for the one-step-ahead forecasts.

Next, for an evaluation at the aggregated level, table 5-27 to table 5-29 demonstrate the average percentages of the superior combination forecasts compared to the best single ones and the forecasting ranking of each combination method over the four forecasting horizons for seven origins measured by MAE, MAPE and RMSE respectively. The highest superior percentages for each comparison are highlighted.

The superior percentages range from 11.70%, which is achieved by the DMSFE ($\alpha = 0.85$) method for forecasting tourism demand from the US judged by MAE, to 30.13%, which is provided by all weighting schemes except the DMSFE ($\alpha = 0.85$) one for the Irish case measured by RMSE. It means that the best forecasts are always produced through the combination forecasting approach for seven origins regardless of weighting scheme, with at least 11.70% composite forecasts being more accurate than the best constituent one as far as one combination method for one market is concerned.

The performance ranking of the same weighting scheme changes greatly for different origin countries. For example, the VACO method, which claims the first place for forecasting tourist arrivals from the Netherlands, and provides good results for the Irish, Spanish and American cases, performs unwell when it comes to the Italian market. The DMSFE ($\alpha = 0.85$) method is the best choice for the French and Italian markets regardless of accuracy measure, but it always stays at the bottom as far as tourism demand from Spain is concerned. The inverse-MAE scheme is the best-performing one for the Irish and Spanish markets, but it performs unsatisfactorily for the French, Italian and Dutch markets.

MAE, MAPE and RMSE shows consistent evaluations regarding the performance ranking of different weighting schemes for the French and Dutch cases. But for other markets, different measures can provide controversial results. For instance, the SA method, which performs very well for forecasting tourism demand from Irish Republic and the US regardless of accuracy measure and ranks number one for the German origin judged by MAE and RMSE, drops to the bottom in the German list measured by MAPE. For the Irish case, the DMSFE ($\alpha = 0.95$) method is the first in position based on RMSE, but it only ranks number five according to MAE and number four judged by MAPE.

For the same origin, the difference among the performance of various combination methods is small (see each row of table 5-27 to table 5-29). For example, for forecasting tourism demand from Irish Republic, more than one method provides the same result when RMSE is used to measure accuracy. The superior percentages generated through all weighting schemes are all between 29% and 30% for the Irish case measured by MAPE. The biggest gap in the superior percentages is seen from the Dutch market judged by RMSE, which is 6.42% between 21.41% gained by the VACO method and 14.99% produced by the SA scheme.

On the other hand, the same weighting scheme can generate quite different superior percentages for different origins (see each column of table 5-27 to table 5-29). For instance, according to MAE, the superior percentage generated by the SA method is 29.63% for the Irish origin, and 12.25% for the American case, and the gap is as large as 17.38%. Similarly, based on RMSE, the highest superior percentage gained by the inverse-MAE method is 30.13% for the Irish case, and the lowest one is 13.40% for the American origin, which means that the greatest difference in the superior percentage is 16.73%.

The percentages of the inferior combination forecasts compared to the worst single ones are zero for seven origins, which means that all composite forecasts are better than the worst constituent ones no matter which weighting scheme or accuracy measure is used. As a result, it is concluded that for seven origins under consideration, forecasting performance can be improved by combining climate econometric models and time series models as far as combining the third group is concerned.

Table 5-24 Superior Percentage and Rank of Each Combination Method at Different Forecasting Horizons for Seven Origins (the Third Combination Group, MAE)

origin	step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse-MAE
France	1	11.75% 6	14.94% 4	16.93% 1	16.33% 2	15.74% 3	12.55% 5
	2	19.92% 6	22.71% 4	24.50% 1	23.71% 2	23.11% 3	22.71% 4
	3	12.15% 5	13.35% 4	13.55% 1	13.55% 1	13.55% 1	11.55% 6
	4	13.75% 6	16.93% 1	16.73% 3	16.73% 3	16.93% 1	14.34% 5
Germany	1	27.89% 1	27.09% 3	26.10% 6	26.29% 5	26.69% 4	27.29% 2
	2	25.50% 4	25.90% 1	25.10% 6	25.50% 4	25.70% 3	25.90% 1
	3	27.29% 3	27.49% 1	26.69% 6	27.09% 5	27.29% 3	27.49% 1
	4	28.49% 1	28.49% 1	27.49% 6	27.89% 5	28.29% 3	28.29% 3
Irish Republic	1	28.88% 2	28.49% 3	28.29% 4	28.29% 4	28.29% 4	29.08% 1
	2	30.08% 1	29.88% 3	29.68% 6	29.88% 3	29.88% 3	30.08% 1
	3	29.88% 1	29.68% 3	29.68% 3	29.68% 3	29.68% 3	29.88% 1
	4	29.68% 1	29.48% 5	29.68% 1	29.68% 1	29.48% 5	29.68% 1
Italy	1	12.55% 3	12.15% 5	13.94% 1	13.15% 2	12.35% 4	11.55% 6
	2	18.33% 4	17.73% 5	20.72% 1	19.92% 2	18.92% 3	16.93% 6
	3	10.96% 1	9.56% 6	10.96% 1	10.16% 3	9.76% 5	9.96% 4
	4	11.55% 4	10.36% 6	12.95% 1	12.15% 2	11.16% 5	11.75% 3
the Netherlands	1	12.75% 6	17.33% 1	17.33% 1	17.33% 1	17.13% 4	14.74% 5
	2	16.53% 6	22.51% 1	22.51% 1	22.51% 1	22.51% 1	18.33% 5
	3	19.92% 5	24.30% 1	23.51% 4	23.71% 3	23.90% 2	19.52% 6
	4	25.70% 6	28.49% 1	28.49% 1	28.49% 1	28.49% 1	26.29% 5
Spain	1	28.49% 2	28.49% 2	28.09% 6	28.29% 4	28.29% 4	28.69% 1
	2	28.09% 3	28.29% 2	27.69% 6	27.89% 5	28.09% 3	28.49% 1
	3	28.88% 5	29.08% 1	28.88% 5	29.08% 1	29.08% 1	29.08% 1
	4	27.89% 6	28.29% 1	28.09% 4	28.29% 1	28.29% 1	28.09% 4
US	1	29.88% 1	28.49% 3	25.30% 6	26.49% 5	27.69% 4	29.68% 2
	2	7.77% 1	7.77% 1	6.37% 6	6.77% 5	7.17% 4	7.77% 1
	3	6.77% 5	6.97% 4	7.97% 1	7.57% 2	7.17% 3	6.18% 6

	4	4.58% 6	6.37% 4	7.17% 1	6.97% 2	6.57% 3	5.18% 5

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

Table 5-25 Superior Percentage and Rank of Each Combination Method at Different Forecasting Horizons for Seven Origins (the Third Combination Group, MAPE)

origin	step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse-MAE
France	1	12.55% 6	18.33% 4	20.12% 1	19.92% 2	19.32% 3	15.14% 5
	2	18.92% 6	22.31% 4	23.71% 1	23.11% 2	22.51% 3	21.31% 5
	3	14.74% 6	20.32% 4	20.52% 3	20.72% 1	20.72% 1	15.74% 5
	4	16.33% 6	20.32% 2	20.52% 1	20.32% 2	20.32% 2	17.93% 5
Germany	1	23.31% 6	24.30% 2	23.90% 4	23.90% 4	24.10% 3	24.90% 1
	2	22.71% 6	25.10% 1	24.30% 5	24.70% 4	24.90% 3	25.10% 1
	3	26.49% 6	27.69% 1	27.29% 3	27.29% 3	27.29% 3	27.49% 2
	4	24.10% 6	26.10% 1	25.30% 4	25.50% 3	25.70% 2	25.30% 4
Irish Republic	1	27.89% 2	27.29% 3	27.09% 4	27.09% 4	27.09% 4	28.29% 1
	2	30.08% 1	29.88% 3	29.68% 4	29.68% 4	29.68% 4	30.08% 1
	3	29.88% 1	29.68% 2	29.68% 2	29.68% 2	29.68% 2	29.68% 2
	4	29.88% 1	29.68% 3	29.68% 3	29.68% 3	29.68% 3	29.88% 1
Italy	1	14.14% 1	11.95% 6	13.55% 2	13.35% 3	12.55% 4	12.55% 4
	2	17.93% 3	16.33% 5	19.32% 1	18.53% 2	17.73% 4	16.14% 6
	3	13.15% 1	9.96% 6	11.55% 3	10.96% 4	10.56% 5	12.15% 2
	4	14.54% 3	13.55% 6	15.54% 1	14.54% 3	13.75% 5	14.94% 2
the Netherlands	1	17.93% 6	21.51% 1	20.92% 4	21.31% 2	21.31% 2	20.52% 5
	2	28.09% 1	27.49% 3	27.49% 3	27.49% 3	27.49% 3	27.89% 2
	3	29.08% 1	28.69% 2	28.09% 5	28.29% 4	28.49% 3	28.09% 5
	4	28.49% 6	29.08% 1	29.08% 1	29.08% 1	29.08% 1	28.88% 5
Spain	1	27.09% 2	26.69% 3	26.10% 6	26.29% 5	26.49% 4	27.49% 1
	2	27.29% 2	26.49% 3	26.29% 5	26.29% 5	26.49% 3	27.49% 1
	3	29.88% 1	29.68% 2	29.68% 2	29.68% 2	29.68% 2	29.68% 2
	4	28.09%	28.49%	28.29%	28.49%	28.49%	28.49%

		6	1	5	1	1	1
US	1	29.68% 1	25.50% 3	23.11% 6	23.71% 5	24.70% 4	29.08% 2
	2	13.75% 1	13.15% 2	11.75% 6	12.35% 5	12.75% 3	12.55% 4
	3	11.75% 4	11.55% 5	13.15% 1	12.75% 2	12.15% 3	10.16% 6
	4	14.74% 1	12.75% 6	14.14% 2	13.55% 3	13.15% 4	13.15% 4

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

Table 5-26 Superior Percentage and Rank of Each Combination Method at Different Forecasting Horizons for Seven Origins (the Third Combination Group, RMSE)

origin	step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse-MAE
France	1	13.75% 6	19.12% 4	21.51% 1	20.72% 2	20.32% 3	17.73% 5
	2	21.31% 6	23.90% 4	25.90% 1	25.50% 2	24.70% 3	23.51% 5
	3	19.72% 6	22.11% 4	23.11% 1	22.91% 2	22.51% 3	20.72% 5
	4	23.11% 6	24.30% 4	24.90% 1	24.50% 3	24.30% 4	24.70% 2
Germany	1	30.08% 1	30.08% 1	29.88% 5	29.88% 5	30.08% 1	30.08% 1
	2	29.68% 1	29.28% 3	29.08% 6	29.28% 3	29.28% 3	29.68% 1
	3	30.08% 1	29.88% 3	29.68% 5	29.68% 5	29.88% 3	30.08% 1
	4	28.49% 1	27.89% 2	27.09% 6	27.49% 5	27.69% 4	27.89% 2
Irish Republic	1	29.88% 1	29.88% 1	29.68% 6	29.88% 1	29.88% 1	29.88% 1
	2	30.08% 1	30.08% 1	30.08% 1	30.08% 1	30.08% 1	30.08% 1
	3	30.28% 1	30.28% 1	30.08% 6	30.28% 1	30.28% 1	30.28% 1
	4	30.28% 1	30.28% 1	30.28% 1	30.28% 1	30.28% 1	30.28% 1
Italy	1	26.69% 2	26.10% 6	26.89% 1	26.69% 2	26.49% 4	26.29% 5
	2	28.09% 1	27.69% 5	27.89% 2	27.89% 2	27.89% 2	27.69% 5
	3	20.52% 6	22.71% 4	23.71% 1	23.51% 2	23.31% 3	22.51% 5
	4	26.89% 6	27.69% 1	27.49% 3	27.49% 3	27.49% 3	27.69% 1
the Netherlands	1	11.55% 6	14.94% 4	15.34% 1	15.34% 1	15.14% 3	12.55% 5
	2	14.34% 6	22.31% 1	21.71% 4	21.91% 3	22.11% 2	17.13% 5
	3	14.14% 6	22.11% 1	21.51% 4	21.91% 2	21.91% 2	15.54% 5
	4	19.92% 6	26.29% 1	26.10% 3	26.10% 3	26.29% 1	23.11% 5

Spain	1	28.88% 3	29.08% 1	28.49% 6	28.69% 4	28.69% 4	29.08% 1
	2	29.28% 1	29.08% 3	28.88% 6	29.08% 3	29.08% 3	29.28% 1
	3	29.68% 1	29.68% 1	29.68% 1	29.68% 1	29.68% 1	29.68% 1
	4	29.68% 1	29.68% 1	29.68% 1	29.68% 1	29.68% 1	29.68% 1
US	1	27.89% 1	22.91% 3	19.92% 6	20.72% 5	21.71% 4	27.29% 2
	2	11.35% 1	9.96% 3	8.57% 6	8.76% 5	9.36% 4	10.96% 2
	3	6.18% 5	6.97% 4	7.97% 1	7.57% 2	7.37% 3	5.78% 6
	4	8.37% 6	10.96% 4	12.35% 1	12.15% 2	11.55% 3	9.56% 5

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

Table 5-27 Superior Percentage and Rank of Each Combination Method for Each Origin (the Third Combination Group, MAE)

origin	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE
France	14.39% 6	16.98% 4	17.93% 1	17.58% 2	17.33% 3	15.29% 5
Germany	27.29% 1	27.24% 2	26.34% 6	26.69% 5	26.99% 4	27.24% 2
Irish Republic	29.63% 2	29.38% 3	29.33% 5	29.38% 3	29.33% 5	29.68% 1
Italy	13.35% 3	12.45% 6	14.64% 1	13.84% 2	13.05% 4	12.55% 5
the Netherlands	18.73% 6	23.16% 1	22.96% 4	23.01% 2	23.01% 2	19.72% 5
Spain	28.34% 5	28.54% 2	28.19% 6	28.39% 4	28.44% 3	28.59% 1
the US	12.25% 2	12.40% 1	11.70% 6	11.95% 5	12.15% 4	12.20% 3

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

Table 5-28 Superior Percentage and Rank of Each Combination Method for Each Origin (the Third Combination Group, MAPE)

origin	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE
France	15.64% 6	20.32% 4	21.22% 1	21.02% 2	20.72% 3	17.53% 5
Germany	24.15% 6	25.80% 1	25.20% 5	25.35% 4	25.50% 3	25.70% 2
Irish Republic	29.43% 2	29.13% 3	29.03% 4	29.03% 4	29.03% 4	29.48% 1
Italy	14.94% 2	12.95% 6	14.99% 1	14.34% 3	13.65% 5	13.94% 4
the Netherlands	25.90% 6	26.69% 1	26.39% 4	26.54% 3	26.59% 2	26.34% 5

Spain	28.09% 2	27.84% 3	27.59% 6	27.69% 5	27.79% 4	28.29% 1
the US	17.48% 1	15.74% 3	15.54% 6	15.59% 5	15.69% 4	16.24% 2

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

Table 5-29 Superior Percentage and Rank of Each Combination Method for Each Origin (the Third Combination Group, RMSE)

origin	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE
France	19.47% 6	22.36% 4	23.85% 1	23.41% 2	22.96% 3	21.66% 5
Germany	29.58% 1	29.28% 3	28.93% 6	29.08% 5	29.23% 4	29.43% 2
Irish Republic	30.13% 1	30.13% 1	30.03% 6	30.13% 1	30.13% 1	30.13% 1
Italy	25.55% 6	26.05% 4	26.49% 1	26.39% 2	26.29% 3	26.05% 4
the Netherlands	14.99% 6	21.41% 1	21.17% 4	21.31% 3	21.36% 2	17.08% 5
Spain	29.38% 2	29.38% 2	29.18% 6	29.28% 4	29.28% 4	29.43% 1
the US	13.45% 1	12.70% 3	12.20% 6	12.30% 5	12.50% 4	13.40% 2

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

5.3.3.2 Comparison over Different Forecasting Horizons

Table 5-30 shows the average percentages of the superior combination forecasts compared to the best single ones over seven origins for four forecasting horizons measured by three measures, which reveals the performance ranking of different combination methods at different forecasting horizons. The ranking of each weighting scheme is also presented in this table with the highest superior percentages being highlighted.

It is shown that for each forecasting length, there are always combination forecasts that are better than the best individual ones, and the superior percentages are all above 19%, which means that at least 19% of composite forecasts are more accurate than the best individual ones no matter which combination method is applied for the one- to four-step-ahead forecasts. The highest superior percentage is 25.41%, which is achieved by the DMSFE ($\alpha = 0.85$) method for the four-step-ahead forecasts measured by RMSE, and the lowest percentage is 19.10%, which is obtained by the inverse-MAE scheme for the third-step-ahead projections judged by MAE.

Compared to the first and second combination groups, the performance ranking of different weighting schemes is more consistent across four forecasting lengths and three accuracy measures. For the first-step-ahead forecasts, the best-performing method is the VACO scheme based on MAE, which is beaten by the inverse-MAE method judged by MAPE and RMSE. As far as the second- to four-step-ahead forecasts are concerned, the DMSFE ($\alpha = 0.85$) method outperforms others regardless of accuracy measure with only one exception: it is overtaken by the DMSFE ($\alpha = 0.90/0.95$) and the VACO methods based on RMSE for the two-step-ahead forecasts. The SA method is the least accurate weighting scheme, taking the last position in eleven out of twelve cases.

For a specific forecasting length, the difference in the superior percentages generated by different combination methods is small (see each row of table 5-30). For example, measured by MAE, the gaps in the superior percentages of the best and worst weighting schemes are 0.68% (between 22.42% of the VACO method and 21.74% of the SA method), 1.48% (between 22.37% of the DMSFE ($\alpha = 0.85$) method and 20.89% of the SA method), 1.08% (between 20.18% of the DMSFE ($\alpha = 0.85$) method and 19.10% of the inverse-MAE method)and 1.28% (between 21.51% of the DMSFE ($\alpha = 0.85$) method and 20.23% of the SA method) for one- to four-step-ahead forecasts respectively.

In addition, for the same combination method, the gap in the superior percentages for different forecasting horizons is small (see each column of table 5-30). For instance, measured by MAE, the greatest divergence in the superior percentages of the inverse-MAE method for the four forecasting horizons is 2.84%, which is between 21.94% for the one-step-ahead forecasts and 19.10% for the third-step-ahead projections.

The percentages of the inferior combination forecasts compared to the worst single ones are zero for all cases, which means that all composite forecasts are better than the worst constituent ones regardless of weighting scheme or accuracy measure for four forecasting lengths. It shows that the least accurate forecasts are not generated through the combination forecasting approach. Therefore, it is concluded that the combination forecasting approach is superior with higher degree of accuracy and lower risk of forecasting failure for one- to four-step-ahead forecasts for the third combination group.

Table 5-30 Superior Percentage and Rank of Each Combination Method for Each Forecasting Horizon (the Third Combination Group)

step	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE
MAE						
1	21.74% 6	22.42% 1	22.28% 4	22.31% 2	22.31% 2	21.94% 5
2	20.89% 6	22.11% 4	22.37% 1	22.31% 2	22.20% 3	21.46% 5
3	19.41% 5	20.06% 3	20.18% 1	20.12% 2	20.06% 3	19.10% 6
4	20.23% 6	21.20% 4	21.51% 1	21.46% 2	21.31% 3	20.52% 5
MAPE						
1	21.80% 6	22.23% 2	22.11% 5	22.23% 2	22.23% 2	22.57% 1
2	22.68% 6	22.97% 4	23.22% 1	23.16% 2	23.08% 3	22.94% 5
3	22.14% 5	22.51% 4	22.85% 1	22.77% 2	22.65% 3	21.86% 6
4	22.31% 6	22.85% 4	23.22% 1	23.02% 2	22.88% 3	22.65% 5
RMSE						
1	24.10% 6	24.59% 3	24.53% 5	24.56% 4	24.62% 2	24.70% 1
2	23.45% 6	24.62% 3	24.59% 4	24.64% 1	24.64% 1	24.05% 5
3	21.51% 6	23.39% 4	23.68% 1	23.65% 2	23.56% 3	22.08% 5
4	23.82% 6	25.30% 4	25.41% 1	25.38% 2	25.33% 3	24.70% 5

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best-performing combination methods for each comparison are highlighted.

5.3.3.3 General Comparison among Various Weighting Schemes

Table 5-31 summarizes the average percentages of the superior combination forecasts generated by each weighting scheme compared to the best single ones over four forecasting horizons and seven origins measured by MAE, MAPE and RMSE, which is a general comparison and shows the forecasting ranking of each combination method. The highest superior percentages in each comparison are highlighted.

It is demonstrated that all superior percentages are above 20%, which means that at least 20% of combination forecasts produced by any weighting scheme are better than the best single projections, suggesting that combing individual forecasts can improve forecasting accuracy irrespective of weighting scheme or accuracy measure. The highest superior percentage of 24.56% is provided by the DMSFE ($\alpha = 0.90$) method measured by RMSE, and the lowest superior percentage of 20.57% is generated by the SA scheme judged by MAE. The performance of the inverse-MAE method is unsatisfactory, and it stays at the fifth position regardless of accuracy measure.

Three accuracy measures show consistent results regarding the performance of different weighting schemes. The DMSFE methods with different discounting factors always rank top three regardless of accuracy measure. Generally, a discounting factor of 0.85 provides the best results, achieving the first place judged by MAE and MAPE; and a discounting factor of 0.95 is inferior to the other two, always staying at the third place. The SA method, again, stays at the bottom whichever measure is used.

The difference between the performances of different weighting schemes is insignificant. The greatest gaps between the superior percentages generated by the best and the worst weighting schemes are 1.02% based on MAE (between 20.57% of the SA method and 21.59% of the DMSFE ($\alpha = 0.85$) method), 0.62% according to MAPE (between 22.23% of the SA method and 22.85% of the DMSFE ($\alpha = 0.85$) method) and 1.34% measured by RMSE (between 23.22% of the SA method and 24.56% of the DMSFE ($\alpha = 0.90$) method).

The percentages of the inferior combination forecasts compared to the worst single ones are zero for all cases, which means that all composite forecasts are better than the worst constituent ones regardless of weighting scheme or accuracy measure. It shows that the least accurate forecasts are not generated through the combination forecasting approach. Therefore, it is concluded that combining individual forecasts are beneficial with higher degree of accuracy and lower risk of forecasting failure as with the third combination group.

Comparisons with the results generated from combining the first (table 5-15) and the second groups (table 5-23) reveal that the best-performing methods vary when the constituent single models are different.

The inverse-MAE and the VACO methods both have chances to be the best scheme when combining all models in the first group or combining traditional econometric and time series models in the second group. But they are outperformed by the DMSFE methods when the constituent models change to be climate econometric and time series models in the third group. On the other hand, the worst-performing one-stage combination method stay unchanged: the SA method is the worst among the one-stage combination methods, ranking number six for two out of three cases as with combining the first group, and it always stays at the bottom as with combining the second and third groups.

Table 5-31 General Comparison among Various Combination Methods (the Third Combination Group, Superior Percentage and Rank)

measure	SA	VACO	DMSFE (.85)	DMSFE (.90)	DMSFE (.95)	inverse MAE
MAE	20.57% 6	21.45% 4	21.59% 1	21.55% 2	21.47% 3	20.75% 5
MAPE	22.23% 6	22.64% 4	22.85% 1	22.79% 2	22.71% 3	22.50% 5
RMSE	23.22% 6	24.47% 4	24.55% 2	24.56% 1	24.54% 3	23.88% 5

Notes: There are two figures presented in each cell with the ones in the first line being the superior percentages and the ones below them being the ranks. Two decimal places are retained for the superior percentages. The superior percentages and ranks of the best combination method are highlighted.

5.3.4 Comparison among Three Combination Groups

To assess whether combining econometric models with different explanatory variables can contribute to more accurate forecasts, the forecasting performance of the three combination sets measured by MAE, MAPE and RMSE are compared. The comparisons are conducted in two ways: the market-specific evaluation and the general assessment.

Firstly, for a market-specific evaluation, the comparison is conducted for each origin and each weighting scheme, and the results are presented in table 5-32 to Table 5-34 with the highest superior percentages of each comparison highlighted. It is shown that the performance ranking of three combination groups varies according to origin market and accuracy measure. In general, combining all models, which include both traditional and climate econometric models as well as time series models produce the best forecasts.

Table 5-32 shows that when MAE is used as accuracy measure, combining all models for the French, German, Spanish and American markets is the best no matter which weighting method is applied. For the Irish origin, the first combination group performs the best when the weights are computed by the SA, VACO and DMSFE methods; and the third combination group, which combines climate econometric and time series models, beats the other two when the inverse-MAE scheme is utilized. For the Italian case, the first combination category outperforms the others when the SA method is applied, and the second group, which combines traditional econometric and time series models is superior when other weighting schemes are used. As far as forecasting tourism demand from the Netherlands is concerned, the first combination group ranks number one when weights are generated by the SA and the inverse-MAE methods; and the third combination group is the first in position when weights are produced by the VACO and DMSFE methods.

Table 5-33 demonstrates that when forecasting performance is judged by MAPE, combining all models behaves the best for the German, Irish, Dutch and Spanish markets for all weighting schemes. For the French origin, when weights are computed by the SA method, the second combination category beats the first combination group, which is the highest in place when other weighting methods are used. For the Italian case, the second combination panel performs the best with the VACO, DMSFE and inverse-MAE schemes being the weighting methods, and it is beaten by the first combination group when the SA method is applied. With respect to the American case, combining traditional econometric and time series models in the second group is always superior compared to the other two groups regardless of combination method.

According to table 5-34, if RMSE is used to measure forecasting performance, the first combination group always produces the best forecasts for the German, Dutch, Spanish and American markets. For forecasting tourism demand from France, the first combination category yields the best projections when the VACO and the DMSFE methods are utilized, which is beaten by the second combination group when the SA and the inverse-MAE schemes are applied. For the Irish case, combining climate econometric and time series models outperforms the other two combination groups for all weighting methods. As far as the Italian market is concerned, combining all models in the first group is the best when weights are computed by the SA and the inverse-MAE methods, and integrating traditional econometric and times series models in the second group is the first in position when the VACO and DMSFE schemes are used.

Table 5-35 shows the results of the general comparison among three combination groups for each weighting method by taking averages across seven origins. It is clear that the first combination category achieves the best results for all cases, which means that including econometric models with different explanatory variables in combination is beneficial. It implies that introducing the climate factor in combination forecasting can contribute to more accurate forecasts.

In the current tourism demand literature, all existing combination forecasting studies consider econometric models with the same economic explanatory variables as constituents. No combination has been applied to econometric models with the climate factor as a demand determinant. This study fills this gap, and the empirical results show that combining econometric models with different explanatory variables as well as time series models can make use of information embedded in different constituent models, which can bring about better combination forecasts. Therefore, including econometric models with different influencing factors in combination are recommended for future studies.

Table 5-32 Comparison among Different Combination Groups for Each Origin (MAE)

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

Note: The superior percentages of the best combination group for each combination method and each origin are highlighted.

Table 5-33 Comparison among Different Combination Groups for Each Origin (MAPE)

panel	France	Germany	Irish Republic	Italy	the Netherlands	Spain	the US
SA							
All	26.39%	29.36%	29.68%	19.79%	28.21%	29.68%	19.24%
Traditional	27.04%	27.14%	28.83%	17.08%	7.47%	26.10%	28.09%
Climate	15.64%	24.15%	29.43%	14.94%	25.90%	28.09%	17.48%
VACO							
All	27.87%	29.55%	29.76%	18.61%	28.32%	29.69%	16.41%
Traditional	26.49%	27.64%	29.33%	22.71%	8.22%	26.94%	27.84%
Climate	20.32%	25.80%	29.13%	12.95%	26.69%	27.84%	15.74%
DMSFE (.85)							
All	28.22%	29.44%	29.78%	20.22%	27.95%	29.79%	17.15%
Traditional	27.04%	27.59%	28.98%	24.15%	6.62%	27.24%	27.59%
Climate	21.22%	25.20%	29.03%	14.99%	26.39%	27.59%	15.54%
DMSFE (.90)							
All	28.14%	29.47%	29.78%	19.69%	28.06%	29.76%	16.94%

Traditional	26.89%	27.59%	29.08%	23.56%	7.02%	27.14%	27.69%
Climate	21.02%	25.35%	29.03%	14.34%	26.54%	27.69%	15.59%
DMSFE (.95)							
All	28.02%	29.51%	29.77%	19.13%	28.19%	29.73%	16.69%
Traditional	26.64%	27.59%	29.18%	23.16%	7.62%	27.09%	27.79%
Climate	20.72%	25.50%	29.03%	13.65%	26.59%	27.79%	15.69%
inverse-MAE							
All	27.81%	29.62%	29.81%	19.67%	28.44%	29.68%	17.84%
Traditional	27.04%	27.69%	29.33%	21.71%	8.72%	26.49%	28.29%
Climate	17.53%	25.70%	29.48%	13.94%	26.34%	28.29%	16.24%

Note: The superior percentages of the best combination group for each combination method and each origin are highlighted.

Table 5-34 Comparison among Different Combination Groups for Each Origin (RMSE)

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

Note: The superior percentages of the best combination group for each combination method and each origin are highlighted.

Table 5-35 General Comparison among Different Combination Groups for Each Weighting Scheme

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

Note: The superior percentages of the best combination group for each combination method are highlighted.

5.4 Which Models to Combine?

The superior combination forecasts compared to the best single ones generated by various weighting schemes can be a combination of different constituents. What are the chances of each individual model to be a component of the superior combination forecasts? Whether more accurate individual forecasts have higher opportunities to constitute the superior composite forecasts? Which type of econometric models are more likely to contribute to the superior combination forecasts, the traditional one or the climate one? To answer these questions, the frequencies and rankings (in terms of individual models' chances to be in the superior combination forecasts) of each individual model in the superior combination forecasts for every origin country are computed and presented in table 5-36 and table 5-37. The highest three frequencies of each comparison are highlighted.

To take all individual models and all weighting schemes into account, the first combination group which combines all 15 models through 12 weighting schemes are studied, and the frequencies shown in the tables are the overall frequencies across the four forecasting horizons measured by three accuracy measures of MAE, MAPE and RMSE for each market.

Firstly, table 5-36 shows the frequencies and rankings of each individual model in the superior combination forecasts generated by different weighting schemes for each origin.

It is revealed that for most weighting schemes, every individual model has a chance to constitute the superior combination forecasts for each origin. The highest frequency of 59.81% is achieved by the SARIMA model for forecasting tourism demand from Italy when combination weights are worked out by the TS-inverse-MAE method. It means that 59.81% of all superior combination forecasts obtained through the TS-inverse-MAE method for the Italian market have the SARIMA model as a constituent. The lowest frequency except zero is 10.32%, which is gained by the climate ADLM for the French case through applying the TS-DMSFE ($\alpha = 0.85$) combination method. It shows that for all superior combination forecasts for the French market generated through the TS-DMSFE ($\alpha = 0.85$) weighting scheme, the chance of the climate ADLM to be a component is 10.32%.

A few cases of zero frequencies are seen when the SA, VACO and TS-SA combination methods are used to compute the weights. For the American case, the ETS and climate TVP models have no chance to be a constituent of the superior combination forecasts if combining weights are generated through the SA method. When weights are determined by the VACO scheme, the same two individual models do not contribute to superior combination forecasts as far as tourism demands from France, Germany, Irish Republic, Italy, and the Netherlands are concerned. Produced by the TS-SA combination method, the superior combination forecasts for all markets except the American one do not have the traditional VAR model as a constituent.

For the same weighting scheme, individual model's frequency to make up the superior combination forecasts changes according to the origin country under consideration. As with the SA method, the three individual models that have the highest-three frequencies are the same for the French and American markets, which are the climate bounds test model, traditional ADLM and climate SD model. The climate bounds test model and traditional ADLM also have high chances to constitute the superior combination forecasts for the German, Irish, Italian, Dutch and Spanish origins, ranking top three together with the traditional VAR model, which is also high in position as with the French case, ranking number four, but only ranks number 12 for the American market. For forecasting tourism demand from the US, the ETS, climate TVP and traditional TVP models have the lowest chances to form the superior composite forecasts; and for other origins, the ETS, climate LI and traditional LI models stay at the bottom of the list.

For combination forecasts generated from the VACO method, the climate bounds test model achieves the highest frequency for the French, German, Irish, Italian and Dutch origins, followed by the traditional ADLM and climate ADLM; and the ETS, climate TVP and traditional TVP models have the lowest frequencies. When it comes to the Spanish and American markets, the traditional VAR model, which only ranks number eleven for the other five origins, take up the first place, followed by the traditional bounds test model, and the models which are the third in position are the traditional TVP model for the Spanish origin and the ETS model for the American case. The individual models staying at the bottom are the same for these two countries, which are the climate ADLM, climate LI and SARIMA models.

As with the DMSFE ($\alpha = 0.85$) weighting scheme, the traditional VAR and traditional bounds test models have high opportunities to form the superior composite forecasts for all markets, joined by the ETS model for the French, German, Irish and Italian cases, and the traditional LI model for the other three origins to rank top three. On the other hand, the climate ADLM and climate LI models have low frequencies to be a component of the superior combination forecasts for all origins, together with the SARIMA model for the French, German, Irish and Italian markets, and the traditional TVP model for other countries to be the last three in place.

As far as the DMSFE ($\alpha = 0.90$) method is concerned, the rankings of individual models' frequencies are similar for the French, German and Irish markets, with the traditional VAR, traditional bounds test and traditional LI models staying at the top of the list, and the climate LI, traditional TVP models and the climate ADLM having the lowest chances. The climate LI model, which stays at the bottom for the French, German and Irish cases, achieves the highest frequency for other countries, followed by the climate SD and traditional SD models for tourism demand from the Netherlands, Spain and the US; and the traditional SD and traditional TVP models for the Italian case. For the Dutch, Spanish and American markets, the climate TVP, traditional ADLM and traditional bounds test models have the lowest chances to be constituents of the superior combination forecasts. With respect to tourism demand from Italy, the traditional ADLM, climate TVP model and climate ADLM stay at the bottom.

For combination forecasts using the DMSFE ($\alpha = 0.95$) scheme, individual models have similar chances to form the superior composite forecasts for the French and German markets, with the climate LI, climate SD and traditional SD models ranking top three, and the traditional ADLM, climate TVP and traditional bounds test models staying at the bottom. For other origins, the traditional SD and traditional TVP models rank top two, succeeded by the climate bounds test model for the Irish market, and the traditional LI model for the Italian, Dutch, Spanish and American cases. And the traditional bounds test, climate ADLM and state space ETS models are the three ones that have the lowest frequencies.

When the inverse-MAE scheme is applied, individual models obtain similar chances and rankings for forecasting tourism demand from Irish Republic, Italy, the Netherlands, Spain and the US, with three time series models of SARIMA, ETS and state space ETS as well as the climate SD models ranking top four; and the climate LI, climate bounds test and traditional VAR models ranking bottom three. For the German case, the ETS and climate SD models also achieve high frequencies, standing at the top of the list with the climate TVP model; while the climate LI and climate bounds test models have low chances, joined by the traditional LI model to be the last three in position. As far as the French market is concerned, the climate bounds test, traditional TVP and traditional SD models achieve the highest chances to constitute the superior combination forecasts, and the climate ADLM, traditional bounds test and state space ETS models have the lowest frequencies.

If combination forecasts are generated from the TS-SA method, individual models gain similar frequencies and rankings for all origins except the American one, with the ETS, state space ETS and the climate SD models having the highest three frequencies, and the traditional VAR, traditional TVP and SARIMA models staying at the bottom of the list. With respect to the American case, the climate ADLM, state ETS and traditional LI models rank top three, and the climate SD, climate TVP and traditional SD models are the lowest three in position.

As with the TS-VACO weighting scheme, for the French, German, Irish, Italian and Dutch markets, the climate ADLM, state space ETS and traditional LI models obtain the first three places, and the climate TVP, climate SD and traditional SD models are at the bottom of the list. The ETS, state space ETS and the traditional LI models are the ones that have the highest chances for the Spanish and American origins, and the climate ADLM and traditional TVP model are the last two in position.

If the TS-DMSFE ($\alpha = 0.85$) method is applied to determine the weights, for the French, German, Irish and Italian cases, the frequency rankings of individual models are similar, with the ETS, state space ETS and traditional LI models taking up the top three positions, and the climate ADLM, traditional TVP model and traditional ADLM claiming the bottom three places. For the Dutch, Spanish and American cases, the climate LI, traditional VAR and traditional TVP models rank top three, and the traditional LI and traditional bounds test models are placed at the bottom.

Regarding combination forecasts utilizing the TS-DMSFE ($\alpha = 0.90$) method, for the French, German and Irish markets, the traditional VAR, traditional TVP and climate LI models rank top three, and the traditional bounds test, traditional LI and ETS models are positioned at the bottom. For other origins, the traditional TVP, SARIMA and traditional SD models have the highest frequencies, and the traditional ADLM, state space ETS and climate VAR models have the lowest frequencies.

If combining weights are computed based on the TS-DMSFE ($\alpha = 0.95$) technique, for the French and German cases, the traditional TVP, SARIMA and traditional SD models rank top three, and the traditional ADLM, state space ETS and climate VAR models are placed at the bottom. For other origins, individual models have comparable chances to construct the superior combination forecasts, with the climate LI, state space ETS and climate ADLM models ranking top three for the Irish case, and the SARIMA model, climate ADLM and climate SD model having the highest three frequencies for the Italian, Dutch, Spanish and American origins. And the climate TVP, traditional bounds test and climate bounds test models are the last three in position.

As far as the TS-inverse-MAE combination method is concerned, for forecasting tourism demand from France, the climate LI model, climate ADLM and SARIMA model are placed at the top, and the climate TVP, traditional bounds test and climate bounds test models are graded at the bottom. The opportunities of individual models to form the superior composite forecasts are similar across other origins, with the top four models being the same as the ETS, SARIMA, climate LI and climate SD models, and the bottom three being the climate ADLM, traditional SD model and traditional ADLM.

When a specific weighting scheme is applied, the difference among the frequencies of the same individual model to be in the superior combination forecasts vary greatly across origins (see each column of table 5-36). The individual models that achieve high frequencies to constitute the superior combination forecasts for some origins may have low chances to be in the better composite projections as far as other origins are concerned.

For instance, when the combining weights are determined by the VACO method, the ETS and climate TVP models have no chance to be components of the superior combination forecasts with respect to tourism demand from France, Germany, Irish Republic, Italy and the Netherlands, but they achieve high frequencies, which are above 50%, for the Spanish and American origins. Similarly, when the inverse-MAE scheme is applied, the frequencies of the climate ADLM, state space ETS and traditional bounds test models in the superior combination forecasts are around 25% for the French case, but for other origins, they are all above 50%. Besides, as with the TS-DMSFE ($\alpha = 0.85$) method, for the French, German, Irish and Italian cases, the frequencies of the climate ADLM are around 11%, and those of the traditional TVP model are around 21%; but for the Dutch, Spanish and American markets, these two models both achieve high frequencies which are above 50%.

In addition, for the same origin, the difference in the frequencies of various individual models can be significant in some cases, but trivial in others (see each row of table 5-36).

For example, when the SA method is applied, the divergence in individual models' chances to form the superior combination forecasts is great for the American case, with the biggest gap between the top-ranking and bottom-ranking models being 53.88%. But for other origins, the differences in individual models' frequencies are insignificant, and the biggest gaps are all less than 8%. In the same way, if the weights are determined based on the TS-VACO scheme, the divergences in individual models' opportunities are moderate for the French, German, Irish, Italian and Dutch markets, with the biggest gaps being less than 11%. On the other hand, for the Spanish and American cases, the differences are considerable, which are 49.01% (between 10.55% of the climate ADLM and 59.56% of the traditional LI model for the Spanish market) and 45.95% (between 10.45% of the climate ADLM and 56.40% of the state ETS model for the American origin) respectively.

Furthermore, combination methods also affect individual models' chances to make up the superior combination forecasts. For the same individual model, the difference among its chances to be a component in the better composite forecasts for the same origin varies according to which weighting scheme is applied.

For instance, for the climate ADLM, its opportunities to construct the superior combination forecasts for tourism demand from France are comparable when the SA, VACO, DMSFE ($\alpha = 0.85$), DMSFE ($\alpha = 0.95$), TS-VACO, TS-DMSFE ($\alpha = 0.90$) and TS-inverse-MAE weighting schemes are utilized, being around 50%; on the other hand, if combining weights are determined by other methods, its frequencies are divergent, ranging from 10.32% when the TS-DMSFE ($\alpha = 0.85$) method is applied to 40.85% if the TS-SA scheme is used. Likewise, as far as the ETS model is concerned, its chances to be in the superior combination forecasts for the American market are diverse among different combination methods: being above 50% when the VACO, DMSFE ($\alpha = 0.90$), inverse-MAE, TS-SA, TS-VACO, TS-DMSFE ($\alpha = 0.85$), TS-DMSFE ($\alpha = 0.95$) and TS-inverse-MAE schemes are used; around 38% as with the DMSFE ($\alpha = 0.85$), DMSFE ($\alpha = 0.95$) and TS-DMSFE ($\alpha = 0.90$) methods; and 0% based on the SA scheme.

Table 5-37 summarizes the general frequencies and rankings of each individual model in the superior composite forecasts for seven origins by taking averages across the results generated through 12 weighting schemes. The frequencies of the individual models which are placed top three for origin are highlighted.

It is shown that all individual models have the chance to constitute the better combination forecasts for all markets under consideration. The highest frequency of 50.77% is seen from the German case, which is achieved by the climate SD model. It means that 50.77% of the superior composite forecasts for tourism demand from Germany have the climate SD model as a constituent. The lowest frequency of 38.90% is obtained by the traditional TVP model for the Irish market, which demonstrates that 38.90% of the more accurate combination projections for tourist arrivals from Irish Republic include the traditional TVP model as a component.

The rankings of different individual models in terms of their frequencies in the superior combination forecasts vary according to origin market.

The climate bounds test, climate SD and traditional LI models achieve high rankings for all origins with the climate SD model performing the best, being the first in position for five out of seven cases. The ETS and climate TVP models obtain high positions as with forecasting tourist arrivals from Spain, finishing in second and third in place respectively. However, the climate TVP model fail to achieve high rankings for other markets, being positioned at the ninth place or below. The SARIMA model has the highest chance to form the superior combination forecasts for the American market, staying at the top of the list. The chances of the climate ADLM, traditional VAR and traditional TVP models to be a component of the superior composite forecasts are relatively low for each origin. The climate LI model, which ranks number five for the Italian case, and the traditional ADLM model, which is the eighth in place for the Irish market, both remain in the bottom half list as far as other origins are concerned.

For the same origin, the difference among each individual model's opportunity to make up the superior combination forecasts is not great (see each row of table 5-37). The gap in the frequencies between the top-placed and bottom-placed models ranges from 8.91% (between 41.18% of the climate ADLM and 50.09% of the SARIMA model) for the American case and 11.64% for the German origin (between 39.13% of the traditional TVP model and 50.77% of the climate SD model).

In addition, regarding one specific individual model, its chances to constitute the superior composite forecasts for different markets are comparable (see each column of table 5-37).

For example, the frequencies of the climate bounds test model in the better combination projections for seven origins are all around 49%, with the lowest one being 48.63% for the Spanish case, and the highest one being 49.70% regarding forecasting tourism demand from France. Similarly, the chances of the traditional bounds test model to comprise the more accurate composite projections range from 44.78% for the American case to 47.31% for the German market.

The relatively large diversities are seen for the ETS, SARIMA and the climate TVP models. The highest frequencies of the ETS and climate TVP models are both seen for the Spanish market, being 49.32% and 49.04% respectively. However, for other origins, their frequencies are no larger than 47%, and the greatest gaps are 4.83% (between 49.32% for the Spanish case and 44.49% for the French market) and 4.68% (between 49.04% for the Spanish case and 44.36% for the Italian origin) respectively. Besides, the SARIMA model performs well for forecasting tourism demand from the US, and it achieves a frequency of 50.09%; while its frequency is 45.96% for the French case, resulting in a gap of 4.13%.

It is revealed that individual models' opportunities to constitute the superior combination forecasts are irrelevant to their forecasting abilities. As shown in table 5-4, the forecasting performance of the climate bounds test, climate SD and traditional LI models, which have high frequencies and high positions in terms of chances to be in the superior combination forecasts for all origins are unsatisfactory irrespective of accuracy measures.

On the other hand, the three time series techniques, which have most powerful forecasting abilities, do not achieve higher opportunities to be a constituent of the superior combination forecasts for most cases. The highest frequency obtained by the state space ETS model is 48.69% for the Irish market, which is less than those obtained by three econometric models including the climate SD, traditional LI and climate bounds test models, which are inferior compared to the state space ETS model in forecasting performance. The ETS model acquires the second highest frequency for the Spanish market, but the frequencies for other origins are only placed at the sixth position or after. Similarly, the SARIMA model only gains a position among the top three models for the American origin, and for other markets, it is outperformed in frequency by many econometric models that are inferior in forecasting ability. Furthermore, the climate ADLM, traditional VAR and traditional TVP models, which forecast most accurately among all econometric models, have the lowest chances to constituent the superior combination forecast for all markets.

A comparison between the opportunities of the climate econometric models and their traditional counterparts shows consistency across seven origins: for each origin, the climate bounds test, climate VAR, climate TVP and climate SD models have higher chances to form the more accurate combination forecasts; while for the ADLM and LI techniques, the traditional specifications achieve higher frequencies.

In addition, the divergence in the frequency of one particular individual model across seven origins is trivial compared to the diverse in its forecasting ability for different origins.

According to table 5-2, the forecasting performance of individual models changes greatly according to the origin under study. For example, the climate bounds test model performs unwell for other origins, but it generates the most accurate individual projections for forecasting tourism demand from France. However, its chances to be a component in the superior combination forecasts among seven markets are similar, although it achieves the highest frequency for the French market. Likewise, the climate SD model only shows its superior forecasting ability for the Irish origin, but it has high chances to construct the more accurate combination forecasts for all markets.

Table 5-36 Frequencies and Rank of Individual Components in Superior Combination Forecasts for Each Weighting Scheme and Each Origin

origin	bounds©	ETS	SARIMA	ADLM©	State ETS	LI©	VAR©	TVP©	SD©	bounds	ADLM	LI	VAR	TVP	SD
	SA														
France	52.91% 1	48.56% 14	51.10% 10	51.95% 7	50.35% 12	45.41% 15	52.10% 5	50.94% 11	52.66% 3	52.04% 6	52.84% 2	49.55% 13	52.32% 4	51.52% 8	51.24% 9
Germany	52.99% 1	48.43% 15	50.99% 9	51.37% 4	49.95% 12	49.27% 13	51.22% 6	50.10% 11	51.17% 7	51.28% 5	51.62% 3	48.92% 14	51.93% 2	50.66% 10	51.03% 8
Irish Republic	52.51% 1	48.89% 15	50.83% 8	51.35% 4	49.97% 12	49.35% 13	51.19% 5	50.07% 11	51.10% 6	51.08% 7	51.51% 3	49.13% 14	51.57% 2	50.56% 10	50.73% 9
Italy	52.65% 1	48.76% 15	50.89% 8	51.34% 4	49.96% 12	49.32% 13	51.20% 5	50.07% 11	51.12% 7	51.13% 6	51.51% 3	49.08% 14	51.67% 2	50.58% 10	50.83% 9
the Netherlands	52.81% 1	48.61% 15	50.95% 8	51.34% 4	49.96% 12	49.29% 13	51.21% 6	50.07% 11	51.13% 7	51.21% 5	51.56% 3	49.02% 14	51.79% 2	50.62% 10	50.92% 9
Spain	52.67% 1	48.73% 14	50.84% 10	51.60% 4	50.14% 12	48.06% 15	51.42% 6	50.44% 11	51.56% 5	51.41% 7	52.08% 2	49.45% 13	51.84% 3	50.90% 9	50.92% 8
US	53.40% 2	0.00% 14	51.52% 7	52.42% 6	50.63% 8	44.23% 10	52.44% 5	0.00% 14	53.19% 3	52.70% 4	53.88% 1	49.82% 9	39.16% 12	24.91% 13	39.95% 11
	VACO														
France	52.14% 1	0.00% 14	51.06% 5	51.36% 3	49.99% 8	49.30% 9	51.02% 7	0.00% 14	51.04% 6	51.09% 4	51.58% 2	49.07% 10	38.46% 11	24.66% 13	38.37% 12
Germany	51.93% 1	0.00% 14	50.86% 7	51.33% 3	50.05% 8	49.23% 10	51.06% 4	0.00% 14	50.97% 6	51.04% 5	51.57% 2	49.35% 9	38.31% 11	24.90% 13	38.04% 12
Irish Republic	51.97% 1	0.00% 14	50.92% 7	51.34% 3	50.02% 8	49.26% 10	51.05% 4	0.00% 14	50.99% 6	51.02% 5	51.55% 2	49.28% 9	38.35% 11	24.84% 13	38.15% 12
Italy	52.04% 1	0.00% 14	50.99% 7	51.35% 3	49.99% 8	49.27% 9	51.04% 5	0.00% 14	50.99% 6	51.05% 4	51.54% 2	49.18% 10	38.40% 11	24.76% 13	38.25% 12
the Netherlands	52.66% 1	0.00% 14	51.03% 7	51.77% 3	50.31% 8	47.77% 10	51.42% 6	0.00% 14	51.59% 4	51.58% 5	52.44% 2	49.61% 9	38.73% 11	25.13% 13	38.45% 12
Spain	50.24% 9	50.53% 4	49.85% 13	49.62% 15	50.47% 5	49.78% 14	50.27% 8	50.00% 12	50.20% 10	50.62% 2	50.04% 11	50.44% 6	50.65% 1	50.53% 3	50.43% 7
US	50.26% 8	50.52% 3	49.89% 13	49.72% 15	50.48% 4	49.85% 14	50.13% 10	49.96% 12	50.16% 9	50.53% 2	50.04% 11	50.37% 6	50.56% 1	50.42% 5	50.36% 7
	DMSFE (.85)														
France	50.31% 8	50.62% 2	49.84% 13	49.68% 15	50.56% 4	49.77% 14	50.17% 9	49.94% 12	50.14% 10	50.60% 3	50.04% 11	50.46% 6	50.66% 1	50.49% 5	50.36% 7
Germany	50.30%	50.59%	49.86%	49.69%	50.54%	49.80%	50.16%	49.95%	50.15%	50.58%	50.04%	50.43%	50.63%	50.47%	50.37%

	8	2	13	15	4	14	9	12	10	3	11	6	1	5	7
Irish Republic	50.28% 8	50.56% 3	49.87% 13	49.70% 15	50.51% 4	49.82% 14	50.15% 10	49.96% 12	50.16% 9	50.56% 2	50.04% 11	50.40% 6	50.60% 1	50.44% 5	50.36% 7
Italy	50.16% 8	50.44% 3	49.90% 13	49.77% 15	50.39% 5	49.90% 14	50.14% 9	49.97% 12	50.12% 10	50.46% 2	50.03% 11	50.37% 6	50.47% 1	50.42% 4	50.34% 7
the Netherlands	50.52% 4	38.74% 8	36.71% 12	36.46% 13	38.10% 11	11.53% 15	38.33% 9	49.91% 7	50.40% 5	51.47% 2	50.15% 6	51.10% 3	51.58% 1	12.92% 14	38.15% 10
Spain	50.56% 4	38.54% 8	36.97% 12	36.77% 13	38.22% 9	11.88% 15	38.04% 11	49.85% 7	50.35% 5	51.27% 2	50.14% 6	50.93% 3	51.37% 1	12.85% 14	38.08% 10
US	50.59% 4	38.65% 8	36.86% 12	36.69% 13	38.29% 9	11.72% 15	38.13% 10	49.81% 7	50.28% 5	51.34% 2	50.16% 6	51.07% 3	51.51% 1	12.88% 14	38.01% 11
	DMSFE (.90)														
France	50.58% 4	38.62% 8	36.90% 12	36.71% 13	38.27% 9	11.76% 15	38.11% 10	49.82% 7	50.31% 5	51.33% 2	50.15% 6	51.03% 3	51.48% 1	12.87% 14	38.04% 11
Germany	50.58% 4	38.58% 8	36.92% 12	36.74% 13	38.25% 9	11.81% 15	38.08% 10	49.84% 7	50.34% 5	51.30% 2	50.14% 6	50.97% 3	51.43% 1	12.86% 14	38.05% 11
Irish Republic	50.37% 4	38.51% 8	36.95% 12	36.79% 13	38.13% 10	11.88% 15	38.05% 11	49.87% 7	50.28% 5	51.26% 2	50.10% 6	51.04% 3	51.33% 1	12.84% 14	38.14% 9
Italy	50.31% 9	50.36% 8	50.44% 4	49.99% 13	50.40% 7	50.54% 1	50.20% 11	49.93% 14	50.44% 5	50.07% 12	49.57% 15	50.23% 10	50.42% 6	50.45% 3	50.47% 2
the Netherlands	50.28% 6	50.18% 10	50.28% 7	50.15% 11	50.20% 9	50.48% 1	50.06% 12	49.80% 15	50.46% 2	50.05% 13	49.89% 14	50.25% 8	50.32% 4	50.32% 5	50.37% 3
Spain	50.27% 6	50.16% 10	50.22% 7	50.13% 11	50.17% 9	50.47% 1	50.09% 12	49.86% 15	50.45% 2	50.05% 13	49.86% 14	50.21% 8	50.35% 4	50.34% 5	50.37% 3
US	50.27% 6	50.17% 10	50.24% 7	50.14% 11	50.18% 9	50.48% 1	50.07% 12	49.83% 15	50.46% 2	50.04% 13	49.87% 14	50.22% 8	50.34% 4	50.33% 5	50.38% 3
	DMSFE (.95)														
France	50.28% 6	50.18% 10	50.26% 7	50.15% 11	50.19% 9	50.48% 1	50.06% 12	49.81% 15	50.46% 2	50.05% 13	49.88% 14	50.24% 8	50.33% 4	50.33% 5	50.38% 3
Germany	50.24% 8	50.23% 9	50.28% 6	50.05% 12	50.24% 7	50.40% 1	50.11% 11	49.92% 14	50.35% 2	50.05% 13	49.79% 15	50.19% 10	50.29% 5	50.31% 4	50.32% 3
Irish Republic	50.65% 3	38.12% 10	38.23% 8	24.89% 15	25.36% 13	38.38% 7	50.30% 5	49.79% 6	38.18% 9	25.15% 14	36.81% 12	50.62% 4	38.04% 11	50.82% 2	51.00% 1
Italy	50.59% 4	38.01% 9	37.96% 10	25.19% 13	25.14% 14	38.31% 7	50.02% 5	49.44% 6	38.31% 8	25.00% 15	37.49% 12	50.64% 3	37.93% 11	50.65% 2	50.89% 1
the Netherlands	50.53% 4	37.93% 10	37.88% 11	25.17% 13	25.14% 14	38.25% 7	50.04% 5	49.57% 6	38.25% 7	24.98% 15	37.46% 12	50.54% 3	37.95% 9	50.66% 2	50.86% 1

Spain	50.56% 4	37.95% 9	37.89% 11	25.19% 13	25.13% 14	38.27% 7	50.02% 5	49.50% 6	38.27% 8	24.98% 15	37.50% 12	50.58% 3	37.94% 10	50.65% 2	50.87% 1
US	50.57% 4	37.98% 9	37.91% 11	25.19% 13	25.13% 14	38.29% 7	50.01% 5	49.47% 6	38.29% 8	24.98% 15	37.50% 12	50.61% 3	37.93% 10	50.65% 2	50.89% 1
	inverse-MAE														
France	50.52% 3	37.91% 10	37.95% 9	25.04% 15	25.22% 13	38.12% 7	50.14% 5	49.76% 6	38.08% 8	25.06% 14	37.26% 12	50.50% 4	37.88% 11	50.57% 2	50.71% 1
Germany	45.11% 14	56.01% 3	55.48% 5	53.06% 7	55.63% 4	43.93% 15	52.45% 8	56.30% 2	56.59% 1	51.03% 9	50.83% 10	45.26% 13	46.05% 12	47.10% 11	53.25% 6
Irish Republic	45.23% 14	57.11% 2	54.72% 3	52.47% 6	57.48% 1	39.91% 15	51.54% 8	53.80% 5	54.39% 4	50.44% 9	50.26% 10	47.47% 11	46.72% 13	47.18% 12	52.40% 7
Italy	45.93% 14	57.45% 1	54.94% 3	51.77% 6	57.21% 2	41.58% 15	51.21% 8	52.94% 5	54.35% 4	50.27% 9	50.01% 10	48.09% 11	47.12% 13	47.58% 12	51.71% 7
the Netherlands	45.67% 14	57.31% 2	54.77% 3	52.00% 6	57.35% 1	41.12% 15	51.29% 8	53.28% 5	54.36% 4	50.31% 9	50.10% 10	47.86% 11	46.95% 13	47.42% 12	51.95% 7
Spain	45.43% 14	57.19% 2	54.73% 3	52.23% 6	57.44% 1	40.58% 15	51.40% 8	53.55% 5	54.36% 4	50.37% 9	50.18% 10	47.66% 11	46.82% 13	47.29% 12	52.17% 7
US	45.35% 14	56.46% 2	54.91% 4	52.65% 6	56.70% 1	41.15% 15	52.15% 8	54.49% 5	55.21% 3	50.55% 9	50.46% 10	46.90% 12	46.83% 13	47.32% 11	52.63% 7
	TS-SA														
France	46.05% 8	57.92% 1	13.70% 13	40.85% 11	57.00% 2	43.25% 9	52.65% 5	56.78% 4	56.92% 3	52.44% 6	51.27% 7	42.99% 10	0.00% 15	11.74% 14	40.14% 12
Germany	45.25% 9	58.17% 2	12.58% 13	40.86% 10	58.42% 1	39.74% 12	51.67% 5	54.05% 4	54.52% 3	50.92% 6	50.43% 7	46.97% 8	0.00% 15	10.78% 14	40.57% 11
Irish Republic	45.87% 9	58.61% 1	12.96% 13	39.99% 11	58.23% 2	41.44% 10	51.26% 5	53.27% 4	54.34% 3	50.76% 6	50.12% 7	47.64% 8	0.00% 15	11.11% 14	39.56% 12
Italy	45.62% 9	58.33% 1	12.72% 13	40.35% 11	58.22% 2	41.01% 10	51.36% 5	53.55% 4	54.39% 3	50.77% 6	50.23% 7	47.38% 8	0.00% 15	10.93% 14	39.93% 12
the Netherlands	45.37% 9	58.24% 2	12.57% 13	40.61% 10	58.33% 1	40.49% 11	51.54% 5	53.84% 4	54.47% 3	50.83% 6	50.31% 7	47.21% 8	0.00% 15	10.81% 14	40.28% 12
Spain	45.54% 9	57.85% 2	13.11% 13	40.67% 11	57.91% 1	40.68% 10	52.30% 5	54.75% 4	55.34% 3	51.31% 6	50.65% 7	46.03% 8	0.00% 15	11.36% 14	40.10% 12
US	47.79% 12	53.66% 4	53.13% 6	56.44% 1	53.80% 3	48.38% 9	48.17% 10	45.97% 14	43.73% 15	51.75% 7	53.48% 5	56.18% 2	47.95% 11	50.94% 8	46.01% 13
	TS-VACO														
France	48.84% 10	53.59% 4	52.82% 5	54.27% 1	53.97% 2	49.10% 9	48.80% 12	45.96% 15	46.74% 14	51.02% 7	51.75% 6	53.82% 3	48.81% 11	50.44% 8	48.24% 13

Germany	48.52% 12	53.93% 4	52.83% 5	54.49% 1	54.48% 2	48.91% 9	48.63% 10	46.16% 15	46.42% 14	51.28% 7	51.76% 6	54.09% 3	48.60% 11	50.37% 8	48.01% 13
Irish Republic	48.63% 12	53.82% 4	52.83% 5	54.43% 1	54.31% 2	48.97% 9	48.68% 10	46.06% 15	46.53% 14	51.17% 7	51.76% 6	54.03% 3	48.67% 11	50.39% 8	48.09% 13
Italy	48.74% 10	53.72% 4	52.83% 5	54.35% 1	54.14% 2	49.04% 9	48.72% 12	46.00% 15	46.63% 14	51.09% 7	51.76% 6	53.93% 3	48.74% 11	50.42% 8	48.17% 13
the Netherlands	47.91% 12	53.50% 4	52.71% 6	55.88% 1	53.87% 3	49.07% 9	48.21% 11	45.99% 14	44.39% 15	51.38% 7	53.25% 5	55.87% 2	48.22% 10	50.70% 8	46.88% 13
Spain	47.02% 6	55.70% 3	54.67% 4	10.55% 15	56.03% 2	48.13% 5	39.62% 9	45.12% 7	40.78% 8	31.61% 11	31.41% 12	59.56% 1	35.33% 10	21.39% 14	27.87% 13
US	48.25% 6	55.91% 3	54.07% 4	10.45% 15	56.40% 1	48.97% 5	40.09% 9	45.32% 7	45.10% 8	29.61% 12	29.29% 13	56.04% 2	35.54% 10	22.00% 14	32.11% 11
	TS-DMSFE (.85)														
France	47.87% 6	56.30% 2	53.92% 4	10.32% 15	56.98% 1	48.82% 5	40.19% 9	45.53% 7	44.74% 8	29.89% 12	29.46% 13	56.28% 3	35.44% 10	21.60% 14	31.69% 11
Germany	47.99% 6	56.15% 3	53.95% 4	10.35% 15	56.76% 1	48.87% 5	40.15% 9	45.44% 7	44.84% 8	29.77% 12	29.41% 13	56.21% 2	35.49% 10	21.69% 14	31.85% 11
Irish Republic	48.15% 6	56.03% 3	54.01% 4	10.38% 15	56.58% 1	48.95% 5	40.13% 9	45.38% 7	44.96% 8	29.69% 12	29.35% 13	56.14% 2	35.50% 10	21.82% 14	31.95% 11
Italy	47.01% 6	55.59% 3	54.04% 4	11.10% 15	56.37% 2	49.13% 5	39.02% 9	45.18% 7	41.47% 8	30.30% 12	31.20% 11	59.30% 1	36.18% 10	21.21% 14	28.93% 13
the Netherlands	50.10% 10	50.08% 12	50.24% 6	50.43% 4	50.27% 5	50.47% 3	50.20% 8	50.18% 9	50.10% 11	49.93% 14	50.06% 13	49.78% 15	50.48% 1	50.48% 1	50.22% 7
Spain	50.09% 11	50.03% 13	50.28% 6	50.45% 4	50.31% 5	50.46% 3	50.14% 8	50.13% 9	50.06% 12	49.90% 14	50.10% 10	49.87% 15	50.47% 1	50.47% 1	50.22% 7
US	50.05% 11	50.03% 13	50.22% 6	50.34% 4	50.25% 5	50.35% 3	50.14% 9	50.15% 8	50.03% 12	49.93% 15	50.12% 10	49.97% 14	50.36% 1	50.36% 1	50.20% 7
	TS-DMSFE (.90)														
France	50.06% 11	50.03% 13	50.24% 6	50.37% 4	50.27% 5	50.39% 3	50.14% 8	50.14% 9	50.04% 12	49.91% 15	50.12% 10	49.94% 14	50.40% 1	50.40% 1	50.21% 7
Germany	50.07% 11	50.03% 13	50.26% 6	50.41% 4	50.29% 5	50.42% 3	50.14% 8	50.13% 9	50.05% 12	49.90% 15	50.11% 10	49.91% 14	50.43% 1	50.43% 1	50.22% 7
Irish Republic	50.10% 10	50.05% 13	50.25% 6	50.45% 4	50.28% 5	50.47% 3	50.17% 8	50.16% 9	50.08% 11	49.93% 14	50.05% 12	49.79% 15	50.48% 2	50.48% 1	50.20% 7
Italy	50.09% 6	37.97% 12	50.78% 2	38.35% 11	12.75% 14	38.44% 10	25.30% 13	50.39% 4	50.18% 5	49.87% 7	12.53% 15	49.64% 8	38.46% 9	51.28% 1	50.64% 3
the Netherlands	50.07% 6	37.91% 12	50.88% 2	38.37% 11	12.76% 14	38.42% 10	25.27% 13	50.32% 4	50.08% 5	49.79% 8	12.54% 15	49.81% 7	38.46% 9	51.27% 1	50.62% 3

Spain	49.99% 7	37.84% 12	50.76% 2	38.23% 11	12.69% 14	38.29% 10	25.25% 13	50.39% 4	50.02% 5	49.85% 8	12.60% 15	50.02% 6	38.32% 9	51.09% 1	50.60% 3
US	50.01% 6	37.88% 12	50.81% 2	38.29% 11	12.71% 14	38.33% 10	25.26% 13	50.35% 4	50.03% 5	49.82% 8	12.57% 15	49.97% 7	38.37% 9	51.16% 1	50.61% 3
	TS-DMSFE (.95)														
France	50.04% 6	37.90% 12	50.86% 2	38.33% 11	12.73% 14	38.38% 10	25.27% 13	50.33% 4	50.06% 5	49.80% 8	12.56% 15	49.89% 7	38.41% 9	51.21% 1	50.63% 3
Germany	50.09% 6	37.96% 12	50.77% 2	38.34% 11	12.75% 14	38.40% 10	25.29% 13	50.34% 4	50.13% 5	49.88% 7	12.51% 15	49.69% 8	38.41% 9	51.21% 1	50.56% 3
Irish Republic	47.11% 15	51.95% 7	52.10% 6	52.38% 3	52.46% 2	53.38% 1	48.01% 12	47.12% 14	52.21% 4	47.18% 13	51.53% 9	52.17% 5	49.05% 10	48.85% 11	51.89% 8
Italy	46.24% 13	52.73% 6	56.40% 1	55.08% 2	53.16% 4	53.20% 3	46.90% 12	44.36% 15	52.86% 5	45.73% 14	51.79% 9	51.91% 8	48.77% 10	47.97% 11	52.51% 7
the Netherlands	46.28% 13	52.29% 6	57.86% 1	54.52% 2	53.21% 5	53.30% 4	46.71% 12	43.04% 15	53.61% 3	45.96% 14	51.41% 9	51.78% 8	48.80% 10	48.45% 11	52.14% 7
Spain	46.24% 13	52.41% 6	57.44% 1	54.72% 2	53.20% 5	53.28% 4	46.67% 12	43.43% 15	53.51% 3	45.91% 14	51.53% 9	51.79% 8	48.79% 10	48.28% 11	52.31% 7
US	46.22% 13	52.58% 6	56.97% 1	54.91% 2	53.19% 5	53.25% 4	46.75% 12	43.85% 15	53.27% 3	45.80% 14	51.65% 9	51.81% 8	48.77% 10	48.11% 11	52.43% 7
	TS-inverse-MAE														
France	46.80% 13	52.22% 6	52.91% 3	52.98% 2	52.64% 4	53.30% 1	47.75% 12	46.57% 15	52.12% 8	46.67% 14	51.86% 9	52.23% 5	48.96% 10	48.55% 11	52.15% 7
Germany	46.11% 8	53.64% 3	53.52% 4	16.37% 15	38.45% 9	54.96% 1	47.72% 6	46.34% 7	53.68% 2	30.75% 12	28.71% 13	31.14% 11	34.55% 10	48.83% 5	24.00% 14
Irish Republic	44.91% 7	55.23% 2	58.59% 1	16.09% 15	40.91% 9	54.65% 3	46.32% 6	42.02% 8	54.38% 4	31.41% 11	28.14% 13	31.30% 12	32.82% 10	47.50% 5	24.76% 14
Italy	44.98% 7	54.43% 4	59.81% 1	14.10% 15	42.10% 8	54.78% 3	46.16% 6	40.48% 9	55.57% 2	33.15% 10	27.47% 13	30.43% 12	32.20% 11	48.15% 5	24.65% 14
the Netherlands	44.93% 7	54.67% 4	59.50% 1	14.70% 15	41.75% 8	54.72% 3	46.12% 6	40.91% 9	55.33% 2	32.66% 10	27.66% 13	30.67% 12	32.37% 11	47.98% 5	24.75% 14
Spain	44.91% 7	54.92% 3	59.04% 1	15.40% 15	41.34% 9	54.69% 4	46.20% 6	41.44% 8	55.00% 2	32.03% 11	27.92% 13	30.95% 12	32.59% 10	47.76% 5	24.73% 14
US	45.76% 7	54.24% 3	54.60% 2	16.90% 15	38.56% 9	54.96% 1	47.41% 6	45.46% 8	53.30% 4	30.31% 12	28.92% 13	31.93% 11	33.94% 10	48.45% 5	24.17% 14

Notes: There are two figures presented in each cell with the ones in the first line being the frequencies and the ones below them being the ranks. Two decimal places are retained for the frequencies. The frequencies and ranks of the top three individual constituent models are highlighted for each combination method and each origin.

Table 5-37 Frequencies and Rank of Individual Components in Superior Combination Forecasts for Each Origin

origin	bounds©	ETS	SARIMA	ADLM©	State ETS	LI©	VAR©	TVP©	SD©	bounds	ADLM	LI	VAR	TVP	SD
France	49.70% 2	44.49% 11	45.96% 7	42.67% 13	45.68% 8	44.01% 12	46.37% 5	45.47% 9	49.44% 3	46.66% 4	44.90% 10	50.50% 1	41.93% 14	39.53% 15	46.01% 6
Germany	49.10% 2	46.14% 8	47.36% 4	41.92% 13	47.15% 6	44.65% 10	46.39% 7	45.71% 9	50.77% 1	47.31% 5	43.91% 11	48.60% 3	41.34% 14	39.13% 15	43.86% 12
Irish Republic	48.82% 3	46.57% 7	46.86% 6	40.85% 14	48.69% 4	44.70% 11	48.07% 5	44.79% 10	49.80% 1	44.97% 9	45.94% 8	49.08% 2	41.09% 13	38.90% 15	43.94% 12
Italy	48.70% 3	46.48% 8	48.48% 4	41.06% 14	46.65% 7	47.04% 5	46.77% 6	44.36% 11	49.70% 1	44.91% 9	42.93% 12	49.18% 2	40.03% 15	42.03% 13	44.78% 10
the Netherlands	48.93% 2	44.95% 9	47.11% 4	43.45% 13	45.11% 8	43.74% 12	46.70% 5	44.74% 10	50.35% 1	46.68% 6	44.74% 11	48.62% 3	41.30% 15	41.40% 14	45.47% 7
Spain	48.63% 5	49.32% 2	47.15% 6	39.63% 15	45.25% 8	43.72% 11	45.95% 7	49.04% 3	49.99% 1	44.94% 9	42.83% 12	48.96% 4	41.20% 13	41.08% 14	44.89% 10
US	49.04% 4	44.84% 6	50.09% 1	41.18% 15	44.69% 9	44.16% 12	45.90% 5	44.56% 10	49.42% 3	44.78% 8	43.16% 13	49.57% 2	44.27% 11	42.29% 14	44.81% 7

Notes: There are two figures presented in each cell with the ones in the first line being the frequencies and the ones below them being the ranks. Two decimal places are retained for the frequencies. The frequencies and ranks of the top three individual constituent models are highlighted for each origin.

5.5 How Many Models to Combine?

For each origin, there exist a forecast that is the most accurate in every comparison based on combination method, forecasting horizon and accuracy measure. And the most accurate forecasts can be combination forecasts of different components, the number of which ranges from 2 to 15. Is there an optimal number of the components in the combination that can lead to the best composite forecasts? Or how many individual models to combine can contribute to the best forecasts? To reveal the answer, the first combination group which combines all 15 models through 12 weighting schemes are examined to take all individual models and all weighting schemes into account. The most accurate forecasts among the 15 individual and 32752 combination candidates are identified in every 144 comparisons (i.e., 12 combination methods, four forecasting horizons and three accuracy measures) for each origin, and the number of their components as well as the proportions that they account for in all 144 best forecasts are summarized.

Table 5-38 shows the distribution of the 144 best forecasts amongst the one- (i.e., no combination) to 15-component combinations for each origin. It is clear that the most accurate forecasts are all combination forecasts, which reinforces the conclusion that combining individual models can improve forecasting accuracy.

The distribution of the number of constituents in the best forecasts varies according to origin market, and the optimal number ranges from two to six. In six out of seven markets, the two-component combination forecasts take up more than half of all the best forecasts, with the proportion ranging from 50.69% for the Italian case to 94.44% for the French market. For all origins, the number of constituents in the best forecasts are below seven, which means that combining more than six individuals cannot bring about the most accurate forecasts.

For the French, Italian and American cases, the best forecasts are comprised of two to four components, and the proportions of the two-component combination forecasts are all above 50%. As far as tourism demand from Germany and Irish Republic are concerned, the number of individuals in the best composite forecasts are between two to five, with two and three being the most frequent ones for the German case, and two and four being the most frequent ones for the Irish origin. As with the Dutch market, the best forecasts are composed of either two or three components, 93.06% of which are made up by the two-component combinations. With respect to the Spanish origin, the number of constituents in the best forecasts ranges from two to six, with the number of four having the highest frequency, followed by the number of two and three.

It indicates that, the most accurate forecasts are always generated by combining two to six individual projections. Although the number of individual models in the combination group is as large as 15, combining less than half of them is enough to generate the best result. However, it does not mean that there is no need to include so many individual models in the combination. More individual models to be combined means more information to be used and more possible combinations. As shown in table 5-35, combining all 15 models are better than combining 9 models in the second or third group, which means that including all single models in the combination panel is beneficial. And discussion in section 5.4 also demonstrates that every model has a chance to contribute to the superior combination forecasts for most cases. The constituents in the best forecasts can be any of the 15 individual models, but the number of the constituents in the best forecasts is not the greater the better.

Table 5-38 Distribution of the Best Forecasts in Terms of Number of Components for Seven Origins (Number and Proportion)

No. of constituents	France	Germany	Irish Republic	Italy	the Netherlands	Spain	the US
1	0	0	0	0	0	0	0
2	136 94.44%	96 66.67%	88 61.11%	73 50.69%	134 93.06%	39 27.08%	118 81.94%
3	4 2.78%	43 29.86%	19 13.19%	67 46.53%	10 6.94%	30 20.83%	19 13.19%
4	4 2.78%	3 2.08%	31 21.53%	4 2.78%	0	64 44.44%	7 4.86%
5	0	2 1.39%	6 4.17%	0	0	2 1.39%	0
6	0	0	0	0	0	9 6.25%	0
7	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0
Total	144 100%	144 100%	144 100%	144 100%	144 100%	144 100%	144 100%

Notes: Except for the cells with the number of zeros, there are two figures presented in each cell with the ones in the first line being the numbers of the components in the best combination forecasts and the ones below them being the corresponding percentages. Two decimal places are retained for the percentages.

5.6 Conclusion

In this chapter, inbound tourism demand to the UK from seven leading markets are forecasted using a wide range of individual and combination forecasting methods, and comprehensive comparisons, which include country-specific, forecasting-horizon-specific and general comparisons regarding the forecasting ability of different approaches are conducted.

As far as the forecasting performance of individual models are concerned, no single model can generate the best forecast in all situations. It is shown that the forecasting accuracy of various single models depends mainly on the origin market under consideration, which highlights the importance of considering the apparent sharp national contrasts among origin countries when investigating inbound tourist arrivals to the UK. In general, non-causal time series techniques are superior compared to causal econometric models in terms of generating more accurate forecasts, and the state space ETS models is the best choice. The climate ADLM performs the best among all econometric models, and the climate LI model is the worst choice.

One possible reason for the forecasting inferiority of the causal econometric models could be the possible inappropriate choice of the measurement of different model variables. For example, tourism demand is measured by tourist arrivals to the UK from seven markets, which ignores both tourists' length of stay and their expenditure while at the destination, and therefore cannot reflect the actual demand for tourism. Besides, tourism prices are measured by CPI of the corresponding country (adjusted by exchange rates) rather than by tourists price indices, as there is no continuous and consistent data. But the baskets of goods consumed by tourists tend to be different to those consumed by the local residents. As a result, the price variables in the model may not exactly capture the movements of tourism prices. Furthermore, the climate attribute of the UK is measured by the UK's TCI. Although the TCI integrates the most relevant climate elements to tourists' experience including the thermal, physical and aesthetic aspects, it may not assign appropriate weights to them. The weights of different elements in TCI are decided subjectively based on expert judgements and meteorological literature, which lacks empirical validation and may not able to reveal the attitudes of inbound travelers to the UK. It is ideal to adjust the weights based on the information collected from a survey, which is conducted to international tourists to the UK regarding their preferences towards different elements of the climate index.

Whether introducing the climate factor can improve the forecasting ability of econometric models also depends on the origin country under consideration. For the Irish and Italian cases, the climate econometric models generate more accurate forecasts compared to their traditional counterparts, and for other origins, the traditional econometric specification, which excludes the climate determinant, performs better. The climate determinant has been found to have significant impact on tourism demand at the 1% significance level for the Irish and Italian markets (refer to table 4-3), which can explain the superior performance of the climate econometric models when forecasting tourism demand from these two countries. The climate factor influences tourism demand significantly and can contribute to better demand forecasts. It suggests that, the characteristics of different markets should be taken into consideration when forecasting tourism demand, and the effect of the climate factor on tourism demand should be estimated firstly. If there exists significant relationship between the climate factor and tourism demand, the climate factor should be considered in the forecasting models.

As with the performance of the combination forecasting approach, it is revealed that combining individual forecasts can improve the forecasting performance in all cases regardless of origin country, forecasting length, accuracy measure, combination group or combination method. The most accurate forecasts are always produced by the combination forecasting approach, and the worst projections are always generated through the individual forecasting approach.

There is no clear-cut evidence showing which combination method is the best. The origin country and forecasting horizon under consideration, the combination method applied, and the accuracy measure used all affect the forecasting performance of different weighting schemes.

Besides, the best-performing methods vary across the three combination groups. In general, the inverse-MAE method performs the best for the first combination group, the VACO method is the best choice for the second group, and the DMSFE schemes rank at the top for the third group. The two-stage combination methods are inferior, and the SA combination method provides the worst results except the two-stage ones. The general inferior performance of the two-stage combination methods compared to the one-stage ones shows that excluding the worst-performing individual forecasts from the combination panel does not contribute to better combination forecasts.

The comparison across three combination groups shows that including all models, which consist of traditional and climate econometric models, as well as time series techniques, can bring about the best combination forecasts. It demonstrates that it is beneficial to introduce the climate factor into combination, and the information embedded in different model specifications can contribute to better forecasts.

With respect to which models to combine, it is revealed that every individual model has a chance to be a component in the superior combination forecasts for most cases, no matter how bad its forecasting ability is. For all markets, the climate bounds test, climate SD and traditional LI models have high opportunities to construct the superior combination forecasts, while the climate ADLM, traditional VAR and traditional TVP models are the ones that have the lowest chances to be in the superior composite projections. The country-specific evaluation shows that, for the same origin, the difference among each individual model's opportunity to make up the superior combination forecasts is not great; and for one individual model, its chances to constitute the superior composite forecasts for different markets are comparable.

In addition, it is revealed that the frequencies of various individual models to constitute the superior combination forecasts is irrelevant to their forecasting performance. The individual models that can forecast more accurately do not have higher chances to be in the combination forecasts.

Regarding how many models to combine to obtain the best forecast, it is shown that the best forecasts in every comparison are always combination forecasts which are comprised of two to six constituents. And for six out seven markets, combining two individual models have the highest possibility to lead to the best forecasts. The large number of individual models available to be combined in the combination group means more possible combination forecasts, which is beneficial as there is more information to be integrated. It does not necessarily mean that the number of constituents in the best forecast is also large.

Chapter 6 Conclusion

6.1 Introduction

The primary aim of this research is to assess whether including the climate factor in econometric models and combination forecasts can improve tourism demand forecasting performance. The thesis starts with a systematic review on tourism demand modelling and forecasting literature, followed by an illustration of the research method applied in this study. The impact of the climate factor on UK inbound tourism demand from seven leading markets are evaluated next. And the forecasting performance of various individual and combination models are assessed and compared. In brief, the empirical results show that the forecasting ability of econometric models can be improved by introducing the climate factor in some cases. And generally, the combination forecasts that are generated through combining traditional and climate econometric models as well as time series techniques are superior to others for all combination methods evaluated by MAE, MAPE and RMSE, which means that including the climate factor in the combination is beneficial.

The rest of this chapter is arranged into three sections. Section 6.2 summarizes of the main findings of previous chapters with the purpose of clarifying the theoretical and practical contributions of the current research. Section 6.3 discusses the limitations of the research and section 6.4 provides the recommendations for future research.

6.2 Summary of The Findings

The current research seeks to understand the value of the climate factor in forecasting tourism demand. Climate is a crucial resource for tourism and a main driver of international tourist flow. However, climate variables are generally neglected when forecasting tourism demand, especially when the combination forecasting approach is applied.

There exist two quantitative forecasting approaches in the current tourism demand literature: the individual forecasting approach and the combination forecasting approach. Individual forecasts are direct projections generated by various single models, which are dominated by the causal econometric models and the non-causal time series techniques. Combination forecasts are indirect ones generated through combining individual models using different combination methods. In the current tourism demand forecasting literature, the influencing factors identified in econometric forecasting models mainly belong to the economic ones, and climate variables are generally neglected. Furthermore, the single econometric models included in the combination forecasting methods are confined to the ones that exclude the climate factor. No combination approach has been applied taking econometric models with different influencing factors as constituent individual models.

In order to fill these research gaps, the current research forecasts UK inbound tourism demand through a wide range of forecasting methods including individual forecasting models, which are diverse in modelling techniques and are different in identified influencing factors, as well as combination forecasting models, which differ in weighting schemes and vary in individual constituents. Comprehensive forecasting comparisons are conducted among various individual and combination forecasts to reveal which forecasting approach is superior and whether including the climate factor in econometric and combination forecasting models is beneficial.

One important purpose of this research is to evaluate the impact of the climate factor on tourism demand through an empirical study on UK inbound tourism industry.

According to the bounds test results, there exist long-run relationships between tourism demand and the identified influencing factors which include income, own price, substitute price and the climate factor (proxied by UK's TCI) in all seven origin-destination pairs. The estimation results show that there exists sharp market divergence regarding the determinants of tourism demand. All influencing factors are estimated with expected signs. Income is proved to be statistically significant for the French, German, Italian and American cases; own price is estimated to be a significant demand determinant for the French, German, Irish and Dutch markets; and substitute price has significant effect on tourism demand for the French, Italian, Spanish and American origins. The estimated coefficients of the climate factor are of positive signs for all countries and are statistically significant in six out of seven cases, which confirms the assumption that better climate condition in the destination contributes to more inbound tourist flows. The statistically significant dummy variables including the 9/11 terrorist attack, the outbreak of the foot-and-mouth disease at the beginning of 2001 and the 2008 financial crisis demonstrate that the UK tourism industry is highly vulnerable to adverse events.

The primary aim of the current research is to explore the value of the climate factor in forecasting tourism demand.

To test the forecasting ability of alternative individual models, the climate factor is introduced to the econometric forecasting models, and the forecasting performance of the climate econometric models, which include the climate factor as a demand determinant, the traditional econometric models, which only consider economic factors and dummy variables as explanatory variables, as well as the time series techniques are compared. The country-specific assessment, the forecasting-horizon-specific evaluation and the general comparison are all made based on three accuracy measures including MAE, MAPE and RMSE. It shows that there is no single model which can forecast the best in all situations, which means that model customization is required to improve the forecasting accuracy over different markets and forecasting horizons. Generally, non-causal time series techniques are better than causal econometric models. Including the climate factor can improve the forecasting ability of econometric models for some markets. It suggests that the impact of the climate factor on tourism demand should be evaluated before forecasting. When there exists significant effect of the climate factor, forecasting tourism demand will benefit by including the climate factor in the econometric models.

To investigate the forecasting performance of alternative combination forecasting methods, the individual models are categorized into three groups: the first group containing all individual models, the second group consisting of the traditional econometric and time series models, and the third groups including the climate econometric and time series models. Combination forecasts are generated through various weighting schemes regarding three groups respectively.

It demonstrates that there are always combination forecasts that are better than the best individual forecasts, and the worst forecasts are always produced by individual forecasting models. It means that the combination forecasting approach is superior to the individual one, as it can improve forecasting accuracy and reduce forecasting failure. Combining all individual models in the first group, which include traditional econometric, climate econometric and time series models produce the best combination forecasts, which means that introducing the climate variable into combination contribute to more accurate projections. The performance of different weighting schemes depends on the origin market under study, the forecasting length under consideration, the accuracy measure used and the combination group under analysis. In general, the two-stage combination method performs unsatisfactorily, but the inverse-MAE method performs very well.

It also reveals that the chances of individual models to be a constituent in the superior combination forecasts are irrelevant to their forecasting abilities. Individual models that can forecast more accurately do not have higher opportunities to constitute the superior combination forecasts. The best forecasts in each comparison are all combination forecasts, and the number of constituents in the best forecasts range from two to six. Since the total number of the individual models in three combination group is either nine or 15, including all single models in the combination group in one combination forecast is not recommended.

The current research contributes to our knowledge on the tourism demand analysis in general and on the tourism industry of the UK on both the theoretical front and the practical front:

1. This research represents the first attempt to investigate combination forecasts which include econometric models with different explanatory variables as constituents, and the climate factor is, for the first time, introduced to the combination forecasting approach;
2. This research represents the first effort to examine combination forecasts with more than ten single models as constituent forecasts in the combination group;
3. This research introduces new combination methods to the tourism demand forecasting literature and tests their forecasting abilities in the empirical study;
4. The empirical results provide new insights into forecasting both in the tourism context and other economic disciplines;
5. This research provides demand elasticity estimates based on quarterly data from 1994Q1 to 2017Q4, which will assist the government and destination managers in matters of policy formulation and implementation;
6. This research empirically proves the existence of a link between the climate factor and inbound tourism demand to the UK.

The results of the current research have implications for future research and for stakeholders in the tourism industry. Firstly, it suggests more research attention to the combination forecasting approach. This research shows the general forecasting superiority of the combination forecasting approach compared to the individual forecasting approach. In the current literature, too much attention is paid to improving the forecasting ability of individual forecasting models and to identifying the best single forecasting model. It has been proven, once again, that no single model can forecast the best in all situations, which makes identifying the best single forecasting model a moving target. But there are always a portion of combination forecasts that are superior to the best single projections. If improving forecasting accuracy is the aim, combination forecasts deserve more study and should be included in forecasting comparisons.

In addition, it shows that the diversity in constituent individual models included in the combination group can contribute to more accurate forecasts. This research paves the way for further empirical investigations on combination forecasts including component models which are diverse not only in modelling techniques but also in model specifications.

Besides, policy-makers in public and private sectors can make use of the elasticity estimates in their planning process. For example, tourist arrivals from France, Germany and the Netherlands are proved to be price-elastic, which means that in order to increase tourism incomes from these markets, policy-makers can implement low-pricing strategy. Tourists from France and the US regard alternative countries as substitute destinations to the UK and they are sensitive to price fluctuations in the competing destinations. Providers of tourism services need to monitor the price strategies of the competing markets and maintain competitive prices to attract consumers from France and the US.

Moreover, government and destination managers can improve the efficiency of their planning exercises by taking into account additional information on climate trends. The climate condition in the destination is proved to have positive impact on international tourism demand: the better the climate condition in the destination is, the more the tourist arrivals are. It implies that climate change is projected to have significant impacts on the distribution of international tourism demand, as physical resources supporting tourism in each country is affected by climate change. For instance, given the current trends of climate change, southern Europe will experience climate conditions that are less favourable to tourists than the current climate conditions in summer. At the same time, countries in northern Europe, which are the countries of origin of many of the current visitors of the Mediterranean, will enjoy better climate conditions in summer, as well as a longer season with good weather. On the other hand, in winter, snow depth in northern Europe is supposed to be seriously affected, which will result in less tourist arrivals. As a result, it is expected that there will be redistribution of international tourism demand both temporally and spatially due to climate change.

Public sectors and businesses in the tourism industry should take the impact of climate into account when forming policies and strategies. Consumer satisfaction is one of the top priorities of service providers, and climate can affect consumer satisfaction in the tourism industry. Resources should be allocated accordingly to meet the high demand in the seasons that the climate is favourable to tourists. At the same time, tourism destinations and businesses should try to reduce tourists' vulnerability to climate change by offering a diverse set of tourism activities, especially in the seasons which have bad climate conditions. Examples include developing all-year-round tourism activities, building venues suitable for various types of tourism activities, promoting indoor tourism activities such as museum visiting in the bad season, advancing less climate-dependent types of tourism, and taking technical measures such as artificial snowmaking and air conditioning to offset the impact of bad climate conditions on tourism demand.

6.3 Limitations of The Current Research

The current research, like other studies, is not without limitations.

Firstly, important determinants may be missing from the causal econometric models, and the measurements of different model variables may be inappropriate.

For instance, no variables for travel costs are included due to unavailability of suitable data. Tourism demand is measured by tourist arrivals, which may fail to represent the actual demand for tourism, as it disregards both tourists' length of stay and their expenditure at the destination. Owing to the lack of continuous and consistent data, leisure traveler's income is proxied by GDP instead of disposable income, and tourism prices are measured by the exchange rate adjusted CPI instead of the exchange rate adjusted tourists price index. Besides, the climate factor is represented by the destination's condition which is measured by the UK's TCI. The climate condition in the origin country and other competing destinations is excluded from this study. Climate is an important tourism resource which can attract tourists to travel abroad or stay at home. The climate difference between the destination and the origin, and the main destination's climate condition relative to those of the competing areas may play a crucial role in shifting tourist flows. And the current research chooses

Mieczkowski's TCI as the tourism climate index, which is the best choice among all existing climate index to represent tourists' attitudes towards climate. But TCI is constructed subjectively based on expert judgements and meteorological literature, which lacks empirical validation and may not be able to reveal the true attitudes of inbound travelers to the UK.

Secondly, there exist limitations in the forecasting and evaluation practice.

Only one- to four-step-ahead forecasts are generated and other forecasting horizons are omitted. The out-of-sample forecasting performance is evaluated based on data from 2015Q1 to 2017Q4, which is quite limited in sample size, using descriptive accuracy measures including MAE, MAPE and RMSE. Statistical tests regarding whether the difference in the accuracy of competing forecasting methods is statistically significant are not conducted. In addition, AI-based individual and combination forecasting methods are not included as they are beyond the researcher's expertise.

6.4 Recommendations for Future Research

The results of the current research have the potential for influencing further research.

The empirical findings of the current research support the application of the combination forecasting approach in the tourism context. An obvious extension is to investigate different combination forecasting methods for more origin-destination pairs and other forecasting horizons. The individual models contained in the combination group can be expanded to other modelling techniques such as the recently introduced BGVAR model and SSA, and the combination methods evaluated can be extended to AI-based weighting schemes. And formal statistical tests such as D-M test, HLN test and Clements & Harvey test can be conducted to evaluate whether combination forecasts are significantly better than single forecasts.

Combining econometric models with different explanatory variables has been examined for the first time in the tourism demand literature, and it has generated good demand forecasts. It deserves more study 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 the influencing factors of the included econometric models. And the value of other factors such as search engine data in improving combination forecasting accuracy can be explored.

The biggest obstacle of popularizing the combination forecasting approach 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. It suggests that a user-friendly software which can produce combination forecasts easily should be made available considering the powerful forecasting ability of the combination forecasting approach. And with the help of the software, combination forecasts should be included in forecasting comparisons and can be used as benchmarks for forecasting evaluation.

Besides, our understanding of tourists' preferences with respect to climate conditions remains very limited. It is supposed that preferences should differ between tourism activities and between tourists from different countries and cultures. More empirical research is required to reveal tourists' attitudes. For example, surveys can be conducted to tourists to different destinations, from different origins or for different types of tourism to collect information regarding their preferences towards different elements of the climate, and different types of tourism climate index can be constructed for country-specific and segment-specific studies.

Bibliography

Agnew, M. and Palutikof, J., 2006. Impacts of short-term climate variability in the UK on demand for domestic and international tourism. *Climate Research*, 31 (1), 109-120.

Akın, M., 2015. A novel approach to model selection in tourism demand modeling. *Tourism Management*, 48, 64-72.

Albaladejo, I. P., González-Martínez, M. I. and Martínez-García, M. P., 2016. Nonconstant reputation effect in a dynamic tourism demand model for Spain. *Tourism Management*, 53, 132-139.

Algieri, B. and Kanellopoulou, S., 2009. Determinants of demand for exports of tourism: An unobserved component model. *Tourism and Hospitality Research*, 9 (1), 9-19.

Amelung, B. and Moreno, A., 2012. Costing the impact of climate change on tourism in Europe: results of the PESETA project. *Climatic Change*, 112 (1), 83-100.

Andrawis, R. R., Atiya, A. F. 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.

Andraz, J. L. M., Gouveia, P. M. D. C. B. and Rodrigues, P. M. M., 2009. Modelling and forecasting the UK tourism growth cycle in Algarve. *Tourism Economics*, 15 (2), 323-338.

Apergis, N., Mervar, A. and Payne, J. E., 2017. Forecasting disaggregated tourist arrivals in Croatia: Evidence from seasonal univariate time series models. *Tourism Economics*, 23 (1), 78-98.

Arellano, M. and Bond, S., 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The review of economic studies*, 58 (2), 277-297.

Artus, J. R., 1972. An econometric analysis of international travel. *Staff papers*, 19 (3), 579-614.

Assaf, A. G., Barros, C. P. and Gil-Alana, L. A., 2011. Persistence in the Short- and Long-Term Tourist Arrivals to Australia. *Journal of travel research*, 50 (2), 213-229.

Assaf, A. G., Li, G., Song, H. and Tsionas, M. G., 2018. Modeling and Forecasting Regional Tourism Demand Using the Bayesian Global Vector Autoregressive (BGVAR) Model. *Journal of Travel Research*, 0047287518759226.

Athanasopoulos, G., Ahmed, R. A. and Hyndman, R. J., 2009. Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, 25 (1), 146-166.

Athanasopoulos, G., Deng, M. F., Li, G. and Song, H. Y., 2014. Modelling substitution between domestic and outbound tourism in Australia: A system-of-equations approach. *Tourism Management*, 45, 159-170.

Athanasopoulos, G. and Hyndman, R. J., 2008. Modelling and forecasting Australian domestic tourism. *Tourism Management*, 29 (1), 19-31.

Athanasopoulos, G., Hyndman, R. J., Song, H. and Wu, D. C., 2011. The tourism forecasting competition. *International Journal of Forecasting*, 27, 822-844.

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

Athanasopoulos, G., Song, H. and Sun, J. A., 2017. Bagging in Tourism Demand Modeling and Forecasting. *Journal of Travel Research*, 0047287516682871.

Balli, F., Balli, H. O. and Jean Louis, R., 2016. The impacts of immigrants and institutions on bilateral tourism flows. *Tourism Management*, 52, 221-229.

Balli, F. and Tsui, W. H. K., 2016. Tourism demand spillovers between Australia and New Zealand: Evidence from the partner countries. *Journal of Travel Research*, 55 (6), 804-812.

Banerjee, A., Dolado, J. J., Galbraith, J. W. and Hendry, D., 1993. Co-integration, error correction, and the econometric analysis of non-stationary data. *OUP Catalogue*.

Bangwayo-Skeete, P. F. and Skeete, R. W., 2015. Can Google data improve the forecasting performance of tourist arrivals? Mixed-data sampling approach. *Tourism Management*, 46, 454-464.

Barry, K. and O'Hagan, J., 1972. An econometric study of British tourist expenditure in Ireland. *Economic and Social Review*, 3 (2), 143-161.

Bates, J. M. and Granger, C. W. J., 1969. The Combination of Forecasts*,* 451.

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.

Bento, J. P. C., 2014. The determinants of international academic tourism demand in Europe. *Tourism Economics*, 20 (3), 611-628.

Blundell, R. and Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87 (1), 115-143.

Bonham, C., Gangnes, B. and Zhou, T., 2009. Modeling tourism: A fully identified VECM approach. *International Journal of Forecasting*, 25 (3), 531-549.

Box, G. E., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M., 2015. *Time series analysis: forecasting and control*. John Wiley & Sons.

Breusch, T. S., 1978. Testing for autocorrelation in dynamic linear models. *Australian Economic Papers*, 17 (31), 334-355.

Breusch, T. S. and Pagan, A. R., 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society*, 1287-1294.

Brida, J. G. and Risso, W. A., 2009. A dynamic panel data study of the German demand for tourism in South Tyrol. *Tourism and Hospitality Research*, 9 (4), 305-313.

Brida, J. G. and Risso, W. A., 2011. Tourism demand forecasting with SARIMA models - the case of South Tyrol. *Tourism Economics*, 17 (1), 209-221.

Buigut, S., 2018. Effect of terrorism on demand for tourism in Kenya: A comparative analysis. *Tourism and Hospitality Research*, 18 (1), 28-37.

Camacho, M. and Pacce, M. J., 2018. Forecasting travellers in Spain with Google's search volume indices. *Tourism Economics*, 24 (4), 434-448.

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

Cang, S., 2014. 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.

Cang, S. and Hemmington, N., 2010. Forecasting UK inbound expenditure by different purposes of visit. *Journal of Hospitality & Tourism Research*, 34 (3), 294-309.

Cao, Z., Li, G. and Song, H., 2017. Modelling the interdependence of tourism demand: The global vector autoregressive approach. *Annals of Tourism Research*, 67, 1-13.

Carson, R. T., Cenesizoglu, T. and Parker, R., 2011. Forecasting (aggregate) demand for US commercial air travel. *International journal of Forecasting*, 27 (3), 923-941.

Cazanova, J., Ward, R. W. and Holland, S., 2014. Habit persistence in air passenger traffic destined for Florida. *Journal of Travel Research*, 53 (5), 638-655.

Cegnar, T. and Matzarakis, A., 2004. Trends of thermal bioclimate and their application for tourism in Slovenia. *Advances in tourism climatology*, 12, 66-73.

Chan, C. K., Witt, S. F., Lee, Y. and Song, H., 2010. Tourism forecast combination using the CUSUM technique. *Tourism Management*, 31 (6), 891-897.

Chan, F. and Lim, C., 2011. Spectral analysis of seasonality in tourism demand. *Mathematics and Computers in Simulation*, 81, 1409-1418.

Chatziantoniou, I., Degiannakis, S., Eeckels, B. and Filis, G., 2016. Forecasting tourist arrivals using origin country macroeconomics. *Applied Economics*, 48 (27), 2571-2585.

Chatziantoniou, I., Filis, G., Eeckels, B. and Apostolakis, A., 2013. Oil prices, tourism income and economic growth: A structural VAR approach for European Mediterranean countries. *Tourism Management*, 36, 331-341.

Chen, K.-Y., 2011. Combining linear and nonlinear model in forecasting tourism demand. *Expert Systems with Applications*, 38 (8), 10368-10376.

Chen, Z. and Haynes, K. E., 2015. Impact of high-speed rail on international tourism demand in China. *Applied Economics Letters*, 22 (1), 57-60.

Cheng, K., 2012. Tourism demand in Hong Kong: income, prices, and visa restrictions. *Current Issues in Tourism*, 15 (3), 167-181.

Cho, V., 2003. A comparison of three different approaches to tourist arrival forecasting. *Tourism management*, 24 (3), 323-330.

Cho, V., 2010. A study of the non-economic determinants in tourism demand. *International Journal of Tourism Research*, 12 (4), 307-320.

Chou, C. M., Hsieh, S. F. and Tseng, H. P., 2014. The crowding-out effects of Chinese tourists on inbound tourism in Taiwan. *Tourism Economics*, 20 (6), 1235-1248.

Choyakh, H., 2008. A model of tourism demand for Tunisia: inclusion of the tourism investment variable. *Tourism Economics*, 14 (4), 819-838.

Chu, F., 2011. A piecewise linear approach to modeling and forecasting demand for Macau tourism. *Tourism Management*, 32 (6), 1414-1420.

Chu, F.-L., 2008a. A fractionally integrated autoregressive moving average approach to forecasting tourism demand. *Tourism Management*, 29 (1), 79-88.

Chu, F.-L., 2008b. Analyzing and forecasting tourism demand with ARAR algorithm. *Tourism Management*, 29 (6), 1185-1196.

Chu, F.-L., 2009. Forecasting tourism demand with ARMA-based methods. *Tourism Management*, 30 (5), 740-751.

Claveria, O., Monte, E. and Torra, S., 2015. Tourism Demand Forecasting with Neural Network Models: Different Ways of Treating Information. *International Journal of Tourism Research*, 17 (5), 492-500.

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

Claveria, O. and Torra, S., 2014. Forecasting tourism demand to Catalonia: Neural networks vs. time series models. *Economic Modelling*, 36, 220-228.

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

Clemen, R. T. and Winkler, R. L., 1986. Combining economic forecasts. *Journal of Business & Economic Statistics*, 4 (1), 39-46.

Clements, M. and Hendry, D., 1998. *Forecasting economic time series*. Cambridge University Press.

Cortes-Jimenez, I. and Blake, A., 2011. Tourism Demand Modeling by Purpose of Visit and Nationality. *Journal of travel research*, 50 (4), 408-416.

Cortes-Jimenez, I., Durbarry, R. and Pulina, M., 2009. Estimation of outbound Italian tourism demand: a monthly dynamic EC-LAIDS model. *Tourism Economics*, 15 (3), 547-565.

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

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

Croes, R., Ridderstaat, J. and Rivera, M., 2017. Asymmetric Business Cycle Effects and Tourism Demand Cycles. *Journal of Travel Research*, 0047287517704086.

Crouch, G. I., 1992. Effect of income and price on international tourism. *Annals of Tourism Research*, 19 (4), 643-664.

Crouch, G. I., 1994. The study of international tourism demand: a review of findings. *Journal of Travel Research*, 33 (1), 12-23.

Cuccia, T. and Rizzo, I., 2011. Tourism seasonality in cultural destinations: Empirical evidence from Sicily. *Tourism Management*, 32 (3), 589-595.

Daniel, A. and Rodrigues, P., 2012. Assessing the impact of shocks on international tourism demand for Portugal. *Tourism Economics*, 18 (3), 617-634.

Deaton, A. and Muellbauer, J., 1980. An almost ideal demand system. *The American economic review*, 70 (3), 312-326.

Dergiades, T., Mavragani, E. and Pan, B., 2018. Google Trends and tourists' arrivals: Emerging biases and proposed corrections. *Tourism Management*, 66, 108-120.

Dickey, D. A. and Fuller, W. A., 1979. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American statistical association*, 74 (366a), 427-431.

Dickey, D. A. and Fuller, W. A., 1981. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica: Journal of the Econometric Society*, 1057-1072.

Diebold, F. X. and Mariano, R. S., 2002. Comparing predictive accuracy. *Journal of Business & economic statistics*, 20 (1), 134-144.

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

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

Divino, J. A. and McAleer, M., 2010. Modelling and forecasting daily international mass tourism to Peru. *Tourism Management*, 31 (6), 846-854.

Divisekera, S., 2009. Ex post demand for Australian tourism goods and services. *Tourism Economics*, 15 (1), 153-180.

Divisekera, S., 2010. Economics of leisure and non-leisure tourist demand: a study of domestic demand for Australian tourism. *Tourism Economics*, 16 (1), 117-136.

Dogru, T., Sirakaya-Turk, E. and Crouch, G. I., 2017. Remodeling international tourism demand: Old theory and new evidence. *Tourism management*, 60, 47-55.

Durbin, J. and Koopman, S. J., 2012. *Time series analysis by state space methods*. Oxford university press.

Durbin, J. and Watson, G. S., 1950. Testing for serial correlation in least squares regression: I. *Biometrika*, 37 (3/4), 409-428.

Durbin, J. and Watson, G. S., 1951. Testing for serial correlation in least squares regression. II. *Biometrika*, 38 (1/2), 159-177.

Dwyer, L., Seetaram, N., Forsyth, P. and King, B., 2014. Is the migration-tourism relationship only about VFR? *Annals of Tourism Research*, 46, 130-143.

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Engle, R. F. and Granger, C. W. J., 1987. Co-integration and error correction: representation, estimation, and testing. *Econometrica*, 55 (2), 251.

Eugenio-Martin, J. L. and Campos-Soria, J. A., 2010. Climate in the region of origin and destination choice in outbound tourism demand. *Tourism Management*, 31 (6), 744-753.

Eugenio-Martin, J. L. and Campos-Soria, J. A., 2011. Income and the substitution pattern between domestic and international tourism demand. *APPLIED ECONOMICS*, 43 (20), 2519-2531.

Eugenio-Martin, J. L. and Campos-Soria, J. A., 2014. Economic crisis and tourism expenditure cutback decision. *Annals of Tourism Research*, 44, 53-73.

Eugenio-Martin, J. L., Martín-Morales, N. and Sinclair, M. T., 2008. The role of economic development in tourism demand. *Tourism Economics*, 14 (4), 673-690.

Falk, M., 2010. A dynamic panel data analysis of snow depth and winter tourism. *Tourism Management*, 31 (6), 912-924.

Falk, M., 2013. Impact of long-term weather on domestic and foreign winter tourism demand. *International Journal of Tourism Research*, 15 (1), 1-17.

Falk, M., 2014. Impact of weather conditions on tourism demand in the peak summer season over the last 50years. *Tourism Management Perspectives*, 9, 24-35.

Falk, M., 2015. The sensitivity of tourism demand to exchange rate changes: an application to Swiss overnight stays in Austrian mountain villages during the winter season. *Current Issues in Tourism*, 18 (5), 465-476.

Fildes, R., Wei, Y. and Ismail, S., 2011. Evaluating the forecasting performance of econometric models of air passenger traffic flows using multiple error measures. *International Journal of Forecasting*, 27 (3), 902-922.

Folgieri, R., Baldigara, T. and Mamula, M., 2017. Artificial Neural Networks-Based Econometric Models for Tourism Demand Forecasting, *Tourism in Southern and Eastern Europe 2017*.

Fourie, J. and Santana-Gallego, M., 2011. The impact of mega-sport events on tourist arrivals. *Tourism Management*, 32 (6), 1364-1370.

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

Fuleky, P., Zhao, Q. X. and Bonham, C. S., 2014. Estimating demand elasticities in non-stationary panels: the case of Hawaii tourism. *Annals of Tourism Research*, 44, 131-142.

Garin-Munoz, T., 2009. Tourism in Galicia: domestic and foreign demand. *TOURISM ECONOMICS*, 15 (4), 753-769.

Gatt, W. and Falzon, J., 2014. British tourism demand elasticities in Mediterranean countries. *Applied Economics*, 46 (29), 3548-3561.

Ghaderi, Z., Saboori, B. and Khoshkam, M., 2016. Does security matter in tourism demand? *Current Issues in Tourism*, 1-14.

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

Gil-Alana, L. A., Mudida, R. and Pérez de Gracia, F., 2014. Persistence, long memory and seasonality in Kenyan tourism series. *Annals of Tourism Research*, 46, 89-101.

Gil-Alana, L. A., Cuñado, J. and Pérez de Gracia, F., 2008. Tourism in the Canary Islands: forecasting using several seasonal time series models. *Journal of Forecasting*, 27 (7), 621-636.

Godfrey, L. G., 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.

Godfrey, L. G., 1991. *Misspecification tests in econometrics: the Lagrange multiplier principle and other approaches*. Cambridge University Press.

Goh, C., 2012. Exploring impact of climate on tourism demand. *Annals of Tourism Research*, 39 (4), 1859-1883.

Goh, C. and Law, R., 2011. The methodological progress of tourism demand forecasting: a review of related literature. *Journal of Travel & Tourism Marketing*, 28 (3), 296-317.

Goh, C., Law, R. and Mok, H. M. K., 2008. Analyzing and Forecasting Tourism Demand: A Rough Sets Approach. *JOURNAL OF TRAVEL RESEARCH*, 46 (3), 327-338.

Gonzalez, P. and Moral, P., 1995. An analysis of the international tourism demand in Spain. *International Journal of Forecasting*, 11 (2), 233-251.

Gounopoulos, D., Petmezas, D. and Santamaria, D., 2012. Forecasting tourist arrivals in Greece and the impact of macroeconomic shocks from the countries of tourists' origin. *Annals of Tourism Research*, 39 (2), 641-666.

Granger, C. W. J., 1989. Invited Review Combining Forecasts--Twenty Years Later. *Journal of Forecasting*, 8 (3), 167-173.

Granger, C. W. J. and Ramanathan, R., 1984. Improved Methods of Combining Forecasts. *Journal of Forecasting*, 3 (2), 197-204.

Granvorka, C. and Strobl, E., 2013. The impact of hurricane strikes on tourist arrivals in the Caribbean. *Tourism Economics*, 19 (6), 1401-1409.

Gray, H. P., 1966. The demand for international travel by the United States and Canada. *International Economic Review*, 7 (1), 83-92.

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

Guizzardi, A. and Mazzocchi, M., 2010. Tourism demand for Italy and the business cycle. *Tourism Management*, 31 (3), 367-377.

Guizzardi, A. and Stacchini, A., 2015. Real-time forecasting regional tourism with business sentiment surveys. *Tourism Management*, 47, 213-223.

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.

Habibi, F., 2016. The determinants of inbound tourism to Malaysia: a panel data analysis. *Current Issues in Tourism*, 1-22.

Hadavandi, E., Ghanbari, A., Shahanaghi, K. and Abbasian-Naghneh, S., 2011. Tourist arrival forecasting by evolutionary fuzzy systems. *Tourism Management*, 32 (5), 1196-1203.

Halicioglu, F., 2010. An econometric analysis of the aggregate outbound tourism demand of Turkey. *Tourism Economics*, 16 (1), 83-97.

Hamilton, J. M., Maddison, D. J. and Tol, R. S., 2005a. Climate change and international tourism: a simulation study. *Global environmental change*, 15 (3), 253-266.

Hamilton, J. M., Maddison, D. J. and Tol, R. S., 2005b. Effects of climate change on international tourism. *Climate research*, 29 (3), 245-254.

Hamilton, J. M. and Tol, R. S., 2007. The impact of climate change on tourism in Germany, the UK and Ireland: a simulation study. *Regional Environmental Change*, 7 (3), 161-172.

Harvey, A. C., 1976. Estimating regression models with multiplicative heteroscedasticity. *Econometrica: Journal of the Econometric Society*, 461-465.

Harvey, A. C., 1990. *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.

Hassani, H., 2007. Singular spectrum analysis: methodology and comparison. *Journal of Data Science* 5, 239-257

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

Hassani, H., Heravi, S. and Zhigljavsky, A., 2013. Forecasting UK industrial production with multivariate singular spectrum analysis. *Journal of Forecasting*, 32 (5), 395-408.

Hassani, H., Webster, A., Silva, E. S. and Heravi, S., 2015. Forecasting U.S. Tourist arrivals using optimal Singular Spectrum Analysis. *Tourism Management*, 46, 322-335.

Hendry, D. F., 1995.*Dynamic econometrics*. Oxford University Press.

Holmen, J. S., 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, C. C., 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International journal of forecasting*, 20 (1), 5-10.

Huss, W. R., 1985. Comparative analysis of company forecasts and advanced time series techniques using annual electric utility energy sales data. *International Journal of Forecasting*, 1 (3), 217-239.

Hyndman, R. J., Koehler, A. B., Snyder, R. D. and Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting*, 18, 439-454.

Jarque, C. M. and Bera, A. K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6 (3), 255-259.

Johansen, S., 1988. Statistical analysis of cointegration vectors. *Journal of Economic Dynamics & Control*, 12 (2/3), 231.

Johansen, S. and Juselius, K., 1990. Maximum likelihood estimation and inference on cointegration –with applications to the demand for money. *Oxford Bulletin of Economics & Statistics*, 52 (2), 169-210.

Johnson, T. J. and Garman, A. N., 2015. Demand for international medical travel to the USA. *Tourism Economics*, 21 (5), 1061-1077.

Jose, V. R. R., and Winkler, R. L. 2008. Simple robust averages of forecasts: Some empirical results. *International Journal of Forecasting*, 24(1), 163-169.

Kalman, R. E., 1960. A new approach to linear filtering and prediction problems. *Journal of basic Engineering*, 82 (1), 35-45.

Khadaroo, J. and Seetanah, B., 2008. The role of transport infrastructure in international tourism development: A gravity model approach. *Tourism management*, 29 (5), 831-840.

Khoshnevis Yazdi, S. and Khanalizadeh, B., 2016. Tourism demand: a panel data approach. *Current Issues in Tourism*, 1-14.

Kilian, L., 1998. Small-sample confidence intervals for impulse response functions. *Review of economics and statistics*, 80 (2), 218-230.

Kim, H., Park, J., Lee, S. and Jang, S., 2012. Do expectations of future wealth increase outbound tourism? Evidence from Korea. *Tourism Management*, 33 (5), 1141-1147.

Kim, J. H., Song, H. and Wong, K. K., 2010. Bias-corrected bootstrap prediction intervals for autoregressive model: new alternatives with applications to tourism forecasting. *Journal of forecasting*, 29 (7), 655-672.

Kim, J. H., Wong, K., Athanasopoulos, G. and Liu, S., 2011. Beyond point forecasting: evaluation of alternative prediction intervals for tourist arrivals. *International Journal of Forecasting*, 27 (3), 887-901.

Kim, S. and Song, H., 1998. Analysis of inbound tourism demand in South Korea: a cointegration and error correction approach. *Tourism Analysis*, 3 (1), 25-41.

Kiviet, J. F., 1995. On bias, inconsistency, and efficiency of various estimators in dynamic panel data models. *Journal of econometrics*, 68 (1), 53-78.

Kling, J. L. and Bessler, D. A., 1985. A comparison of multivariate forecasting procedures for economic time series. *International Journal of Forecasting*, 1 (1), 5-24.

Koenker, R., 1981. A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*, 17 (1), 107-112.

Kožić, I., 2014. Detecting international tourism demand growth cycles. *Current Issues in Tourism*, 17 (5), 397-403.

Kulendran, N., 1996. Modelling quarterly tourist flows to Australia using cointegration analysis. *Tourism Economics*, 2 (3), 203-222.

Kulendran, N. and Dwyer, L., 2012. Modeling seasonal variation in tourism flows with climate variables. *Tourism Analysis*, 17 (2), 121-137.

Kulendran, N. and King, M. L., 1997. Forecasting international quarterly tourist flows using error-correction and time-series models. *International Journal of Forecasting*, 13 (3), 319-327.

Kulendran, N. and Witt, S. F., 2003. Forecasting the demand for international business tourism. *Journal of Travel Research*, 41 (3), 265-271.

Kulendran, N. and Wong, K. K. F., 2009. Predicting quarterly Hong Kong tourism demand growth rates, directional changes and turning points with composite leading indicators. *Tourism Economics*, 15 (2), 307-322.

Kulendran, N. and Wong, K. K. F., 2011. Determinants versus composite leading indicators in predicting turning points in growth cycle. *Journal of Travel Research*, 50 (4), 417-430.

Kuo Hsiao, I., Chang, C., Huang, B., Chen, C. and McAleer, M., 2009. Estimating the impact of avian flu on international tourism demand using panel data. *Tourism Economics*, 15 (3), 501-511.

Kuo Hsiao, I., Chen, C., Tseng, W., Ju, L. and Huang, B., 2008. Assessing impacts of SARS and Avian Flu on international tourism demand to Asia. *Tourism Management*, 29 (5), 917-928.

Kwack, S. Y., 1972. Effects of income and prices on travel spending abroad, 1960 III-1967 IV. *International Economic Review*, 245-256.

Lee, C.-K., Song, H.-J. and Mjelde, J. W., 2008. The forecasting of International Expo tourism using quantitative and qualitative techniques. *Tourism Management*, 29 (6), 1084-1098.

Lee, K., 2011a. Estimating demand elasticities for intra-regional tourist arrivals to Hong Kong - the 'bounds' testing approach. *Applied Economics Letters*, 18 (17), 1645-1654.

Lee, K., 2011b. Forecasting long-haul tourism demand for Hong Kong using error correction models. *Applied Economics*, 43 (4/6), 527-549.

Lemke, C., and Gabrys, B. 2010. Meta-learning for time series forecasting and forecast combination. *Neurocomputing*, 73(10-12), 2006-2016.

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

Li, G., Song, H. and Witt, S., 2002. Econometric analysis and forecasts of international tourism demand in Thailand.

LI, G., Song, H. and Witt, S. F., 2005. Recent Developments in Econometric Modeling and Forecasting. *Journal of Travel Research*, 44 (1), 82-99.

Li, H., Goh, C., Hung, K. and Chen, J. L., 2018a. Relative climate index and its effect on seasonal tourism demand. *Journal of Travel Research*, 57 (2), 178-192.

Li, H., Song, H. and Li, L., 2016. A Dynamic Panel Data Analysis of Climate and Tourism Demand: Additional Evidence. *Journal of Travel Research*, 1, 14.

Li, H. Q., Park, S. Y. and Seo, J. H., 2011. Quantile elasticity of international tourism demand for South Korea using the quantile autoregressive distributed lag model. *Tourism Economics*, 17 (5), 997-1015.

Li, S., Chen, T., Wang, L. and Ming, C., 2018b. Effective tourist volume forecasting supported by PCA and improved BPNN using Baidu index. *Tourism Management*, 68, 116-126.

Li, S. and Song, H., 2013. Economic impacts of visa restrictions on tourism: a case of two events in China. *Annals of Tourism Research*, 43, 251-271.

Li, X., Pan, B., Law, R. and Huang, X., 2017. Forecasting tourism demand with composite search index. *Tourism management*, 59, 57-66.

Lim, C., 1997. Review of international tourism demand models. *Annals of Tourism Research*, 24 (4), 835-849.

Lim, C., 1999. A meta-analytic review of international tourism demand. *Journal of Travel Research*, 37 (3), 273-284.

Lim, C., Chang, C. and McAleer, M., 2009a. Forecasting h (m) otel guest nights in New Zealand. *International journal of hospitality management*, 28 (2), 228-235.

Lim, C. and McAleer, M., 2001a. Cointegration analysis of quarterly tourism demand by Hong Kong and Singapore for Australia. *Applied Economics*, 33 (12), 1599-1619.

Lim, C. and McAleer, M., 2001b. Forecasting tourist arrivals. *Annals of Tourism Research*, 28 (4), 965-977.

Lim, C. and McAleer, M., 2002. Time series forecasts of international travel demand for Australia. *Tourism Management*, 23 (4), 389-396.

Lim, C., McAleer, M. and Min, J. C. H., 2009b. ARMAX modelling of international tourism demand. *Mathematics and Computers in Simulation*, 79, 2879-2888.

Lim, C., Min, J. C. and McAleer, M., 2008. Modelling income effects on long and short haul international travel from Japan. *Tourism Management*, 29 (6), 1099-1109.

Lim, C. and Wang, Y., 2008. China's post-1978 experience in outbound tourism. *Mathematics and Computers in Simulation*, 78 (2), 450-458.

Lin, S., Liu, A. and Song, H., 2015. Modeling and forecasting Chinese outbound tourism: an econometric approach. *Journal of Travel & Tourism Marketing*, 32 (1/2), 34-49.

Lin, T.-P. and Matzarakis, A., 2011. Tourism climate information based on human thermal perception in Taiwan and Eastern China. *Tourism Management*, 32 (3), 492-500.

Lin, V. S., Goodwin, P. and Song, H., 2014. Accuracy and bias of experts' adjusted forecasts. *Annals of Tourism Research*, 48, 156-174.

Lise, W. and Tol, R. S., 2002. Impact of climate on tourist demand. *Climatic change*, 55 (4), 429-449.

Liu, A. and McKercher, B., 2014. The impact of visa liberalization on tourist behaviors—The case of China outbound market visiting Hong Kong. *Journal of Travel Research*, 0047287514564599.

Liu, T.-M., 2014. Analysis of the economic impact of meteorological disasters on tourism: the case of typhoon Morakot's impact on the Maolin National Scenic Area in Taiwan. *Tourism Economics*, 20 (1), 143-156.

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

Loeb, P. D., 1982. International travel to the United States: an econometric evaluation. *Annals of Tourism Research*, 9 (1), 7-20.

Long, W., Liu, C. and Song, H., 2018. Pooling in tourism demand forecasting. *Journal of Travel Research*, 0047287518800390.

Lorde, T., Li, G. and Airey, D., 2015. Modeling Caribbean Tourism Demand an Augmented Gravity Approach. *Journal of Travel Research*, 0047287515592852.

Lorde, T. and Moore, W., 2008. Co-movement in tourist arrivals in the Caribbean. *Tourism Economics*, 14 (3), 631-643.

Ma, E., Liu, Y., Li, J. and Chen, S., 2016. Anticipating Chinese tourists arrivals in Australia: A time series analysis. *Tourism Management Perspectives*, 17, 50-58.

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

Maddison, D., 2001. In search of warmer climates? The impact of climate change on flows of British tourists. *Climatic change*, 49 (1-2), 193-208.

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

Makridakis, S., Wheelwright, S. C. and Hyndman, R. J. 2008. *Forecasting methods and applications*. John Wiley & Sons.

Mangion, M. L., Cooper, C., Cortés-Jimenez, I. and Durbarry, R., 2012. Measuring the effect of subsidization on tourism demand and destination competitiveness through the AIDS model: an evidence-based approach to tourism policymaking. *Tourism Economics*, 18 (6), 1251-1272.

Massidda, C. and Etzo, I., 2012. The determinants of Italian domestic tourism: a panel data analysis. *Tourism Management*, 33 (3), 603-610.

Massidda, C., Etzo, I. and Piras, R., 2015. Migration and inbound tourism: an Italian perspective. *Current Issues in Tourism*, 18 (12), 1152-1171.

Massidda, C. and Mattana, P., 2013. A SVECM analysis of the relationship between international tourism arrivals, GDP and trade in Italy. *Journal of Travel Research*, 52 (1), 93-105.

Massidda, C. and Piras, R., 2015. Does internal migration affect Italian domestic tourism? A panel data analysis. *Tourism Economics*, 21 (3), 581-600.

Matzarakis, A., Mayer, H. and Iziomon, M. G., 1999. Applications of a universal thermal index: physiological equivalent temperature. *International Journal of Biometeorology*, 43 (2), 76-84.

McKercher, B., Chan, A. and Lam, C., 2008. The Impact of Distance on International Tourist Movements. *Journal of Travel Research*, 47 (2), 208-224.

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.

Moore, W. R., 2010. The impact of climate change on Caribbean tourism demand. *Current Issues in Tourism*, 13 (5), 495-505.

Morabito, M., Cecchi, L., Modesti, P. A., Crisci, A., Orlandini, S., Maracchi, G. and Gensini, G. F., 2004. The impact of hot weather conditions on tourism in Florence, Italy: the summer 2002–2003 experience. *Advances in tourism climatology*, 12, 158-165.

Moreno, A. and Amelung, B., 2009. Climate change and tourist comfort on Europe's beaches in summer: a reassessment. *Coastal Management*, 37 (6), 550-568.

Morgan, R., Gatell, E., Junyent, R., Micallef, A., Özhan, E. and Williams, A., 2000. An improved user-based beach climate index. *Journal of Coastal Conservation*, 6 (1), 41-50.

Morley, C., Rosselló, J. and Santana-Gallego, M., 2014. Gravity models for tourism demand: theory and use. *Annals of Tourism Research*, 48, 1-10.

Morley, C. L., 2009. Dynamics in the specification of tourism demand models. *Tourism Economics*, 15 (1), 23-39.

Newbold, P. and Granger, C. W., 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.

Nelson, L. A., Dickey, D. A. and Smith, J. M., 2011. Estimating time series and cross section tourism demand models: mainland United States to Hawaii data. *Tourism Management*, 32 (1), 28-38.

Nowman, K. B. and Dellen, S. v., 2012. Forecasting overseas visitors to the UK using continuous time and autoregressive fractional integrated moving average models with discrete data. *Tourism Economics*, 18 (4), 835-844.

Onafowora, O. and Owoye, O., 2012. Modelling international tourism demand for the Caribbean. *Tourism Economics*, 18 (1), 159-180.

Ongan, S. and Gozgor, G., 2018. Tourism demand analysis: The impact of the economic policy uncertainty on the arrival of Japanese tourists to the USA. *International Journal of Tourism Research*, 20 (3), 308-316.

Otero-Giráldez, M. S., Álvarez-Díaz, M. and González-Gómez, M., 2012. Estimating the long-run effects of socioeconomic and meteorological factors on the domestic tourism demand for Galicia (Spain). *Tourism Management*, 33 (6), 1301.

Ouerfelli, C., 2008. Co-integration analysis of quarterly European tourism demand in Tunisia. *Tourism Management*, 29 (1), 127-137.

Page, S., Song, H. and Wu, C., 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.

Pan, B. and Yang, Y., 2017. Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research*, 56 (7), 957-970.

Park, S., Lee, J. and Song, W., 2017. Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data. *Journal of Travel & Tourism Marketing*, 34 (3), 357-368.

Peng, B., Song, H., Crouch, G. I. and Witt, S. F. 2015. A meta-analysis of international tourism demand elasticities. *Journal of Travel Research*, 54(5), 611-633.

Pesaran, M. H. and Timmermann, A., 1992. A simple nonparametric test of predictive performance. *Journal of Business & Economic Statistics*, 10 (4), 461-465.

Pesaran, M. H., Schuermann, T. and Weiner, S. M., 2004. Modeling regional interdependencies using a global error-correcting macroeconometric model. *Journal of Business & Economic Statistics*, 22 (2), 129-162.

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

Pesaran, M. H., Shin, Y. and Smith, R. P., 1999. Pooled Mean Group Estimation of Dynamic Heterogeneous Panels*,* 621.

Pham, T. D., Nghiem, S. and Dwyer, L., 2017. The determinants of Chinese visitors to Australia: A dynamic demand analysis. *Tourism Management*, 63, 268-276.

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

Pintassilgo, P., Rosselló, J., Santana-Gallego, M. and Valle, E., 2016. The economic dimension of climate change impacts on tourism the case of Portugal. *Tourism Economics*, 22 (4), 685-698.

Provenzano, D., 2015. A dynamic analysis of tourism determinants in Sicily. *Tourism Economics*, 21 (3), 441-454.

Ramos, F. F. R., 2003. Forecasts of market shares from VAR and BVAR models: a comparison of their accuracy. *International Journal of Forecasting*, 19 (1), 95-110.

Ramsey, J. B., 1969. Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society. Series B (Methodological)*, 350-371.

Ridderstaat, J., Oduber, M., Croes, R., Nijkamp, P. and Martens, P., 2014. Impacts of seasonal patterns of climate on recurrent fluctuations in tourism demand: evidence from Aruba. *Tourism Management*, 41, 245-256.

Rodríguez, X. A., Martínez-Roget, F. and Pawlowska, E., 2012. Academic tourism demand in Galicia, Spain. *Tourism Management*, 33 (6), 1583-1590.

Rosselló-Nadal, J., Riera-Font, A. and Cárdenas, V., 2011. The impact of weather variability on British outbound flows. *Climatic Change*, 105 (1-2), 281-292.

Saayman, A. and Botha, I., 2017. Non-linear models for tourism demand forecasting. *Tourism Economics*, 23 (3), 594-613.

Saayman, A. and Saayman, M., 2008. Determinants of inbound tourism to South Africa. *Tourism Economics*, 14 (1), 81-96.

Saayman, A. and Saayman, M., 2015. An ARDL bounds test approach to modelling tourist expenditure in South Africa. *Tourism Economics*, 21 (1), 46-66.

Saha, S. and Yap, G., 2014. The moderation effects of political instability and terrorism on tourism development: A cross-country panel analysis. *Journal of Travel Research*, 53 (4), 509-521.

Santana-Jiménez, Y. and Hernández, J. M., 2011. Estimating the effect of overcrowding on tourist attraction: The case of Canary Islands. *Tourism Management*, 32, 415-425.

Santos, G. E. d. O., 2009. Forecasting tourism demand by disaggregated time series - empirical evidence from Spain. *Tourism Economics*, 15 (2), 467-472.

Schiff, A. and Becken, S., 2011. Demand elasticity estimates for New Zealand tourism. *Tourism Management*, 32 (3), 564-575.

Seetanah, B., Durbarry, R. and Ragodoo, J. F. N., 2010. Using the panel cointegration approach to analyse the determinants of tourism demand in South Africa. *Tourism economics*, 16 (3), 715-729.

Seetaram, N., 2010. Use of dynamic panel cointegration approach to model international arrivals to Australia. *Journal of Travel Research*, 49 (4), 414-422.

Seetaram, N., 2012a. Estimating demand elasticities for Australia's international outbound tourism. *Tourism Economics*, 18 (5), 999-1017.

Seetaram, N., 2012b. Immigration and international inbound tourism: empirical evidence from Australia. *Tourism Management*, 33 (6), 1535-1543.

Seetaram, N., Forsyth, P. and Dwyer, L., 2016. Measuring price elasticities of demand for outbound tourism using competitiveness indices. *Annals of Tourism Research*, 56, 65-79.

Seetaram, N., Song, H. and Page, S. J., 2014. Air passenger duty and outbound tourism demand from the United Kingdom. *Journal of Travel Research*, 53 (4), 476-487.

Seo, J. H., Park, S. and Yu, L., 2009. The analysis of the relationships of Korean outbound tourism demand: Jeju Island and three international destinations. *Tourism Management*, 30 (4), 530-543.

Seo, J. H., Park, S. Y. and Boo, S., 2010. Interrelationships among Korean outbound tourism demand: Granger causality analysis. *Tourism Economics*, 16 (3), 597-610.

Sessions, D. N. 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., 2009. Effect of seasonality treatment on the forecasting performance of tourism demand models. *Tourism Economics*, 15 (4), 693-708.

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

Silva, E. S., Ghodsi, Z., Ghodsi, M., Heravi, S. and Hassani, H., 2017. Cross country relations in European tourist arrivals. *Annals of Tourism Research*, 63, 151-168.

Sinclair, M. T., Blake, A. and Sugiyarto, G., 2003. The economics of tourism. *Classic reviews in tourism*, 22, 54.

Smeral, E., 2009. The Impact of the Financial and Economic Crisis on European Tourism. *JOURNAL OF TRAVEL RESEARCH*, 48 (1), 3-13.

Smeral, E., 2010. Impacts of the world recession and economic crisis on tourism: forecasts and potential risks. *Journal of Travel Research*, 49 (1), 31-38.

Smeral, E., 2012. International tourism demand and the business cycle. *Annals of Tourism Research*, 39 (1), 379-400.

Smeral, E., 2014. Forecasting international tourism with due regard to asymmetric income effects. *Tourism Economics*, 20 (1), 61-72.

Song, H., Gao, B. Z. and Lin, V. S., 2013. Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system. *International Journal of Forecasting*, 29, 295-310.

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

Song, H., Kim, J. H. and Yang, S., 2010a. Confidence intervals for tourism demand elasticity. *Annals of Tourism Research*, 37 (2), 377-396.

Song, H. and Li, G., 2008. Tourism demand modelling and forecasting--a review of recent research. *Tourism Management*, 29 (2), 203-220.

Song, H., Li, G., Witt, S. F. and Athanasopoulos, G., 2011a. Forecasting tourist arrivals using time-varying parameter structural time series models. *International Journal of Forecasting*, 27, 855-869.

Song, H. Y., Lin, S. S., Witt, S. F. and Zhang, X. Y., 2011b. Impact of financial/economic crisis on demand for hotel rooms in Hong Kong. *Tourism Management*, 32 (1), 172-186.

Song, H., Li, G., Witt, S. F. and Fei, B., 2010b. Tourism demand modelling and forecasting: how should demand be measured? *Tourism Economics*, 16 (1), 63-81.

Song, H. and Lin, S., 2010. Impacts of the financial and economic crisis on tourism in Asia. *Journal of Travel Research*, 49 (1), 16-30.

Song, H. and Witt, S. F., 2000. *Tourism demand modelling and forecasting: Modern econometric approaches*. Routledge.

Song, H. and Witt, S. F., 2003. Tourism forecasting: the general-to-specific approach. *Journal of Travel Research*, 42 (1), 65-74.

Song, H. and Witt, S. F., 2006. Forecasting international tourist flows to Macau. *Tourism management*, 27 (2), 214-224.

Song, H., Witt, S. F. and Jensen, T. C., 2003a. Tourism forecasting: accuracy of alternative econometric models. *International Journal of Forecasting*, 19(1), 123-141.

Song, H., Witt, S. F. and Li, G., 2003b. Modelling and forecasting the demand for Thai tourism. *Tourism Economics*, 9 (4), 363-387.

Song, H., Witt, S. F. and Li, G., 2009. *The advanced econometrics of tourism demand* [online]. New York: Routledge.

Song, H., Witt, S. F. and Zhang, X. 2008. Developing a Web-based tourism demand forecasting system. Tourism Economics, 14(3), 445-468.

Song, H. and Wong, K. K., 2003. Tourism demand modeling: A time-varying parameter approach. *Journal of Travel Research*, 42 (1), 57-64.

Song, H. Y., Witt, S. F., Wong, K. F. and Wu, D. C., 2009. An empirical study of forecast combination in tourism. *Journal of Hospitality & Tourism research*, 33 (1), 3-29.

Song, H. Y., Witt, S. F. and Zhang, X. Y., 2008. Developing a Web-based tourism demand forecasting system. *Tourism Economics*, 14 (3), 445-468.

Stepchenkova, S. and Eales, J. S., 2011. Destination image as quantified media messages: the effect of news on tourism demand. *Journal of Travel Research*, 50 (2), 198-212.

Stock, J. and Watson, M., 2003. How did leading indicator forecasts perform during the 2001 recession?

Stock, J. H. and Watson, M. W., 1998. *A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series*. National Bureau of Economic Research.

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

Su, Y. and Lin, H., 2014. Analysis of international tourist arrivals worldwide: the role of world heritage sites. *Tourism Management*, 40, 46-58.

Sun, X., Sun, W., Wang, J., Zhang, Y. and Gao, Y., 2016. Using a Grey–Markov model optimized by Cuckoo search algorithm to forecast the annual foreign tourist arrivals to China. *Tourism Management*, 52, 369.

Teräsvirta, T., Van Dijk, D. and Medeiros, M. C., 2005. Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination. *International Journal of Forecasting*, 21 (4), 755-774.

Tol, R. S. and Walsh, S., 2012. The impact of climate on tourist destination choice. *Economic and Social Research Institute (ERSI) Working paper*, (423).

Tsui, W. H. K. and Balli, F., 2017. International arrivals forecasting for Australian airports and the impact of tourism marketing expenditure. *Tourism Economics*, 23 (2), 403-428.

Tsui, W. H. K. and Fung, M. K. Y., 2016. Causality between business travel and trade volumes: Empirical evidence from Hong Kong. *Tourism Management*, 52, 395-404.

Tsui, W. H. K., Gilbey, A., Gow, H. and Ozer Balli, H., 2014. Forecasting of Hong Kong airport's passenger throughput. *Tourism Management*, 42, 62-76.

Turner, L., Kulendran, N. and Fernando, H., 1997. The use of composite national indicators for tourism forecasting. *Tourism Economics*, 3 (4), 309-317.

Tveteras, S., 2014. Non-stop flights and tourist arrivals. *Tourism Economics*, 20 (1), 5-20.

Untong, A., Ramos, V., Kaosa-Ard, M. and Rey-Maquieira, J., 2014. Thailand's long-run tourism demand elasticities. *Tourism Economics*, 20 (3), 595-610.

Untong, A., Ramos, V., Kaosa-Ard, M. and Rey-Maquieira, J., 2015. Tourism demand analysis of Chinese arrivals in Thailand. *Tourism Economics*, 21 (6), 1221-1234.

Valadkhani, A. and O'Mahony, B., 2015. Dynamics of Australia's tourism in a multimarket context. *Annals of Tourism Research*, 55, 173-177.

Vanegas, M., Sr., 2013. Co-integration and error correction estimation to forecast tourism in El Salvador. *Journal of Travel & Tourism Marketing*, 30 (6), 523-537.

Vergori, A. S., 2012. Forecasting tourism demand: the role of seasonality. *Tourism Economics*, 18 (5), 915-930.

Vergori, A. S., 2017. Patterns of seasonality and tourism demand forecasting. *Tourism Economics*, 23 (5), 1011-1027.

Wan, S., Wang, S. and Woo, C., 2013. Aggregate vs. disaggregate forecast: case of Hong Kong. *Annals of Tourism Research*, 42, 434-438.

Wan, S. K. 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.

White, R., 2007. Immigrant-trade links, transplanted home bias and network effects. *Applied Economics*, 39 (7), 839-852.

Wickens, M. R. and Breusch, T. S., 1988. *Dynamic specification, the long run and the estimation of transformed regression models*. *Economic Journal*, 98 (390), 189-205.

Witt, S. F., 1980a. An abstract mode–abstract (destination) node model of foreign holiday demand. *Applied Economics*, 12 (2), 163-180.

Witt, S. F., 1980b. An econometric comparison of UK and German foreign holiday behavior. *Managerial and Decision Economics*, 1 (3), 123-131.

Witt, S. F. and Witt, C. A., 1995. Forecasting tourism demand: A review of empirical research. *International journal of Forecasting*, 11 (3), 447-475.

Wong, K. K., Song, H. and Chon, K. S., 2006. Bayesian models for tourism demand forecasting. *Tourism Management*, 27 (5), 773-780.

Wong, K. K. F., Song, H., Witt, S. F. and Wu, D. C., 2007. Tourism forecasting: To combine or not to combine? *Tourism Management*, 28 (4), 1068-1078.

Wu, C., Li, G. and Song, H., 2011. Analyzing tourist consumption: a dynamic system-of-equations approach. *Journal of Travel Research*, 50 (1), 46-56.

Wu, D. C., Li, G. and Song, H., 2012. Economic analysis of tourism consumption dynamics. A Time-varying Parameter Demand System Approach. *Annals of Tourism Research*, 39, 667-685.

Wu, D. C., Song, H. and Shen, S., 2017. New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management*, 29 (1), 507-529.

Yan, Y., Zhang, H. and Ye, H., 2014. Assessing the impacts of the high-speed train on tourism demand in China. *Tourism Economics*, 20 (1), 157-169.

Yang, C., Lin, H. and Han, C., 2010. Analysis of international tourist arrivals in China: the role of World Heritage Sites. *Tourism Management*, 31 (6), 827-837.

Yang, X., Pan, B., Evans, J. A. and Lv, B., 2015. Forecasting Chinese tourist volume with search engine data. *Tourism Management*, 46, 386-397.

Yang, Y., Liu, Z. and Qi, Q. Y., 2014a. Domestic tourism demand of urban and rural residents in China: does relative income matter? *Tourism Management*, 40, 193-202.

Yang, Y., Pan, B. and Song, H. Y., 2014b. Predicting Hotel Demand Using Destination Marketing Organization's Web Traffic Data. *Journal of Travel Research*, 53 (4), 433-447.

Yap, G. and Allen, D., 2011. Investigating other leading indicators influencing Australian domestic tourism demand. *Mathematics and Computers in Simulation*, 81, 1365-1374.

Yorucu, V. and Mehmet, O., 2011. The bounds-test approach for co-integration between international tourist arrivals, per capita income and cost of living: the case of All Cyprus. *Applied Economics Letters*, 18 (13/15), 1327-1331.

Zhang, C., Singh, A. and Yu, L., 2013. Does it Matter? Examining the Impact of China's Vacation Policies on Domestic Tourism Demand. *Journal of Hospitality & Tourism Research*, 1096348013503993.

Zhang, H. Q. and Kulendran, N., 2016. The Impact of Climate Variables on Seasonal Variation in Hong Kong Inbound Tourism Demand. *Journal of Travel Research*, 0047287515619692.

Zhou-Grundy, Y. and Turner, L. W., 2014. The Challenge of Regional Tourism Demand Forecasting: The Case of China. *Journal of Travel Research*, 53 (6), 747-759.

Zhu, L., Lim, C., Xie, W. and Wu, Y., 2018. Modelling tourist flow association for tourism demand forecasting. *Current Issues in Tourism*, 21 (8), 902-916.

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.

Önder, I., 2017. Forecasting tourism demand with Google trends: Accuracy comparison of countries versus cities. *International Journal of Tourism Research*, 19 (6), 648-660.

Appendix 1

Table A-1 summarizes quantitative tourism demand modelling and forecasting studies published from 2008 to 2018. It presents the study theme, the region/regions under consideration, the data type and frequency and the methodology applied. The data types include pure time series data and panel data, and weekly (W), monthly (M), quarterly (Q) and yearly data are all covered in these studies. Different models employed in each study are shown under the methodology column.

Table A-1 Summary of Tourism Demand Modeling and Forecasting Publications from 2008 to 2018

	Theme	Region[5]	Data	Methodology
Assaf et al. (2018)	Interdependence of tourism demand in a region	Nine countries in Southeast Asia	1985Q1-2014Q2	BGVAR
Buigut (2018)	The effect of terrorism on tourism demand	Kenya (I) from 27 developed and 34 emerging countries	Panel: 2010Q1-2013Q4	Dynamic panel
Camacho and Pacce (2018)	Forecasting tourist arrivals using Google search volume indices	Spain (I)	2007M7-2016M1	Dynamic factor model
Dergiades et al. (2018)	Forecasting tourism demand using web-based search intensity indices	Cyprus (I)	2004M1-2015M12	VAR
Li et al. (2018a)	Relative climate index and seasonal tourism demand	Hong Kong (O) to 13 mainland China cities	Panel: 2006Q1-2011Q4	Partial adjustment panel
Li et al. (2018b)	Tourism demand forecasting based on Baidu index	Beijing and Hainan (I)	2011M1-2016M12	PCA-ADE-BPNN, PCA-VAR, PCA-BPNN, VAR, ARIMA
Long et al. (2018)	Pooling in tourism demand forecasting	342 cities in China (D)	Panel: 2005-2013	Pooled OLS, OLS and Naïve 2
Ongan and Gozgor (2018)	The impact of economic policy uncertainty on tourism demand	US (I) from Japan	1996Q1-2015Q1	Cointegration analysis
Wan and Song (2018)	Forecasting turning points in tourism growth	Hong Kong (I)	1996Q1-2017Q1	Logistic models, combination forecasting,
Zhu et al. (2018)	Forecasting tourism demand taking dependence among different origins into account	Singapore (I)	1995Q1-2013Q4	Copula based approach: copula-ADLM, copula-ECM
Apergis et al. (2017)	Compare the forecasting performance of four univariate seasonal time series models	20 Croatian counties	1998M1-2014M7	SARIMA, SARIMA with Fourier transformation, ARAR, SARFIMA
Athanasopoulos et al. (2017)	Bootstrap aggregation in tourism demand modelling and forecasting	Australia (I)	1981Q1-2012Q3	ADLM and bootstrap forecasting
Cao et al. (2017)	The interdependence of tourism demand in a global level	24 major countries across the world	1994Q1-2011Q4	GVAR
Croes et al. (2017)	Business and tourism demand cycle	Aruba (I) and Barbados (I)	1970-2015	2SLS
Dogru et al. (2017)	Remodelling of international tourism demand	Turkey (I) from nine countries	Panel: 2003Q2-2012Q4; 2003M2-2012M12	FMOLS
Folgieri et al. (2017)	Comparing forecasting accuracy	Croatia (I & D)	2007M1-2012M12	Machine learning model, linear regression

5 I is for inbound tourism demand, O stands for outbound tourism demand and D is for domestic tourism demand.

	of AI and econometric model			
Hassani et al. (2017)	Parametric and nonparametric techniques for forecasting tourism demand	Several European countries (I)	2000M1-2013M12	MA, WMA, ARIMA, ETS, NN, TBATS, ARFIMA, SSA
Li et al. (2017)	Forecasting tourism demand with composite search index	Beijing (D)	2011M1-2015M7 (converted to weekly) Baidu: weekly	Index created in GDFM ADLM
Pan and Yang (2017)	Forecasting weekly hotel demand with big data	Charleston, US (I)	2006W1-2015W30	ARIMAX, MSDR model
Park et al. (2017)	Forecasts of tourism demand to South Korea from Japan using google trends data	South Korea (I)	2004M1-2015M10	ARIMA, ARIMAX, ETS
Pham et al. (2017)	The determinants of tourism demand from China to Australia	Australia (I)	1991-2014	Partial adjustment model
Saayman and Botha (2017)	Compare the forecasting ability of linear and non-linear methods	South Africa (I)	2000M1-2012M12	Season naïve, SARIMA, BSM, STAR, SSA
Silva et al. (2017)	Forecasting tourism demand using MSSA and the cross country relations in tourism demand	10 European countries (I)	2000M1-2013M12	SARIMA, ETS, SSA and MSSA
Tsui and Balli (2017)	Forecasting airport arrivals using time series technique	Australia (I)	2006M1-2012M9	SARIMA, SARIMAX, GARCH, EGARCH
Vergori (2017)	Patterns of seasonality and tourism demand forecasting	Austria, Finland, Portugal and Netherlands (I)	1990M1-2014M12	SARIMA
Önder (2017)	Comparing tourism demand forecasting accuracy of cities and countries using google trends	Vienna, Barcelona, Austria and Belgium (I)	2008M1-2014M12	ADLM, ETS, naïve 1
Albaladejo et al. (2016)	Non-constant reputation effect	Spain (I, D)	Panel: 2000-2013, 17 autonomous communities	Nonlinear simple dynamic panel estimated
Balli et al. (2016)	The impacts of immigrants and institutions	OECD (O) to middle- to low- income countries	Panel: 1995-2010, 34 OECD origins to 52 middle- to low- income countries	Dynamic panel
Claveria et al. (2016)	Disaggregate vs. aggregate forecasts with machine learning models	Spain (I)	1999:M1-2014:M3	machine learning models: SVR, GPR, RBF, MLP; ARMA
Chatziantoniou et al. (2016)	Forecasting tourist arrivals using origin country macroeconomics	Greece(I)	2003:M1-2013:M6	SARIMA, SARIMAX
Ghaderi et al. (2016)	Security and tourism demand	74 countries (I)	Panel: 2006-2012; 29 developed and 45 developing countries	Simple dynamic panel
Habibi (2016)	Determinants of tourism demand to Malaysia	Malaysia (I)	Panel: 2000-2012; from 33 origins	Simple dynamic panel
Li et al. (2016)	The impact of climate on seasonal tourism demand	China (I) from HK	Panel: 2006Q1-2011Q4, 19 major cities in China	Simple dynamic panel
Ma et al. (2016)	Forecasting tourists arrivals to Australia	China (O)	1991:M1-2015:M9	SARIMA
Pintassilgo et al. (2016)	The economic dimension of climate change impacts on tourism	Portugal(I)	Panel: 1995-2010, 178 origins	World gravity model Input-output model
Seetaram et al. (2016)	Present a price competitiveness index	Australia (O)	Panel: 1990-2008, Australians to 45	Dynamic panel cointegration (DOLS)

	(PCI) as proxy for price in tourism demand models		destinations	
Sun et al. (2016)	Presented a new model: CMCSGM	China (I) from 8 major markets including Korea, Japan, Russia, US, Malaysia and Philippines	1997-2012	GM, MC: CMCSGM, MCSGM, SGM, MCGM and GM
Tsui and Fung (2016)	Causality between business travel and trade volumes	HK(I) from mainland China, Taiwan and US	2002Q1-2012Q4	EG-ECM, Granger causality test
Yazdi and Khanalizadeh (2016)	Determinants of tourism demand	US (I)	Panel: 1995-2014; 14 origins	Panel ADLM
Zhang and Kulendran (2016)	Impact of climate variables on seasonal variation in tourism demand	HK(I) from mainland China, Taiwan, South Korea and Japan	1997Q1-2012Q4	ADLM
Akin (2015)	Present a novel approach to model selection	Turkey (I) from top 10 source markets	2001:M1-2011:M12	SARIMA, SVR, MLP
Balli and Tsui (2016)	Tourism demand spillovers between Australia and New Zealand	Australia and New Zealand (I) from 7 origin	2000:M1-2012:M12	14 bivariate GARCH models
Bangwayo-Skeete and Skeete (2015)	Use Google Trends' search query data-based Mixed-Data Sampling to forecast tourism demand	5 Caribbean countries: Jamaica, Bahama, Dominican Republic, St. Lucia and Cayman Islands(I) from UK, US and Canada	2004:M1-2012:M12, weekly data for search engine data	AR-MIDAS (Mixed-data Sampling) vs. SARIMA, AR
Claveria et al. (2015)	Compare tourism demand forecasting using 3 types of ANN	Catalonia (I)	2001:M1-2012:M7	MLP, RBF, and Elman network
Chen and Haynes (2015)	Impact of high-speed rail on international tourism demand in China	China (I)	Panel: 1997-2012, 21 countries as origins	Dynamic panel
Falk (2015)	The sensitivity of tourism demand to exchange rate changes	Austria (I) from Switzerland	Panel: winter season from 2006-07 to 2011-12, 63 Australian ski resorts,	Static panel and panel ECM estimated by PMG, long-run parameters estimated by ML
Guizzardi and Stacchini (2015)	Include business sentiment indicators (BSI) from business surveys in models forecasting demand	Rimini, Italy (I)	1987-2012 Four-monthly data (Jan-Apr, May to Aug, and Sep-Dec)	Basic and augmented naïve models, BSM, CSM
Gunter and Önder (2015)	Compare time series and causal models forecasting monthly tourist arrivals to Paris	Paris(I) from US, Germany, Italy, UK and Japan	2003:M1-2012:M12	Naïve 1, CI (EC-ADLM), ETS, Classical VAR, Bayesian VAR, ARMA, TVP
Hassani et al. (2015)	Using SSA to forecast tourism demand	USA(I)	1996:M1-2012:M11	SSA, ARIMA, ETS, NN (1 hidden layer feed forward NN)
Johnson and Garman (2015)	Medical travel demand	USA (I) from all countries except Mexico and Canada	Cross section data covering all countries except Mexico and Canada, collected during a two-week frame in April 2012	Linear regression with cross-sectional data
Lin et al. (2015)	Forecasting Chinese outbound tourism demand	China (O) total and 11 selected destinations	1985-2011	CI: ADLM bounds test
Lorde et al. (2015)	Caribbean tourism demand modeling using augmented gravity approach	18 Caribbean countries (I) from US, Europe, Canada and the Caribbean	Panel: 1980-2008, 18 Caribbean destinations	Simple dynamic panel Model
Massidda and Piras (2015)	Migration and Italian domestic tourism demand	Italy (D)	Panel: 20 Italian regions over 1987-2010	Panel ECM estimated by three techniques: DFE, PMG and MG
Massidda et al. (2015)	Migration and Italian inbound tourism demand	Italy (I)	Panel: 2005-2011, 65 countries as origins	Dynamic panel estimated by system GMM

Provenzano (2015)	Tourism determinants in Sicily	Sicily, Italy (I & D)	1999-2007	Gravity model
Saayman and Saayman (2015)	ADLM bounds test approach to model tourist expenditure in South Africa	South Africa (I)	2003Q1-2010Q4	CI: ADLM bounds test
Untong et al. (2015)	Tourist arrivals from China to Thailand	China (O) to Thailand	1988:M1-2013:M13 and 1988-2013	Seasonality analysis: X12-ARIMA Long-run relationship: SR using bootstrapping approach; Forecasting: ETS, ARIMA, GM
Valadkhani and O'Mahony (2015)	Dynamics of Australia's tourism in a multimarket context	Australia (I) from NZ, Japan, UK, US and China	1991:M1-2014:M9	Five variable VAR
Yang et al. (2015)	Improve forecasting accuracy by using search engine data	Hainan, China (I)	2006:M6-2010:M12; Baidu: 2006:M6-2013:M9	ARIMAX vs. ARMA
Athanasopoulos et al. (2014)	Substitution between domestic and outbound tourism in Australia	Australia (D, O)	2000Q1-2010Q3	AIDS, EC-AIDS (relative ratios of purchasing power parity index as price variable)
Bento (2014)	International academic tourism demand in Europe	31 countries in Europe (I)	Panel: 2000-2010, 31 sampling countries	Dynamic panel data model estimated by GMM
Cang (2014)	Compare linear and nonlinear combination forecasts	UK(I)	1993:Q1-2007:Q4	Naïve1, Naïve2, ETS, SARIMA, SVR, SA, VACO, DMSFE, MLP, RBF
Cazanova et.al. (2014)	Habit persistence in air travel to Florida and the short- and long-run adjustments in air traffic	Florida (I) from US origins	1995:M1-2006:M12	DAP-PAM (domestic air passenger traffic partial adjustment model), SA, and SUR-AR1
Chou et al. (2014)	Crowding-out effect of Taiwan's opening up to Chinese tourists policy	Taiwan (I) from Japan, Malaysia, USA, Canada, UK, Korea, Singapore, Australia and Philippines	2001:Q1-2011:Q3	Linear regression (estimation method: Monte Carlo simulation)
Claveria and Torra (2014)	Forecasting tourism demand to Catalonia	Catalonia (I)	2001:M1-2008:M1	ARIMA, SETAR, MLP
Dwyer et al. (2014)	Linkage between migration and tourism	Australia (I) from 29 countries	Cross-sectional: 1991 census data, 2006 census data	Linear regression with cross-sectional data
Eugenio-Martin and Campos-Soria (2014)	Economic Crisis and tourism expenditure cutback decision	EU-27 (O)	Cross-sectional: Micro-data from a survey conducted in 2009 to EU-27 households; macro-data at the origin region	Simultaneous semi-ordered bivariate probit model
Falk (2014)	Climate's impact on tourism demand in Austria in the peak summer season from 1960-2012	Austria (I) from 12 visitor countries and (D)	Domestic: 1960-2012; Inbound:1967-2012 (only for July and August)	SR, GRM, ADLM-ECM (estimated by PMG)
Fuleky et al. (2014)	Demand elasticities in non-stationary panels	Hawaii (I) from US Mainland	Panel: 1993:Q1-2012:Q1, 49 states in US as origins	Panel data model Use CCE-MG estimator to deal with cross-sectional dependence panel regression
Gatt and Falzon (2014)	Britain tourism demand elasticities in Mediterranean countries	UK (O) to Mediterranean countries	1963-2009	AIDS (use recursive estimates to study the stability of estimated elasticities)
Gil-Alana et al. (2014)	Persistence, long memory and seasonality in Kenyan tourism series	Kenya (I & O)	1975:Q1-2011:Q4	SARFIMA
Kožić (2014)	Spectrum analysis of tourism demand growth cycles	International	1960-2010	SA

Lin et al. (2014)	Accuracy and bias of experts' adjusted forecasts	HK (I) from 6 origin	1985:Q1-2010:Q4	integration of statistical and judgmental forecasting ADLM-ECM, Naïve 1
Liu (2014)	The impact of meteorological disasters on tourism demand	Taiwan Maolin National Scenic Area (I)	2001:M1-2010:M12	WB-ECM
Liu and McKercher (2014)	The impact of visa liberalization on tourist behaviors	China (O) to HK	1998-2003, 2004-2012	Nonparametric data analysis
Ridderstaat et al. (2014)	Seasonal patterns of climate's impact on recurrent fluctuations in Aruba's tourism demand (climate variables as pull and push factors)	Aruba (I) from US, and Venezuela	Panel: 1986M1-2011M12, US and Venezuela as origins	FE and RE static panel data models
Saha and Yap (2014)	Impacts of political instability and terrorism on tourism development	139 countries (I)	Panel: 1999-2009 139 destinations	Fixed and random effect panel data model estimated by OLS
Seetaram et al. 2014	The impact of air passenger duty on outbound tourism demand of UK	UK (O)	1994:Q4-2010:Q4	CI: ADLM bounds test
Smeral (2014)	asymmetric income effect on demand across business cycles	USA, Canada, Japan, Australia and EU-15 (O)	1980-2010	Linear regression (dependent variable: real tourism imports, RTI)
Su and Lin (2014)	World heritage sites' (WHSs) impact on tourism demand	66 countries (I)	Panel: 2006-2009, 66 countries as destinations	Static panel data models estimated by pooled OLS
Tsui et al. (2014)	Forecast HK airport's passenger throughput	HK (I) from 11 principal origins	SARIMA: 1993:M1-2011:M8; ARIMAX: 2001:M1-2010:M11	SARIMA, ARIMAX
Tveteras (2014)	Non-stop flights and demand	Peru (I)	Panel: 2004-2009, 75 origins	Dynamic panel data model
Untong et al. (2014)	Long-run tourism demand elasticities	Thailand (I) from 11 origin countries	1985-2009, 1985:M1-2009:M12	TVP-LRM estimated by DOLS
Yan et al. (2014)	High-speed train's impact on demand	China (D): Hubei, Hunan and Guangdong (I)	2008:M1-2011:M12	ARMAX
Yang et al. (2014a)	The effect of relative income on tourism demand	China (D)	Panel: Urban: 1996-2007, 35 major cities; Rural: 2000-2007, 30 provinces	Panel multilevel model
Yang et al. (2014b)	Predicting hotel demand using web traffic volume data	Charleston, South Carolina, US (I)	1st week of 2006-16th week of 2011	ARMAX VS. ARMA
Zhou-Grundy and Turner (2014)	Intra country regional forecasting in China	31 regional provinces in China (I)	1994-2007	BSM, TVP, Naïve
Chatziantoniou et al. (2013)	Oil price shock's impact on tourism income and economic growth	4 oil-importing European Mediterranean countries: France, Italy, Spain and Greece (I)	2000:M1-2012:M12	SVAR (structural VAR)
Falk (2013)	Determinants of domestic and inbound tourism demand of Austrian ski resorts	Austria (D, I)	Panel: Winter season: 1986-87 to 2007-08, 28 Austrian ski resorts	Dynamic panel
Granvorka and Strobl (2013)	Hurricane strikes' impact on demand	26 countries/territories in the Caribbean (I)	Panel: 2003-2008, 26 countries/territories In the Caribbean	Static panel data model
Li and Song (2013)	Economic impacts of visa restrictions on tourism demand	China (I) from 10 origin	1989 Tian'an Men Square Incident 2008 Beijing Olympic Games	CGE (incorporating the demand elasticities obtained from econometric models to improve the reliability of CGE model)
Massidda and	Relationship between	Italy (I)	1987Q1-2009Q4	

Mattana (2013)	international tourism arrivals, GDP and trade in Italy			SECM
Song et al. (2013)	Combing statistical and judgmental forecasts via a web-based tourism demand forecasting system	HK (I) including tourist arrivals, total and sectional tourist expenditures, and demand for hotel rooms From both long- and short-haul markets	1985Q1-2009Q4 1985-2009	integration of statistical and judgmental forecasts ADLM; Judgmental: Delphi method (two round surveys conducted to 21 PGRs and 5 staffs)
Vanegas, Sr. (2013)	Modeling tourism demand in El Salvador	El Salvador (I) from 5 origins	1987-2010	CI: ADLM, JML-ECM
Wan et al. (2013)	Aggregate vs. disaggregate forecasts by origin	HK (I) from 8 origin	2002:M8-2011:M10	SARIMA
Zhang et al. (2013)	The impact of China's vacation policies on domestic tourism demand	China (D)	Panel: 2001-2010, 29 Chinese originating cities	Dynamic panel
Daniel and Rodrigues (2012)	The impact of shocks on tourism demand in Portugal	Portugal (I) from Germany, Spain, France, the Netherlands and the UK	1979:Q1-2009Q:4	VAR, VECM
Athanasopoulos and Silva (2012)	Multivariate exponential smoothing for forecasting tourist arrivals	Australia (I) and New Zealand (I) from 11 source markets	1980:M1-2007:M6	VLTS, seasonal naïve, ETS and SARIMA
Cheng (2012)	Tourism demand in Hong Kong: income, prices, and visa restrictions	HK(I) from top 3: China, Taiwan and Japan	1973-2006 1984-2006 for China	EG-ECM
Goh (2012)	Impact of climate on tourism demand of HK	HK (I) from Japan, China, US and UK	1984:M8-2011:M12	JML-ECM
Gounopoulos et al. (2012)	The impact of macroeconomic shocks from origin on tourism demand	Greece (I) from UK, US, France, Germany, Italy and Netherlands	1977:M1-2009:M12	VAR, ARIMA, ETS, Double ETS
Kim et al. (2012)	Explore wealth effect on tourism demand	Korea (O)	1989:Q1-2009:Q4	SR
Kulendran and Dwyer (2012)	The link between seasonal variation in tourism demand and climate variation	Australia (I) from UK, US, Japan and NZ	1975:Q3-2009:Q3	Dynamic regression with ARCH estimated by ML
Mangion et al. (2012)	Quantify the effect of tourism policy on demand and consequently on destination competitiveness	UK (o) to Malta, Spain and Cyprus	1973-2004	Dynamic AIDS
Massidda and Etzo 2012	The determinants of Italian domestic tourism demand	Italy (D)	Panel: 1998-2007, 20 Italian regions	Dynamic panel estimated by GMM
Nowman and Dellen (2012)	Forecasting performance of continuous time model with discrete data	UK (I)	1986:M4-2010:M10	Continuous time model vs. ARIMA and ARFIMA
Onafowora and Owoye (2012)	International demand for Caribbeans	Bahamas, Barbados, Jamaica and St Lucia (I)	1970-2004	CI: ADLM bounds test
Otero-Giráldez et al. 2012	Long-run effects of socioeconomic and meteorological factors on tourism demand	Galicia (Spain) (I)	1999:M1-2010:M12	CI: ADLM bounds test (bootstrapping for confidence intervals)
Page et al. (2012)	The impacts of the Global Economic Crisis and Swine Flu on tourism demand	UK (I) from 14 major markets	1993:Q1-2009:Q2	TVP (forecasting under 2 scenarios to separate the impacts of economic crisis and swine flu
(Rodríguez et al. 2012)	Academic tourism demand in Galicia, Spain	Galicia, Spain (I)	Panel: 2001-2009, students from 36 countries registered in three Galician	Dynamic panel

			universities	
Rossello and Santana (2012)	Climate change and global international tourism	178 countries (I & O)	Panel: 1995-2010 178 countries	pooled OLS (gravity model)
Seetaram(2012a)	Demand elasticities for Australia's outbound tourism	Australia (O)	Panel: 1991-2008, 47 destinations	Panel data co-integration
Seetaram(2012b)	Immigration and the inbound tourism of Australia	Australia (I) from 15 main markets	Panel: 1980-2008 for 15 origins	Simple dynamic panel data model (difference GMM CLSDV)
(Smeral 2012)	Investigate asymmetric income and price effects on tourism demand across business cycle	US, Canada, Australia, Japan, the EU-15 (O)	1978-2009 for Australia, Canada and EU-15; 1985-2009 for Japan and US	SR (with dummy variables to separate the business cycle in two phases)
Tol and Walsh 2012	The impact of climate on tourist destination choice	182 countries	Panel: 1995-2009	Pooled OLS (gravity model)
(Vergori 2012)	Forecasting tourism demand: the role of seasonality	Province of Lecce, Italy (I)	1988:M1-2005:M12	SARIMA VS. ARIMA
Wu et al. (2012)	TVP-EC-AIDS analysis of tourism consumption dynamics	HK (I) from top 4 markets, namely China, Japan, Taiwan and US shopping, hotel accommodation and meals outside hotels	1984-2008 1993-2008 for Mainland China	TVP-EC-AIDS
Andrawis et al. (2011)	Present a different combination strategy concerning long- and short-term forecasts	Egypt (I) from 33 markets and the aggregate tourist arrivals	1993:M1-2007:M12	ETS based on ML, Combination methods: SA, VACO, INV-MES, RANK, least squares estimation, Shrinkage method, Geometric mean, Harmonic mean, a method based on testing the performance difference, HIER
Assaf et al. (2011)	Persistence in the short- and long-term tourism arrivals to Australia	Australia (I) classified with intended length of stay	1991:M1-2009:M1	SARIMA, SAR
Athanasopoulos et al. (2011)	A comprehensive forecasting competition		366 monthly data 427 quarterly data 518 yearly data	ARIMA, state space ETS, Forecast Pro, Theta method, Damped trend, Naïve; ADLM-ECM, TVP, VAR,
Cang (2011)	Nonlinear combination model and forecasting accuracy	UK (I)	1993:Q3-2007:Q4	Naïve 1, Naïve 2, SARIMA, ETS, SVR (with different dimension of inputs from 4 to 8), MLP, SA, VACO and DMSFE
Carson et al. (2011)	Aggregate vs. disaggregate forecast of US commercial air travel	179 busiest airports in US	1990:M1-2004:M12	Simple dynamic
Chan (2011)	Spectrum analysis of seasonality in tourism demand in NZ	NZ (I) from Australia and USA	1980:M1-2007:M12	SA
Chang and McAleer (2011)	Interdependence of international tourism demand and volatility	Indonesia, Malaysia, Singapore and Thailand (I)	1997:M1-2009:M7	CCC, VARMA-GARCH, VARMA-AGARCH
Chen (2011)	Integrating linear and nonlinear modeling techniques to forecast tourism demand	Taiwan (O)	1998:M1-2009:M6	Naïve, ETS, ARIMA, Naïve_BPNN, Naïve_SVR, ETS_BPNN, ETS_SVR, ARIMA_BPNN and ARIMA_SVR
Chu (2011)	Present a piecewise linear model in forecasting tourism demand	Macau (I)	1991:M1-2007:M12	Piecewise linear model vs. AR, SARIMA, ARFIMA
Cortes-Jimenez and Blake (2011)	Tourism demand modeling by travel	UK (I) from 7 origins 4 purposes	1994:Q1-2006:Q3	CSM

	purpose and nationality			
Coshall and Charlesworth (2011)	A management-oriented approach to combine forecasting models	UK (O) to 18 most popular destinations in Europe	1976: Q1-2004:Q3	Statistically-based combine, Goal programming combine, ETS, univariate volatility model, dynamic regression, Naïve 2
(Cuccia and Rizzo 2011)	Tourism seasonality in cultural destinations	Sicily, Italy (I)	1998:M1-2006:M12	Seasonality measured by Census-X12
(Deng and Athanasopoulos 2011)	Modeling tourism demand using a spatial-temporal approach	Australia (D, I)	Domestic: 1998:Q1-2008:Q2; Inbound: 1999:Q1-2008:Q4	Dynamic spatial panel Origin-Destination flow model
(Eugenio-Martin and Campos-Soria 2011)	Income and the substitution pattern between domestic and international tourism demand	EU-15 (D, O)	Cross-sectional: A survey conducted in 1997 to 16183 households	Seemingly unrelated bivariate probit model
Fildes et al. (2011)	Compare models forecasting airline traffic	UK (I & O) from/to US, Canada, Germany, Sweden and Italy	1961-2002; Panel data: 1961-2002 for all origin-destination pairs	ADLM, Pooled ADLM (estimated by SUR), TVP, VAR, AR, ETS, Naïve1, Naïve 2, PcGive Automatic econometric model selection
Fourie and Santana-Gallego (2011)	Impact of mega-sport event on tourism demand	169 countries as destination and 200 as origin	Panel: 1995-2006, for 33,800 pairs	Static panel data model
Hadavandi, Ghanbari, Shahanaghi and Abbasian-Naghneh (2011)	Present a hybrid AI model to develop a Mamdani-type fuzzy rule-based system to forecast tourist arrivals	Taiwan (I) from HK, US and Germany	1989-2003	GFS, ANFIS, GM, Markov residual modified model, FTS
Kim et al. (2011)	Evaluation of alternative intervals forecasts	HK (I) from Australia, China, UK, US as well as Asia, Europe and total; Australia (I) from Germany, NZ, UK, US as well as Europe and total	1985:M1-2008:M5 1980:M:1-2007:M6	AR, AR using the bias-corrected bootstrap, SARIMA, BSM, innovation state space models for ETS
Kulendran and Wong (2011)	Determinants versus composite leading indicators in predicting turning points in growth cycle	HK (I)	1981:Q1-2008:Q4	Logistic and Probit regression (with economic determinants vs. with LIM)
Lee (2011a)	Demand elasticities for inbound tourism demand of HK	HK (I) from 4 short-haul markets: China, Taiwan, Japan and Australia	1959-1999	CI: ADLM bounds test
Lee (2011b)	Forecasting tourism demand for HK, Assess the application of PI-LC ECM in improving forecasting performance	HK (I) from long-haul markets: US, UK, and Germany, as well as short-haul markets: China, Taiwan and Japan	1959-1999	PI-LC ECM (permanent income-life cycle hypothesis ECM) ARIMA, Naïve
Li et al. (2011)	Quantile elasticity of international demand for South Korea	South Korea (I) from US and Japan	1980:M11-2005:M12	QADL
Lin et al. (2011)	Determinants of Taiwan's international tourism demand	Taiwan (I) from Japan, HK, USA	1971:M1-2008:M12	TFM with calendar effects
Nelson et al. (2011)	Investigate factors affecting tourist numbers to Hawaii from US mainland	Hawaii from 49 states in US mainland	Panel and cross-sectional: 1993-2006, 49 states	Panel regression and cross-sectional regression
Rosselló-Nadal et al. (2011)	The impact of weather variability on British outbound flows	UK(O)	1980M1-2009M1	Transfer function models
Santana-Jiménez, and Hernández	The effect of overcrowding on	5 islands of Canary Islands (I) from UK and	Panel: 1990-2005, 5 islands from UK; 5	Panel data model

(2011)	tourism demand	Germany	islands from Germany	
Schiff and Becken (2011)	Demand elasticities for New Zealand	NZ (I) both tourist arrivals and on-the-ground expenditure per arrival	1997-2007	SR
Shen et al. (2011)	Combination forecasts vs. individual forecasts	UK (O) to 7 destinations	1984Q1-2004Q4	RE-ADLM, WB-ECM, JML-ECM, VAR, TVP, seasonal Naïve, SARIMA, combination: SA, VACO, discounted MSFE, Granger-Ramanathan regression method, the shrinkage method and the TVP forecast combination method
Song et al. (2011a)	Present TVP-STSM	HK (I) from China, South Korea, UK, US	1985:Q1-2008:Q4	BSM, CSM, TVP, Naïve 1, Naïve 2, SARIMA, ADLM, TVP-STSM
Song et al. (2011b)	Impact of financial crisis on demand for hotel rooms in HK	HK (I)	1998:Q1-2008:Q4	ADLM bounds test (generate interval estimates and forecasts)
Stepchenkova and Eales (2011)	Destination image as quantified media messages: the effect of news on tourism demand	Russia (I) from UK	1993-2007	SR
Wu et al. (2011)	Analyzing tourist expenditure pattern using dynamic system-of-equations approach	HK (I) from 8	1984-2006	EC-AIDS
Yap and Allen (2011)	Leading indicators of Australian domestic tourism demand	Australia (D)	Panel: 1999Q1-2007Q4,	Dynamic panel estimated by 3SLS (3 stage least squares)
Yorucu and Mehmet (2011)	Bounds test approach for co-integration relationships in tourism demand of Cyprus	Cyprus(I)	1960-2006	CI: ADLM bounds test
Álvarez-Díaz and Rosselló-Nadal (2010)	Whether including meteorological variables can improve predictive power	UK (O) to Balearic Islands	1981:M12-2006:M12	Causal artificial neural network vs. ARIMA and autoregressive neural network
Cang and Hemmington (2010)	Forecasting UK inbound expenditure by visit purpose	UK (I)	1993:Q1-2006:Q4	SARIMA, ETS, Naïve 2
Chan et al. (2010)	Forecast combination using CUSUM technique (apply a quadratic programming approach to determine the combination weights)	HK (I) from top 10 markets	1984:Q1-2004:Q2	6 weighting methods combination: SA, Fixed weighting, Rolling Window, Controlled Weighting, Highest Weighting and Hybrid Method; individual models: ADLM, EG-ECM, VAR
Cho (2010)	Non-economic determinants of tourism demand	Asia, America, Europe and Oceania	Cross-sectional: 4 continents	Cross-sectional regression
Divino and McAleer (2010)	Model the growth rate and volatility in daily international tourist arrivals to Peru	Peru (I)	Daily: 1/1/1997-28/2/2007	GARCH, Exponential GARCH, GJR, ARMA, AR
Divisekera (2010)	Australia's domestic demand	Australia (D)	1998:Q1-2007:Q1	AIDS incorporating seasonality
Eugenio-Martin and Campos-Soria (2010)	Climate in the origin and tourists' destination choice	European Union (O)	Cross-sectional: Data collected form a survey conducted in 1997 with 16183 households	Bivariate probit model
Falk (2010)	The impact of snow depth on winter tourism	Austria (O)	Panel: 1986/87-2005/06, 28 Austrian ski resorts	Panel ECM

Gil-Alana (2010)	Degree of persistence of inbound demand time series in Canary Islands	Canary Island (I)	1997:M1-2008:M7	SARFIMA
Guizzardi and Mazzocchi (2010)	Model the effect of business cycle on tourism demand	Italy (I&D)	1985:Q1-2004:Q4	BSM; CSM vs. Naïve 2
Halicioglu (2010)	Short- and long-run elasticities of outbound demand	Turkey (O)	1970-2005	CI: ADLM bounds test
Kim et al. (2010)	Bias-corrected bootstrap prediction intervals for AR models	HK (I) from Canada, UK and US	1990:M1-2006:M12	AR
Moore (2010)	Impact of climate change on Caribbean tourism demand	18 Caribbean islands	Panel: 1980-2004, 18 Caribbean islands	Static fixed effect panel and panel ECM
Seetanah et al. (2010)	Determinants of demand in South Africa	South Africa (I) from 38 origin countries	Panel: 1985-2000, 38 origin countries	Panel data model
Seetaram (2010)	Model tourism demand using dynamic panel data co-integration technique	Australia (I) from 10 main markets	Panel 1991-2007, from 10 origin	Simple dynamic panel data model (GMM, CLSDV), panel cointegration test
Seo et al. (2010)	Interrelationship among Korean outbound demand	South Korea (O) to 7 countries	1993:M1-2006:M6	VAR
Smeral (2010)	Impacts of the world recession and economic crisis on tourism	Australia, Canada, US, Japan and EU-15 (O)	1977-2008	ECM
Song et al. (2010a)	Confidence intervals for tourism demand elasticities	HK (I) from 9 markets	1985:Q1-2006:Q4	CI: ADLM bounds test (bias-corrected bootstrap to construct confidence intervals for demand elasticities)
Song et al. (2010b)	How should demand be measured?	HK (I)	1981-2006	ADLM with 4 different demand measure
Song and Lin (2010)	Impact of financial and economic crisis on tourism in Asia	Asia (I, O)	1980-2008: I 1984-2008: O	CI: ADLM bounds test (generates interval estimates and forecasts)
Yang et al. (2010)	The role of WHSs in inbound tourism demand in China	26 provinces in China (I) from 9 countries	Panel: 2000-2005, 26 provinces from 9 origins	Static panel
Algieri and Kanellopoulou (2009)	Determinants of tourism demand	France, Greece, Spain and Australia (I)	1985:M1-2006:M1	CSM (structural time series model with ECM)
Andraz et al. (2009)	Relationship between economic cycles and tourist flows	Algarve (I) from UK	1987:M1-2005:M12	Diffusion index model vs. ARMA and AR
Athanasopoulos et al. (2009)	Hierarchical forecasts for Australian domestic tourism demand	Australia (D)	1998:Q1-2006:Q4	State space ETS (disaggregate with aggregate)
Bonham et al. (2009)	Fully identified approach to model and forecast tourism demand	Hawaii (I)	1980:Q1-2005:Q4	VECM
Brida and Risso (2009)	German demand for tourism in South Tyrol	South Tyrol, Italy (I) from Germany	Panel: 1987-2007, 116 tourism destinations of South Tyrol	Dynamic panel
Chu (2009)	Forecast tourism demand with ARMA-based models	HK, Japan, Korea, Taiwan, Singapore, Thailand, Philippines, Australia and NZ (I)	1975:M1-2006:M12 For Australia and Philippines: 1980:M1-2006:M12 1975Q1-2006Q4	SARIMA, ARMA, ARFIMA
Cortes-Jimenez et al. (2009)	Italian outbound tourism demand	Italy (O)	1996:M1-2005M12	Restricted LAIDS vs. EC-LAIDS Estimated by iterative SUR
Coshall (2009)	Combining volatility and exponential	UK (O) to 12 destinations	1976Q1-2007Q3	ARIMA-volatility model, TGARCH, EGARCH,

	smoothing forecasting models			ETS
Divisekera (2009)	Demand for Australian tourism goods and services	Australia (I) from 10 source markets	1996Q1-2006Q4	AIDS:
Garin-Munoz (2009)	Main determinants of demand	Galicia (D&I)	Panel: 1999-2006, domestic for 17 regions and inbound from 24 countries	Panel data model
Kulendran and Wong (2009)	Tourism demand growth rates, directional changes and turning points	HK (I) from Australia, Japan, UK and USA	1975:M3-2003:M12	LIM vs. ARIMA
Kuo et al. (2009)	Avian flu's impact on global and Asian tourism	12 countries or regions in Asian, Europe and Africa (I)	Panel: 2004:M1-2006:M12, 12 countries or regions as destinations	Static and dynamic fixed effect panel data model
Lim et al. (2009a)	Forecasting h(m)otel guest nights in NZ	NZ (I)	1997:M1-2006:M12	ARMA
Lim et al. (2009b)	ARMAX modeling of Japan outbound tourism	Japan (O) to NZ and Taiwan	1980:Q1-2004:Q2	ARMAX
Morley (2009)	How to model dynamics in tourism demand?	Australia (I) From UK, US, Japan and NZ	1980-2001	Simple dynamic, diffusion model (include quadratic functions of previous demand as terms), ARIMAX, ECM, ADLM, and Naïve 1
Santos (2009)	Forecast using data disaggregate d by origins	Spain (I)	1997:M1-2008:M12	SARIMA
Seo et al. (2009)	Determinants of the relationship among Korean tourism demand for Jeju and 3 other international island destinations	Korea (o)	1980:M4-2006:M6	MGARCH, VECM
Shen et al. (2009)	Seasonality treatment and forecasting accuracy	UK (O)	1984:Q1-2004:Q4	Seasonal Naïve model, BSM, SARIMA, reduced ADLM, WB-ECM, JML-ECM, VAR, TVP and CSM
Smeral (2009)	The impact of the financial and economic crisis on European Tourism	EU-15 (O)	1978-2007	ECM
Song et al. (2009)	Comparison of combination and individual forecasts of tourism demand	HK (I)	1994:Q1-2003:Q1	SA, DMSFE and VACO; ADLM, EG-ECM and VAR; ARIMA
Wang (2009)	The impact of crisis events and macroeconomic activity on Taiwan's tourism demand	Taiwan (I)	1996:Q1-2006:Q2	ADLM
Athanasopoulos and Hyndman (2008)	Modelling and forecasting Australian domestic tourism	Australia (D)	1998:Q1-2005:Q2	Dynamic regression, ETS, ETSX (innovations state space models with exogenous variables)
Choyakh (2008)	Including tourism investment variable	Tunisia (I) from Italy, France, Germany and UK	1962-2005	JLM-ECM
Chu (2008a)	Using ARFIMA to forecast tourism demand	Singapore (I)	1977:M7-2004:M11	ARFIMA, naïve 1, naïve 2, ARIMA, Linear regression, CP, SW
Chu (2008b)	Forecasting tourism demand with ARAR algorithm	9 destinations in Asian-Pacific	1975:M1-2006:M12; 1975:Q1-2006:Q4	ARAR vs. SARIMA
Eugenio-Martin et al. (2008)	The role of economic development level in demand modeling	Australia, France, Germany, Japan, Spain, UK and USA (O) to 208 countries all over the world	Panel: 1985-2002, all possible pairs of origin-destination between 7 origin countries to 208 countries	Fixed and random effect panel data model FGLS
Gil-Alana et al.	Forecasting tourism	Canary Islands (I)	1992:M1-2005:M12	

(2008)	demand using seasonal time series models			SARFIMA
Goh et al. (2008)	A rough sets approach to forecast tourism demand	HK(I) from US and UK	1987:M1-2002:M7	SR with rough sets data mining (leisure time index)
Khadaroo and Seetanah (2008)	The role of transport infrastructure in tourism demand	28 countries as both origin and destination globally	Panel: 1990-2000, 28 countries as both origin and destination globally	Dynamic panel
Kuo et al. (2008)	The impact of SARs and avian flu on international tourism demand	China, HK, Singapore and Taiwan, Indonesia and Vietnam (I)	Time series and panel: SARS: 2001:M1-2004:M12 Avian flu: 2002:M10-2006:M9	Time series: ARMA, ARMAI Panel: dynamic panel estimated by difference GMM
Lee et al. (2008)	Predict number to an international tourism Expo held in Korea in 2012	Korea (I)	1990:Q1-2005:Q4	SARIMA, willingness-to-visit survey Delphi method
Lim and Wang (2008)	China's outbound tourism demand	China (O) to Australia	1984-2004	ARIMA
Lim et al. (2008)	Income effects on long and short haul international travel from Japan	Japan (O) to NZ and Taiwan	1980-2004	ARMAX
Lorde and Moore (2008)	Convergence and stability in arrivals in Caribbean	22 Caribbean countries(I)	Panel: 1977:M1-2002:M12, 22 Caribbean countries	Panel data unit root test
McKercher et al. (2008)	Impact of distance on international tourist movements	Departing visitor share from 41 major source markets to 146 destinations	Cross-sectional: market share for 1915 origin-destination pairs in 2002	Cross-sectional regression
Ouerfelli (2008)	Co-integration analysis of quarterly tourism demand in Tunisia	Tunisia (I)	1981:Q1-2004:Q4	JML-ECM, BSM
Saayman and Saayman (2008)	Determinants of inbound tourism demand to South Africa	South Africa (I)	1993:Q1-2004:Q4	VECM
Shen et al. (2008)	Combination forecasts vs. individual forecasts	UK (O) leisure demand for US	1984:Q1-2004;Q4	RE-ADLM, WB-ECM, JML-ECM, VAR, TVP, seasonal Naïve, SARIMA, combination: SA, VACO, discounted MSFE method
Song et al. (2008)	Developing a web-based tourism demand forecasting system	HK (I)	1985:Q1-2006:Q4	Integration of statistical and judgmental forecasts: VAR, Delphi forecasting (two round surveys conducted to 21 PGRs and 5 staffs)

Note: I stands for inbound tourism demand, O stands for outbound tourism demand and D is for domestic tourism demand.

Appendix 2

The programs for generating the combination forecasts and conducting forecasting comparison and analysis are written in Matlab 2018a and presented here. There are three programs altogether, and they are for combining three groups of different individual models respectively. These three programs are similar with different individual inputs. For each program, every origin market is treated individually with the same codes and different data inputs. For each origin market, each combination method is studied separately. Refer to 3.8 for a detailed illustration of the programs.

Programs for Combination Forecasts

Combining the First Group

```
cols=['LFR1_B','LFR1_E','LFR1_SAR','LFR1_SD','LFR1_SE','LFR1_T','LFR1_V','LFR1_LI','LFR1_RA','LFR01_B','LFR01_SD','LFR01_T','LFR01_V','LFR01_LI','LFR0
1_RA','L_FR']
f1 = csvread('H:\201803-new work\individuals\france.csv',2,1);
file = f1;
fstep1 = file(find(file(:,17)==1),:);
fstep2 = file(find(file(:,17)==2),:);
fstep3 = file(find(file(:,17)==3),:);
fstep4 = file(find(file(:,17)==4),:);
all_result1_4=[];
all_combs1_4={};
%[step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse single_median_mae
single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
all_result = [];
all_combs={};
format long g
all_ct=0;
for step=1:4
    if step==1
        m1 = fstep1;
    elseif step==2
        m1=fstep2;
    elseif step==3
        m1=fstep3;
    else
        m1=fstep4;
    end
maes=[];
mapes=[];
rmses=[];
for k =1:15
    forcast_value=m1((length(m1)-8+1):length(m1),k);
    real_value=m1((length(m1)-8+1):length(m1),16);
    mae_error = mean(abs(forcast_value - real_value));
    mape_error = mean(abs(forcast_value - real_value)./real_value);
    rmse_error = (sum((forcast_value - real_value).^2)/length(real_value)).^0.5;
    maes(k)=mae_error;
    mapes(k)=mape_error;
    rmses(k)=rmse_error;
end
single_min_mae = min(maes);
single_min_mape = min(mapes);
single_min_rmse = min(rmses);
single_median_mae = median(maes);
single_median_mape = median(mapes);
single_median_rmse = median(rmses);
single_max_mae = max(maes);
single_max_mape = max(mapes);
single_max_rmse = max(rmses);
trim_type = 0;
alpha=0;
best_mae_comb=[];
best_mape_comb=[];
best_rmse_comb=[];
min_best_mae_comb=100000000000;
min_best_mape_comb=10000000000;
min_best_rmse_comb=10000000000;
for k =1:15
    A=1:15;
    combs=nchoosek(A,k);
    for type=1:4
        switch type
            case 1 % SA
                ct = 0;
                for row=1:1:size(combs,1)
                    ct=ct+1;
                    all_ct=all_ct+1;
                    comb = combs(row,:);
                    forcast_value=mean(m1((length(m1)-8+1):length(m1),comb),2);
                    real_value=m1((length(m1)-8+1):length(m1),16);
                    mae_error = mean(abs(forcast_value - real_value));
                    mape_error = mean(abs(forcast_value - real_value)./real_value);
                    rmse_error = (sum((forcast_value - real_value).^2)/length(real_value)).^0.5;
                    if mae_error < min_best_mae_comb
                        min_best_mae_comb = mae_error;
                        best_mae_comb = comb;
                    end
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse
single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
            case 2 %
                ct = 0;
                for row=1:1:size(combs,1)
                    ct=ct+1;
                    all_ct=all_ct+1;
                    comb = combs(row,:);
                    new_forecasts=[];
                    for forecast_row =(length(m1)-8+1):length(m1)
                        errors=[];
                        weights=[];
                        history_data=m1(1:forecast_row-1,comb);
                        for comb_ind=1:length(comb)
                        weights(comb_ind)=1/sum((history_data(:,comb_ind) -m1(1:forecast_row-1,16)).^2);
                        end
                        new_weights = weights/sum(weights);
                        new_forecast = sum(new_weights.*m1(forecast_row,comb));
                        new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                    end
                    real_value=m1((length(m1)-8+1):length(m1),16);
                    mae_error = mean(abs(new_forecasts' - real_value));
                    mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                    rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse
single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
        case 3 %
            for alpha =[0.85 0.9 0.95]
                ct = 0;
                for row=1:1:size(combs,1)
                    all_ct=all_ct+1;
                    ct=ct+1;
                    comb = combs(row,:);
                    new_forecasts=[];
                    for forecast_row =(length(m1)-8+1):length(m1)
```

```
                        errors=[];
                        weights=[];
                        history_data=m1(1:forecast_row-1,comb);
                        for comb_ind=1:length(comb)
                            hist_weights=alpha.^[forecast_row-1:-1:1];
                            weights(comb_ind)=1/(sum((history_data(:,comb_ind) -m1(1:forecast_row-1,16))'.^2 .* hist_weights));
                        end
                        new_weights = weights/sum(weights);
                        new_forecast = sum(new_weights.*m1(forecast_row,comb));
                        new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                    end
                    real_value=m1((length(m1)-8+1):length(m1),16);
                    mae_error = mean(abs(new_forecasts' - real_value));
                    mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                    rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
            end
        case 4 %
            ct = 0;
            alpha=0;
            for row=1:1:size(combs,1)
                ct=ct+1;
                all_ct=all_ct+1;
                comb = combs(row,:);
                new_forecasts=[];
                for forecast_row =(length(m1)-8+1):length(m1)
                    errors=[];
                    weights=[];
                    history_data=m1(1:forecast_row-1,comb);
                    for comb_ind=1:length(comb)
                        weights(comb_ind)=1/mean(abs((history_data(:,comb_ind) -m1(1:forecast_row-1,16))./m1(1:forecast_row-1,16)));
                    end
                    new_weights = weights/sum(weights);
                    new_forecast = sum(new_weights.*m1(forecast_row,comb));
                    new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                end
                real_value=m1((length(m1)-8+1):length(m1),16);
                mae_error = mean(abs(new_forecasts' - real_value));
                mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                res
                all_result(ct,:)=res;
                all_combs{ct}=comb;
                all_result1_4(all_ct,:)=res;
                all_combs1_4{all_ct}=comb;
            end
            end
        end
    end
end
% TWO STEP
trim_type = 1;
alpha=0;
[sA index]=sort(rmses,'descend');
top1=find(index==1);
top2=find(index==2);
top3=find(index==3);
A=1:15;
A((A==top1)|(A==top2)|(A==top3))=[];
for k =1:length(A)
    combs=nchoosek(A,k);
    for type=1:4
        switch type
            case 1 %
                ct = 0;
                for row=1:1:size(combs,1)
                    ct=ct+1;
                    all_ct=all_ct+1;
                    comb = combs(row,:);
                    forcast_value=mean(m1((length(m1)-8+1):length(m1),comb),2);
                    real_value=m1((length(m1)-8+1):length(m1),16);
                    mae_error = mean(abs(forcast_value - real_value));
                    mape_error = mean(abs(forcast_value - real_value)./real_value);
                    rmse_error = (sum((forcast_value - real_value).^2)/length(real_value)).^0.5;
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    all_ct
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
            case 2 %
                ct = 0;
                for row=1:1:size(combs,1)
                    ct=ct+1;
                    all_ct=all_ct+1;
                    comb = combs(row,:);
                    new_forecasts=[];
                    for forecast_row =(length(m1)-8+1):length(m1)
                        errors=[];
                        weights=[];
                        history_data=m1(1:forecast_row-1,comb);
                        for comb_ind=1:length(comb)
                            weights(comb_ind)=1/sum((history_data(:,comb_ind) -m1(1:forecast_row-1,16)).^2);
                        end
                        new_weights = weights/sum(weights);
                        new_forecast = sum(new_weights.*m1(forecast_row,comb));
                        new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                    end
                    real_value=m1((length(m1)-8+1):length(m1),16);
                    mae_error = mean(abs(new_forecasts' - real_value));
                    mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                    rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
                size2=size(all_combs1_4);
            case 3 %
                for alpha= [0.85 0.9 0.95]
                    ct = 0;
                    for row=1:1:size(combs,1)
                        all_ct=all_ct+1;
                        ct=ct+1;
                        comb = combs(row,:);
                        new_forecasts=[];
                        for forecast_row =(length(m1)-8+1):length(m1)
                            errors=[];
                            weights=[];
                            history_data=m1(1:forecast_row-1,comb);
                            for comb_ind=1:length(comb)
                                hist_weights=alpha.^[forecast_row-1:-1:1];
                                weights(comb_ind)=1/(sum((history_data(:,comb_ind) -m1(1:forecast_row-1,16))'.^2 .* hist_weights));
                            end
                            new_weights = weights/sum(weights);
                            new_forecast = sum(new_weights.*m1(forecast_row,comb));
                            new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                        end
                        real_value=m1((length(m1)-8+1):length(m1),16);
                        mae_error = mean(abs(new_forecasts' - real_value));
                        mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                        rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                        res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                        res
                        all_result(ct,:)=res;
                        all_combs{ct}=comb;
                        all_result1_4(all_ct,:)=res;
                        all_combs1_4{all_ct}=comb;
                    end
                end
        %   case 4
                ct = 0;
                alpha=0;
                for row=1:1:size(combs,1)
                    ct=ct+1;
                    all_ct=all_ct+1;
                    comb = combs(row,:);
                    new_forecasts=[];
```

```
                    for forecast_row =(length(m1)-8+1):length(m1)
                        errors=[];
                        weights=[];
                        history_data=m1(1:forecast_row-1,comb);
                        for comb_ind=1:length(comb)
                            weights(comb_ind)=1/mean(abs((history_data(:,comb_ind) -m1(1:forecast_row-1,16))./m1(1:forecast_row-1,16)));
                        end
                        new_weights = weights/sum(weights);
                        new_forecast = sum(new_weights.*m1(forecast_row,comb));
                        new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                    end
                    real_value=m1((length(m1)-8+1):length(m1),16);
                    mae_error = mean(abs(new_forecasts' - real_value));
                    mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                    rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
            end
        end
end
end
smaes = [];
smapes= [];
srmses= [];
ct=0;
for step=1:4
    ct=ct+1;
    resn = all_result1_4(find(all_result1_4(:,1)==step & all_result1_4(:,2)==1 ),:);
    mae =resn(1:15,7)';
    mape=resn(1:15,8)';
    rmse=resn(1:15,9)';
    smaes(ct,:)=mae;
    smapes(ct,:)=mape;
    srmses(ct,:)=rmse;
end
csvwrite ('smaes.csv',smaes)
csvwrite ('smapes.csv',smapes)
csvwrite ('srmses.csv',srmses)
lens = [];
l=0;
for step =1:4
    for trim_type=[0 1]
        for type=[1 2 3 4]
            for alpha= [0 0.85 0.9 0.95]
                mae_index0  = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,7)<= all_result1_4(:,10));
                mape_index0 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,8)<= all_result1_4(:,11));
                rmse_index0 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,9)<= all_result1_4(:,12));
                mae_index1  = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,7)>= all_result1_4(:,16));
                mape_index1 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,8)>= all_result1_4(:,17));
                rmse_index1 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,9)>= all_result1_4(:,18));
                l=l+1;
                lens(l,:)=[step  trim_type  type  alpha  length(mae_index0)  length(mape_index0)  length(rmse_index0)  length(mae_index1) length(mape_index1) length(rmse_index1) ];
            end
        end
    end
end
bmaes = [];
bmapes= [];
brmses= [];
ct=0;
for step=1:4
    ct=ct+1;
    step_lens = lens(find(lens(:,1)==step),:);
    mae = step_lens(:,5)';
    mape= step_lens(:,6)';
    rmse= step_lens(:,7)';
    bmaes(ct,:) = mae;
    bmapes(ct,:)= mape;
    brmses(ct,:)= rmse;
end
wmaes = [];
wmapes= [];
wrmses= [];
ct=0;
for step=1:4
    ct=ct+1;
    step_lens = lens(find(lens(:,1)==step),:);
    mae = step_lens(:,8)';
    mape= step_lens(:,9)';
    rmse= step_lens(:,10)';
    wmaes(ct,:) = mae;
    wmapes(ct,:)= mape;
    wrmses(ct,:)= rmse;
endcsvwrite('bmaes.csv',bmaes)
csvwrite('bmapes.csv',bmapes)
csvwrite('brmses.csv',brmses)
csvwrite('wmaes.csv',wmaes)
csvwrite('wmaeps.csv',wmapes)
csvwrite('wrmses.csv',wrmses)
 for trim_type=[0 1]
        for type=[1 2 3 4]
            for alpha= [0 0.85 0.9 0.95]
                ct=ct+1;
                single_in_combs2=zeros(1,15);
                mae_index2  = find(all_result1_4(:,4)==type  & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,7)<= all_result1_4(:,10));
                mape_index2 = find(all_result1_4(:,4)==type  & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,8)<= all_result1_4(:,11));
                rmse_index2 = find(all_result1_4(:,4)==type  & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,9)<= all_result1_4(:,12));
                index2=[mae_index2 ;mape_index2;rmse_index2];
for item = index2'
                    comb = all_combs1_4{item};
                    for tind = 1:length(comb)
                        ti=comb(tind);
                        single_in_combs2(ti)=single_in_combs2(ti)+1    ;
                    end
                end
sum2= size(index2,1);
                single_in_combs2 =single_in_combs2/sum2;
                all_single_in_combs2(ct,:)=single_in_combs2;
            end
        end
    end
all_single_in_combs20 = all_single_in_combs2;
all_single_in_combs20(any(isnan(all_single_in_combs20)'),:) = [];
best_result_mae=[];
best_comb_mae={};
bst_ct_mae=0;
best_result_mape=[];
best_comb_mape={};
bst_ct_mape=0;
best_result_rmse=[];
best_comb_rmse={};
bst_ct_rmse=0;
for step =1:4
    best_result=[];
    bst_comb =[];
    for trim_type=[0 1]
        for type=[1 2 3 4]
            for alpha= [0 0.85 0.9 0.95]
                tmp_all_result1_4   =   all_result1_4(find(all_result1_4(:,1)==step  &  all_result1_4(:,2)==k  &  all_result1_4(:,4)==type  & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type ),:);
                if length(tmp_all_result1_4)>0
                    bst_mae_index   = find(all_result1_4(:,1)==step & all_result1_4(:,2)==k & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,7)== min(tmp_all_result1_4(:,7)));
                    bst_mape_index = find(all_result1_4(:,1)==step & all_result1_4(:,2)==k & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,8)== min(tmp_all_result1_4(:,8)));
                    bst_rmse_index = find(all_result1_4(:,1)==step & all_result1_4(:,2)==k & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,9)== min(tmp_all_result1_4(:,9)));
                    bst_mae = all_result1_4(bst_mae_index,:);
                    bst_comb_mae = all_combs1_4{bst_mae_index};
                    bst_mape = all_result1_4(bst_mape_index,:);
                    bst_comb_mape = all_combs1_4{bst_mape_index};
                    bst_rmse = all_result1_4(bst_rmse_index,:);
                    bst_comb_rmse = all_combs1_4{bst_rmse_index};
                    bst_ct_mae=bst_ct_mae+1;
                    best_result_mae(bst_ct_mae,:)=bst_mae;
                    best_comb_mae{bst_ct_mae}=bst_comb_mae;
```

```
                        bst_ct_mape=bst_ct_mape+1;
                        best_result_mape(bst_ct_mape,:)=bst_mape;
                        best_comb_mape{bst_ct_mape}=bst_comb_mape;
                        bst_ct_rmse=bst_ct_rmse+1;
                        best_result_rmse(bst_ct_rmse,:)=bst_rmse;
                        best_comb_rmse{bst_ct_rmse}=bst_comb_rmse;
                    end
                end
            end
        end
end
best_result = [best_result_mae;best_result_mape;best_result_rmse];
best_single_count = tabulate(best_result(:,2));
```

Combining the Second Group

```
cols=['LFR1_E','LFR1_SAR','LFR1_SE','LFR0_B','LFR0_SD','LFR0_T','LFR0_V','LFR0_VE','LFR0_RA','L_FR']
f1 = csvread('H:\201803-new work\individuals\france0.csv',2,1);
file = f1;
fstep1 = file(find(file(:,11)==1),:);
fstep2 = file(find(file(:,11)==2),:);
fstep3 = file(find(file(:,11)==3),:);
fstep4 = file(find(file(:,11)==4),:);
all_result1_4=[];
all_combs1_4={};
all_result =[];
all_combs={};
format long g
all_ct=0;
 for step=1:4
    if step==1
        m1 = fstep1;
    elseif step==2
        m1=fstep2;
    elseif step==3
        m1=fstep3;
    else
        m1=fstep4;
    end
maes=[];
mapes=[];
rmses=[];
for k =1:9
    forcast_value=m1((length(m1)-8+1):length(m1),k);
    real_value=m1((length(m1)-8+1):length(m1),10);
    mae_error = mean(abs(forcast_value - real_value));
    mape_error = mean(abs(forcast_value - real_value)./real_value);
    rmse_error = (sum((forcast_value - real_value).^2)/length(real_value)).^0.5;
    maes(k)=mae_error;
    mapes(k)=mape_error;
    rmses(k)=rmse_error;
end
single_min_mae = min(maes);
single_min_mape = min(mapes);
single_min_rmse = min(rmses);
single_median_mae = median(maes);
single_median_mape = median(mapes);
single_median_rmse = median(rmses);
single_max_mae = max(maes);
single_max_mape = max(mapes);
single_max_rmse = max(rmses);

trim_type = 0;
alpha=0;
best_mae_comb=[];
best_mape_comb=[];
best_rmse_comb=[];
min_best_mae_comb=100000000000;
min_best_mape_comb=10000000000;
min_best_rmse_comb=10000000000;
for k =1:9
    A=1:9;
    combs=nchoosek(A,k);
    for type=1:4
        switch type
            case 1 % SA
                ct = 0;
                for row=1:1:size(combs,1)
                    ct=ct+1;
                    all_ct=all_ct+1;
                    comb = combs(row,:);
                    forcast_value=mean(m1((length(m1)-8+1):length(m1),comb),2);
                    real_value=m1((length(m1)-8+1):length(m1),10);
                    mae_error = mean(abs(forcast_value - real_value));
                    mape_error = mean(abs(forcast_value - real_value)./real_value);
                    rmse_error = (sum((forcast_value - real_value).^2)/length(real_value)).^0.5;
                    if mae_error < min_best_mae_comb
                        min_best_mae_comb = mae_error;
                        best_mae_comb = comb;
                    end
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse
single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
            case 2 %
                ct = 0;
                for row=1:1:size(combs,1)
                    ct=ct+1;
                    all_ct=all_ct+1;
                    comb = combs(row,:);
                    new_forecasts=[];
                    for forecast_row =(length(m1)-8+1):length(m1)
                        errors=[];
                        weights=[];
                        history_data=m1(1:forecast_row-1,comb);
                        for comb_ind=1:length(comb)
                            weights(comb_ind)=1/sum((history_data(:,comb_ind) -m1(1:forecast_row-1,10)).^2);
                        end
                        new_weights = weights/sum(weights);
                        new_forecast = sum(new_weights.*m1(forecast_row,comb));
                        new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                    end
                    real_value=m1((length(m1)-8+1):length(m1),10);
                    mae_error = mean(abs(new_forecasts' - real_value));
                    mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                    rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse
single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
        case 3 %
            for alpha =[0.85 0.9 0.95]
                ct = 0;
                for row=1:1:size(combs,1)
                    all_ct=all_ct+1;
                    ct=ct+1;
                    comb = combs(row,:);
                    new_forecasts=[];
                    for forecast_row =(length(m1)-8+1):length(m1)
                        errors=[];
                        weights=[];
                        history_data=m1(1:forecast_row-1,comb);
                        for comb_ind=1:length(comb)
                            hist_weights=alpha.^[forecast_row-1:-1:1];
                            weights(comb_ind)=1/(sum((history_data(:,comb_ind) -m1(1:forecast_row-1,10))'.^2 .* hist_weights));
                        end
                        new_weights = weights/sum(weights);
                        new_forecast = sum(new_weights.*m1(forecast_row,comb));
                        new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                    end
                    real_value=m1((length(m1)-8+1):length(m1),10);
                    mae_error = mean(abs(new_forecasts' - real_value));
                    mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                    rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse
single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
            end
        case 4 %
            ct = 0;
            alpha=0;
```

```
        for row=1:1:size(combs,1)
            ct=ct+1;
            all_ct=all_ct+1;
            comb = combs(row,:);
            new_forecasts=[];
            for forecast_row =(length(m1)-8+1):length(m1)
                errors=[];
                weights=[];
                history_data=m1(1:forecast_row-1,comb);
                for comb_ind=1:length(comb)
                    weights(comb_ind)=1/mean(abs((history_data(:,comb_ind) -m1(1:forecast_row-1,10))./m1(1:forecast_row-1,10)));
                end
                new_weights = weights/sum(weights);
                new_forecast = sum(new_weights.*m1(forecast_row,comb));
                new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
            end
            real_value=m1((length(m1)-8+1):length(m1),10);
            mae_error = mean(abs(new_forecasts' - real_value));
            mape_error = mean(abs(new_forecasts' - real_value)./real_value);
            rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
            res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
            res
            all_result(ct,:)=res;
            all_combs{ct}=comb;
            all_result1_4(all_ct,:)=res;
            all_combs1_4{all_ct}=comb;
        end
    end
  end
end
lens = [];
i=0;
single_in_combs_mae=zeros(1,9);
single_in_combs_mape=zeros(1,9);
single_in_combs_rmse=zeros(1,9);
best_result_mae=[];
best_comb_mae={};
bst_ct_mae=0;
best_result_mape=[];
best_comb_mape={};
bst_ct_mape=0;
best_result_rmse=[];
best_comb_rmse={};
bst_ct_rmse=0;
for step =1:4
    best_result=[];
    bst_comb =[];
        for type=[1 2 3 4]
            for alpha= [0 0.85 0.9 0.95]
                mae_index0   = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type    & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,7)<= all_result1_4(:,10));
                mape_index0 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type    & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,8)<= all_result1_4(:,11));
                rmse_index0 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type    & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,9)<= all_result1_4(:,12));
                mae_index1   = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type    & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,7)<= all_result1_4(:,13));
                mape_index1 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type    & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,8)<= all_result1_4(:,14));
                rmse_index1 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type    & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,9)<= all_result1_4(:,15));
                mae_index2   = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type    & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,7)>= all_result1_4(:,16));
                mape_index2 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type    & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,8)>= all_result1_4(:,17));
                rmse_index2 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type    & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,9)>= all_result1_4(:,18));
                i=i+1;
                lens(i,:)=[step  trim_type  type  alpha  length(mae_index0)  length(mape_index0)  length(rmse_index0)  length(mae_index1) length(mape_index1) length(rmse_index1)   length(mae_index2) length(mape_index2) length(rmse_index2) ];
            end
        end
    end
bmaes = [];
bmapes= [];
brmses= [];
ct=0;
for step=1:4
    ct=ct+1;
    step_lens = lens(find(lens(:,1)==step),:);
    mae = step_lens(:,5)';
    mape= step_lens(:,6)';
    rmse= step_lens(:,7)';
    bmaes(ct,:) = mae;
    bmapes(ct,:)= mape;
    brmses(ct,:)= rmse;
end
mmaes = [];
mmapes= [];
mrmses= [];
ct=0;
for step=1:4
    ct=ct+1;
    step_lens = lens(find(lens(:,1)==step),:);
    mae = step_lens(:,8)';
    mape= step_lens(:,9)';
    rmse= step_lens(:,10)';
    mmaes(ct,:) = mae;
    mmapes(ct,:)= mape;
    mrmses(ct,:)= rmse;
end
wmaes = [];
wmapes= [];
wrmses= [];
ct=0;
for step=1:4
    ct=ct+1;
    step_lens = lens(find(lens(:,1)==step),:);
    mae = step_lens(:,11)';
    mape= step_lens(:,12)';
    rmse= step_lens(:,13)';
    wmaes(ct,:) = mae;
    wmapes(ct,:)= mape;
    wrmses(ct,:)= rmse;
end
csvwrite('bmaes.csv',bmaes)
csvwrite('bmapes.csv',bmapes)
csvwrite('brmses.csv',brmses)
csvwrite('mmaes.csv',mmaes)
csvwrite('mmaeps.csv',mmapes)
csvwrite('mrmses.csv',mrmses)
csvwrite('wmaes.csv',wmaes)
csvwrite('wmaeps.csv',wmapes)
csvwrite('wrmses.csv',wrmses)
```

Combining the Third Group

```
cols=['LFR1_B','LFR1_E','LFR1_SAR','LFR1_SD','LFR1_SE','LFR1_T','LFR1_V','LFR1_VE','LFR1_RA','L_FR']
f1 = csvread('H:\201803-new work\individuals\france1.csv',2,1);
file = f1;
fstep1 = file(find(file(:,11)==1),:);
fstep2 = file(find(file(:,11)==2),:);
fstep3 = file(find(file(:,11)==3),:);
fstep4 = file(find(file(:,11)==4),:);
all_result1_4=[];
all_combs1_4={};
all_result =[];
all_combs={};
format long g
all_ct=0;
 for step=1:4
    if step==1
        m1 = fstep1;
    elseif step==2
        m1=fstep2;
    elseif step==3
        m1=fstep3;
    else
        m1=fstep4;
    end
maes=[];
mapes=[];
rmses=[];
for k =1:9
    forcast_value=m1((length(m1)-8+1):length(m1),k);
    real_value=m1((length(m1)-8+1):length(m1),10);
    mae_error = mean(abs(forcast_value - real_value));
    mape_error = mean(abs(forcast_value - real_value)./real_value);
    rmse_error = (sum((forcast_value - real_value).^2)/length(real_value)).^0.5;
    maes(k)=mae_error;
    mapes(k)=mape_error;
    rmses(k)=rmse_error;
end
```

```
single_min_mae = min(maes);
single_min_mape = min(mapes);
single_min_rmse = min(rmses);
single_median_mae = median(maes);
single_median_mape = median(mapes);
single_median_rmse = median(rmses);
single_max_mae = max(maes);
single_max_mape = max(mapes);
single_max_rmse = max(rmses);

trim_type = 0;
alpha=0;
best_mae_comb=[];
best_mape_comb=[];
best_rmse_comb=[];
min_best_mae_comb=100000000000;
min_best_mape_comb=10000000000;
min_best_rmse_comb=10000000000;
for k =1:9
    A=1:9;
    combs=nchoosek(A,k);
    for type=1:4
        switch type
            case 1 % SA
                ct = 0;
                for row=1:1:size(combs,1)
                    ct=ct+1;
                    all_ct=all_ct+1;
                    comb = combs(row,:);
                    forcast_value=mean(m1((length(m1)-8+1):length(m1),comb),2);
                    real_value=m1((length(m1)-8+1):length(m1),10);
                    mae_error = mean(abs(forcast_value - real_value));
                    mape_error = mean(abs(forcast_value - real_value)./real_value);
                    rmse_error = (sum((forcast_value - real_value).^2)/length(real_value)).^0.5;
                    if mae_error < min_best_mae_comb
                        min_best_mae_comb = mae_error;
                        best_mae_comb = comb;
                    end
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse
single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
            case 2 %
                ct = 0;
                for row=1:1:size(combs,1)
                    ct=ct+1;
                    all_ct=all_ct+1;
                    comb = combs(row,:);
                    new_forecasts=[];
                    for forecast_row =(length(m1)-8+1):length(m1)
                        errors=[];
                        weights=[];
                        history_data=m1(1:forecast_row-1,comb);
                        for comb_ind=1:length(comb)
                        weights(comb_ind)=1/sum((history_data(:,comb_ind) -m1(1:forecast_row-1,10)).^2);
                        end
                        new_weights = weights/sum(weights);
                        new_forecast = sum(new_weights.*m1(forecast_row,comb));
                        new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                    end
                    real_value=m1((length(m1)-8+1):length(m1),10);
                    mae_error = mean(abs(new_forecasts' - real_value));
                    mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                    rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse
single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
        case 3 %
            for alpha =[0.85 0.9 0.95]
                ct = 0;
                for row=1:1:size(combs,1)
                    all_ct=all_ct+1;
                    ct=ct+1;
                    comb = combs(row,:);
                    new_forecasts=[];
                    for forecast_row =(length(m1)-8+1):length(m1)
                        errors=[];
                        weights=[];
                        history_data=m1(1:forecast_row-1,comb);
                        for comb_ind=1:length(comb)
                            %errors=(history_data(:,comb_ind) -m1(1:forecast_row-1,10)).^2;
                            hist_weights=alpha.^[forecast_row-1:-1:1];
                            weights(comb_ind)=1/(sum((history_data(:,comb_ind) -m1(1:forecast_row-1,10))'.^2 .* hist_weights));
                        end
                        new_weights = weights/sum(weights);
                        new_forecast = sum(new_weights.*m1(forecast_row,comb));
                        new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                    end
                    real_value=m1((length(m1)-8+1):length(m1),10);
                    mae_error = mean(abs(new_forecasts' - real_value));
                    mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                    rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                    res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse
single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                    res
                    all_result(ct,:)=res;
                    all_combs{ct}=comb;
                    all_result1_4(all_ct,:)=res;
                    all_combs1_4{all_ct}=comb;
                end
            end
        case 4 %
            ct = 0;
            alpha=0;
            for row=1:1:size(combs,1)
                ct=ct+1;
                all_ct=all_ct+1;
                comb = combs(row,:);
                new_forecasts=[];
                for forecast_row =(length(m1)-8+1):length(m1)
                    errors=[];
                    weights=[];
                    history_data=m1(1:forecast_row-1,comb);
                    for comb_ind=1:length(comb)
                        weights(comb_ind)=1/mean(abs((history_data(:,comb_ind) -m1(1:forecast_row-1,10))./m1(1:forecast_row-1,10)));
                    end
                    new_weights = weights/sum(weights);
                    new_forecast = sum(new_weights.*m1(forecast_row,comb));
                    new_forecasts(forecast_row-(length(m1)-8+1)+1)=new_forecast;
                end
                real_value=m1((length(m1)-8+1):length(m1),10);
                mae_error = mean(abs(new_forecasts' - real_value));
                mape_error = mean(abs(new_forecasts' - real_value)./real_value);
                rmse_error = (sum((new_forecasts' - real_value).^2)/length(real_value)).^0.5;
                res = [step k row type alpha trim_type mae_error mape_error rmse_error single_min_mae single_min_mape single_min_rmse
single_median_mae single_median_mape single_median_rmse single_max_mae single_max_mape single_max_rmse];
                res
                all_result(ct,:)=res;
                all_combs{ct}=comb;
                all_result1_4(all_ct,:)=res;
                all_combs1_4{all_ct}=comb;
            end
        end
    end
end
end
end
lens = [];
i=0;
single_in_combs_mae=zeros(1,9);
single_in_combs_mape=zeros(1,9);
single_in_combs_rmse=zeros(1,9);
best_result_mae=[];
best_comb_mae={};
bst_ct_mae=0;
best_result_mape=[];
best_comb_mape={};
bst_ct_mape=0;
best_result_rmse=[];
best_comb_rmse={};
bst_ct_rmse=0;
for step =1:4
    best_result=[];
    bst_comb =[];
```

```
        for type=[1 2 3 4]
            for alpha= [0 0.85 0.9 0.95]
                mae_index0   = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,7)<= all_result1_4(:,10));
                mape_index0 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,8)<= all_result1_4(:,11));
                rmse_index0 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,9)<= all_result1_4(:,12));
                mae_index1   = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,7)<= all_result1_4(:,13));
                mape_index1 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,8)<= all_result1_4(:,14));
                rmse_index1 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,9)<= all_result1_4(:,15));
                mae_index2   = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,7)>= all_result1_4(:,16));
                mape_index2 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,8)>= all_result1_4(:,17));
                rmse_index2 = find(all_result1_4(:,1)==step & all_result1_4(:,4)==type   & all_result1_4(:,5)==alpha & all_result1_4(:,6)==trim_type & all_result1_4(:,9)>= all_result1_4(:,18));
                i=i+1;
                lens(i,:)=[step   trim_type   type   alpha   length(mae_index0)   length(mape_index0)   length(rmse_index0)   length(mae_index1) length(mape_index1) length(rmse_index1)   length(mae_index2) length(mape_index2) length(rmse_index2) ];
            end
        end
    end
end
bmaes = [];
bmapes= [];
brmses= [];
ct=0;
for step=1:4
    ct=ct+1;
    step_lens = lens(find(lens(:,1)==step),:);
    mae = step_lens(:,5)';
    mape= step_lens(:,6)';
    rmse= step_lens(:,7)';
    bmaes(ct,:) = mae;
    bmapes(ct,:)= mape;
    brmses(ct,:)= rmse;
end
mmaes = [];
mmapes= [];
mrmses= [];
ct=0;
for step=1:4
    ct=ct+1;
    step_lens = lens(find(lens(:,1)==step),:);
    mae = step_lens(:,8)';
    mape= step_lens(:,9)';
    rmse= step_lens(:,10)';
    mmaes(ct,:) = mae;
    mmapes(ct,:)= mape;
    mrmses(ct,:)= rmse;
end
wmaes = [];
wmapes= [];
wrmses= [];
ct=0;
for step=1:4
    ct=ct+1;
    step_lens = lens(find(lens(:,1)==step),:);
    mae = step_lens(:,8)';
    mape= step_lens(:,9)';
    rmse= step_lens(:,10)';
    wmaes(ct,:) = mae;
    wmapes(ct,:)= mape;
    wrmses(ct,:)= rmse;
end
csvwrite('bmaes.csv',bmaes)
csvwrite('bmapes.csv',bmapes)
csvwrite('brmses.csv',brmses)
csvwrite('wmaes.csv',wmaes)
csvwrite('wmaeps.csv',wmapes)
csvwrite('wrmses.csv',wrmses)
csvwrite ('mmaes.csv',mmaes)
csvwrite ('mmapes.csv',mmapes)
csvwrite ('mrmses.csv',mrmses)
```