The Hong Kong Polytechnic University

School of Hotel and Tourism Management

Improving the Design of Tourism Demand Forecasting Support System

Zixuan Gao

A thesis submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

04/2015

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

______________________________ (Signed)

Zixuan Gao (Name of student)

ABSTRACT

Accurate tourism demand forecast is the foundation of all tourism-related businesses. As a particular type of decision support system, forecasting support systems (FSS) have been widely applied in tourism demand forecasting in recent years. One of the typical characteristics of existing tourism demand forecasting support systems (TDFSS) is the combination of statistical and judgmental forecasting techniques. A review of recent studies in this area shows that most studies on the development of TDFSS focus on the improvement of statistical forecasting methods. The effectiveness of human participation in the forecasting process is largely neglected, especially the influence of forecasters' cognitive bias on forecast accuracy during the judgmental forecasting process when using TDFSS.

Focusing on three typical cognitive biases (desire bias, anchoring bias, and overconfidence bias) in the literature of judgmental forecasting, this study represents the first attempt to identify the influence of these three cognitive biases on the judgmental forecasting of tourism demand and how they affect forecast accuracy. The second purpose of this study is to propose a systematic debiasing model that is able to effectively reduce the forecast error associated with the identified cognitive biases and can be easily implemented in the design of TDFSS. The proposed debiasing model comprises two parts: cognitive bias detection and debiasing. In the first part, potential cognitive biases involved in forecasters' judgmental forecasts can be detected with a series of post-hoc tests. Based on the typical design features of FSS, both informative guidance and suggestive guidance are used as the debiasing strategies in the second part of the model.

To test its effectiveness and related hypotheses, the proposed debiasing model has been implemented in the design of the Hong Kong tourism demand forecasting support system (HKTDFS). A two-stage laboratory experiment using HKTDFS and the empirical data of international tourist arrivals to Hong Kong from 10 destination-origin (D-O) pair markets was conducted. The experiment proceeded in three sessions and 75 qualified forecasters agreed to participate. Ultimately, 68 participants provided qualified data for further analysis.

The results show that 14 of 21 hypotheses are supported, one is partially supported, and the remaining six are rejected. Generally, the three cognitive biases examined are common in judgmental forecasting of tourism demand and contribute significantly to forecast error. Both performance feedback (PF) and system-suggested forecasts are effective in eliminating the influence of cognitive bias on forecast accuracy. In the design of TDFSS, these two debiasing strategies should be used in dealing with different cognitive biases. To be specific, PF should be provided to forecasters when desired outcome-related cognitive biases are detected; system-suggested forecasts should be recommended to replace forecasters' judgmental forecasts when forecasters anchor their judgmental forecast on the statistical forecast or the latest observation of the forecasting series. In extreme cases, when system-suggested forecasts are not available, keeping statistical forecasts unchanged is the backup strategy when forecasters anchor their judgmental forecasts on statistical forecast; Naïve I forecast is the backup strategy when forecasters anchor on the latest observation of the forecasting series. These results provide evidence to further revise the debiasing model in order to improve the design of TDFSS.

Keywords: forecasting support system, tourism demand, judgmental forecasting, cognitive bias, desire bias, anchoring bias, overconfidence, debiasing, decision guidance

ACKNOWLEDGEMENTS

I would like to express my warmest and sincere gratitude to my chief supervisor Prof. Haiyan Song for his continuous support of my study. His invaluable guidance always be a lantern in the dark when I facing difficulties in my research and thesis writing. The completion of this thesis would even been impossible without his persistent help and encouragement throughout the whole life periods of my Ph.D. study.

My sincere thanks also go to the faculty members and my colleagues in the school of Hotel and Tourism Management (SHTM) for their insightful comments and suggestions on my study. Special thanks to Prof. Rob Law for his guidance and kind suggestions at the starting stage of my research; thanks to Dr. Alan Wang and Dr. Kam Hung for their great support in the phase of data collection. Thanks to my fellow research mates in SHTM: Dr. Yong Chen, Dr. Lily Sun, Dr. Rosanna Leung, Emmy, and Louisa, and Daniel for the helpful discussions and bright ideas related to my research. I would also like to express my gratitude to my BoE Chair Dr. Karin Weber and my external examiners Prof. Honggang Xu from Sun Yat-sen University (China) and Dr. Bing Pan from the College of Charleston (US). Their comments are valuable for me to further improve my thesis.

My appreciations also goes to researchers from oversea universities and colleagues who provided patient guidance and direction on the improvement of my research design during the period of attachment programme in the United Kingdom. My special thanks to Prof. Gang Li from the University of Surrey, and Prof. Robert Filds, Dr. Sven F. Crone, Dr. Fotios Petropoulos, Dr. Nikolaos Kourentzes, and Dr Gokhan Yildirim from Lancaster University.

Furthermore, thanks to my beloved friends' support during my Ph.D. life. Special thanks to Ms. Baijing Song, Mr. Louis Shih, and Dr. Eve Ren who have been giving their consistent support in many different ways.

Lastly, I would like to express my deepest appreciation to my wife, my parents, and all relatives who shared their love and wholehearted support in my four-year research period. You are truly my spiritual pillars and this thesis is for all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATION

Abbreviation	Meaning	Page
ADL	Autoregressive Distributed Lag model	12
AIC	Akaike Information Criterion	119
ANOVA	Analysis of Variance	132
APE	Absolute Percentage Error	123
ARIMA	Autoregressive Moving-Average model	9
CF	Corrected judgmental Forecast	200
DBMS	Database Management System	10
DSR	Design Science Research	21
DSS	Decision Support System	3
ES	Exponential Smoothing	7
ETS	An automatic exponential smoothing approach	115
FSS	Forecasting Support System	II
GDC	General Design Cycle	21
GUI	Graphical User Interface	8
HKTDFS	Hong Kong Tourism Demand Forecasting System	II
HR	Human Resource	41
IE	Internet Explorer	148
IS	Information Systems	21
MAPE	Mean Absolute Percentage Error	12
MdAPE	Median Absolute Percentage Error	5
OLS	Ordinary Least Squares	124
PC	Percentage Change	94
PE	Percentage Error	94
PF	Performance Feedback	99
RBEFS	Rule-Based Expert Forecasting System	4
RMSPE	Root Mean Square Prediction Error	12
SFTIS	Statistical and Forecasting Tourism Information System	4
TDFSS	Tourism Demand Forecasting Support System	II
TIFIS	Theta Intelligent Forecasting Information System	4
TourMIS	Tourism Marketing Information System	4
UNWTO	UN World Tourism Organization	114
VAR	Vector Autoregressive model	12

1 INTRODUCTION

1.1 Background of the study

Tourism demand forecasting is a major research topic in academia and the tourism industry. From a microeconomic perspective, accurate tourism demand forecasts can help tourism businesses effectively establish investment and marketing strategies. For example, forecasts of tourist arrivals are important for hoteliers, tour operators, and airline companies because of the perishability of tourism products or services such as unused hotel rooms and unfilled airline seats, which cannot be stockpiled for future use (Archer, 1987). From a macroeconomic perspective, tourism demand forecasting is essential for the government and the tourism industry in formulating tourism development policies at the regional, national, or even global level. For example, accurate tourism demand forecasts can help local or national government to achieve its full tourism potential, as well as maximize the potential contribution of tourism to employment, small business development, income, and foreign exchange earnings (Burger, Dohnal, Kathrada, & Law, 2001). As stated by Song, Witt, and Zhang (2008), accurate tourism demand forecasts are crucial for all practitioners in the tourism industry.

Studies on tourism demand forecasting in recent decades have mainly focused on the development of quantitative forecasting techniques and their application in practice (Petropoulos, Patelis, Metaxiotis, Nikolopoulos, & Assimakopoulos, 2003). Numerous time series and econometric modeling and forecasting methods have been developed, improved, and tested in the past few decades. A comprehensive literature review conducted by Li, Song, and Witt, (2005) presented about 420 studies on this topic published between 1960 and 2002. One of the most recent review studies in tourism demand forecasting was conducted by Song and Li (2008). They reviewed 121 studies on tourism demand analysis and forecasting techniques published between 2000 and 2006. All of the reviewed studies focused on methodological developments, forecast competition, combination, and integration. According to their study, 38 methods covering time series models, causal econometric models, and other quantitative forecasting models were applied in tourism demand forecasting.

However, researchers face two problems in tourism demand forecasting: first, selecting the optimal model among various statistical forecasting methods or combining several of these methods is a complex task because tourism demand is sensitive to various special events; and high uncertainty is usually involved in tourism demand forecasting. Second, tourism demand is highly sensitive to a group of influencing factors, especially to certain special events (such as financial crises, the World Cup, the Olympic Games, etc.). Some foreseeable special events contribute to the irregular component of forecasts and cannot be captured by statistical forecasting methods; nevertheless, they can be captured by forecasters, who are able to use their knowledge and expertise in the domain (Fildes, Goodwin, & Lawrence, 2006). To overcome the limitations of statistical forecasting techniques, researchers have also integrated statistical forecasting techniques with judgmental forecasting techniques (Mules & McDonald, 1994; Tideswell, Mules, & Faulkner, 2001). Statistical and judgmental forecasting methods are usually integrated by adjusting statistical forecasts based on forecasters' domain knowledge, which has been proved to effectively improve the accuracy of tourism demand forecast (Baggio & Corigliano, 2008; Tideswell et al., 2001). With the help of advanced information technologies, a new research area in the integration of statistical and judgmental forecasting techniques, the forecasting support system (FSS), has been developed. This research area has expanded rapidly in recent decades and is expected to satisfy the two spheres of tourism demand forecasting. As a particular type of decision support system (DSS), FSS involves the statistical forecasting model and the procedures for forecasters' judgmental forecasting (Fildes et al., 2006). The adoption of FSS in forecasting tasks aims to provide forecasters or managers with effective solutions based on accurate forecasts (Keen & Morton, 1978). Drawing on Armstrong's (2001) definition, FSS is a set of procedures (typically computer-based) that allow forecasters to easily access, organize, and analyze a range of information. It also enables forecasters to incorporate judgmental adjustment, so that statistical and judgmental forecasting techniques are expected to be integrated more effectively in FSS. As mentioned by Croce and Wöber (2011), FSS is particularly meaningful (i) to facilitate access to data that is relevant to the forecast, (ii) to provide quantitative techniques for forecasting, (iii) to allow the storing of both statistical and judgmental forecasts, (iv) to provide feedback on forecast accuracy, and (v) to store information about user behavior in using FSS.

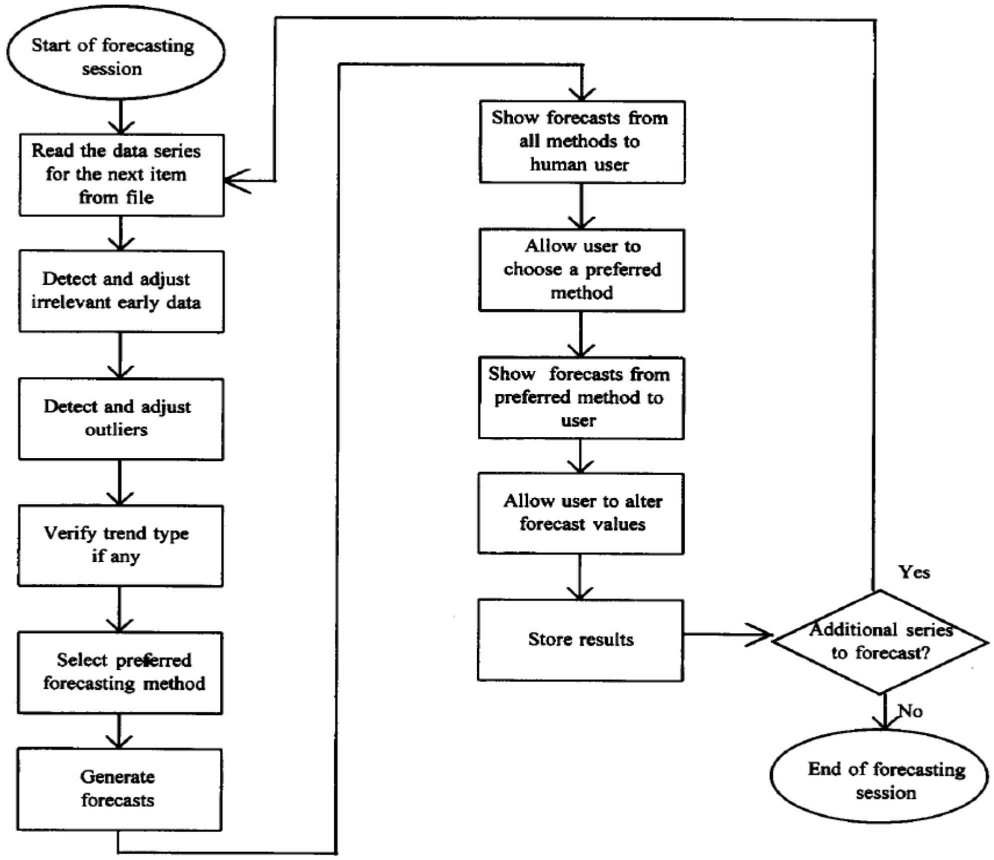
From the perspective of application, FSS can be classified as general system and specific system (Nikolopoulos & Assimakopoulos, 2003). The former is focused on general market demand forecasting, which mainly relies on general market research information, such as macroeconomic statistics; the latter is designed for specific organizations' forecasting needs. Forecasting using specific FSS mainly relies on organizations' internal market information (e.g., internal information for executive boards and customer behavior) and specific external market information (e.g., markets and environment, competition in the industry). Therefore, the design features of FSS and the processes of using FSS differ between these two types of FSS. General FSS can be further categorized according to the utilization area. For example, some general FSS are designed without any industrial features; these FSS can be used to forecast market demand in different industries. In contrast, some FSS are designed with specific industrial features; these FSS can be used to forecast market demand in certain industries. The current study focuses on general system development in the area of tourism demand forecasting, so the term "industrial FSS" is used in this study to reflect industrial-based general FSS in order to avoid confusion with general systems used without an industrial base.

1.2 Features of typical FSS

In order to identify the features of recently developed industrial FSS in the field of tourism demand, five of the most up-to-date and widely adopted systems are reviewed in this section. The first two systems reviewed are general systems that can be used in tourism demand forecasting: Vokurka, Flores, and Pearce's (1996) Rule-Based Expert Forecasting System (RBEFS), and Nikolopoulos and Assimakopoulos' (2003) Theta Intelligent Forecasting Information System (TIFIS). The other three systems are industrial FSS specifically designed for tourism demand forecasting: the web-based Statistical and Forecasting Tourism Information System (SFTIS), the Tourism Management Information System (TourMIS), and the Hong Kong Tourism Demand Forecasting System (HKTDFS).

Following Collopy and Armstrong (1992) and Pearce's (1995) studies, RBEFS contains a set of automatic functions for statistical modeling and forecasting, including detection and adjustment of outliers and irrelevant historical data; detection of the functional form and various statistical characteristics of the target variable; model selection and parameter estimation. Each function can be automatically processed by the system with pre-defined rules; then the results are further analyzed and adjusted by forecasters if necessary. If this is case, the component of statistical forecasting in RBEFS is entirely automated based on the characteristics of historical data (Figure 1-1). Empirical studies on the accuracy of data feature detection have observed differences between RBEFS and human judgment. Further examination of forecast accuracy was carried out in an experiment using 126 annual time series from the M1-competition data[1]. The median absolute percentage error (MdAPE) of the experiment revealed that the forecasts automatically produced by RBEFS cannot outperform the forecasts produced by the user adjustment process (Vokurka et al., 1996).

[1] The M1-competition compared forecasts from a table of techniques for 1001 real life time series of economic and financial indicators (micro, macro and demographic), including 181 at the annual frequency, 203 at the quarterly frequency, and 617 at the monthly frequency.

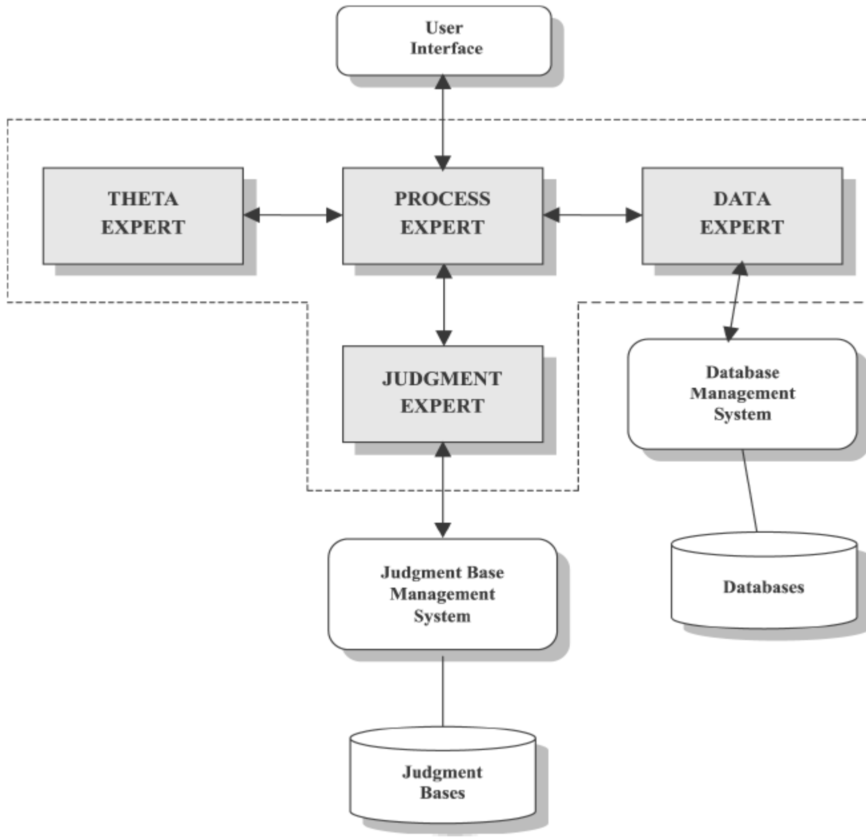


Source: Vokurka et al. (1996, p500)

Figure 1-1 The system architecture of RBEFS

The core parts of TIFIS have been further packaged into four "EXPERTS" (Figure 1-2). PROCESS EXPERT is designed to maintain the problem solving scheme and controls other components; "DATA EXPERT" and "THETA EXPERT" are components related to statistical forecasting, covering database management and statistical forecasting, respectively. "JUDGMENT EXPERT" is designed to provide judgmental adjustment (Lee & Yum, 1998). Compared with traditional FSS, the advantage of TIFIS is its adoption of the Theta method in statistical forecasting, which performs best in M3-competition[2] (Makridakis & Hibon, 2000). Furthermore, the statistical forecasting process in TIFIS is totally automated; judgmental forecasting is also supported by ARBA, a semi-automated function to automatically identify outliers, level shifts, changes in basic trends, and unusual latest observations (Adya, Collopy, Armstrong, & Kennedy, 2001). According to the result of an experiment using 3,003 series from the M3-competition data, TIFIS performs well in forecasting monthly and microeconomic data (Nikolopoulos & Assimakopoulos, 2003); and performs comparably to simple exponential smoothing (ES) with drift in annual time series forecasting (Hyndman & Billah, 2003).

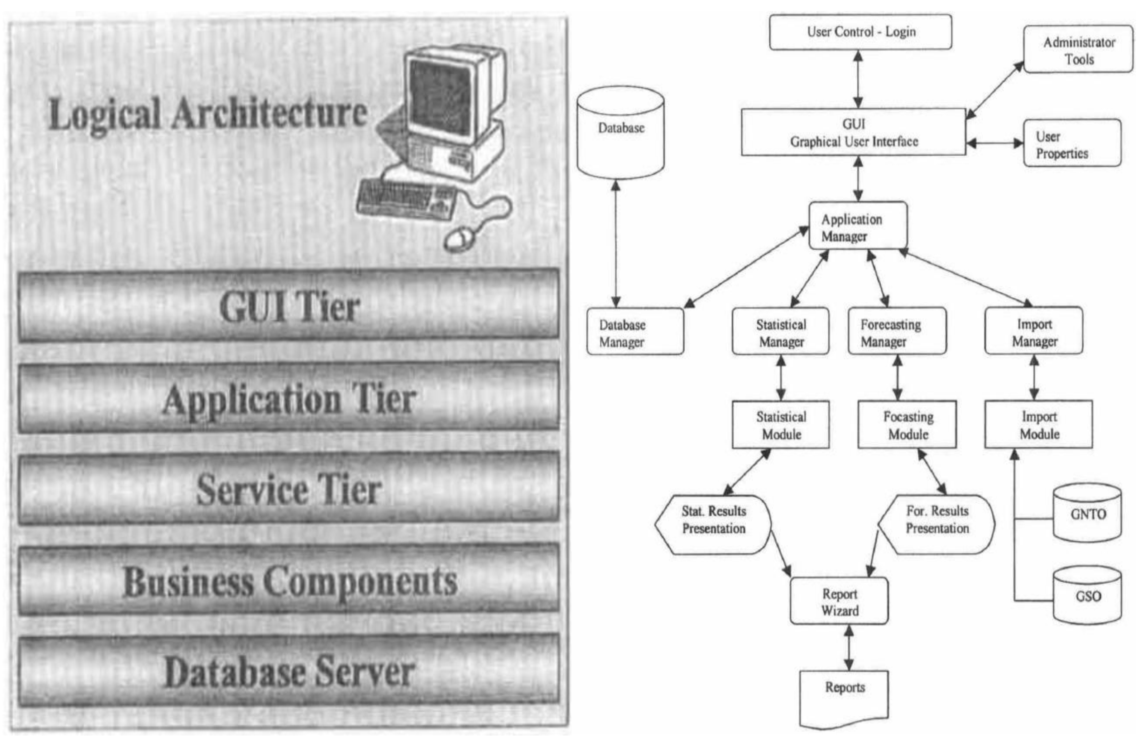
[2] M3-competition analyzes 24 forecasting methods based on 3003 real life time series.



Source: Nikolopoulos & Assimakopoulos (2003, p714)

Figure 1-2 The system architecture of TIFIS

SFTIS adopts a five-tier architecture, which contains a Graphical User Interface (GUI) tier, an application tier, service tier, business tier, and database server (Figure 1-3(a)). Four general features of FSS have been further divided into specific application components. These components are deployed in different tiers of the system, while seven core components of the system are deployed in the application tier (Figure 1-3(b)). The database manager is the connection between a database and specific applications requested by the application manager. The statistical manager and modules are in charge of historical data analysis and statistical modeling; then statistical forecasts are generated by the modeling results and historical data in the forecasting manager and modules. As highlighted by Petropoulos et al. (2003), a two-stage econometric modeling and forecasting method has been developed and applied to these four components, which is one of the two characteristics that most distinguish it from traditional FSS. The other is the design of import manager and module. These two components enable the system to communicate with external data sources and keep the internal database updated automatically. An empirical test using Greek inbound tourism data between 1980 and 1999 showed that the system forecast has better forecast accuracy than the four benchmark methods, including Naïve 1 and 2, ES, and Autoregressive Integrated Moving Average (ARIMA).

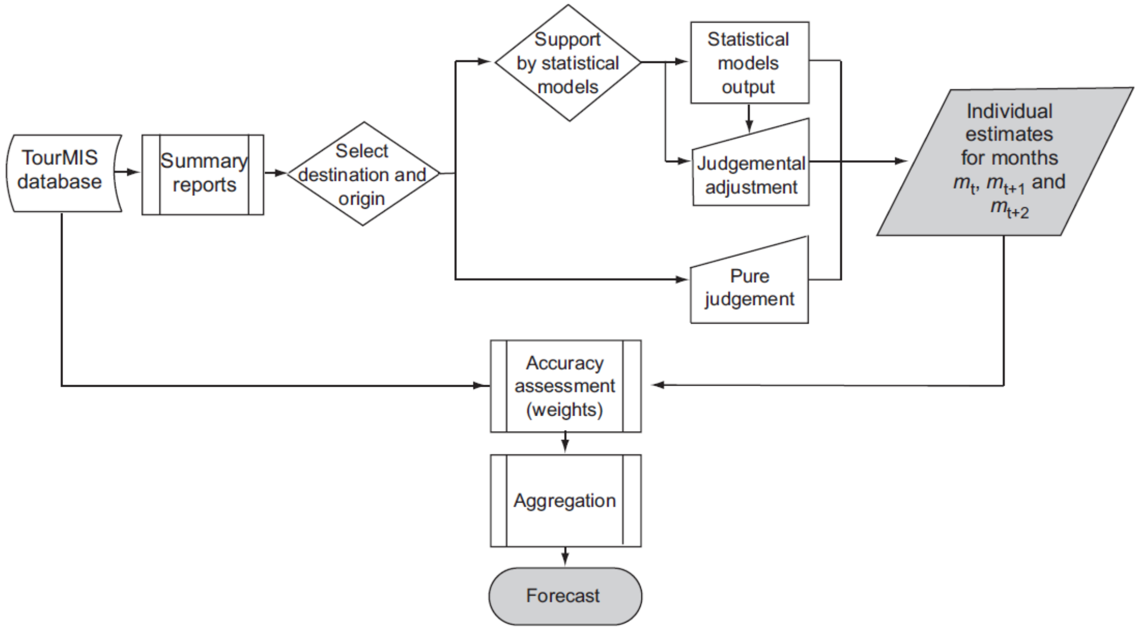


(a) (b)

Source: Petropoulos et al. (2003, p25-26)

Figure 1-3 The system architecture of SFTIS

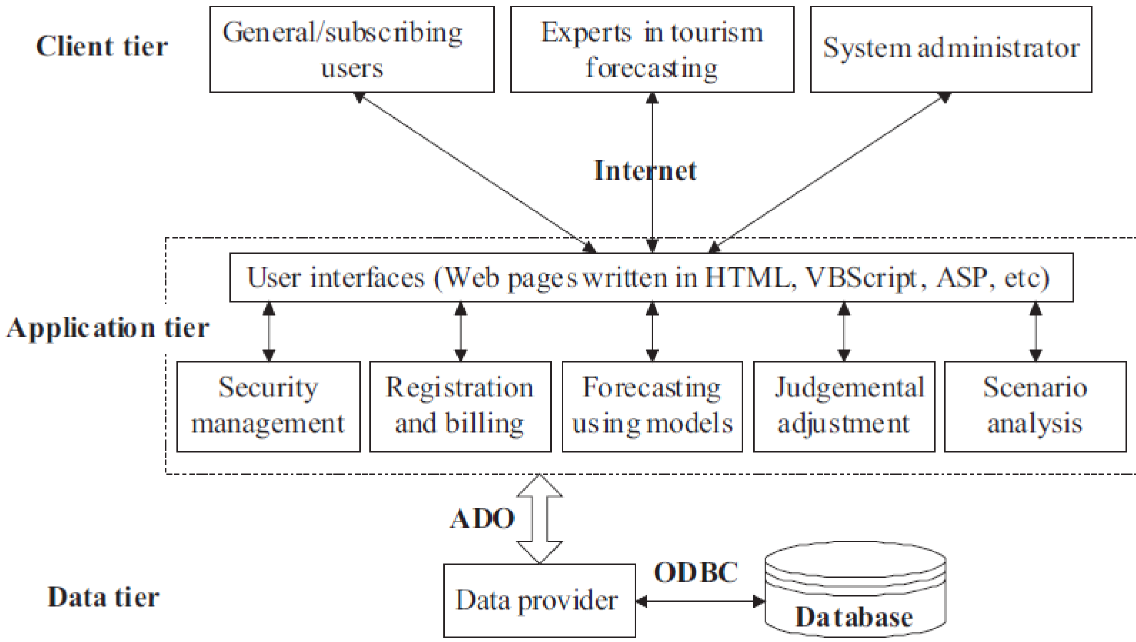
TourMIS is an industrial FSS designed for Austrian tourism demand forecasting. The first version of this system was designed in early 1986 (Mazanec, 1986). After several rounds of updates, the latest version was released in 2011 (Croce & Wöber, 2011). In TourMIS, four main components are strictly designed to implement the four features of FSS: (i) a database, supported by a SQL database management system (DBMS); (ii) two statistical forecasting models (Naïve 2 and Winters' ES); (iii) components for both pure judgmental forecasting and judgmental adjustment of statistical forecasts; and (iv) components for comparing forecasts with real outcomes (Table 1-4). Besides the four main components, some notable components are also designed to supplement forecasters' forecasting process. First of all, the characteristics of data used for forecasting are first clarified in the summary reports in order to help forecasters better understand the historical data. Second, Delphi survey technologies have been applied to judgmental forecasting with the aim of reducing individual forecasting errors. Third, comparing results between forecasts and real outcomes for each individual forecaster is further quantified as the accuracy of their adjustments, which are used as weights in forecast aggregation to generate the final forecasts. According to Wöber (2003), TourMIS is widely used by domestic and foreign governments, tourism research organizations, accommodation suppliers, tour operators, travel agencies, restaurants, private persons, and other tourism industry stakeholders. Unfortunately, the system performance of TourMIS has received only limited research interest.



Source: Croce & Wöber (2011, p718)

Figure 1-4 The architecture of TourMIS

Another notable FSS in tourism demand forecasting is HKTDFS, which was first released in 2008 and has been updated several times (www.tourismforecasting.net). This system is designed with the purpose of forecasting Hong Kong tourism demand, indicated by tourist arrivals, tourist expenditures, and demand for hotel rooms. The potential users of HKTDFS include government offices responsible for tourism policymaking, business executives in tourism-related sectors, planning and marketing agencies, consultancy firms focused on tourism, and various tourism-specific research/education organizations (Song et al., 2008). As the latest system in tourism demand forecasting, many advanced information technologies and forecasting methods are applied in HKTDFS. First of all, it is a web-based FSS with all functional components deployed in three tiers: client tier, application tier, and data tier (Figure 1-5). Since all functional components are developed in the application tier, users from the client tier will not be bothered by system installation and updates. Furthermore, advanced econometric modeling techniques are applied in the statistical forecasting component. The database is able to provide historical data of all tourism demand indicators, as well as the historical data on the influencing factors of Hong Kong tourism demand (such as income and price variables), and the information on the special events that significantly influence the demand for Hong Kong tourism. Based on the rich database, the system is able to perform complex econometric modeling and forecasting exercises, such as the Vector-autoregression (VAR) and Autoregressive Distributed Lag (ADL) models. Finally, HKTDFS is able to provide scenario analysis based on the estimated parameters in the econometric models. Scenario analysis extends the forecasting ability of HKTDFS by providing both point and interval forecasts, and helps forecasters to further predict the potential situation of the tourism market. According to the report published by HKTDFS in 2011, quarterly forecasts of tourist arrivals to Hong Kong from 14 main source markets between 2010 quarter 4 and 2011 quarter 2 are fairly accurate with an average of mean absolute percentage error (MAPE) around 5.21% and an average Root Mean Square Prediction Error (RMSPE) around 6.40% (http://www.tourismforecasting.net/ hktdfs/home/project/newsDetail.jsp?id=d6c64bec-89a1-4d54-9aab-3da2d7de0547).



Source: Song et al., (2008, p457)

Figure 1-5 The system architecture of HKTDFS

According to the features of the five FSS reviewed above, it can be suggested that a general FSS should be structured around four key components: database, statistical forecasting, judgmental adjustment, and comparison of forecasts and real outcomes at regular intervals. According to Figure 1-6, the database is used to store all necessary data for forecasting, including the historical data of the target variables and their influencing factors, the historical forecasts, and domain knowledge. The database is also the foundation of the three other key components. In statistical forecasting, historical data is used to produce forecasts using quantitative forecasting techniques, such as ES techniques (Hyndman & Khandakar, 2008), ARIMA models, and causal econometric techniques (Song & Witt, 2000). The results of statistical forecasting provide baseline forecasts for the next key component of FSS – judgmental adjustment. Compared with statistical forecasting, judgmental adjustment is usually designed as a qualitative forecasting process that aims to improve forecast accuracy using forecasters' domain knowledge (Fildes et al., 2006). According to the forecasting literature, the application of user-system interaction mainly concentrates on this component (Asimakopoulos, 2008; Fildes et al., 2006). F-R comparison is used to compare historical forecasts with real outcomes when multi-step ahead forecast is conducted. This component is developed in FSS to further adjust forecasts when new real outcomes on the target variable are available (Fildes et al., 2006; Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009).

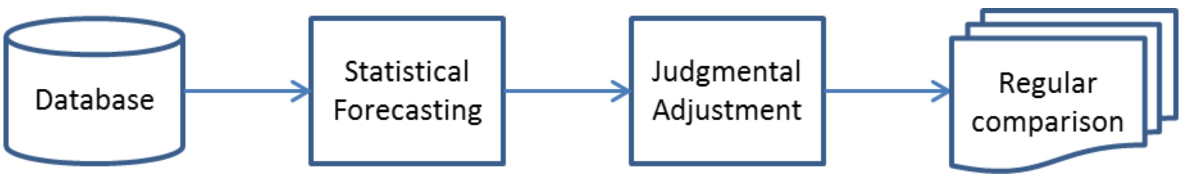


Figure 1-6 General architecture of FSS

1.3 Problem statement

Overall, the design features of FSS are reflected by its functional components. The developments of the functional components vary from one FSS to another; a variety of system architectures have also been proposed, as we can see from the five FSS reviewed. From developers' perspective, FSS is an integration of technologies; its functional components are therefore considered enablers for forecasters' use (Asimakopoulos & Dix, 2013). In that sense, developers pay much attention to the enhancement of systems' functionality by adopting various technologies. However, the performance of system functionality does not only depend on the technology, but also on the process of use in practice. According to recently developed FSS and related studies, the process of using

FSS has been neglected or it is simply assumed that an ideal process is used. However, FSS usage in practice often departs from such ideals (Fildes et al., 2006).

1.3.1 Ideal vs actual use of FSS

Many studies reveal that mismatches always exist between developers' understanding of how an FSS should be used and forecasters' actual use; and these mismatches significantly influence the implementation of systems' functionality (Asimakopoulos & Dix, 2013; Goodwin, Lee, Fildes, Nikolopoulos, & Lawrence, 2007). In statistical forecasting, for example, some FSS provide a group of models for forecasters to select, or ask forecasters to define the parameter values of a specific model (Lawrence, Goodwin, & Fildes, 2002; Wöber, 2003). From FSS developers' point of view, forecasters' expertise is valuable for model selection and parameter identification; therefore, more user participation in statistical forecasting can benefit forecast accuracy. This understanding depends on the ideal condition that forecasters are skillful in statistical forecasting. However, empirical studies have revealed that forecasters, especially those from the industry, always lack training in forecasting techniques (Fildes & Beard, 1992; Fildes et al., 2006). They often select the default parameter values or sub-optimal models to produce statistical forecasts, and make large judgmental adjustments in order to make forecasts more reliable. Unfortunately, unreliable statistical forecasts with large-scale adjustment usually perform poorly compared with those produced based on well-fitting statistical forecasts in the first place (Goodwin, Fildes, Lawrence, & Nikolopoulos, 2007). In addition, judgmental adjustment would partially duplicate the role of statistical forecasting when forecasters lack formal training in statistical methods. On the one hand, some FSS adopt econometric modeling techniques in statistical forecasting, and the influencing factors involved in the econometric modeling process have already been counted in statistical forecasting process. If forecasters have a poor understanding of the econometric modeling procedures, there is a risk that the influencing factors could be double counted in the judgmental adjustments (Fildes et al., 2006). On the other hand, the danger of double counted bias also exists when a regression model is used to produce statistical forecasts. That is, if a certain excluded variable is collinear with another variable that has been used in the model, the latter would act as a proxy for the former. This means that, to some extent, the effects of the excluded variable would already have been taken into account by the regression model and further adjustment for the effects of the excluded variable would be double counted (Goodwin, 2002).

Besides over-optimism about forecasters' skills, some studies have also revealed that existing FSS are designed upon the premise that forecasters play an ideal role in the entire forecasting process. According to Fildes et al. (2006), a typical demand variable to be forecast consists of three components: regular patterns based on the historical data or on its relationships with the influencing factors; irregular components arising from foreseeable interruptions; and unpredictable noise. Theoretically, statistical modeling and forecasting techniques should be used to predict the regular patterns. Irregular but foreseeable interruptions that cannot be captured by statistical methods would be captured by forecasters using their domain knowledge. It is therefore not necessary for forecasters to predict regular patterns; instead, they should fully participate in identifying foreseeable events that significantly influence the target variable. In addition, they should adjust forecasts based on reliable information of such interruptions in the judgmental forecasting process. However, in practice, forecasters always go beyond their ideal role when using FSS. According to Goodwin and Fildes (1999), forecasters attempt to forecast regular patterns as well as noise. Lim and O'Connor (1995) and Sanders and Ritzman (2001) also pointed out that people usually get confused with these two components and make unnecessary or damaging adjustments to the statistical forecasts.

Furthermore, some characteristics of forecasters also make it difficult to operate FSS in an ideal way. First, forecasters' interest is usually confined to a few recent observations, which are mainly determined by short-term patterns, and some current ongoing or recently completed events. When forecasters are able to select the statistical model, they attempt to choose the model that fits well with the short-term trends and ignores the long-term trends. As explained by Goodwin et al. (2007), recalling many events and circumstances that shaped history would put too great a load on memory; therefore there would be a natural bias towards a few recent observations. Lim and O'Connor's (1995) study further revealed that such limitation arises from various human information-processing inadequacies.

Second, forecasters usually use extra-model information in judgmental forecasting without carefully assessing its reliability. An experiment conducted by Goodwin (2000b) revealed that adjustments based on untrustworthy information significantly damage forecast accuracy. According to Goodwin et al. (2007), forecasters in a pharmaceutical company attempted to make adjustments on statistical forecasts even if they did not possess any extra information. When unreliable information is available, forecasters usually make small changes to statistical forecasts in order to hedge their bets; however, empirical studies reveal that such small changes damage the accuracy of adjustment in most cases (Fildes et al., 2006; Fildes et al., 2009).

The third typical characteristic of forecasters (especially the experts in a domain) is that they are overconfident (Arkes, 2001; Fildes et al., 2006). McNees (1992) examined the accuracy of economic forecasts made by 22 economists over a period of 11 years. A summary of the forecast accuracy in previous rounds was given to the economists in each forecasting round. As a result, these economists were continuously overconfident in their adjustments, even when warned in advance against overconfidence. In Goodwin and Fildes' (1999) study, forecasters apparently tended to base the entire forecast on judgmental forecasts. Statistical forecasts were completely ignored, even when the regular pattern of the underlying time series was accurately predicted. This tendency is considered the illusion of control in decision-making literature (Kottemann, Davis, & Remus, 1994). In Goodwin and colleagues' (2007) study, for example, forecasters in supply-chain companies frequently adjusted statistical forecasts in order to establish their ownership of both the forecasts and the forecasting process. Ashton (1990) explained this flawed cognitive mechanism as self-efficacy: "a tendency to overestimate one's own ability and the poor relation between self-assessments of ability and actual performance could contribute to subjects' reluctance to rely heavily on decision aids" (p. 163).

1.3.2 Debiasing

Drawing on the above discussion, the advanced functional capabilities of FSS are usually used ineffectively in practice because of forecasters' cognitive bias. Such bias is unavoidable in the process of user-system interaction. Therefore, one of the topics in future study of FSS development should be the improvement of the effectiveness of the system usage by debiasing forecasts in FSS usage. The term "debiasing" in this sense refers to a means of systematically overcoming forecasters' judgmental bias. From this perspective, there are two major research gaps in improving the effectiveness of FSS, which have been pointed out by Fildes et al. (2006): when to intervene judgmentally, and how to carry out such interventions. Fildes et al. (2006) also suggested that, compared with restrictiveness, decision guidance is a broad approach to effectively calibrate forecasters' judgmental bias. Regarding the DSS literature, decision guidance refers to the degree to which, and the manner in which, the system guides its users in constructing and executing the decision-making process by assisting them in choosing and using its operators (Silver, 1991). Guidance can be delivered informatively or suggestively. Informative guidance enables FSS to provide unbiased and relevant information without offering suggestions, while suggestive guidance enables FSS to suggest forecasters with one or several actions to be executed next.

In previous studies, informative guidance has been found to be effective when applied in decision-making and the design of DSS (Montazemi, Wang, Khalid Nainar, & Bart, 1996; Singh, 1998). However, informative guidance has only been found to be effective in statistical model selection (Gönül, Önkal, & Lawrence, 2006; Parikh, Fazlollahi, & Verma, 2001); its application in forecast calibration has only been studied twice in the FSS literature (Goodwin & Fildes, 2001; Goodwin, Fildes, Lawrence, & Stephens, 2011). In the design of tourism demand FSS, informative guidance is only available in HKTDFS. It is used to explain the performance of statistical models by displaying core diagnostic statistics. With this informative guidance, forecasters are able to make decisions on whether statistical forecasts are reliable. However, there is no study on the effectiveness of information guidance in debiasing forecasters' judgmental adjustments to be found in the literature of tourism demand FSS development.

Similar to information guidance, suggestive guidance has been proved to be effective in the design of DSS (Montazemi et al., 1996; Singh, 1998). Some studies have also suggested that integrating informative and suggestive guidance can effectively improve decision-making, which is better than applying either of them separately (Gregor & Benbasat, 1999; Kasper, 1996; Singh, 1998). In FSS development, suggestive guidance was only verified to be effective in statistical model selection (Goodwin et al., 2011; Parikh et al., 2001). Yet, little study on the application of suggestive guidance in forecast calibration has been found in the literature. The integration of suggestive and informative guidance in judgmental adjustment is also left unexplored.

Having identified the limitations of previous research, this study focuses on evolving FSS design by incorporating the systematic debiasing mechanism in the module of judgmental forecasting (Figure 1-7). In the context of tourism demand forecasting, this study aims to enhance forecasters' ability to find appropriate times to conduct judgmental forecasts based on their domain knowledge, experience, and other supportive information; and to apply accurate judgmental interventions when appropriate.

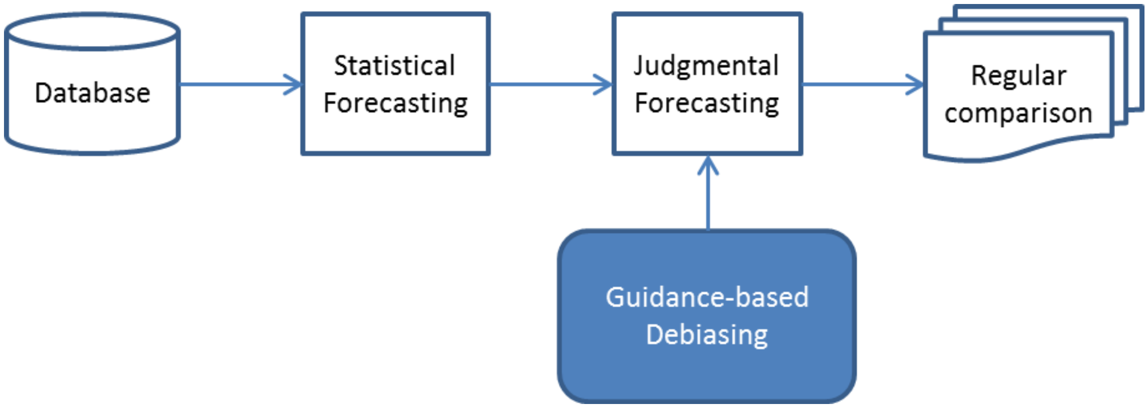


Figure 1-7 Guidance-based debiasing in judgmental forecasting of tourism demand

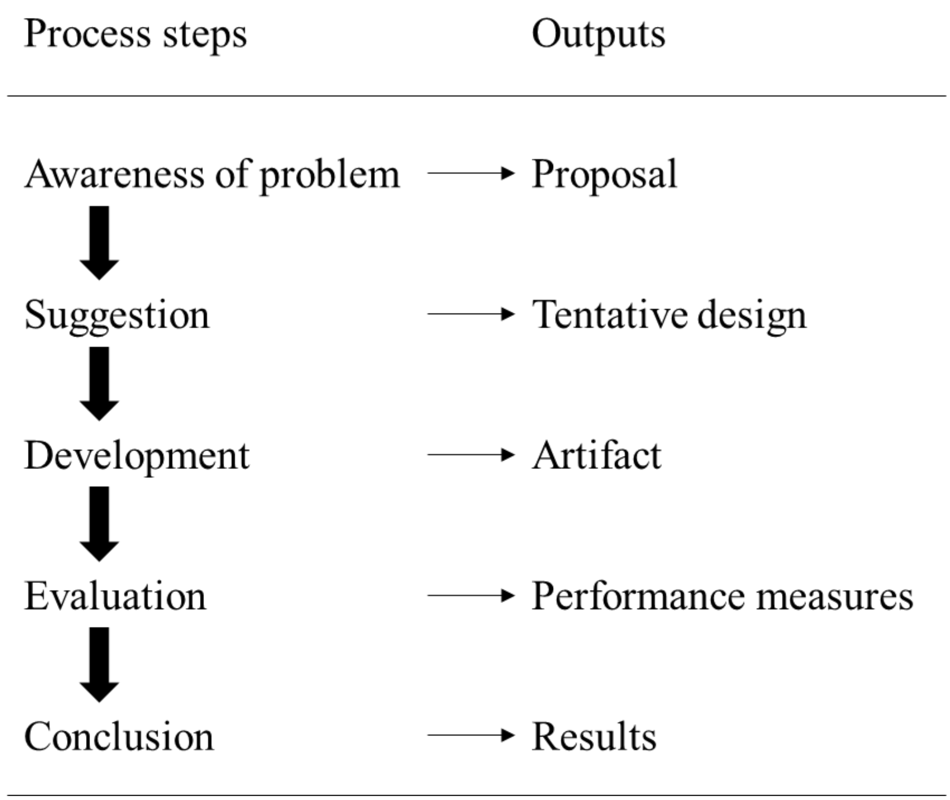
1.4 Research objectives

The research objectives identified in this study are:

- To identify the most common cognitive biases in judgmental forecasting literature and their contributions to judgmental forecast error;
- To propose a debiasing model that is able to effectively reduce judgmental forecast error caused by the common cognitive biases;
- To test the effectiveness of the proposed debiasing model in FSS in the context of tourism demand forecasting.

In order to achieve the identified research objectives, this study is conducted following a design science research (DSR) framework that has been widely adopted in the development of information systems (IS) and DSS (Arnott, 2006; Hevner & Chatterjee, 2010; Vaishnavi & Kuechler, 2004). This framework is based on Takeda, Veerkamp, and Yoshikawa's (1990) analysis of the general design cycle (GDC) and has been applied to DSR by Vaishnavi and Kuechler (2007). In this framework, the research process has been structured in five steps: awareness of problem, suggestion (solution), development, evaluation, and conclusion. The outputs in each step are proposal, tentative design, artifact, performance measures, and results, respectively (Figure 1-8).

Awareness of the problem. In the first step of the DSR framework, specific research problems and the motivation of research should be clarified. The output of this step is a general proposal with specific problems to be addressed, and the research objectives. It is also useful to atomize the research problems conceptually so that the solution can easily capture their complexity (Hevner & Chatterjee, 2010). In this study, awareness of the problem has already been addressed by the problem statements of recent studies on FSS development, which focused on the mismatches between the system artifact and the actual use of the system. Specifically, system developers' assumption of ideal system usage is not valid in practice; forecasters' cognitive bias may significantly affect judgmental adjustment of statistical forecasts and further damage the accuracy of the final forecast; and also the recently developed FSS artifacts lack effective debiasing strategies against forecasters' cognitive bias when judgmentally adjusting statistical forecasts.



Source: Hevner & Chatterjee (2010, p.27)

Figure 1-8 Design Science Research Framework

Suggestion (Solution). The suggestion step aims to propose the solutions that are possible and feasible to solve the identified problems. The resources required in this step include knowledge of the state of the problems, rational solutions, and possibly their efficacy (Hevner & Chatterjee, 2010). The output in the suggestion step, a tentative design, is constructed upon the newly proposed functionalities and is intimately connected with the proposal produced in the first step. In the current study, this step starts with a comprehensive review of the literature relating to the theories of cognitive bias, debiasing, decision guidance, and background knowledge of tourism demand forecasting. Rational solutions are then generated from the perspective of these four groups of knowledge. In order to avoid any confusion with “suggestive guidance” in the guidance-based forecast debiasing model, this step is henceforth referred to as “Solution.”

Development. The tentative design is further developed in the Development step and most of the actual design takes place here. Creativity is required when combining existing knowledge and well-defined problem definitions into an artifact in order to solve problems. The generated artifact of DSR may be rather abstract in nature, such as constructs, models, or methods (March & Smith, 1995). In this study, a conceptual model of guidance-based forecast debiasing and accompanying hypotheses are developed based on the relevant theory, research, and the problem domain in the Development step.

Evaluation. Once the artifact is developed, it is necessary to evaluate it according to the criteria identified in the research objectives. This evaluation is conducted by testing the hypotheses developed and the conceptual model in order to identify the extent to which the research objectives are achieved. The results of hypotheses testing reflect the performance of the artifact and are the output of the Evaluation step. Furthermore, deviations from expectations, as well as the additional information gained in the Development step, are used as feedback to revise the tentative design and the artifact. Depending on the nature of the artifact, evaluation can be carried out in many forms, such as action research, controlled experiments, simulation, case study, proof, or other appropriate methods (Hevner & Chatterjee, 2010; Petter, 2007). In the current research, an experiment is conducted to evaluate the effectiveness of the guidance-based forecast debiasing model in the context of tourism demand forecasting, due to its high internal validity and control (Petter, 2007; Whitley, 1996). The Hong Kong tourism market is selected as the research context and the tentative model of forecast debiasing is applied to improve the design of HKTDFS. This step focuses on the examination of the effectiveness of the proposed conceptual model to reduce forecasters' cognitive bias when using the revised HKTDFS for tourism demand forecasting.

Conclusion. The final step of the DSR framework is to summarize the knowledge developed from the study. In the current study, the conclusion reflects the experiment results back to the problems and concludes whether the model can effectively reduce forecasters' cognitive bias in judgmental forecasting. The deviations from expectations are used to revise the tentative design and the original debiasing model. The specified research framework in this study is shown in Figure 1-9.

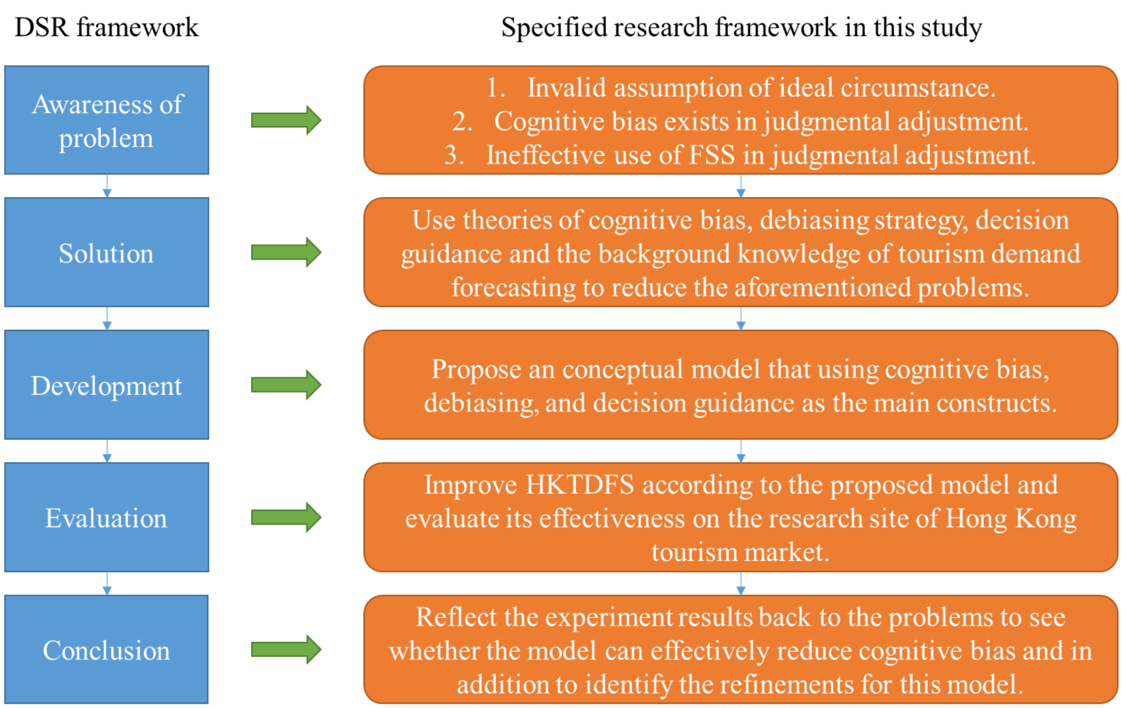


Figure 1-9 Research framework

1.5 Significance of the study

1.5.1 Theoretical contribution

Traditional FSS research usually considers FSS as a technological artifact. The development of FSS mainly focuses on the innovation or the improvement of its functional capability. For this purpose, a variety of advanced methods in statistical and judgmental forecasting have been adopted in the design of FSS with the presumption that FSS would be used in ideal conditions. However, the forecast effectiveness of FSS can be significantly affected by forecasters' judgmental bias. Recent studies have highlighted that the ideal assumption is far from reality, due to the dynamic circumstances and forecasters' different backgrounds (Asimakopoulos & Dix, 2013; Fildes et al., 2006; Goodwin et al., 2007; Orlikowski, 2000). Advanced functional capability cannot improve forecast accuracy if forecasters' judgmental bias is not calibrated. Therefore, forecast debiasing is crucial for improving the effectiveness of system usage. This study focuses on improving the capability of forecast debiasing in FSS development; in particular, to enhance forecasters' ability to recognize the appropriate time for adjustment and to conduct accurate interventions.

Furthermore, decision guidance is one of the general approaches that has the potential to improve the forecast effectiveness of FSS (Fildes et al., 2006), and to effectively support users' decision-making. In FSS development, the usefulness of decision guidance is only examined in statistical modeling and forecasting; however, there is a lack of recent literature proposing a systematic approach of decision guidance in forecast debiasing. The current study aims to fill this gap by proposing a guidance-based forecast debiasing model in FSS design.

Finally, the forecast debiasing model is proposed separately from other features of FSS. Therefore, it can be easily embedded into any existing FSS framework or applied in any innovative FSS framework in the future.

1.5.2 Practical contribution

The effectiveness of the proposed conceptual model of forecast debiasing in judgmental forecasting is tested by a FSS prototype, which is developed based on the HKTDFS platform. The experiment of this study is grounded upon Hong Kong tourism demand forecasting. Therefore, the verified conceptual model can be used to improve the design and forecast accuracy of HKTDFS. Furthermore, the guidance-based forecast calibration model is proposed separately from other features of FSS. The application of this model is not only limited to the design of tourism demand FSS; it can also be applied in other FSS developments that contain the feature of judgmental forecasting.

1.6 Structure of this thesis

According to the DSR framework, the structure of this study can be mapped as Figure 1-10. Chapter 1 introduces the background of tourism demand forecasting and the problems of previous studies in the design of tourism demand FSS. According to the identified problems, the research objectives and their potential contributions in this study are further identified.

Chapter 2 provides an extensive literature review of understandings and solutions of the research problems in extant research. It starts from the theory of cognitive bias and those of its specifications that can significantly influence forecasters' decision-making in judgmental forecasting. Then the findings of previous studies about the forms of cognitive bias and how they occur in the context of tourism demand forecasting are summarized. Next, the theory and methods of debiasing that focus on the identified cognitive biases are reviewed, followed by the theory of decision guidance in FSS development.

Chapter 3 describes the conceptual model of forecast debiasing based on the theories and concepts reviewed from the literature. In order to achieve the research objectives, this model covers the concepts related to cognitive bias, debiasing, decision guidance, and tourism demand forecasting. The construction of this model is built upon a series of hypotheses, which need to be tested.

Chapter 4 is devoted to the methodology of hypotheses testing and the refinement of the conceptual model. A two-stage laboratory experiment is designed for data collection. The experiment is focused on Hong Kong tourism demand forecasting and a prototype of FSS according to the proposed conceptual model is developed using the platform of HKTDFS. A group of forecasters with experience either in FSS usage or in the area of tourism demand forecasting are invited to forecast Hong Kong tourism demand using the prototype. The forecasts and the forecasters' decisions made in each step of judgmental forecasting are observed and recorded.

Chapter 5 reveals the results of the hypotheses testing using both parametric and non-parametric statistical methods. The deviations from the hypotheses indicate invalid parts of the debiasing model, which mainly focuses on the hypotheses regarding the debiasing strategies of desire bias and overconfidence bias. Further discussion based on the testing results, as well as the revision of the proposed debiasing model, are provided in Chapter 6.

Chapter 7 clarifies the utility and novelty of the final debiasing model, as well as the rigor of its development. Further study deriving from the conclusion of the current study is also identified at the end of this chapter.

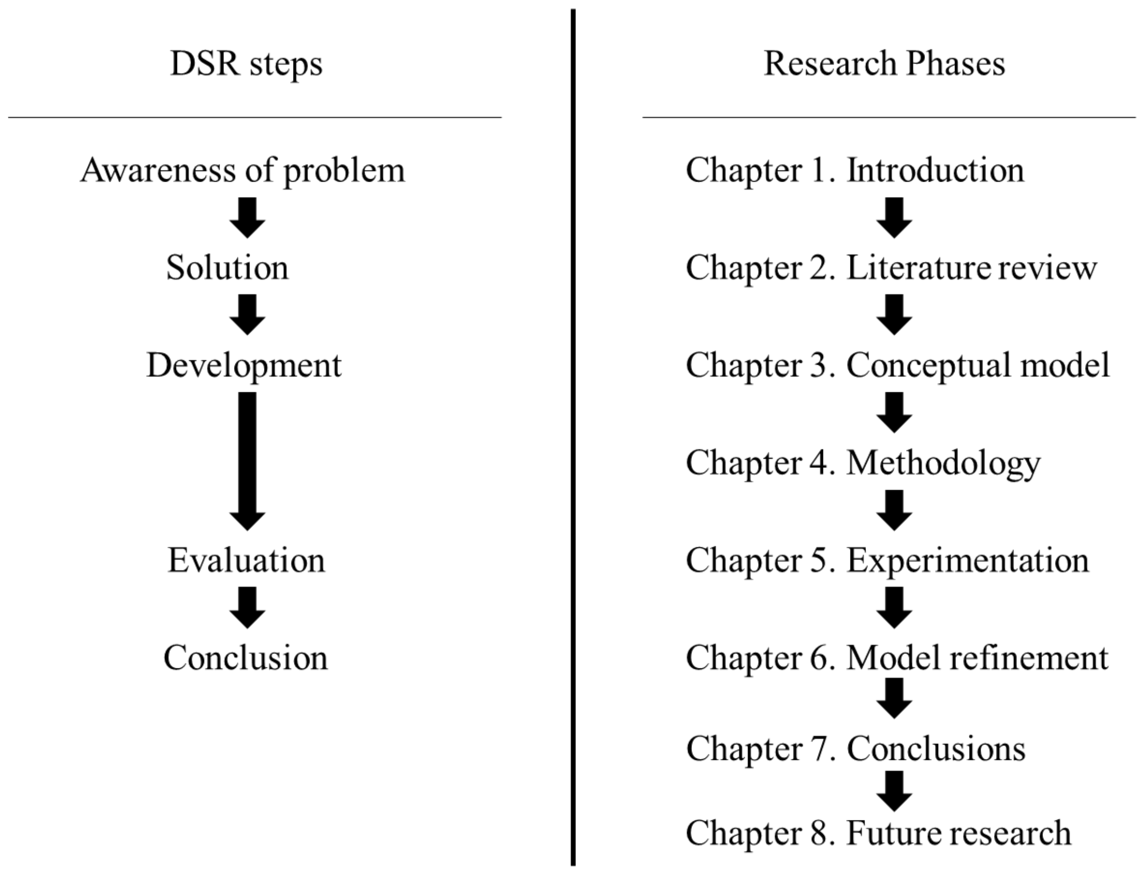


Figure 1-10 Research framework

1.7 Summary

This chapter first explained the background of tourism demand FSS and the features of five example designs. The problems of the existing FSS were then discussed, followed by the research gaps identified in the literature. In particular, the lack of a systematic approach to calibrating forecasters' judgmental bias was highlighted. Key areas of research interest fall into proposing a guidance-based forecast debiasing model in order to improve forecasters' ability to make accurate adjustment at appropriate times. Three research objectives were identified and the DSR framework was chosen to structure the current study. At the end of this chapter, potential contributions of this study were highlighted, followed by an outline of the structure of the study.

2 LITERATURE REVIEW

2.1 Description of cognitive bias

Cognitive bias, which is also called judgmental bias or decision bias, is cognitive or mental behavior that prejudices decision quality (Arnott, 2006). It is commonly viewed as predictable deviations from rational decision-making in DSS research (Arnott, 1998). Many cognitive biases have been identified in the literature of decision theory (Arnott, 2006; Ralph, 2011). Some early works, such as Tversky and Kahneman (1974), Remus and Kottemann (1986), and Hogarth (1987), provide the foundation upon which to identify cognitive biases, as well as to classify them from the perspectives of human heuristic, decision-making, and IS development. Based on these early works, recent studies (e.g., Arnott, 2006; Carter, Kaufmann, & Michel, 2007; Ralph, 2011) have further supplemented identification, as well as the taxonomy of cognitive bias.

2.1.1 Cognitive biases in Tversky and Kahneman's heuristic principles

The first codification of cognitive bias theory was published by Tversky and Kahneman (1974), who classified 12 cognitive biases according to the judgmental heuristic principles of availability, anchoring, and representativeness.

2.1.1.1 Availability

The heuristic principle of availability indicates that people evaluate the subjective probabilities of events by the degree to which similar events or instances are available in their memory. To illustrate, one can assess the probability of a new couple visiting the Maldives for their honeymoon by recalling such occurrences among one's acquaintances.

Similarly, one can assess the probability that a given tourism destination will move to the expansion stage in its life cycle by comparing the market share and the amount of visitors in certain similar destinations in the past (McKercher, 1995). However, availability is affected by factors other than probability or frequency. One such factor is the retrievability of instances (Taylor & Thompson, 1982). When the frequencies of several events are judged by the availability of their instances, the events whose instances are easily retrieved will appear more numerous than others whose instances are less retrievable. For example, in Tversky and Kahneman's (1973) study, four groups of subjects heard different lists of names of both sexes. The men in two of the lists were relatively more famous than the women; the opposite was the case in the other two lists. Then the subjects were asked to judge whether the frequency of male names was higher than that of female names in the list they heard. The subjects in each group erroneously judged that the sex with more famous names was the more numerous, even if the actual situation was quite the opposite.

The effectiveness of a search set is another factor that influences the availability heuristic (Bazerman & Moore, 2008). When judging the probability or the frequency of an event, people usually judge the frequency of possible circumstances in which such events could occur. However, search sets differ across different tasks, which may cause serious bias in judgment. For example, when asked whether there are fewer words starting with "r" than there are words with "r" as the third letter, people approach such a question by recalling words with both characteristics and assessing the relative frequency. Since it is relatively easier to think of words starting with "r," people usually think that words starting with "r" are more numerous. Indeed, the frequency of the possible circumstances in which an event may happen cannot indicate its frequency.

Sometimes one has to assess the frequency or the probability of an event whose instances are not in one's memory but can be imagined according to certain given rules. In this case, the imaginability of instances influences people's availability heuristic (Taylor & Thompson, 1982). Typically, people tend to generate several instances that can be easily found and assess the frequency or the probability of these instances. However, the ease with which the instances can be found does not always reflect the actual frequency of the event. Considering a group of 10 examiners who are qualified to form an examination committee, the size of which can be either three or eight members. Without computation, which size of the committee gains the most variety? According to the binomial coefficients C_3^{10} and C_8^{10}, any committee of eight members constructs a unique non-committee of two members, and the correct answer is the committee with three members. However, without calculation, people can only evaluate the frequency of two kinds of committee by mentally constructing committees with three and eight members. Committees with few members are more countable than committees of many members. So the large committees will appear more numerous than smaller committees if frequency is assessed by imaginability or accountability (Tversky & Kahneman, 1974).

Correlation between events is another factor that influences the availability heuristic. According to Tversky and Kahneman (1974), people's judgment of the frequency at that two events co-occur is typically based on the strength of the associative bound between them. One is likely to conclude that the events have been frequently paired if the association between them is strong or that strong associates between two events will occur together frequently. As a result, people have at their disposal the availability heuristic for estimating the frequency of co-occurrences according to the ease with which the relevant mental operations of retrieval, construction, or association can be performed. However, such estimation of correlation results in systematic errors. To illustrate, Chapman and Chapman's (1967) experiment provides evidence of the existence of bias when two events co-occur. They provided a group of subjects with information about several hypothetical mental patients, including individual clinical diagnosis and a drawing of a person produced by each patient. Then the subjects were asked to estimate the frequency with which each diagnosis (e.g., suspiciousness) had been accompanied by various characteristics of the drawing (e.g., peculiar eyes). The result showed that the subjects markedly overestimated the co-occurrence of suspiciousness and peculiar eyes. Such an illusory correlation effect was strongly resistant to contradictory instances, even when the correlation between symptom and diagnosis was actually negative.

2.1.1.2 Anchoring effect

When making estimates, people usually start from an initial value that can be achieved by a partial computation or the formulation of the problem. However, different starting points lead to different estimates, which indicates that adjustments based on initial values are insufficient (Slovic & Lichtenstein, 1971). Two demonstrations were given by Tversky and Kahneman (1974). In one, the subjects were asked to estimate the percentage of African countries in the United Nations by judging and adjusting a randomly selected starting point. The result showed that the median estimates were 25% and 45% when subjects received 10% and 65% as starting points, respectively. Thus, payoffs for accuracy have nothing to do with the anchoring effect when a starting point is given. Another interesting demonstration was conducted to illustrate that the anchoring effect also exists when a starting point is not available. Two groups of high school students were asked to estimate the results of two numerical expressions within five seconds. One expression was 8*7*6*5*4*3*2*1 and the other was 1*2*3*4*5*6*7*8. Since the time was limited, the subjects can only perform a few steps of computation, such as the multiplication of first three or four numbers in the expressions. Then they had to extrapolate and adjust the incomplete computation. Because the result of the first few numbers of multiplication is higher in the first expression (descending sequence) than in the second (ascending sequence), the median estimate for the former and the latter expressions were 2,250 and 512, respectively; the correct answer is 40,320.

Furthermore, the anchoring effect also exists in comparison of conjunctive and disjunctive events. According to Cohen, Chesnick, and Haran (1972), people tend to overestimate the probability of conjunctive events and underestimate the probability of disjunctive events. In Bar-Hillel's (1973) study, a comparison of simple and conjunctive events, and between simple and disjunctive events, further verified this conclusion. In her study, a simple event is designated as drawing a red marble from a bag containing 50% red marbles and 50% white ones; a conjunctive event is designated as successfully drawing a red marble seven times with replacement from a bag containing 90% red marbles and 10% white marbles; a disjunctive event is designated as drawing a red marble at least once in seven successive tries with replacement from a bag containing 10% red marbles and 90% white marbles.

The subjects were asked to bet between two comparisons and the result showed that most subjects bet on the less likely events. The majority of subjects preferred to bet on the conjunctive event (the probability of which is 0.48) rather than the simple event (the probability of which is 0.50), and also preferred to bet on the simple event rather than the disjunctive event (the probability of which is 0.52).

2.1.1.3 Representativeness

People rely on the representativeness heuristic when assessing the degree to which an event represents a class of events, or the degree to which one event resembles another. This heuristic process significantly influences the judgment of probability. For example, when A is similar to B, the probability that A originates from B is considered to be high and vice versa. However, such a principle leads to serious errors in people's judgment because certain factors which have major effects on judgments of probability, but do not influence judgments of similarity or representativeness, would be ignored under the representativeness heuristic. One such factor is the prior probability (base-rate frequency) of outcomes. In Kahneman and Tversky's (1973) experiment, prior probabilities were neglected if people evaluated probability by representativeness. People considered prior probabilities correctly only if they had no information about representativeness, or valuable information was given; prior probabilities were ignored when a description was introduced, even when such a description was totally uninformative about the judgment of probability. In this case, the representativeness heuristic results in a cognitive bias of insensitivity to the prior probability of outcomes. Besides the insensitivity to base rate frequency, the factors that may cause cognitive biases from the perspective of judgmental heuristics include insensitivity to sample size (Sedlmeier & Gigerenzer, 1997), misconceptions of chance (Wagenaar, 1988), insensitivity to predictability (Joram & Read, 1996), illusion of validity (Kahneman & Tversky, 1973), and misconceptions of regression (Joyce & Biddle, 1981). Detailed explanation of such factors can be found in Tversky and Kahneman's (1974) study and the cognitive biases regarding these factors are briefly described in Table 2-1.

Table 2-1 Tversky and Kahneman's classification of cognitive biases

Cognitive Biases	Heuristic	Description
Retrievability of instances	Availability	People assess the frequency or the probability of an event by the ease with which instances or occurrences can be recalled.
Ineffective search set	Availability	People assess the frequency or the probabilities of an event by assessing the frequency of possible contexts in which such event may occur.
Limited imaginability	Availability	People tends to imagine several instances that can be easily found and assess the frequency or the probability of an event according to their imagines.
Illusory correlation	Availability	People estimate the frequency of co-occurrences by the ease with which the relevant mental operations can be performed.
Overestimated conjunctive problems	Anchoring effect	Probability is often overestimated in compound conjunctive problems.
Underestimated disjunctive problems	Anchoring effect	Probability is often underestimated in compound disjunctive problems
Insensitivity to base rate frequency	Representativeness	Prior probability of outcomes tend to be ignored when other data are available.
Insensitivity to sample size	Representativeness	The sample size is often ignored when judging its representative power.
Misconceptions of chance	Representativeness	Chance is a process in which deviations are diluted, rather than a self-correcting process for restoring the equilibrium.
Insensitivity to predictability	Representativeness	People's prediction is both insensitive to the reliability of the evidence and insensitive to the expected accuracy of the prediction.
The illusion of validity – Redundancy	Representativeness	People tends to have great confidence in forecasting based on redundant input variables.
Misconceptions of regression	Representativeness	Events will tend to regress towards the mean on subsequent trials is often not allowed for in judgment.

Tversky and Kahneman's (1974) study on cognitive biases from the judgmental heuristic perspective remains influential. However, this study has been criticized for two major problems. One is the core concept of judgmental heuristics. When cognitive biases can be experimentally identified, judgmental heuristics as theoretical explanations for people's decision-making are actually untestable (Arnott, 1998). The other problem is that the identification of cognitive biases from judgmental heuristic principles cannot cover all the cognitive biases identified since 1974, including those identified by Tversky and Kahneman themselves. On the other hand, some biases (e.g., framing) are likely to span all three heuristics.

2.1.2 Cognitive biases in Remus and Kottemann's information system development

Based on Tversky and Kahneman's (1974) study, Remus and Kottemann (1986) further extended the study of cognitive bias from an IS perspective. They identified 22 biases and classified them into three levels. At the highest level, biases are classified in association with two steps of decision-making: information acquisition and information processing. In each step, the sources of cognitive biases are further identified as the second level of classification. The basic level is the specification of cognitive bias regarding the sources (Table 2-2).

2.1.2.1 Information acquisition biases

Information acquisition indicates the way in which information related to decision-making is delivered to the user of IS. Sources of cognitive bias in this step are irrelevant information, information display, selective perception, and frequency.

Table 2-2, Remus and Kottemann's classification of cognitive biases

Steps of decision-making (Level 1)	Source of cognitive biases (Level 2)	Specified cognitive biases (Level 3)
Information acquisition	Irrelevant information	Irrelevant information
	Information display	Type Format Order and logic Context
	Selective perception	Information filter Expectation Confirmation of attitude
	Frequency	Recall Base rate error Illusory correlation Redundancy
Information processing	Heuristics	Similarity Rules of thumb Anchoring and adjustment Inconsistency
	Misunderstanding of statistical properties of data	Misunderstanding of change Small sample Gambler's fallacy
	Search strategies	Search strategies
	Conservatism	Conservatism
	Extrapolation	Extrapolation

Irrelevant information may reduce the quality of decision-making. Koester and Luthans (1979) found that people who are inexperienced in the use of IS or DSS are more likely to be influenced by computer-generated information that is irrelevant to their task. In Collopy and Armstrong's (1992) study of expert system development, irrelevant early observations of a timer series were considered to influence the features of data, such as trends, outliers, and level discontinuity, and to damage the accuracy of data features' detection. Ebert (1972) found that, besides the irrelevant early data involved in the target variable, irrelevant factors that were considered in schedulers' decision-making significantly hindered the performance of IS usage. One of the latest studies on the feedback process (Ernst, 2013) showed that learning from feedback in decision-making can be impaired when relevant feedback is combined with irrelevant feedback.

Some properties of information display can also cause cognitive biases when decision makers use IS:

- The type of information – the information acquired through the human-system interaction has more impact on the system user than just the information itself (Schwenk, 1986). For example, one recent study (Nettelhorst, Brannon, & Trey Hill, 2013) revealed that over-reliance on case history information or base rate neglect may bias individuals' judgment.
- The format of display – the format of information displayed to system users affects the decisions they make. Some empirical studies have revealed that summarized presentations, such as statistics, tables, and graphs, can effectively reduce cognitive bias compared with the display of raw data (Benbasat & Dexter, 1985; Remus, 1984), while the choice of graphical summaries or tabular summaries depends on the level of environmental stability (Remus & Kottemann, 1986; Tullis, 1981).
- The order and logic of display – the order in which information is presented affects the load of certain information in the decision maker's judgment. Similarly, decision makers may neglect alternatives when one set of information seems to

capture the majority of possibilities. It has been observed in some studies that the first (primacy effect) and the last (recency effect) pieces of data may be overvalued in decision-making (Kirs, Pflughoeft, & Kroeck, 2001; Moskowitz, Schaefer, & Borcherding, 1976).

- Context – when evaluating the variability of information, especially a data series, decision makers may be biased by the absolute value of the data points and the sequence in which they are presented. Hogarth and Makridakis' (1981) study revealed that cognitive biases involved in people's judgment are linked with the environment, including people's schema, actions, outcomes, as well as the feedback to the schema.

Selective perception, as the third source of cognitive bias in the information acquisition step of decision-making, means that people selectively perceive and partially remember the information they received. First, people filter information according to their experience (Egeth, 1967). Decision makers tend to pay particular attention to the information related to the areas in which they have expertise. For example, when applying DSS in human resource (HR) management, it has been observed that HR managers' perception of salary increases were biased towards their own single increase experience; while, the base rate increases provided by the system were ignored (Kydd, 1989). Second, people's expectations can bias perceptions (Makridakis, Hibon, & Moser, 1979). When decision makers are reviewing information, parts of which are contrary to their expectations, the probability that they remember incongruent pieces of information inaccurately is increased. Bennett's (1982) study of college teaching evaluation revealed that the evaluations of female instructors were biased in some ways because they failed to meet students' gender-appropriate expectations. Third, people seek information that is consistent with their own attitudes (Batson, 1975). When decision makers have some prejudices about a problem, they tend to seek information that can be used to confirm their prejudices. In Brannon, Tagler, and Eagly's (2007) experiments, the subjects preferred attitudinally consistent information in their decision-making, and this phenomenon became more pronounced when certain attitudes were strongly held.

Cognitive biases identified from the source of frequency are the same as defined by Tversky and Kahneman (1974). Such cognitive biases include: (i) recall (the retrievability of instances); (ii) base rate error (insensitivity to base rate frequency); (iii) illusory correlation; and (iv) frequency to imply strength of relationship (redundancy). The description of these cognitive biases will not be duplicated here.

2.1.2.2 Information processing biases

After the information is obtained, the second step of decision-making is to process the information in two ways: system data analysis and user's judgment. From the system user's perspective, the sources of cognitive bias involved in this step include heuristics, misunderstanding of the statistical properties of data, limited search strategies, conservatism in decision-making, and inability to extrapolate growth processes.

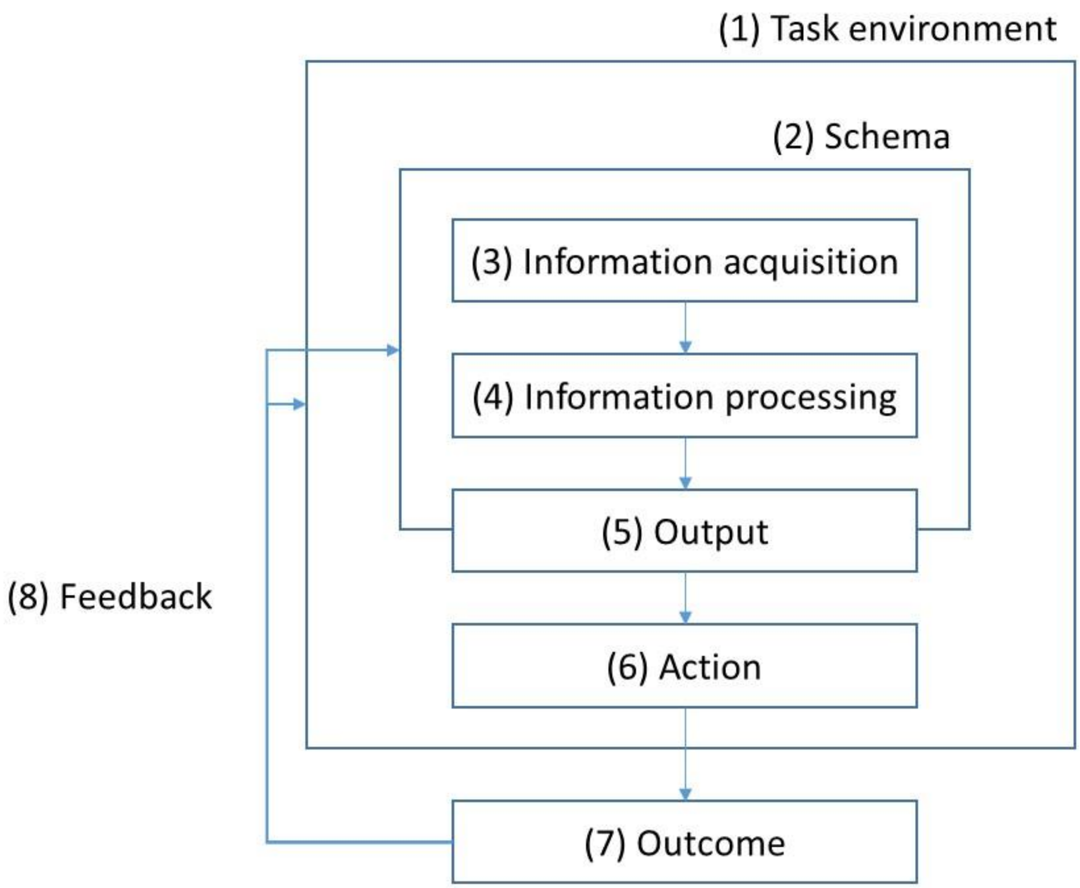
In Remus and Kottemann's (1986) study, heuristics is considered as a source of cognitive bias in a specific situation of information overload and the decisions about reducing the information are made according to three or four crucial factors. Based on those factors, decisions are made using heuristics; however, some built-in biases come from certain heuristics, such as structuring problems based on experience, rules of thumb, anchoring and adjustment, and inconsistency in the use of heuristics. First, decision makers usually try to find the best fit between the new problem they face and the old problems they have solved. Once a match is found, decision makers make decisions by slightly altering the solution of the old problem without rational evaluation (Kahneman & Tversky, 1972). Schulster (2004) mentioned in a study of time-critical decision-making in spaceflight operations that, based on experience and expertise, decision makers try to find the "best fit" between the new problem and their catalogue of mental models with the main purpose of avoiding having to analyze each problem from scratch. Second, rules of thumb as a cognitive bias indicates that a set of rules followed by the decision maker in his/her prior experience will be used again in solving similar problems since such rules proved satisfactory the last time. According to Hartvigsen (1992), rules of thumb are widely used as a financial ratio analysis technique in banking; however, the decision made using this technique is biased, especially when comparing failed firms with non-failed ones in bank lending. Third, decision makers' inconsistent use of heuristics also biases their decision-making. Some studies have compared people's judgmental heuristics and heuristics based on regression (e.g., Fildes et al., 2009; Remus, 1977). The results of these studies show that people's judgmental heuristics cannot outperform regression and the main reason for decision makers' failure is inconsistency of heuristics. The last cognitive bias arising from people's heuristics is anchoring and adjustment, as defined by Tversky and Kahneman (1974).

When statistical data analysis is adopted in information processing, four cognitive biases may arise from misunderstanding of the statistical properties of data. First, people usually mistake random variations of data for persisting change (Langer, 1977). Observations of a series of data higher or lower than the mean value may bias decision makers' judgment and make them believe that an upward or downward trend is emerging. In two experiments, Lopes and Oden (1987) showed that naive subjects and statistically sophisticated subjects have trouble distinguishing between random and non-random events. Furthermore, sampling is the foundation of reducing the perceived complexity of a task. According to the law of large numbers, a relatively large sample will be highly representative of the population from which it is drawn. However, cognitive bias arises when this law is held for small samples (Tversky & Kahneman, 1971). To illustrate small sample bias, Tversky and Kahneman (1974) conducted an experiment with two hospitals: one is large, with 45 babies born each day on average; the other is small, with 15 babies born each day on average. Within a period of one year, the two hospitals recorded the number of days on which more than 60% of newborn babies were boys. The subjects were asked to vote on which hospital had more boys. According to sampling theory, the probability of such an event in a small sample (the small hospital) is much larger than in a large one (the large hospital). Unfortunately, the majority of the subjects judged the probability of more than 60% of births being boys to be equal in both hospitals. Rabin (2002) further examined the small sample bias involved in agents' decision-making in the stock market, and found that agents tended to overestimate the precision of unreliable signals from small samples. Third, gambler's fallacy is identified as a cognitive bias in which peoples' judgment regarding future events is based on the occurrence of past events. For example, most people erroneously believe that a black on the roulette wheel is due after observing a long run of reds (Tversky & Kahneman, 1974); or over-estimate the probability of a fair coin turning a heads when there have been a number of consecutive tails in prior tosses (Rabin, 2002). Their judgments are biased by erroneously believing that gambling or similar events are self-correcting processes in which a deviation from the mean induces a regression in order to restore the equilibrium, when the events within such processes are actually independent.

In Remus and Kottemann's (1986) study, another three sources of cognitive biases are search strategies, conservatism, and extrapolation. Search strategies and conservatism were first identified by Tversky and Kahneman (1974); the bias of extrapolation was identified for the first time by Remus and Kottemann. Extrapolation bias focuses on a specific situation in which the problem is that of extrapolating a time series, and exponential growth is observed in certain series. In this case, people may underestimate the outcomes of the growth process, regardless of how many data points are presented in the time series (Timmers & Wagenaar, 1977; Wagenaar & Timmers, 1978). In Levy and Tasoff's (2012) experiment, the subjects were asked to estimate the value of two assets with given interest rates. The result shows that assets were underestimated twice as often as they were overestimated.

2.1.3 Cognitive biases in Hogarth's model of judgment

Hogarth (1987) proposed another comprehensive classification in early studies of cognitive bias according to his model of human judgment (Figure 2-1).



Source: Hogarth (1987, p207)

Figure 2-1 Hogarth's conceptual model of judgment

This model is established upon the relationships between three main entities: the person who makes the judgment, the task environment, and the actions resulting from the judgment. A person's decision-making occurs within a task environment (box 1 in Figure 2-1). When the person is making a decision, he/she is represented by a schema (box 2 in Figure 2-1) within which the operations of decision-making are decomposed into the three operations of the decision-making process: information acquisition, information processing, and output (boxes 3–5 in Figure 2-1). In the conceptual model, output has been drawn as the interface between the task environment and the schema. In some cases, output is considered to be contiguous with action (box 6 in Figure 2-1) which occurs within the task environment. Subsequently, the action leads to an outcome (box 7 in Figure 2-1) of the persons' decision-making, and this outcome can further feed back into the person's schema, as well as the environment in which the action takes place. For example, consider that a waiter who provides service to both young and old people in a restaurant (task environment) and knows from experience that young customers do not tip as generously as old customers (information acquisition – memory). At a particular point, there is an age distribution of customers in the restaurant (information acquisition – features of the task) since the waiter believes that his efforts serving young customers are unlikely to be rewarded as highly as those serving old customers (schema). After the information acquired from both memory and the features of the task are processed in the waiters' mind, the result of his judgment is the quality of service he gives to customers in different age groups (output): he devotes much attention to old customers but serves young customers poorly (action). The outcomes of his action is that the tips given by young customers are small, which reinforces the waiter's notion that young customers do not tip generously (feedback to the schema). The young customers, who received poor service from the waiter, may not come to the restaurant frequently, therefore affecting the age distribution of customers (feedback to the task environment). As shown in this example, cognitive biases first occur in the acquisition of information from both the task characteristics and memory. Then the rules a person chooses for information processing can induce biases. Third, the manner in which the judgment or the choice is expressed can be biased. Finally, the interpretation of the significance of outcomes, as well as the learning relations for predictive activity, can induce biases. In addition, biases can also occur as a result of interactions between the different stages of decision-making (Table 2-3).

Table 2-3 Hogarth's cognitive bias

Main entities of decision-making	Source of bias	Cognitive bias
Information presentation	Memory	Limited recall from memory
		Limited ability on prediction
	Task characteristics	Order effects
		Availability heuristic
		Selective perception
		Form of information presentation
	Interaction of memory and task characteristics	Confirmation of expectation
		Ignorance of base-rate information
		Biased causal framework
Information processing	Memory	Habit of judgment
		Availability of experience
	Task characteristics	Effects of task variables
		Inconsistent criteria of judgment
Output	Probability estimation	Sensitive to the scale
Feedback	Learning	Misinterpretation of chance and cause
		Gambler's fallacy
		Illusory of correlation
		Hindsight
		Misplaced confidence

2.1.3.1 Information acquisition

In Hogarth's (1987) model of judgment, there are two sources of information that can be accessed by decision makers: memory and task characteristics. Therefore, the cognitive biases in information acquisition can be identified according to the two sources of information and their interactions.

Cognitive biases as the functions of memory include the ease with which information is recalled from memory and people's limited ability to make predictions. Bias of recall is the same as identified in Remus and Kottemann's (1986) study, which indicates that people use recall to predict the frequency of an event according to the extent to which the event is well publicized. Concrete information is considered to be more salient in memory than abstract information. In a vividness study, Taylor and Thompson (1982) revealed that vividly presented information was more persuasive and had greater impact on judgments than non-vividly presented information. An experiment (Hogarth, 1987) tested people's intuitions of the relative frequency of causes of death. The report showed that well-publicized causes, like homicide, cancer, and tornado, were overestimated, while asthma and diabetes were underestimated. Furthermore, the bias of people's predictive ability means that people tend to judge the probability of an event by its frequency rather than its relative frequency. Given two companies, for example, one of which has successfully marketed 10 innovations in the past five years while the other has successfully marketed six, some people may believe that the former company has been more successful without consideration of how many innovations both companies have attempted in total (Remus & Kottemann, 1986).

Cognitive biases as the functions of task characteristics include: (i) order effects, (ii) availability heuristic, (iii) selective perception, and (iv) the form of information presentation. The bias of order effects indicates that the first or the last in a series of items dominates people's judgment. It is similar to the bias of order and logic of display defined in Remus and Kottemann's (1986) study. Furthermore, the availability heuristic focuses on identifying the bias induced by limited information about task characteristics, and the definition of this bias aims to avoid any confusion with limited recall from memory. A classic example of this bias is the "Eureka" effect: the understanding of a problem appears suddenly and its solution seems to be smoothly processed. Thus, the person experiencing such a "Eureka" moment is convinced that the solution is true, even if it is actually based on only partial understanding of the problem (Topolinski & Reber, 2010).

The bias of selective perception is similar to the bias of information filter in Remus and Kottemann's (1986) study. The adoption of a particular background can determine which part of a reference is to be considered in decision-making; therefore, the selection of particular information can exert important influences on people's choices. Keil, Depledge, and Rai (2007) developed and tested an escalation decision model that incorporates problem recognition, escalation of commitment of failing courses of action, and the cognitive bias of selective perception. Their result revealed that selective perception significantly affects both problem recognition and escalation. In addition, the form of information presentation can also affect people's judgment. It covers five cognitive biases in Tversky and Kahneman (1974) and Remus and Kottemann's (1986) studies: the manner and order of presentation, context effects, the logic of data display, information overload, and redundant information display.

The interaction of memory and task characteristics can also bias people's judgment. First, dysfunctional judgment can be caused by taking expectations for reality. Expectations usually come from a summary of information from one's memory. Such as the example of the restaurant waiter who expects high rewards (tips) from old customers and low rewards from young customers because of his memory of previous experience. Based on expectations, people tend to seek information from the task environment that is consistent with those expectations, rather than seek conflicting evidence. So, the waiter will seek out old customers in the restaurant and put more effort into serving them in order to make the reality conform with their expectations (Remus & Kottemann, 1986). Second, the base rate bias is the same as defined in Tversky and Kahneman (1974) and Remus and Kottemann (1986), which reflects that people ignore the base rate information when specific information is presented and becomes salient in people's judgment. Third, people's causal framework for thinking about the current situation of a task guides their interpretation of information. If the framework is biased, the acquisition of particular information is also biased. The first empirical evidence for the framing bias in decision-making is Tversky and Kahneman's (1981) study of Asian disease problems. More evidence has been found to confirm the existence of framing bias and its negative effect on the quality of decision-making (Camerer, Babcock, Loewenstein, & Thaler, 1997; Camerer, Loewenstein, & Rabin, 2004; McNeil, Pauker, Sox Jr, & Tversky, 1982).

2.1.3.2 Information processing

As cognitive biases arise from information acquisition, biases in information processing are also classified by memory and task characteristics. Memory biases in processing include people's habit of judgment and the availability of experience. The habit in decision-making is the same as the rule of thumb in Remus and Kottemann's (1986) study, which indicates a set of rules followed by the decision maker in his/her prior experience that will be used again to solve similar problems. However, a previous decision or a rule to guide towards a decision carries no guarantee of future success, even if successful in the past. Plenty of empirical studies in the field of finance and economics have revealed that people's habitual frames contribute to the underestimation of risks in the financial markets (Ferguson, 2008; Fisher & Malde, 2011). Furthermore, habit also affects the choice of decision rules. That is, specific personal experience can make specific rules more available to one person than to others who do not have such experience. However, as Remus and Kottemann (1986) emphasized, one's expertise cannot be generalized if it is influenced by other factors or the circumstances of the problem change.

The cognitive biases resulting from task characteristics in information processing include the effects of task variables and inconsistent criteria of judgment. The variables that can bias information processing are various, such as the amount of information available (Gelardi, 2010), time pressures (Perrin, Barnett, & Walrath, 1993), and inconsistent or missing information (Perrin, Barnett, Walrath, & Grossman, 2001). The major bias in information processing is the lack of consistency in people's judgment. First, the criteria or rules that people follow in their judgments may be inconsistent. Remus and Kottemann (1986) found that this is so across a series of cases. The validity of judgments based on such fallibility is debatable. Sjöberg (1982) found that people become increasingly inconsistent as the amount of information increases, which results in unreliable confidence in establishing the quality of decision-making. Second, people's decisions may also be inconsistent with their criteria. Some studies (e.g., Armstrong, 2001; Armstrong & Collopy, 1994) have revealed that the inconsistency between people's decisions and their criteria depends on the way in which information is presented.

2.1.3.3 Output

Output biases are triggered by the way that people express their judgment. Hogarth (1987) did not develop a classification of cognitive biases arising from output, but offered an example of such bias. Considering probability estimation, the scale used to measure people's responses is of great importance in probability assessment. The type of scale, such as relative or absolute scales, linear or logarithmic scales, can significantly influence people's judgment. Therefore, cognitive bias may arise from the output if the scale of measurement is not properly designed. A special case study of scale bias was conducted by Hageman (2010), which focused on examining people's confidence when using tax DSS. The result showed that participants were overconfident in their operations when a large scale of response was designed (e.g., five or more errors), and that cognitive bias was significantly decreased when a 100-point scale measurement was used.

2.1.3.4 Feedback

The last component of Hogarth's judgmental model is feedback, which supports people's learning from the outcome of their judgment. Thus, the cognitive bias arising from feedback mainly focuses on the effects on learning. Feedback biases are mainly caused by the unobservable outcomes associated with the total range of options. In addition, delayed feedback and specific events that significantly affect outcomes can also bias people's judgment. Five general biases caused in the feedback of judgment have been identified in Hogarth's judgmental model. The first is misinterpretation of "chance" and "cause." According to Hogarth's (1987) definition, "cause" specifically indicates that an event is persistently changed by the effect of a factor (e.g., a sales increase is the result of a special advertising effort); “chance” refers to a combination of unidentifiable or random factors (e.g., a sales increase is the result of a combination of several unstable circumstances). The cognitive bias caused by misinterpretation of chance and cause arises when people mistake an unidentifiable or random process for a persisting change, which leads to erroneous causal attributions. A typical instance of this bias is the failure to understand the effects of regression (Einhorn & Hogarth, 1981; Joyce & Biddle, 1981). Furthermore, when “cause” is reflected by a person’s skill and the task involves both skill and chance, people tend to attribute good outcomes to skill and bad outcomes to chance (Hogarth, 1987). When a poor decision leads to a good outcome, a false feeling of control over the judgment situation arises, though the outcomes probably result from chance. In Langer’s (1975) study, such cognitive bias is named “the illusion of control.”

The second and third biases of learning are gambler’s fallacy and illusory correlation, which are defined by Remus and Kottemann (1986). When estimating probabilities, people tend to confuse random events with independent events. When two events have co-occurred several times in the past, the probability of them occurring together in the future can be overestimated even if the circumstances of one event change.

Hindsight bias is the fourth cognitive bias arising from learning, which happens in retrospect and leads to overestimation of the probability that an event will happen. For example, an IS produces an estimation and the system user claims “I predicted this.” However, Remus and Kottemann (1986) mentioned that people have short memories concerning their prior uncertainties, which limits their ability to imagine alternative explanatory schemes for the past. Indeed, hindsight bias reduces people’s ability to learn from past events. Ofir and Mazursky (1997) examined people's learning from surprising outcomes and concluded that hindsight bias exists and significantly influences people's judgment and that such bias is diminished only if an outcome highly surprised the participants.

Misplaced confidence is the last cognitive bias arising from people's learning and should be identified from both overconfidence and lack of confidence. Overconfidence indicates that people often overestimate their ability to solve difficult or novel problems. Both empirical and experimental evidence have revealed that substantial overconfidence is common in decision-making (Arnott, 1998; Yates, Lee, & Shinotsuka, 1996). However, some studies have also revealed that experts usually show less overconfidence than naïve subjects (e.g., Bazerman & Moore, 2008; Camerer & Johnson, 1997). In Önkal, Yates, Simga-Mugan, and Öztin's (2003) study of foreign exchange rate forecasting, for example, experts tended to underestimate the accuracy of their forecasts. The main reason for misplaced confidence is that people interpret outcomes without fully understanding the characteristics of the task structure. Many studies have also suggested that misplaced confidence (usually overconfidence) is a complex human behavior caused by several other cognitive biases, including adjustment, confirmation, hindsight, recall and similarity biases (Keren, 1997; Russo & Schoemaker, 1992; Yates et al., 1996).

2.1.4 Recent studies of cognitive biases

Many studies have been conducted based on the abovementioned research in cognitive bias identification, and can be classified in two groups. The first group of studies aims to extend the identification of cognitive bias in human decision-making from the perspectives of economics, psychology, and organizational behavior. For example, Bazerman (1998) proposed a new cognitive bias, escalation of commitment, in judgmental decision-making, according to which decision makers tend to increase the commitment of resources to a decision even it is known to have been incorrect in the past. Such a bias may be caused by competition, and a desire to win the competition, and make decision makers feel that they are close to the attainment of their goals. Tversky and Kahneman (1982) further extended their study of judgmental bias in the judgment of probability and identified subset bias, which acts contrary to conjunction bias. With subset bias, people tend to judge that a conjunction or subset is more probable than any of its sets when, in reality, the opposite is true according to probability theory (Briggs & Krantz, 1992). Samuelson and Zeckhauser (1988) identified status quo bias in a series of experiments. When participants were facing important real decisions (e.g., to keep or change a health plan or a retirement program), they disproportionately stuck with the status quo regardless of its relative superiority or inferiority to other options. Todd and Gigerenzer (2007) also identified this bias and named it default bias in their study. In psychology research, VandenBos (2007) also considered bandwagon effect a type of cognitive bias, which describes individuals' tendency to align themselves or their stated options with the majority opinions they perceive in social or political situations. Carter et al. (2007) further extended Arnott's study by systematically reviewing studies in the fields of economics, psychology, and managerial decision-making published between 1933 and 2006. Their study collected 76 cognitive biases and has been considered the most comprehensive collection of cognitive biases up to the present.

Given the large number of cognitive biases identified in the literature, the relationships between them are quite complex, some of them overlap (e.g., success bias is related to illusory control), while causal relationships also exist between them (e.g., overconfidence is a complex bias arising from anchoring and adjustment, confirmation, hindsight, recall, and similarity biases). A problem therefore arises in using cognitive bias theory in the development of DSS: how to mutual-exclusively and exhaustively identify the presence of cognitive biases according to a given decision-making task? To solve this problem, some studies have also focused on the taxonomy of cognitive bias. Indeed, the cognitive biases identified in early studies (Hogarth, 1987; Remus & Kottemann, 1986; Tversky & Kahneman, 1974) have been categorized from different perspectives. Several other studies have provided other categorizations (Bazerman, 1998; Keren, 1990; Ralph, 2011). However, these classifications are mainly based on subjective groupings and inevitably suffer from two weaknesses: a lack of systematic methodologies in creating the categorizations, and a lack of mutual exclusivity and exhaustiveness (Arnott, 1998; Carter et al., 2007). For example, Tversky and Kahneman (1974) classified 12 biases according to three judgmental heuristic principles they proposed, which are not comprehensive since some major biases (e.g., biases regarding persistence) are missing. Hogarth (1987) classified 19 biases according to his model of human judgment; however, some biases can be found to straddle multiple steps of human decision-making (e.g., recall bias in both information acquisition and information processing). Some of the most recent studies on the taxonomy of cognitive bias were conducted by Arnott (1998, 2006). According to an exhaustive review of the literature, Arnott identified 37 cognitive biases in human decision-making, and proposed a six-group taxonomy, which is considered the most mutually-exclusive and exhaustive classification of cognitive biases. Arnott's taxonomy is thus adopted in the current research:

1) Memory biases. This group of biases covers those occurring in the storage and recall of information stage, so they can be regarded as the lowest or the deepest level of cognitive biases. The cognitive biases involved in this group include hindsight, imaginability, recall, search, similarity, and testimony.
2) Statistical biases. This group of biases is concerned with the general human tendency to process information contrary to the normative principles of statistics and probability theory. They include biases of base rate, chance, conjunction, correlation, disjunction, sample, and subset.
3) Confidence biases. This group of biases acts to influence decision makers' confidence. The consequence of confidence biases in judgmental decision-making is the curtailing of the search for new information related to the task. The cognitive biases in this group include completeness, control, confirmation, desired outcome, overconfidence, redundancy, selectivity, success, and test.
4) Adjustment biases. This group of biases align with Kahneman, Slovic, and Tversky's (1982) definition of anchoring and adjustment heuristics, which include anchoring, conservatism, reference, and regression.
5) Presentation biases. This group of biases should not be simply considered as produced by the display of data. Regarding Hogarth's (1987) model of human judgment, presentation biases are observed to arise from both output and feedback of information processing, and include framing, linear, mode, order, and scale.

6) Situation biases. This group of biases are produced by the way in which people respond to the general decision situation; they are considered to represent the highest level of bias abstraction. The cognitive biases involved in this group include attenuation, complexity, escalation, habit, inconsistency, and rule.

2.2 Empirical studies of cognitive biases

Based on Arnott's (2006) taxonomy of cognitive bias, a further literature review focuses on identifying the common cognitive biases in people's judgmental forecasting. Three leading databases were used in this study: *Science Direct*, *EBSCOhost*, and *Google Scholar*. The following features of the reviewed studies were collected in order to fully describe the cognitive biases occurring in practice:

- Whether the forecasting task is time series forecasting or not;
- Whether the forecasters are experts, non-experts, or a mixed group of both;
- Description of forecasters' cognition that biases their judgments;
- Specific cognitive biases and their classification according to Arnott's taxonomy;
- Debiasing strategy and the objectives of debiasing.

The preliminary search was based on various combinations of the key words "judgmental forecasting," "judgmental adjustment," "cognitive biases," and "cognitive error." Studies that either clearly identified cognitive biases according to Arnott's taxonomy, or identified cognitive biases that have the same meaning as found in Arnott's taxonomy, were further filtered. The search of the preliminary key words yielded 145 relevant articles, of which 55 were qualified for review. These studies were all published between 1977 and 2013.

Thirty-nine of them identified cognitive biases in time series forecasting and the other 16 focused on probability forecasting. In terms of participants, 16 studies examined experts' cognitive bias; another 33 focused on non-experienced forecasters; and six studies were conducted with the purpose of comparing experienced and inexperienced forecasters' cognitive behaviors. In terms of the research fields and the data used for forecasting, 16 studies were found in the fields of finance and economics; another 16 were related to business and marketing; four focused on supply chain management; and eight covered other research areas, including education, health care, sports, and weather forecasting. The other 11 studies were experimental, using artificial data or data relating to general knowledge. Most of the studies reviewed only identified cognitive biases; 15 proposed or examined particular debiasing strategies after cognitive biases were identified. Further details on the articles reviewed are shown in Table 2-4.

Table 2-4 Cognitive biases in judgmental forecasting research

Studies	*Topic*	*Time series*	*Participants*	*Performance*	*Cognitive Sources*	*Taxonomy*	*Debiasing Strategy*	*Debiasing objectives*
Andersson, Edman and Ekman (2005)	Sport	No	Both	Soccer experts were overconfident, because their extensive knowledge of soccer would make them sensing a certain degree of control and thus they overestimated their forecasting ability.	Control	Confidence	N/A	N/A
Andreassen (1990)	Stock Price	Yes	Non-experts	Forecasters consider factors that altered the salience of the price change information from low to high.	Regression	Adjustment	N/A	N/A
Andreassen and Kraus (1990)	Judgmental extrapolation	Yes	Non-experts	Censorship may have dampened the predicted pattern by making the tracking scores less positive rather than less negative; The high regressive forecasts implied that rises usually lead to increased initiation of sell	Regression	Adjustment	N/A	N/A

Studies	*Topic*	*Time series*	*Participants*	*Performance*	*Cognitive Sources*	*Taxonomy*	*Debiasing Strategy*	*Debiasing objectives*
				trades, which are likely to dampen the previous positive change.				
Aukutsionek and Belianin (2001)	Marketing; Human Resource Management ; Investment	No	Experts	Leaving aside investment forecasts, all but one point for all indicators fell outside the allowed rectangles, which indicating bad calibration and significant overconfidence.	Overconfi dence	Confidence	N/A	N/A
Ayton, Pott and Elwakili (2007)	Human Behavior	No	Non-experts	People over-predicted the impact of events on their emotions, in particular they believed that events will have an impact for a longer interval than they actually do; They overestimated the impact of negative events and hence the durability bias in affective forecasting results.	Overconfi dence, Desired Outcome	Confidence	N/A	N/A
Batchelor (2007)	Economics	Yes	Experts	A bias towards optimism in the consensus forecast was inevitable as rational forecasters learn about the new trend; Biases toward optimism and pessimism were presented at both long and short horizons.	Herding	Confidence , Group forecasting	N/A	N/A
Batchelor and Dua (1992)	Economics	Yes	Experts	When revising forecasts, forecasters gave too much weights to their own past forecasts.	Conservati sm, Overconfi dence	Adjustment , Confidence	N/A	N/A
Benson and Önkal (1992)	Feedback & Training	No	Non-experts	Overforecast exists in both control and experiment group; The provision of only outcome, covariance and resolution feedback was not sufficient to improve forecasting performance; The provision of calibration feedback resulted in improved forecasting performance.	Habit	Situation	Feedback	Habit
Bolger and Önkal-Atay (2004)	Stock Price	No	Non-experts	Results showed that forecasts were initially overconfident but improved significantly after receiving feedback; Interval forecasts were initially manifested overconfidence, which was significantly reduced after forecasters received feedback about their performance.	Overconfi dence	Confidence	Performanc e feedback	Overconfid ence
Buehler, Messervey and Griffin (2005)	Collaborativ e Planning	No	Non-experts	Predictions generated through group discussion were more optimistic than those generated individually; Group discussion	Overconfi dence, Herding	Confidence , Group forecasting	N/A	N/A

Studies	Topic	Time series	Participants	Performance	Cognitive Sources	Taxonomy	Debiasing Strategy	Debiasing objectives
				heightened participants' tendency to focus primarily on factors promoting successful task completion, and to enhance their optimistic outlook.				
De Bondt (1993)	Stock Price; Exchange Rates	Yes	Both	Forecasters were optimistic in bull markets and pessimistic in bear markets; If a large price increase is predicted, the subjective probability distribution of future prices is left-skewed, recognizing a possible decline; and vice versa; After an 'up' week in a bear market, the subjects were more willing to see a turnaround than after a 'down' week.	Underconfidence, Desired Outcome, Regression	Bias in judgmental forecasting, Confidence, Adjustment	N/A	N/A
Du and Budescu (2007)	Stock Price	Yes	Non-experts	Participants showed underconfidence at 50% confidence level, but overconfidence at 90% confidence level; Participants raised (lowered) the point estimates but biased their confidence intervals downward (upward) to hedge the potential for price declines (rises).	Overconfidence, Underconfidence	Confidence, Bias in judgmental forecasting	N/A	N/A
Eggleton (1982)	Costs of product	Yes	Non-experts	Forecasts for independent series have been found to lie between the mean and the last data point; Forecasts for untrended series are too high.	Selectivity, Anchoring	Confidence, Adjustment	N/A	N/A
Eroglu and Croxton (2010)	Sales of product	Yes	Both	The level of optimism bias decreased with greater variability in statistical forecast errors and increased with age; Openness to experience increased optimism bias, whereas agreeableness decreased it; Individuals who were high in conscientiousness and agreeableness and low on extraversion were more prone to anchoring bias; The level of overreaction bias is significantly affected by a forecaster's education level.	Desire, Anchoring, Overconfidence	Confidence, Adjustment	N/A	N/A
Fildes (1991)	Economics	Yes	Experts	Individuals restricted their attention to only one or two sources which are thought to be more relevant to the task.	Selectivity	Confidence	Optimally process information sources	Selectivity

Studies	*Topic*	*Time series*	*Participants*	*Performance*	*Cognitive Sources*	*Taxonomy*	*Debiasing Strategy*	*Debiasing objectives*
Fildes et al. (2009)	Supply Chain Management	Yes	Experts	The retailer's positive adjustments that applied to system forecasts are too high, which indicated an optimism bias; The smaller adjustments often damaged accuracy; Positive adjustments were much less likely to improve accuracy; Wrong direction adjustment suggesting a general bias towards optimism;	Conservatism, Desired Outcome	Confidence, Adjustment	The results of Blattberg–Hoch approach and error bootstrapping, Avoiding small adjustments, Avoiding wrong-sided adjustments)	Conservatism, Desired Outcome
Glaser, Langer, Reynders and Weber (2007)	Stock Price	Yes	Non-experts	Practical expertise made practitioners overconfident or made them behave as if they were overconfident due to institutional reasons; Framing effect is significant in the experiment and even stronger when participants did not receive additional return information; The degree of mean reverting expectations is always higher in the price forecast mode.	Overconfidence, Framing, Regression	Confidence, Presentation, Adjustment	N/A	N/A
Goldfarb, Stekler and David (2005)	Economics	Yes	Experts	Representative forecasters were more optimistic than FED's analysis of the economic forecasting; The use of the 1973–1974 experience as the basis for the later forecasts contributed to their inaccuracy.	Desired Outcome, Overconfidence, Similarity	Confidence, Memory	N/A	N/A
Goodwin (2000a)	Forecast combination	Yes	Both	There was nothing to be gained by combining judgment with statistical forecasts.	Chance	Statistical	Correct than combine	Chance
Goodwin (2000b)	Sales of product	Yes	Non-experts	Individuals tended to overreact to random movements in the data.	Chance, Overconfidence	Confidence, Statistical	Making statistical forecasts as default, Reason of adjustment	Chance
Goodwin et al. (2007)	Marketing	Yes	Non-experts	Forecasters tended to examine a small number of forecasting methods and often miss the method that provided the best fit to past data; Forecasters who missed the well-fitting statistical method tended to make large judgmental adjustments and further damaged the forecast accuracy.	Control, Test	Confidence	Training, best fit statistical model to past data, avoiding making substantial adjustments.	Control, Overconfidence, Test
Goodwin et al. (2013)	Economics	Yes	Non-experts	The participants tended to narrow the provided	Base rate, Chance,	Statistical, Confidence	N/A	N/A

Studies	Topic	Time series	Participants	Performance	Cognitive Sources	Taxonomy	Debiasing Strategy	Debiasing objectives
				intervals, demonstrating overconfidence; They rarely placed total trust in the forecasts provided by the system; Forecasters had a tendency to perceive each new period as a special case, and hence, to consider it as unrelated to the base-rates provided; When people had the option of following optimal advice, they performed a simulated intensive care of control even if they were assured to be the best possible advice available.	Control, Overconfidence			
Gu and Xue (2007)	Stock Price & Earning	Yes	Experts	Forecast errors were always negative on average but more negative at the extreme ends (especially the lower end) of earning changes, suggesting systematic forecast optimism; Forecast errors that increased for the lower quartile and decreased for the upper quartile suggested underreaction to extreme bad news and overreaction to extreme good news.	Desired Outcome, Chance	Confidence, Statistical	N/A	N/A
Harvey (1995)	Sales of product	Yes	Non-experts	Subjects tended to underestimate the trends of time series; Casual adjustments with the aim of compensating for omitted variables in statistical model were prone to a double-counting bias in practice.	Desired Outcome, Redundancy	Confidence	N/A	N/A
Harvey and Bolger (1996)	Sales of product	Yes	Non-experts	The negative constant error values with both presentation formats indicate the presence of an overforecasting bias in both the experiments; Forecasters made forecasts for trended series by anchoring on the last data point and then making an adjustment away from it to take the trend into account; Trend-damping was much greater when data were presented in tabular format; The inconsistency effect for untrended series was greater with a graphical presentation.	Anchoring, Desired Outcome, Mode, Inconsistency	Adjustment, Presentation, Confidence, Situation	N/A	N/A

Studies	***Topic***	***Time series***	***Participants***	***Performance***	***Cognitive Sources***	***Taxonomy***	***Debiasing Strategy***	***Debiasing objectives***
Harvey and Harries (2004)	Sales of product	Yes	Non-experts	People put too much weights on their own initial opinions and not enough on the new opinions provided by their advisors.	Conservatism	Adjustment	N/A	N/A
Jain, Bearden and Filipowicz (2013)	Sport	No	Non-experts	The poorer performance of the depressed participants associated with the later forecasting stages was related to their greater tendency to neglect base rates; When people were asked to make predictions on political or social issues, their predictions were often coincided with their own preferences; Depressed participants paid more attention to cues which were not predictive of the situation; or found the task overwhelmingly more difficult than did the nondepressed, such that they could not use the truly predictive cues appropriately (which then led to overestimation).	Base Rate, Desired Outcome, Test	Statistical, Confidence	N/A	N/A
Koriat, Lichtenstein and Fischhoff (1980)	General Knowledge	No	Non-experts	There is a tendency to disregard evidence inconsistent with (contradictory to) the chosen answer; Asking subjects to write a supporting reason did not affect their calibration (presumably because they were already thinking of these reasons), whereas asking them to write a contradicting reason did.	Selectivity	Confidence	Supporting and(or) contradiction reasons	Desired Outcome
Lawrence and O'Connor (1992)	Business	Yes	Non-experts	More information would lead not to a better but to a worse forecast; Forecasters failed to estimate the amount of adjustment when the adjustment direction was correct according to the slope of time series; Most of the errors in the judgemental forecasts were caused by human inconsistency rather than from a systematic bias.	Anchoring, Redundancy, Inconsistency	Situation, Confidence, Adjustment	N/A	N/A
Lawrence, O'Connor and Edmundson (2000)	Marketing	Yes	Both	Graph is more accurate than table in the short-term, but table is more accurate than graph in the long-term.	Mode	Presentation	Combination method using graph for the short-term and table for the long-term.	Mode

Studies	*Topic*	*Time series*	*Participants*	*Performance*	*Cognitive Sources*	*Taxonomy*	*Debiasing Strategy*	*Debiasing objectives*
Lawrence et al. (1985)	Sales of product	Yes	Experts	Differences between over-stocking costs and under-stocking costs may lead to forecast bias, though the direction seemed to depend on the organisational circumstances.	Complexity	Situation	N/A	N/A
Lee, Goodwin, Fildes, Nikolopoulos and Lawrence (2007)	Marketing	Yes	Non-experts	Human memory limitations leads to the fact that only a small sample of past cases may be recalled; while, the details of the cases were usually recalled incorrectly; The more 'unusual' the case was, the more likely people were to remember it and to recall; some unsuitable cases recalled in such manner may hamper the recollection of other more suitable cases.	Recall, Testimony, Search	Memory	Memory+Similarity+Adaptation support	Recall, Testimony, Search
Lim and O'Connor (1994)	Sales of product	Yes	Non-experts	The weighting people placed on high reliable reference cues was far from optimal, but favouring their own initial judgement. Subjects' adjustment strategy was to anchor on their initial forecast with a weight of about 2/3 and then to adjust according to their perception of reliability of the reference forecast provided.	Conservatism, Anchoring	Confidence, Adjustment	N/A	N/A
Lim and O'Connor (1996)	Sales of product	Yes	Non-experts	People relied too heavily on their initial forecasts compared with the optimal model, and did not seem to learn over time to modify their conservative behaviour.	Conservatism	Adjustment	N/A	N/A
Löffler (1998)	Stock Price & Earning	Yes	Experts	Information contained in the lagged consensus was systematically neglected, which could be explained either by rational boasting or overconfidence; Analysts were too conservative when revising their estimates due either to rational stubbornness or to underreact to new information.	Overconfidence, Conservatism	Confidence, Adjustment	N/A	N/A
Mathews and Diamantopoulous (1990)	Sales of product	Yes	Experts	The forecasters had a direct interest in the outcome and hence preferred some outcomes to others, which lead to optimism biases.	Desired Outcome	Confidence	N/A	N/A

Studies	*Topic*	*Time series*	*Participants*	*Performance*	*Cognitive Sources*	*Taxonomy*	*Debiasing Strategy*	*Debiasing objectives*
Murphy and Daan (1984)	Weather Forecasting	No	Non-experts	Overforecasting, such as a strong tendency for forecast probabilities to exceed observed relative frequencies, for all events, periods and forecasters was observed in the experiment; Overforecasting was reflected in a rapid deterioration in the skill of the forecasts as a function of lead time.	Chance	Statistical	Feedback	Chance
O'Connor, Remus and Griggs (1993)	Sales of product	Yes	Non-experts	The subjects tended to focus only on the position of the last data point on the series in relation to their last forecast; The subjects were trying to read too many signals into a series as it changed; as a consequence, they overreacted to each new value of the series as it was revealed to them.	Chance, Anchoring	Statistical, Adjustment	N/A	N/A
O'Connor, Remus and Griggs (1997)	Time Series Forecasting	Yes	Non-experts	People tended to damp trends and underestimate the steepness of trends in noisy series. Little or no dampening for flat series, underforecasting for up series and overforecasting for down series were identified.	Conservatism	Adjustment	Training on trial series, Feedback	Conservatism
O'Connor, Remus and Griggs (2001)	Judgmental Confidence	Yes	Non-experts	People generally estimated asymmetric confidence intervals where the point forecast was not the midpoint of the estimated interval, and that many of these intervals were grossly skewed.	Underconfidence	Bias in judgmental forecasting	N/A	N/A
Önkal, Yates, Simga-Mugan and Öztin (2003)	Economics	Yes	Both	On average, the differences between expectations and actual matches strongly implicated a particular kind of overconfidence; There was a tendency for forecasts to be slightly underconfident when the professionals were making one-day-ahead predictions; Apparent overconfidence could transform itself into underconfidence depending on when and how forecasters must articulate their confidence; Data formats can greatly affect how accurately forecasters make their	Overconfidence, Underconfidence, Mode	Confidence, Presentation	N/A	N/A

Studies	Topic	Time series	Participants	Performance	Cognitive Sources	Taxonomy	Debiasing Strategy	Debiasing objectives
				predictions and the processes by which they arrive at those predictions.				
Reimers and Harvey (2011)	Sales of product	Yes	Non-experts	Participants' forecasts were significantly higher when the series were described as profits than when they were described as losses; The participants' estimates were biased towards the final observation; People brought to their judgments an inherent bias towards forecasting assuming a moderate degree of positive autocorrelation.	Framing, Anchoring, Correlation	Adjustment, Presentation, Statistical	N/A	N/A
Russo and Schoemaker (1992)	General Knowledge	No	Experts	Of the 2,000-plus individuals to whom a ten-question quiz was given out using 90 percent confidence intervals, less than 1 percent were not overconfident.	Anchoring, Hindsight, Imaginability, Confirmation, Overconfidence	Memory, Confidence, Adjustment	Accelerated feedback, Counter argumentation, Paths to trouble, Paths to the future (Scenario analysis), Awareness alone	Overconfidence
Schnaars and Topol (1987)	Sales of product	Yes	Both	Scenarios did not reduce confidence in the forecasts but had the oppositive effect; Scenario adjustments on stable series were generally more inaccurate than the adjustments without scenarios.	Overconfidence	Confidence	N/A	N/A
Schustack and Sternberg (1981)	General Knowledge	No	Non-experts	The subjects favored confirming information over disconfirming information; They undervalued evidence presented in a negative form relative to information in a positive form; They ignored the base rates of occurrence of outcomes; They were insensitive to notions of sample size and proportionality.	Confirmation, Framing, Desired Outcome, Base rate, Sample	Confidence, Presentation, Statistical	N/A	N/A
Soll and Mannes (2011)	Sport	No	Non-experts	The subjects weighted an opinion about 20 percentage points higher when it was their own, controlling for beliefs and confidence.	Conservatism	Conservatism	N/A	N/A
Song, Boulier and Stekler (2007)	Sport	No	Experts	Addition information provided to experts decreased the forecast accuracy.	Complexity	Situation	N/A	N/A

Studies	***Topic***	***Time series***	***Participants***	***Performance***	***Cognitive Sources***	***Taxonomy***	***Debiasing Strategy***	***Debiasing objectives***
Sterman (1988)	Energy	Yes	Experts	Conservatism appeared for the more distant forecast horizon.	Conservatism	Adjustment	N/A	N/A
Timmers and Wagenaar (1977)	Time Series Forecasting	Yes	Non-experts	When the trend was not linear, judgmental extrapolations became significantly biased due to different beliefs about the nature of the series being forecast.	Selectivity	Confidence	N/A	N/A
Tyebjee (1987)	Sales of product	No	Experts	People who were more deeply involved in a planning exercise were more optimistic about the outcome of the plan than those who were less involved; After engaging in planning activities, the planners considered the uncontrollable environment to be more favorable than before they began the planning task; The bias was resulted from a regression-to-the-mean phenomenon in which low values never have the chance to regress upwards.	Overconfidence, Control, Regression	Confidence, Adjustment	N/A	N/A
Welch, Bretschneider and Rohrbaugh (1998)	Time Series Forecasting	Yes	Non-experts	Over 20% of the individual participants were identified as producing a cue weighted at least 0.15 on the last available data point (Mean 50.10) with a sizable recency effect; Participants in the experimental condition lacking explicit statistical information did not receive a precise indication of periodic growth from a graphic display of time series data; The poor performance of participants' forecasts linked to their inconsistency and incorrect use of cues.	Order, Mode, Inconsistency	Presentation, Situation	Statistical information about long-term levels and trends	Order, Mode, Inconsistency
Wilkie and Pollock (1996)	Economics	Yes	Non-experts	The subjects gave a fairly low mean response, which indicated the underconfidence shown in the performance.	Underconfidence	Confidence	N/A	N/A
Yaniv (2004)	General Knowledge	No	Non-experts	The weights on advice were too low, suggesting that respondents' evaluations of their own knowledge were exaggerated overall; The weight of advice decreased as its distance	Desired Outcome, Anchoring	Confidence, Adjustment	Suggestive feedback, group forecasting	Overconfidence, Anchoring

Studies	Topic	Time series	Participants	Performance	Cognitive Sources	Taxonomy	Debiasing Strategy	Debiasing objectives
				from the initial opinion increased.				
Yaniv (2011)	Education	No	Non-experts	The framing effect was evident in the overall individual responses, as well as in the homogeneous and heterogeneous groups; Homogeneous groups enhanced the framing effect, whereas heterogeneous groups lowered it.	Framing	Presentatio n	N/A	N/A
Zhang (2006)	Business	Yes	Experts	The negative forecast error is consistent with the prior literature and suggests optimism in the overall sample; A negative (positive) forecast error in the bad-news (good-news) subsample indicated that the analyst underreact to the new information due to a conservatism bias.	Overconfi dence, Conservati sm	Confidence , Adjustment	N/A	N/A

A summary of the identified cognitive biases and their frequency of occurrence in the articles reviewed is shown in Table 2-5. As a result, a new cognitive bias – Herding – which differs from Arnott's taxonomy, has been observed in some reviewed studies. Furthermore, two commonly identified confidence biases and another two commonly identified adjustment biases have been identified: overconfidence, desired outcomes, anchoring, and conservatism.

2.2.1 Herding bias

Herding indicates that some forecasters give too much weight to the forecasts generated by other forecasters, leading to excessive concentration of forecasts in group forecasting (Batchelor, 2007; Buehler, Messervey, & Griffin, 2005). Herding is considered a separate bias in the new category of cognitive bias, named group forecasting bias.

Table 2-5 Summary of cognitive biases

	No. of Studies		%	
Reviewed studies	**55 (total)**		**100%**	
Confidence biases	36		65%	
Overconfidence		18		33%
Desired Outcome		7		13%
Adjustment biases	20		36%	
Conservatism		11		20%
Anchoring		8		15%
Statistical biases	9		16%	
Presentation biases	8		15%	
Situation biases	6		11%	
Memory biases	3		5%	
Group Forecasting bias (Herding)	2		4%	

"_" indicates the most identified cognitive biases in the category.

Batchelor (2007) found that institutions that made macroeconomic forecasts showed significant bias towards optimism in consensus forecast. He concluded that herding bias is inevitable in group forecasting because rational forecasters learn the new trend of others' opinions. However, herding bias is not the same as anchoring bias. The latter explains that forecasters tend to adjust their decision based on an initial position, which could be the forecasts published by other forecasters or certain baseline forecasts (e.g., statistical forecasts). Herding bias focuses on a forecaster's attitude to accept others' forecasts without forming his/her own opinion. Information cascade theory can explain how herding bias arises when forecasts are made sequentially by different forecasters. According to Easley and Kleinberg (2010), people tend to abandon their own opinion in favor of inferences based on earlier people's actions. Each published forecast becomes part of the next forecaster's information set; later forecasts are biased towards the early forecasts.

Another theory to explain herding bias is incentive concavity theory, which assumes that the rewards for making accurate but bold forecasts are smaller than the penalties for making inaccurate but bold forecasts. Therefore, herding bias is highly associated with the rewards of accurate forecast and the penalties of inaccurate forecast in the situation of group forecasting with less experienced forecasters involved in the forecasting task. Lamont (2002) assumed that less experienced forecasts herd more than experienced forecasts. His assumption is supported by analyzing the GNP forecasts published by a group of US forecasters in several issues of *BusinessWeek*: less experienced forecasters do produce fewer extreme predictions. Buehler and colleagues' (2005) experiments in project planning prediction made by non-experienced forecasters revealed that predictions generated by the collaborative forecasting approach were more over-optimistic than those generated by individuals. They concluded that collaborative forecasting heightened non-experts' tendency to focus primarily on factors promoting successful task completion, and enhanced their optimistic outlook. Since the background theory of herding bias differs from those of other biases in Arnott's taxonomy, it has been identified as a separate cognitive bias in this study.

Group forecasting is widely adopted as a qualitative forecasting method in the tourism demand forecasting literature. Some existing TDFSS also provide group forecasting functions (Song, Gao, & Lin, 2013; Wöber, 2003). One of the mainstream uses of group forecasting is to combine the forecasting opinions from different sectors related to the tourism industry. Kibedi's (1981) study investigated a group of experts including both academic researchers and practitioners in the tourism industry to predict the relationships between various environments, economic development, and tourism demand. The result shows a general agreement on the influence of various tourism-related environments on tourism demand, as well as the significant contribution of tourism on the economic development of most countries. To forecast Hawaii's visitor arrivals and maximum visitor accommodation, Liu (1988) invited experts from two different sectors to participate in a group forecasting, including both local experts and experts from major overseas suppliers. Their results show a high forecast reliability because of a general consistency between two sectors' opinions. The other mainstream uses of group forecasting is to supplement quantitative forecasts. Tideswell et al. (2001), for example, proposed an integrative approach to combine statistical forecasts with group adjustments. In their study, a group forecasting (a quasi-Delphi survey) process was conducted to further adjust the statistical forecasts generated by timer series models. The empirical results showed that this approach performed well for the South Australia's international markets. However, there is a lack of further investigation about whether, or to what extent, group forecasting participants' opinions are biased by herd effect in tourism demand forecasting.

2.2.2 Commonly identified confidence biases

Overconfidence and desire bias are the most commonly identified confidence biases, occurring in almost half of the reviewed studies. According to Arnott's (1998) definition, overconfidence bias indicates decision makers' overestimation of their ability to solve difficult problems.

In judgmental forecasting, overconfidence bias indicates forecasters' irrational confidence in their forecast accuracy. Three features of overconfidence bias in judgmental forecasting can be identified. First, forecasters who exhibit overconfidence usually over rely on new cues about the factors which can significantly influence the forecasting task. Ayton, Pott, and Elwakili (2007) examined the influence of upcoming events on people's emotions, showing that people always overestimate the impact of events on their emotions, especially when they believe that events will have an impact for a longer period than they actually do. New information in a forecasting task usually helps forecasters construct scenarios, the paths of which the forecasting series might take. In an ideal situation, scenarios can help forecasters better understand the possible outcomes and improve their decision quality. However, Schnaars and Topol (1987) revealed that forecasters always focus on a single, favored scenario instead of the entire set. As the available information is always limited, it is impossible for forecasters to collect complete information about a forecasting task. Any new information received by forecasters would lead to inexplicable overweighting and, at the same time, underweighting the unavailable information. Second, overconfidence is also commonly observed when the participants are asked to make interval forecasts. Forecasters prefer to narrow the intervals in which forecasts might fall when they feel overconfident. In Goodwin, Sinan Gönül, and Önkal's (2013) study, participants tended to narrow both upper and lower limits of forecasts for flat series, and to raise the lower limit for downward trending series. A wider verification of overconfidence bias in interval forecasts was conducted by Russo and Schoemaker (1992), who examined over 2,000 individuals' cognitive behavior and found that more than 99% of forecasters were overconfident. Moreover, if forecasters are asked to evaluate their confidence intervals at different levels (e.g., to evaluate their confidence at the 50% and 90% confidence intervals), forecasters' confidence positively correlates to the level of confidence interval. Du and Budescu (2007) revealed that the smaller the confidence interval forecasters are asked to predict (e.g., 90%), the higher the overconfidence bias they express. Third, the seriousness of the impact of overconfidence bias on judgmental forecast accuracy is negatively correlated to forecasters' experience and education level. Overconfidence bias is more frequently observed among non-experts or inexperienced forecasters (e.g., Aukutsionek & Belianin, 2001; Bolger & Önkal-Atay, 2004). Eroglu and Croxton's (2010) research further revealed that an individual with a college education or a postgraduate degree is less likely to present overconfidence bias.

Desire bias indicates that the probability of a desired outcome being irrationally increased. Forecasters with rich forecasting experience are more likely to establish an expectation before judgmental forecasting. This expectation usually comes from previous forecasting experience and will bias judgments because the forecaster ignores the differences between the previous experience and the forecasting task. Goldfarb, Stekler, and David (2005) revealed that the use of forecasters' previous forecasting experience as the basis for later forecasts results in inaccuracy in economic forecasting. When judging the probability of an event's occurrence, people tend to judge the availability of event contexts or the class of the event in their memory. The search set of event contexts and classes of event significantly influences this heuristic process, thus biasing judgmental forecasts with inefficient search sets (Tversky & Kahneman, 1974). Desired outcome is also affected by forecasters' expectations. For example, Mathews and Diamantopoulous (1990) found that sales forecasters who have a direct interest in a certain outcome always prefer that outcome to others, which leads to optimism adjustments. Fildes and colleagues (2009) also revealed that experienced retailers' sales forecasts were biased by their desired sales targets, which were always set too high and frequently caused wrong-direction adjustments. From the point of view of the availability heuristic, people tend to estimate event probability using their imagination when they cannot find the relevant class of event in their memory (Tversky & Kahneman, 1974). However, human imagination is always restricted. The occurrence of an event will be judged more probable if it is easily imagined. Thus, any cue that constructs forecasters' expectations will bias their judgment towards a desired outcome.

Although overconfidence and desired outcome have been mentioned in the literature of tourism demand forecasting, they are not considered to be specific cognitive biases and have not attracted enough attention from researchers. De Menezes and Vieira (2008) conducted a series of interviews to predict passengers' willingness to pay extra to avoid penalties when changing tickets. They found that the willingness to pay is quite low comparing with their real payment decision, which further reveals that people are prone to overestimate their own ability in a number of settings due to overconfidence. In their study of travel demand, Hubers and Lyons (2013) concluded that forecasters had a tendency to presume that certain development outcomes are more likely than others; but they did not discuss further the influence of such desired outcomes on forecasters' judgment.

2.2.3 Commonly identified adjustment biases

Conservatism and anchoring bias are the most frequently identified adjustment biases, in one-third of the studies reviewed. Conservatism indicates that forecasters tend to overweight their own forecasts and prefer to keep their initial forecasts unchanged, even when significant new cues are available. In the studies reviewed, conservatism bias is exhibited by both experienced and non-experienced forecasters. For example, Batchelor and Dua (1992) examined the judgmental forecasts produced by a number of US institutional forecasters. They measured forecasters' conservatism by a statistical method and concluded that the participants gave too much weight to their own past forecasts, ignoring new information. Harvey and Harries' (2004) study of non-experienced students generated similar findings. Participants gave more weight to their own forecasts than the opinions of advisers. Conservatism bias also seems more significant in time series forecasting, because nine out of 10 reviewed studies identifying conservatism bias were based on time series variables. According to the representativeness heuristic, forecasters always believe that events will tend to regress towards the mean in the forecasting period. So they prefer to dampen the trend of the historical data. However, the illusion of regression may bias forecasts when the trend of a time series is significant but difficult to observe (Harvey & Bolger, 1996).

Furthermore, a common pattern in judgmental forecasting is to begin with an initial position and then adjust based on it. In such a pattern, forecasts may be biased by the anchoring effect (anchoring bias), which indicates that the adjustments from an initial point of forecast are usually in the right direction but insufficient (Arnott, 2006). Most of the time, forecasters have one anchor in their judgments, which is usually the last data point of the forecasting series. The reviewed studies with anchoring bias show that many forecasters only focus on, or at least overweight, the last data point (O'Connor, Remus, & Griggs, 1993; Reimers & Harvey, 2011). Harvey and Bolger (1996) further revealed that when the forecasting series are trended, people with anchoring bias prefer to anchor on the last data point before taking the trend into account. Sometimes, forecasters have more than one anchor and their adjustments are usually positioned between anchors. In Eggleton's (1982) experiment, participants express anchoring bias towards the mean value of the series and the last data point, and their forecasts have been found to lie between these two anchors. Adjustment and anchoring heuristics can be an appropriate forecasting strategy only if continuous feedback is available; otherwise, the amount of adjustment is usually insufficient. Some studies also identified that forecasts are anchored on forecasters' initial opinions (Lim & O'Connor, 1994; Russo & Schoemaker, 1992; Yaniv, 2004). In this case, the effects of anchoring bias are the same as the effects of conservatism bias. So forecasts biased by anchoring on forecasters' initial opinions are classified as displaying conservatism bias in this study.

Adjustment biases are not widely observed in the literature of tourism demand forecasting. One of the few studies regarding conservatism bias was conducted by Hubers and Lyons (2013). Some of their participants' judgmental forecasts for travel demand were primarily based on their past experience, which biased their understanding of the available information. Another interesting study regarding anchoring bias was conducted by Ankomah and Crompton (1992). They investigated the cognitive distances of several airlines, asking a group of forecasters to use judgmental forecasts to predict the distance. The result shows that some forecasters overestimated the length of airlines by misplacing their mental markers at the point where the slant of the line departs from the horizontal, resulting in a cognitive lengthening of the target lines.

As a result, the research gaps identified from the empirical studies about cognitive bias can be summarized as below:

1. Group forecasting method has been widely applied in tourism demand forecasting and the design of TDFSS. However, herding bias and its influence on forecast accuracy have not been examined fully; and no effective strategy has been developed to successfully detect herding bias.
2. Overconfidence and desired outcome have been mentioned in tourism demand forecasting studies but have not attracted enough attention from researchers. There is a lack of study of the situations in which these two cognitive biases occur.
3. Studies about the impact of conservatism and anchoring effects on tourism demand forecasting are quite limited; and no function in the design of TDFSS has been developed to detect these two adjustment biases.

2.3 Empirical studies on debiasing strategies

Besides the commonly identified cognitive biases, some of the reviewed studies also proposed debiasing strategies to reduce the forecasting errors caused by forecasters' irrational cognition (Table 2-4). Generally, the Delphi method is considered an effective strategy to reduce herding bias; the reviewed studies offer three other debiasing strategies to reduce the influence of the four commonly identified cognitive biases on forecast accuracy: optimal forecasting process, training, and feedback.

2.3.1 The Delphi method

The Delphi method is usually conducted through several rounds of survey. In the first round, the panel members in a Delphi survey need to provide independent responses to the forecasting task. In the following rounds, participants are asked to further revise their forecasts according to a summary of all responses collected from the previous round. The participants never meet or communicate with each other during the entire process, guaranteeing their anonymity. The Delphi method is better than traditional group forecasting techniques at avoiding the problems of specious persuasion, undue influence of recognized experts, unwillingness to abandon publicly expressed views, and the bandwagon effect whereby participants are reluctant to state views at odds with a developing consensus (Frechtling, 2012). Herding bias is thus effectively reduced. Early studies of tourism demand forecasting using the Delphi method focused on judgmental forecast of the growth of air traffic and visitor arrivals (English & Kernan, 1976; Liu, 1988). Recently, the application of the Delphi method in this field has been extended to predict tourist expenditure (Landeta, 2006), hotel room demand (Yüksel, 2007), and the impacts of natural and anthropogenic catastrophic events on the tourism industry (Cunliffe, 2002).

2.3.2 Optimal forecasting process

Optimally designed forecasting processes can effectively reduce the occurrence of desired outcome and conservatism biases in judgmental forecasting. Both the tourism demand forecasting and general forecasting literatures suggest that the combination of statistical and judgmental forecasting methods can generally improve forecast performance by reducing the risk of complete forecast failure (Armstrong, 2001; Song & Li, 2008). Based on this principle, a group of studies have compared the cognitive biases occurring in different combination strategies and proposed optimal forecasting processes. Goodwin (2000a) compared three forecast combination methods: (i) statistical correction of judgmental biases; (ii) simple average of judgmental and statistical forecasts; and (iii) correction of judgmental biases followed by combination. Goodwin's result shows the robustness of correcting judgmental bias using Theil's optimal linear correction in reducing forecast error. Error bootstrapping is also suitable for correcting judgmental forecasts biased by forecasters' cognition (Fildes et al., 2009). In another study, Goodwin (2000b) proposed an approach that takes statistical forecast as a starting point and asks the reason for adjustment. This has been verified to be a good way to reduce forecast error caused by overconfidence bias. Taking statistical forecasts as a default aims to remind forecasters of the base rate forecasts and to help them avoid overreaction when new information is available. Asking the reason for adjustments can help forecasters examine their necessity and avoid unnecessary adjustments. Furthermore, small and wrong-side adjustment is usually caused by forecasters' irrational cognition, which should be avoided in the forecasting process (Fildes et al., 2009).

As a result, an optimal forecasting process as a debiasing strategy involves both the order of forecasting (e.g., statistical correction of judgmental biases, and taking statistical forecasts as a default) and the restrictiveness of forecasters' judgmental behaviors (e.g., asking the reason for adjustments, and avoiding small and wrong-side adjustments). The application of an optimal forecasting process in the literature of tourism demand forecasting can be widely identified, especially when statistical and judgmental forecasting processes are combined in the forecasting task (Song & Li, 2008; Witt & Witt, 1995). The general order of forecasting is to produce statistical forecasts of tourism demand, followed by judgmental adjustments using statistical forecasts as a default.

However, a default forecast given to forecasters can also be considered an anchor to bias forecasters' judgment. Statistical forecasts are usually produced either by time series models or by econometric models based on a long-term sample of the historical data. Statistical forecasts thus mainly reflect long-term trends and the seasonal pattern of the forecasting task, as well as its sensitivity to changes in the influencing factors during the forecasting period. Meanwhile, the short-term trends and the potential influence of special events in the forecasting period cannot be captured by statistical forecasting methods (Eroglu & Croxton, 2010; Fildes et al., 2006). Anchoring on statistical forecasts may therefore cause insufficient adjustment. One of the characteristics of tourism demand is its high sensitivity to a variety of special events, including mega activities (such as the Olympic Games and the World Cup), disasters (such as bird flu, earthquake, and tsunami), and terrorist attack (such as 9/11), etc. However, such special events and the sensitivity of tourism demand to such events in the forecasting period cannot be captured by statistical forecasts. Insufficient adjustment caused by anchoring bias may therefore lead to severe error in tourism demand forecasting. In that case, the barrier to using statistical forecasts as the default in the design of TDFSS is the question of how to detect the potential bias of anchoring and how to reduce such bias during the adjustment process. Furthermore, as Goodwin (2000a) suggested, statistical methods can be used to correct forecasters' judgmental bias, but the empirical research on statistical correction of judgmental biases in the area of tourism demand forecasting is quite limited.

Considering the restrictiveness of forecasters' judgmental behaviors, some studies have adopted or suggested scenario analysis as a judgmental forecasting strategy to forecast tourism demand (Prideaux, Laws, & Faulkner, 2003; Yeoman & McMahon-Beattie, 2005). As Hubers and Lyons (2013) emphasized, tourism demand comprises high uncertainty and scenario analysis embraces rather than conceals uncertainty in the prediction of tourism demand. Using scenario analysis, different or even contrasting depictions of possible futures can be fairly judged according to identified principal drivers of scenarios. In the design of TDFSS, scenario analysis is widely adopted as a judgmental forecasting strategy. For example, scenario analysis functions are developed in the HKTDFS in order to help system users fully recognize the uncertainty caused by the changes of several drivers of tourism demand in both the short and long term (Song et al., 2008, 2013).

Another suggestion to restrict forecasters' behavior in judgmental forecasting is to avoid adjustment in the wrong direction (Fildes et al., 2009). The direction of adjustment depends on the position of statistical forecasts, judgmental adjustments, and the real outcomes. The right direction of adjustment is identified when the adjustment and the real outcome are on the same side of the statistical forecast; wrong-direction adjustment occurs when judgmental adjustment and the real outcome are located on the opposite sides of the statistical forecast. The direction of adjustment must be captured by ex-post evaluation when the real outcomes of the forecasting task are available. In the design of TDFSS, however, debiasing strategies are designed with the aim of reducing forecasters' cognitive bias during their judgmental forecasting processes; the right (or wrong) direction of adjustment must be detected before the real outcome is available. A key problem is how to detect the right (or wrong) direction of adjustment during the use of TDFSS. Experimental evidence suggests that, when forecasters' adjustments are made on the basis of events, the accuracy of adjustment is improved only if the information about the event is reliable (Goodwin & Fildes, 1999; Lim & O'Connor, 1996). Fildes et al. (2009) tried to eliminate potential wrong-side adjustments in their experiment using an FSS; the result improved by 50%. However, the generalizability and effectiveness of such a method in the design of TDFSS has not been tested. Furthermore, forecasters' forecasting habits may also influence the direction of adjustment. If a forecaster always expressed optimism in the previous forecasting exercises, for example, the probability that he/she makes adjustment in the wrong direction is increased even if negative but reliable information is available. For individual use of TDFSS, is it reasonable to estimate the wrong direction of adjustment based on one's previous forecasting exercises? In group forecasting, is it reasonable to use the direction of mean (or median) adjustments, or the majority adjustment direction, as the right direction? Further empirical studies are necessary in order to provide reliable answers to these questions.

2.3.3 Training

Among the reviewed studies, two articles proposed that training can also reduce the biases of overconfidence and conservatism (Goodwin et al., 2007; O'Connor, Remus, & Griggs, 1997). This is therefore considered the second debiasing strategy and is normally used with the third debiasing strategy, “feedback” in judgmental forecasting. Training means conducting trial forecasts using artificial data before the real forecasting task. The main purpose of training is to support forecasters’ learning about the forecasting process and data features. For example, the two studies mentioned used artificial data in both trial and real forecasts. In each study, artificial data for the training and the real forecasting task were generated following the same rules. Features of trial time series that forecasters learned in the training can be directly applied in the real forecasting task in order to reduce cognitive bias. Goodwin and colleagues (2007) used an FSS to support participants’ judgmental forecasting, and proposed training in order to reduce cognitive bias by familiarization with the system. However, the application of training in tourism demand forecasting, as well as the influence of training using artificial data before the real forecasting task, have not been found in recent literature.

One of the features of the studies using training as a debiasing strategy is that the series used for both training and real forecasting are all artificial data with similar features of trend, seasonality, variation, and inflation (Goodwin et al., 2007; O’Connor et al., 1997). Trial forecasting using artificial data can thus help forecasters better understand the features of the data used in real forecasting and finally reduce the cognitive biases in their judgments. In real forecasting tasks, however, the features of real time series may not be well simulated by artificial data. In tourism demand forecasting especially, the features of tourism demand are usually unpredictable because of their high sensitiveness to special events. If the features identified from the artificial data differ from real tourism demand, learning from training may seriously bias forecasters' judgment. Therefore, training based on artificial data may not be as helpful as expected in real tourism demand forecasting. Indeed, another source of information that can also help forecasters better understand the features of forecasting tasks, or help forecasters familiar with the FSS, is forecasting performance in the previous forecasting seasons (Fildes et al., 2009); unfortunately, the benefits of forecasters' previous forecasting performance to forecast accuracy in tourism demand forecasting and the design of TDFSS have not been widely studied.

2.3.4 Feedback

Feedback, the third debiasing strategy, is widely suggested with the intention of fostering learning. More than half of the reviewed studies that proposed debiasing strategies suggested using feedback to reduce conservatism, overconfidence, and desired outcome. Regarding content, feedback can be classified as outcome feedback, task feedback, process feedback, and performance feedback (Balzer, Doherty, & O'Connor, 1989; Benson & Önkal, 1992). With the exception of task feedback, these feedback types are widely adopted in the judgmental forecasting literature.

Outcome feedback refers to simple information about the latest outcome of the forecasting task or the realization of a previously predicted event. In the probability forecasting literature, outcome feedback has been proven ineffective in forecast calibration because of the limited information it provides (Benson & Önkal, 1992). Also, because outcome feedback cannot provide information about the key relationships in the environment, it is not suggested for use in probability forecasting (Fischer, 1982). In the time series forecasting literature, outcome feedback is considered the historical data of a forecasting series, and the effectiveness of outcome feedback dependents on the mode of data display (Lawrence, 1983). For example, Lawrence, Edmundson, and O'Connor (1985) showed that graphical display is more helpful in the short term because it clearly reflects the influence of short-term interruptions; tabular display is more helpful in the long term because it concentrates attention on long-term trends. One strategy to provide outcome feedback is to combine these two modes of display and highlight their usefulness in different durations. Therefore, outcome feedback is more useful in reducing mode bias than reducing the four commonly identified cognitive biases in judgmental forecasting.

Process feedback can be evidence perceived by the forecaster, the appropriate forecasting strategy, or information about the forecasts themselves. Yaniv (2004) conducted a series of experiments to evaluate the effects of process feedback on forecast accuracy, one of which is to evaluate the accuracy improvements when forecasters receive feedback about appropriate forecasting strategy (advisor's forecasts) and their initial forecasts. The results show that forecasters learned from this feedback and their adjustments generally resulted in 20% improvement of forecast accuracy compared with their initial forecasts. Lee, Goodwin, Fildes, Nikolopoulos, and Lawrence (2007) examined the contribution of three process feedbacks in time series forecasting, including cases similar to the forecasting task (memory support), similarity evaluation of the retrieved cases (similarity support), and the estimated changes of independent variables (adaptation support). They proved that providing the conjunctive support of these three feedback types significantly improved forecast accuracy in some conditions and at least had no harmful effect on forecasting accuracy in other conditions.

PF is information about performance in forecasters' previous forecast exercises, which normally relates to their previous forecast accuracy. In the studies reviewed, PF is the main feedback debiasing strategy. Simply providing the forecast error of one's previous forecasting exercises in later forecasting tasks is the main application. In Murphy and Daan's (1984) study on weather forecasting, participants' forecasts in the first year were compared with the real outcomes; the forecast accuracy was used as PF for forecasters' second-year forecasting. Both the accuracy and the reliability of forecasts improved in the second year. Bolger and Önkal-Atay (2004) studied the benefits of PF in interval forecasting. The forecasters who expressed overconfidence in their earlier forecasting exercise reduced their cognitive bias based on feedback about their performance, and later forecasts were significantly more accurate. Besides previous forecast error, some studies have also suggested statistical methods for PF. Fildes et al. (2009), for example, used both the forecasts and the forecast errors of forecasters' previous forecasting exercises to regress the real outcome by an error bootstrapping method. Then an optimal forecast based on the estimated relationship between forecast, forecast error, and real outcome was provided as PF intended to minimize forecasters' cognitive bias. PF is usually applied with the training strategy. O'Connor et al. (1997) revealed that providing forecast error in the trial forecast can effectively diminish forecasters' cognitive bias in the real forecast.

PF is the most popular form of feedback provided in the application of the Delphi method in tourism demand forecasting (Garrod & Fyall, 2005; Sheldon & Var, 1985). In order to present the general consensus and the range of participants' opinions, PF in a Delphi survey focuses on the summary of forecasts from all participants rather than any individual.

Moreover, the content of the feedback does not reflect forecast accuracy but certain descriptive statistics, such as the mean, median, standard deviation, skewness, and kurtosis of participants' forecasts in the previous round. The provision of well-organized PF after each round usually shortens the process, which is controlled within four to six rounds of survey before a convergence of group opinion is achieved. Many studies in tourism demand forecasting have highlighted that the success of Delphi forecasting comes from effective PF provided in each survey round (Lee & King, 2009; Liu, 1988; Spenceley, 2008; Yong, Keng, & Leng, 1989). The first application of PF in the design of TDFSS was TourMIS (Croce & Wöber, 2011; Wöber, 2003). Since the year 2000, the system has provided statistical forecasts for the Austrian tourism market with experts' adjustment for system users' consideration. For the first time, a standard online Delphi survey for tourism demand forecasting was developed as a function in HKTDFS. PF on forecasters' adjustments in each round of survey was available during the whole survey session. Besides PF, HKTDFS also provides process feedback for system users' reference in their judgmental forecasting processes, which include the goodness of fit, the statistical significance of coefficients in the statistical forecasting model, and other key statistics of economic modeling. However, there is a lack of further investigation of the influence of PF on forecast accuracy in existing TDFSS.

As a result, the research gaps identified from the empirical studies on debiasing strategies can be summarized as below:

1. Regarding the forecasting process that uses statistical forecasts as a default, forecaster may anchor on statistical forecasts and underweight the influence of

short-term trends and special events when forecasting tourism demand. However, few studies have focused on how to detect the anchoring bias and how to reduce such bias in forecasters' judgmental forecasting process.

2. Judgmental adjustment based on statistical forecasts is widely adopted as the optimal forecasting process in the tourism demand forecasting literature; nonetheless, this process cannot effectively reduce forecasters' anchoring bias in the stage of judgmental adjustment. Statistical correction of cognitive bias after forecasters' judgmental forecasting has not been thoroughly researched in the areas of tourism demand forecasting and the design of TDFSS.
3. As a strategy to restrict forecasters' judgmental adjustment based on statistical forecasts, it is suggested to avoid wrong-side adjustment. However, it is difficult to detect whether the direction of adjustment is right or wrong before the real outcome is available. In tourism demand forecasting and the design of TDFSS, how to detect the right (or wrong) direction of adjustment is little explored.
4. Training based on artificial data may not be helpful, and may even generate a negative influence on forecast accuracy in real forecasting tasks. Forecasting performance (e.g., forecast error) in previous forecasting sessions is made valuable with the aim of better understanding the features of the forecasting task, or the aim of becoming familiar with the FSS, but this practice has not been widely studied.
5. PF is a widely adopted debiasing strategy in tourism demand forecasting and the design of TDFSS. However, there is a lack of study of whether, or to what extent, it contributes as a debiasing strategy.

2.4 Summary

Following the review of cognitive bias theory, the current study adopts Arnott's (2006) taxonomy of cognitive bias, which classified 37 cognitive biases in six categories. Further review of empirical studies in judgmental forecasting, tourism demand forecasting, and the design of TDFSS identifies three research gaps in cognitive bias identification and five research gaps in debiasing strategy development. In the next chapter, the conceptual model and research hypotheses are developed based on the identified research gaps.

3 CONCEPTUAL FRAMEWORK

Research hypotheses are now developed in response to the research gaps identified in the literature review chapter. A conceptual debiasing framework in the design of TDFSS is created to guide the rest of this study.

3.1 Research hypotheses

Among the three research gaps relating to cognitive bias identification, herding bias is the only one connected to group cognition; the other two research gaps focus on individual cognitions. The methods used to investigate group cognition and individual cognition are quite different. Generally, investigating group cognition is much more difficult than investigating individual cognition, for two reasons. First, individual cognition can be identified by investigating single forecasts; the unit of investigation is therefore individual. Group cognition is identified from group decision-making and the research unit is therefore group, rather than individual. Large research samples and a complex investigation process are therefore required. In Buehler et al. (2005), for example, a class of undergraduate students were invited to conduct group decisions. The data about group cognition were collected through two independent experiments and at least two stages (sessions) of investigations were designed in each experiment. In Batchelor's (2007) study, the herding bias was identified by investigating a group of institutions, such as international or national banks, business corporations, trade associations, or research institutes. In each of the investigated institutions, 40–60 professional forecasters were involved, representing a huge sample target. Second, much more time is spent investigating group cognition than individual cognition. Buehler et al. (2005) spent a semester (four months) identifying the features of group cognition; Batchelor's (2007) study ran "over a number of years." As a result, investigation of group cognition requires a complex research design with a large number of participants, and also needs to be conducted over a long time period. In order to maximize the contribution of this study with limited resources, it is reasonable to focus on the identified research gaps relating to individual cognition rather than group cognition.

Conservatism bias, one of the most widely identified individual cognitive biases in the literature, has been widely identified in probability estimation (DuCharme, 1970; Harvey & Harries, 2004; Soll & Mannes, 2011; Yaniv, 2004). In judgmental forecasting of time series data, Poulton (1994) argued that conservatism is just an example of under-adjustment from forecasters' mental anchor; so anchoring bias is a more general account that subsumes the expression of conservatism. According to the theory of behavioral decision-making, some other studies have also argued that it is difficult to distinguish anchoring, conservatism, and egocentrism in people's decision-making (Harvey, 2007; Krueger, 2003; Svenson, 1981). Tourism demand forecasting mainly uses time series data. The current study focuses on investigating the occurrence of commonly identified cognitive biases in tourism demand forecasting, as well as proposing systematic debiasing strategies for TDFSS development; it is not the purpose of this study to investigate controversial cognitive biases in time series forecasting from other cognitive or heuristic behaviors. Therefore, the current study will not investigate conservatism bias or corresponding debiasing strategies. As a result, the current study focuses on filling the research gaps relating to overconfidence bias, desire bias, and anchoring bias, as well as relevant debiasing strategies, in the design of TDFSS. Generally, cognitive biases in the use of TDFSS should be reduced in two stages: cognitive bias detection and debiasing.

3.1.1 Stage I: cognitive bias detection

As reviewed in the literature, the error of tourism demand forecasting can be generally described as:

$$Err = f(S, D, A, O)$$

where *Err* indicates the error of tourism demand forecasting; *S* indicates statistical forecast; *D* indicates the forecaster's desired outcome; *A* indicates the anchoring bias; and *O* indicates the forecaster's confidence in his/her adjustment. Therefore, the forecast error of judgmental adjustment can be divided into five components:

$$Err = \alpha + \beta_1 S + \beta_2 D + \beta_3 A + \beta_4 O + \mu \tag{3.1}$$

where α is a constant term; $\beta_1 S$ indicates the component of statistical forecast bias in the final percentage error of forecast (*Err*); $\beta_2 D$, $\beta_3 A$, and $\beta_4 O$ indicate the components of desired outcome error, anchoring error, and overconfidence error, respectively; and μ is the error term, which is assumed to be a white noise series. In order to keep the measures of forecast errors and cognitive biases on the same level, percentage error (PE) or percentage change (PC) are usually adopted.

Based on the above forecast error equation, four sets of hypotheses can be developed as below.

1) According to Equation (3.3), statistical forecast bias *S* is measured by

$$S = \frac{Q - SF}{Q} * 100\%$$

where Q and SF indicate the real outcome of tourist arrivals and the statistical forecast, respectively. According to Fildes et al. (2006), statistical forecast is robust to estimate regular patterns of time series, such as long-term trends, seasonality, and stable relationships with explanatory variables, while S is mainly caused by short-term trends or the influence of special events that have not been considered in statistical forecasting methods. In such a case, statistical forecast would be biased and statistical forecast error is unavoidable. Ideally, forecasters' expertise and domain knowledge contribute to the identification of such interruptions in judgmental forecasting. Therefore, this component of forecast error should be removed from the adjusted forecasts and not be incorporated into the model of cognitive bias detection. In other words, S_t should not significantly influence the final forecast error after judgmental forecasting. Thus, it is assumed that:

H1. Statistical forecast bias has no influence on the final forecast error after judgmental forecasting.

2) Desire bias reflects forecasters' wishful thinking and the importance of the desired outcome to the forecaster. According to the review of empirical studies, desired outcome is usually generated based on forecasters' expectations without influence from other available information (Arnott, 2006; Hogarth, 1987; Olsen, 1997). Therefore, desire bias (D) can be measured as:

$$D = \frac{Q - DO}{Q} * 100\%$$

where Q denotes the real outcome of tourist arrivals and DO denotes the desired outcome.

As a common identified cognitive bias in judgmental forecasting literature, desire bias is expected to be observed in tourism demand forecasting. Such a bias also contributes to the final forecast error. Thus,

H2a. Desire bias exists in forecasters' judgmental forecasting.
H2b. Desire bias contributes to the final forecast error.

3) As a common judgment pattern, anchoring bias usually occurs when an initial position is given, even if the initial position (anchor) is wrongly determined and forecasters are aware of this (Arnott, 1998). According to the literature, the most popular anchors in judgmental forecasting are the statistical forecast and the latest observation of the task time series (Eggleton, 1982; Eroglu & Croxton, 2010; O'Connor et al., 1993). Therefore, it is believed that the adjustments based on the statistical forecast and the historical data of tourism demand may cause anchoring bias in forecasters' judgmental forecasts when the latest data is provided. Furthermore, desire bias is one of the cognitive biases to be investigated in this study. The identification of one's desired outcome is driven by one's domain knowledge and expertise, instead of any mental anchor. However, once the desired outcome is provided, it is easily used as an initial value for further adjustments, and further adjustments may be biased by the desired outcome as an anchor. Therefore, it is believed that anchoring bias may occur regarding the availability of the statistical forecast, the last outcome of the historical data, and the desired outcome in tourism demand forecasting. Following the method

of measuring anchoring bias in Eroglu and Croxton's (2010) study, three types of anchoring bias can be measured as:

$$\begin{cases} A_S = \frac{F-SF}{SF} \\ A_D = \frac{F-DO}{DO} \\ A_L = \frac{F-LO}{LO} \end{cases} \tag{3.2}$$

where A_S, A_D, and A_L denote the biases anchoring on the statistical forecast, the desired outcome, and the latest observation of tourism demand, respectively; F denotes the judgmental forecast of tourism demand; and SF and LO denote the statistical forecast and the latest observation of the historical data, respectively.

In order to incorporate these three types of anchoring bias into Equation (3.1), three dummy variables are developed:

$$I_S = \begin{cases} 1, if \min(|F-SF|, |F-DO|, |F-LO|) = |F-SF| \\ 0, otherwise \end{cases}$$

$$I_D = \begin{cases} 1, if \min(|F-SF|, |F-DO|, |F-LO|) = |F-DO| \\ 0, otherwise \end{cases}$$

$$I_L = \begin{cases} 1, if \min(|F-SF|, |F-DO|, |F-LO|) = |F-LO| \\ 0, otherwise \end{cases}$$

where I_S, I_D, and I_L denote the occurrence of anchoring bias on the statistical forecast, the desired outcome, and the latest observation, respectively. Therefore, Equation (3.1) can be further revised by:

$$Err = \alpha + \beta_1 S + \beta_2 D + \beta_3 I_S A_S + \beta_4 I_D A_D + \beta_5 I_L A_L + \beta_6 O + \mu \tag{3.3}$$

When the judgmental forecast of tourism demand for one tourism market is made by a forecaster, only one type of anchoring bias would normally be identified with one exception that the judgmental forecast is located at the middle point of two starting points (e.g., $F = \frac{SF+LO}{2}$). In that occasion, both statistical forecast anchor and latest observation anchor are identified ($I_S = I_L = 1$). Three types of anchoring bias are expected to be observed in tourism demand forecasting and can contribute to the final forecast error. Therefore,

H3a. Anchoring bias in <u>statistical forecast</u> is unavoidable in forecasters' judgmental forecasting.

H3b. Anchoring bias in <u>desired outcome</u> is unavoidable in forecasters' judgmental forecasting.

H3c. Anchoring bias in the <u>latest observation</u> is unavoidable in forecasters' judgmental forecasting.

H3d. Anchoring bias in <u>statistical forecast</u> contributes to the final forecast error.

H3e. Anchoring bias in <u>desired outcome</u> contributes to the final forecast error.

H3f. Anchoring bias in the <u>latest observation</u> contributes to the final forecast error.

4) Overconfidence bias indicates that "the ability to answer difficult or novel questions is often over-estimated" (Arnott, 2006). In the reviewed empirical studies, overconfidence bias can be measured by the difference between (i) the confidence level given by a forecaster after his/her adjustment is made and (ii) the percentage of correct adjustment (Bolger & Önkal-Atay, 2004). In practice, confidence level can be measured by five-point Likert scales and then transformed into the form of a percentage of confidence (confidence score divided by five). The percentage of correct adjustment is reflected by the proportion of adjusted forecasts that are more accurate than statistical forecasts. Therefore, overconfidence bias (O) is measured as:

$$O_i = Confidence_i - \frac{\sum_{i=1}^{m} I_i}{m}, \quad I = \begin{cases} 1, if\,|pe(F_i)| < |pe(SF_i)| \\ 0, if\,|pe(F_i)| \geq |pe(SF_i)| \end{cases} \tag{3.4}$$

where $Confidence_i$ indicates the forecaster's confidence percentage about his/her adjustment of the forecast of tourism demand series i; m indicates the number of tourism demand series to be forecast in a forecasting season; I indicates the amount of correct adjustments; $pe(F_i)$ indicates the percentage error of adjustment; and $pe(S_i)$ indicates the percentage error of statistical forecast.

As a commonly identified cognitive bias in the judgmental forecasting literature, overconfidence bias is expected to be observed in tourism demand forecasting and to contribute to the final forecast error. Thus,

H4a. Overconfidence is unavoidable in forecasters' judgmental forecasting.
H4b. Overconfidence bias contributes to the final forecast error.

3.1.2 Stage II: anchoring bias reduction

According to Equation (3.3) and the hypotheses developed in the stage of cognitive bias detection, forecast error can be decomposed into different components, each of which is driven by a cognitive bias or statistical forecast bias. Assuming that forecasters' cognitive behavior will remain unchanged between two closed forecasting seasons, the detected cognitive errors made by the forecaster, as well as the statistical forecast bias, in the previous forecasting season can be used to indicate the forecaster's cognitive bias in the next forecasting season. Fildes et al. (2006) emphasized that two general strategies can be applied for debiasing in the design of FSS: informative guidance and suggestive guidance. An FSS could provide the system user different kinds of information relating to the forecasting task, including one's forecasting performance using the system in the past (PF), as informative guidance. In this study, the information about forecasters' cognitive bias detected in the previous forecasting season can be provided as PF to forecasters in the form of mean forecast error. The revised forecasts based on such performance guidance are expected to be more accurate since the error of related cognitive bias should be excluded from the revised forecasts. Suggestive guidance means that the FSS automatically produces the suggested forecasts and directly provides them to forecasters for their decision-making. When a forecaster's judgmental forecasts in the previous forecasting season and the real outcomes of tourist arrivals are available, the system can calculate his/her cognitive biases following a set of pre-defined algorithms. Based on the assumption that a forecaster's cognitive behavior would not present significant changes between two closed forecasting seasons, the system would combine the detected cognitive bias with the forecaster's new judgmental forecasts in the following forecasting season. In this study, three sets of algorithms are proposed to reduce each kind of cognitive bias in tourism demand forecasters' judgmental forecasting process. The accuracy of the suggested forecasts produced by such algorithms are compared with forecasters' non-aided judgments, and their revisions according to the PF. System-suggested forecasts are expected to perform better than forecasters' judgmental forecasts. Regarding these two debiasing strategies, three sets of hypotheses can be developed in Stage II.

5) According to the desire bias detected in the previous forecasting season, the extent to which such bias influences the adjustment can be measured by the corresponding forecast error. Therefore, the mean forecast error caused by a forecaster's desired outcome, or so-called desire error, in the previous forecasting season is expected to be used as PF in the following forecasting season. According to Equation (3.3), the percentage error caused by a forecaster's desired outcome is measured by the

component $\beta_2 D$; therefore, the mean desire error in one's previous forecasting season can be calculated according to the following equation:

$$Mean(Desire\ Error) = \frac{1}{m}\sum_{i=1}^{m} \beta_2 D_i * Q_i \tag{3.5}$$

Thus,

H5a. Feedback of the mean desire error in a forecaster's previous forecasting season reduces desire bias in the following forecasting season.

Desire error (Err_D), the difference between the real outcome and the desired outcome, is calculated as:

$$Err_D = Q - DO$$

If Err_D is predictable, the ideal forecast of $\hat{Q}$ should be:

$$\hat{Q} = DO + \widehat{Err}_D$$

Based on the assumption of unchanged cognitive behavior in two closed forecasting seasons, $\widehat{Err}_D$ can be estimated by the mean of desire error produced by a forecaster in his/her previous forecasting season:

$$\hat{Q}_{i,t+1} = DO_{i,t+1} + \widehat{Err}_{D_{t+1}} = DO_{i,t+1} + Mean(Err_{D_t})$$

where i indicates the tourism demand series to be forecast. According to Equation (3.5), the system-suggested forecasts regarding a forecaster's desire bias (D^S), which is actually the ideal forecast of $\hat{Q}_{i,t+1}$ with desire error eliminated, can be further specified as:

$$D^S_{i,t+1} = \hat{Q}_{i,t+1} = DO_{i,t+1} + \frac{1}{m}\sum_{i=1}^{m} \beta_2 D_{i,t} * Q_{i,t} \tag{3.6}$$

The hypothesis related to this suggestive guidance is developed as:

H5b. Suggested forecast with desire error correction is the most accurate adjustment, better than a forecaster's unaided adjustment and the adjustment based on the corresponding PF.

6) According to the three types of anchoring bias detected in the previous forecasting season, the extent to which such bias influences the adjustment can be measured by the corresponding forecast error. Therefore, the mean forecast error caused by a forecaster's anchoring bias, or so-called anchoring error, in the previous forecasting season is expected to be used as PF in the following forecasting season. The mean anchoring error in a forecaster's previous forecasting season can be calculated according to the following equation:

$$\begin{cases} Mean(A_S) = \frac{1}{N(I_{S_i}=1)} \sum_{i=1}^{N(I_{S_i}=1)} Anchoring\ Error_{S_i} = \frac{1}{N(I_{S_i}=1)} \sum_{i=1}^{N(I_{S_i}=1)} \beta_3 I_{S_i} A_{S_i} * Q_{i,t} \\ Mean(A_D) = \frac{1}{N(I_{D_i}=1)} \sum_{i=1}^{N(I_{D_i}=1)} Anchoring\ Error_{D_i} = \frac{1}{N(I_{D_i}=1)} \sum_{i=1}^{N(I_{D_i}=1)} \beta_4 I_{D_i} A_{D_i} * Q_{i,t} \\ Mean(A_L) = \frac{1}{N(I_{L_i}=1)} \sum_{i=1}^{N(I_{L_i}=1)} Anchoring\ Error_{L_i} = \frac{1}{N(I_{L_i}=1)} \sum_{i=1}^{N(I_{L_i}=1)} \beta_5 I_{L_i} A_{L_i} * Q_{i,t} \end{cases} \quad (3.7)$$

where $N(I_{S_i} = 1)$, $N\left(I_{D_i} = 1\right)$, and $N(I_{L_i} = 1)$ indicate the number of a forecaster's adjustments to all tourism demand series that close to the statistical forecast, the desired outcome, and the latest observation, respectively. Thus:

H6a. Feedback of the mean anchoring error in the statistical forecast in the previous forecasting season reduces anchoring bias in the following forecasting season.

H6b. Feedback of the mean anchoring error in the desired outcome in the previous forecasting season reduces anchoring bias in the following forecasting season.

H6c. Feedback of the mean anchoring error in the latest observation in the previous forecasting season reduces anchoring bias in the following forecasting season.

After Equation (3.3) is estimated using the forecaster's judgmental forecasts and the real outcomes of tourist arrivals in the previous forecasting season, the estimated coefficients are also valid to estimate the influence of statistical forecast error (β_1), desire bias (β_2), three types of anchoring bias ($\beta_3, \beta_4, \beta_5$), and overconfidence bias (β_6) on the final forecast error in the following forecasting season. Therefore,

$$Err_{i,t+1} = \alpha + \beta_1 S_{i,t+1} + \beta_2 D_{i,t+1} + \beta_3 I_S A_{S_{i,t+1}} + \beta_4 I_D A_{D_{i,t+1}} + \beta_5 I_L A_{L_{i,t+1}} + \beta_6 O_{i,t+1} + \mu_{i,t+1} \quad (3.8)$$

The percentage error associated with three anchors can be extracted from Equation 3.8 as below:

$$Err_{i,t+1}(A) = \beta_3 I_S A_{S_{i,t+1}} + \beta_4 I_D A_{D_{i,t+1}} + \beta_5 I_L A_{L_{i,t+1}}$$

As suggestive guidance to eliminate anchoring bias, the system-suggested forecasts in this situation should ideally be given with no percentage error associated with anchoring bias ($Err_{t+1}(A)$ =0). Therefore, the percentage error of (system-suggested) forecasts with anchoring bias eliminated should be given as:

$$Err'_{i,t+1} = \alpha + \beta_1 S_{i,t+1} + \beta_2 D_{i,t+1} + \beta_6 O_{i,t+1} + \mu_{i,t+1} \quad (3.9)$$

The following equation is a combination of Equation (3.8) and Equation (3.9):

$$Err_{i,t+1} - Err'_{i,t+1} = \beta_3 I_S A_{S_{i,t+1}} + \beta_4 I_D A_{D_{i,t+1}} + \beta_5 I_L A_{L_{i,t+1}} \quad (3.10)$$

Also because

$$Err_{i,t+1} - Err'_{i,t+1} = \frac{Q_{i,t+1} - F_{i,t+1}}{Q_{i,t+1}} - \frac{Q_{i,t+1} - F\_A^S_{i,t+1}}{Q_{i,t+1}} = \frac{F\_A^S_{i,t+1} - F_{i,t+1}}{Q_{i,t+1}}$$

where $F_{i,t+1}$ indicates forecasters' unaided judgmental forecast of series *i* at time *t*+1; $F\_A_{i,t+1}^{S}$ indicates system-suggested forecasts with anchoring bias eliminated. Equation (3.10) can be further transferred into the following form:

$$\frac{F\_A_{i,t+1}^{S}-F_{i,t+1}}{Q_{i,t+1}} = \beta_3 I_S A_{S_{i,t+1}} + \beta_4 I_D A_{D_{i,t+1}} + \beta_5 I_L A_{L_{i,t+1}} =$$

$$\begin{cases} \beta_3 \frac{F_{i,t+1}-SF_{i,t+1}}{SF_{i,t+1}}, if\ I_{S_{i,t+1}} = 1 \\ \beta_4 \frac{F_{i,t+1}-DO_{i,t+1}}{DO_{i,t+1}}, if\ I_{D_{i,t+1}} = 1 \\ \beta_5 \frac{F_{i,t+1}-LO_{i,t+1}}{LO_{i,t+1}}, if\ I_{L_{i,t+1}} = 1 \end{cases}$$

Also because the system-suggested forecasts are considered as ideal forecasts of $\hat{Q}_{i,t+1}$ with anchoring error eliminated, $F\_A_{i,t+1}^{S}$ can be calculated as below:

$$F\_A_{i,t+1}^{S} = \hat{Q}_{i,t+1} = \begin{cases} \frac{F_{i,t+1}}{1-\beta_3 \frac{F_{i,t+1}-SF_{i,t+1}}{SF_{i,t\_1}}}, if\ I_{S_{i,t+1}} = 1 \\ \frac{F_{i,t+1}}{1-\beta_4 \frac{F_{i,t+1}-DO_{i,t+1}}{DO_{i,t+1}}}, if\ I_{D_{i,t+1}} = 1 \\ \frac{F_{i,t+1}}{1-\beta_5 \frac{F_{i,t+1}-LO_{i,t+1}}{LO_{i,t+1}}}, if\ I_{L_{i,t+1}} = 1 \end{cases} \tag{3.11}$$

The hypotheses related to this suggestive guidance are developed below:

H6d. System-suggested forecast with <u>statistical forecast anchor</u> correction is the most accurate forecast, better than unaided forecast and adjustment based on the corresponding PF.

H6e. System-suggested forecast with <u>desire anchor</u> correction is the most accurate forecast, better than unaided forecast and adjustment based on the corresponding PF.

H6f. System-suggested forecast with <u>latest observation anchor</u> correction is the most accurate forecast, better than unaided forecast and adjustment based on the corresponding PF.

7) According to the overconfidence bias detected in the previous forecasting season, the extent to which such bias influences the adjustment can be measured by the corresponding forecast error. Therefore, the mean forecast error caused by forecasters' overconfidence, or so-called overconfidence error, in the previous forecasting season is expected to be used as PF in the following forecasting season. The mean overconfidence error in the previous forecasting season can be calculated according to the following equation:

$$Mean(O) = \frac{1}{m}\sum_{i=1}^{m} Overconfidence\ Error_i = \frac{1}{m}\sum_{i=1}^{m} \beta_3 O_i * Q_{i,t} \qquad (3.12)$$

Thus,

H7a. Feedback of the mean overconfidence error in the previous forecasting season reduces overconfidence bias in the following forecasting season.

According to Equation (3.8), the percentage error associated with overconfidence bias can be extracted as:

$$Err_{i,t+1}(OVE) = \beta_6 O_{i,t+1} = \beta_6 * (Confidence_{i,t+1} - Mean(O)_{t+1}) \qquad (3.13)$$

As suggestive guidance to eliminate overconfidence bias, the system-suggested forecasts in this situation should ideally be given with no percentage error of $Err_{i,t+1}(OVE)$. Therefore, $Err_{i,t+1}(OVE)$ should be equal to zero and the percentage error of system-suggested forecasts regarding the elimination of overconfidence bias should be:

$$Err''_{i,t+1} = \alpha + \beta_1 S_{i,t+1} + \beta_2 D_{i,t+1} + \beta_3 I_S A_{S_{i,t+1}} + \beta_4 I_D A_{D_{i,t+1}} + \beta_5 I_L A_{L_{i,t+1}} + \mu_{i,t+1} \qquad (3.14)$$

Combing Equation (3.8) with Equation (3.14) we can see that

$$Err_{i,t+1} - Err''_{i,t+1} = \beta_6 O_{i,t+1} \tag{3.15}$$

Also,

$$Err_{i,t+1} - Err''_{i,t+1} = \frac{Q_{i,t+1} - CF_{i,t+1}}{Q_{i,t+1}} - \frac{Q_{i,t+1} - F\_O^S_{i,t+1}}{Q_{i,t+1}} = \frac{F\_O^S_{i,t+1} - CF_{i,t+1}}{Q_{i,t+1}}$$

where $CF_{i,t+1}$ indicates the forecaster's judgmental forecast of tourism demand series *i* at time *t*+1 with desire bias and anchoring bias corrected but overconfidence bias still involved; $F\_O^S_{i,t+1}$ indicates system-suggested forecasts with overconfidence bias eliminated. Equation (3.15) can therefore be transferred as

$$\frac{F\_O^S_{i,t+1} - CF_{i,t+1}}{Q_{i,t+1}} = \beta_6 O_{i,t+1} = \beta_6 * (Confidence_{i,t+1} - Mean(O)_{t+1}) \tag{3.16}$$

Based on the assumption that a forecaster's cognitive behavior in the closed two forecasting season is not significantly different, his/her mean overconfidence bias in the current forecasting season can be estimated by his/her mean overconfidence bias in the previous forecasting season. Therefore,

$$Mean(\hat{O})_{t+1} = Mean\ (O)_t = \frac{1}{m}\sum_{i=1}^{m} I_{i,t}$$

$F\_O^S_{i,t+1}$ is produced as an ideal forecast of $\hat{Q}_{i,t+1}$ with overconfidence error eliminated. Therefore, $F\_O^S_{i,t+1}$ can be calculated as:

$$F\_O^S_{i,t+1} = \hat{Q}_{i,t+1} = \frac{CF_{i,t+1}}{1 - \beta_6 (Confidence_{i,t+1} - \frac{1}{m}\sum_{i=1}^{m} I_{i,t}} \tag{3.17}$$

The hypothesis related to this suggestive guidance is developed as:

H7b. System-suggested forecast with overconfidence error corrected is the most accurate forecast, better than comparing the statistical forecast, unaided forecast, and the adjustment based on the PF of the mean overconfidence error in the previous forecasting season.

3.2 A conceptual debiasing framework

According to the hypotheses developed in the previous section, a conceptual debiasing framework is proposed with the aim of improving the accuracy of judgmental forecasting in the design of TDFSS (Figure 3-1).

The left side of this framework describes the source of cognitive bias. In judgmental forecasting, cognitive bias is mainly produced by the historical data of tourist arrivals, statistical forecasts at the forecasting point, forecasters' expectations or desired outcome for the market, and their confidence in judgmental forecasts. The middle part represents the leading indicators of five cognitive biases. To test the existence of statistical forecast error after judgmental forecasting, the forecast error associated with statistical forecast is also identified in this part. For each kind of cognitive bias, forecasters' unaided judgmental forecast, the PF-based revision, and system-suggested forecast are identified. Hypotheses developed in the processes of cognitive bias detection and debiasing construct the relationships between each component of the framework.

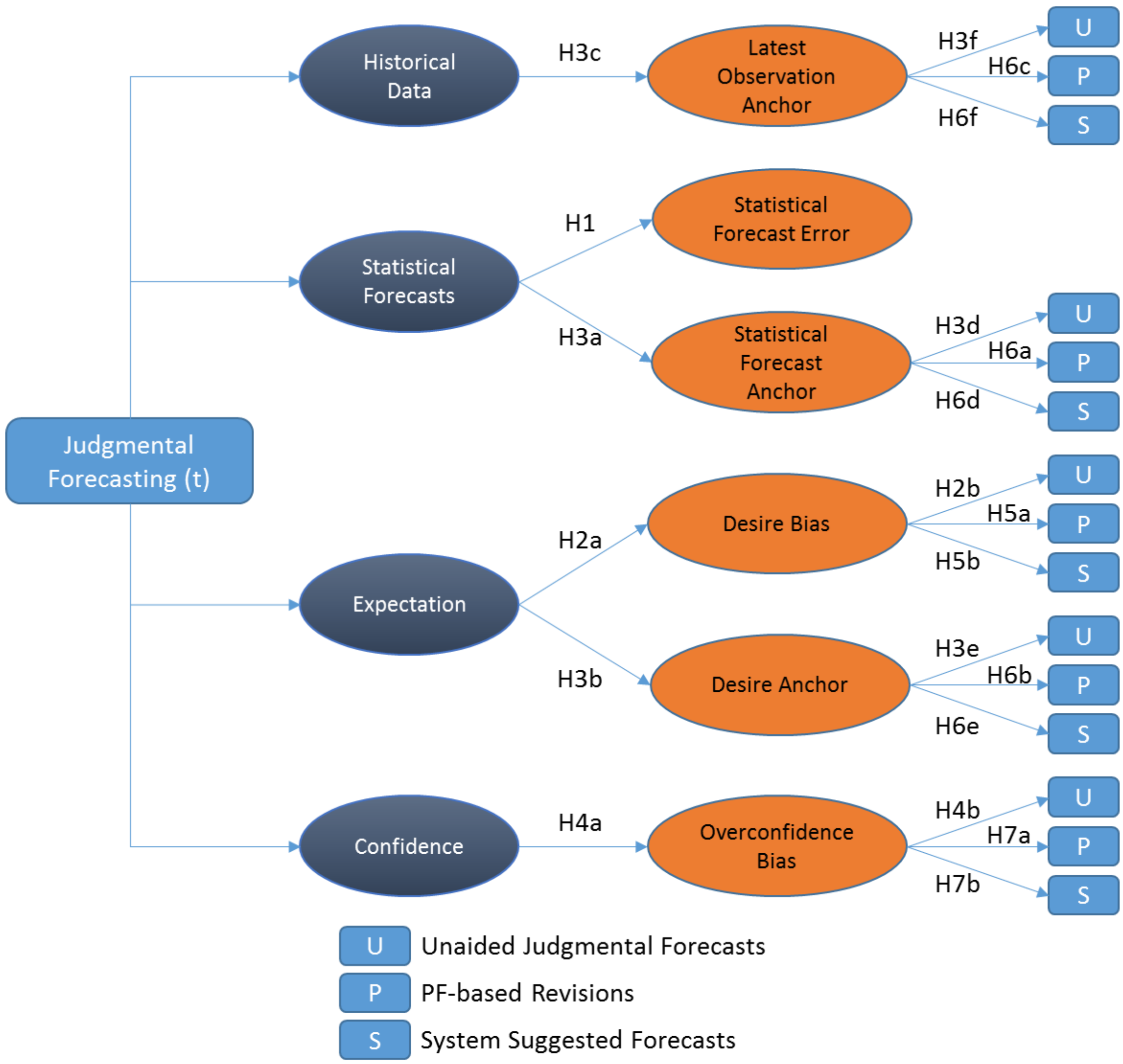


Figure 3-1 A conceptual debiasing framework in the design of TDFSS

3.3 Summary

This chapter has focused on the development of hypotheses regarding cognitive bias and debiasing strategies in the design of TDFSS. As a result, 21 hypotheses in seven groups have been developed to be tested in this study. The first hypothesis is about the influence of statistical forecast bias on the final forecast error, which has nothing to do with cognitive bias. However, it is the foundation upon which to construct the model of cognitive bias detection. It is therefore proposed as the first hypothesis to be tested before others. The following 10 hypotheses are about cognitive bias detection and the last 10 are about the effectiveness of the proposed debiasing strategies.

Based upon the hypotheses developed, a conceptual debiasing framework in the design of TDFSS was proposed at the end of this chapter. To verify the proposed conceptual framework, the rest of this study focuses on testing the proposed hypotheses.

4 EXPERIMENT DESIGN

The hypotheses and the conceptual framework developed in the previous chapter focus on how to reduce forecasters' cognitive bias in the use of TDFSS. This study aims to test the developed hypotheses and to verify the proposed conceptual framework. Generally, a research method can proceed following two approaches: an inductive approach or a deductive approach (Haag, Häggman, & Mattsson, 2010). In an inductive approach, the research starts with the collection of empirical data without any expectations and ends with conclusions and theories drawn from the findings. Following a deductive approach, the research starts from specific theories and concludes with empirical data. In other words, a deductive approach starts by creating expectations, then proceeds with the collection of empirical data and verification of whether the empirical data support the expectations (Jacobsen, 2002). In a deductive approach, the expectations are expressed by hypotheses, which can be empirically tested in specific cases (Patel & Davidson, 2003). A deductive research approach is adopted in this study.

The selection of a qualitative or quantitative research method also depends on the nature of the study and the results expected (Patel & Davidson, 2003). According to the proposed conceptual debiasing framework for TDFSS development, this study has to generate a conclusion on whether the methods of cognitive bias detection and debiasing can effectively reduce forecasters' cognitive bias in their use of TDFSS. The performance of judgmental forecasting with and without debiasing needs to be compared and analyzed. Therefore, a quantitative approach based on a series of experiments is deemed the most appropriate way to fulfill the research purpose because the hypotheses developed in this model can be judged from the differences between aided and unaided adjustments.

Based on the above, the data for hypotheses testing in this study is appropriate for collection from experiment. In this study, the experiment simulates regular tourism demand forecasting tasks in reality. Because the influence of conservatism bias on tourism demand forecasting is not investigated in this study, the experiment focuses on one-step-ahead forecasting of tourism demand. Therefore, Equation (3.3) can be further specified as:

$$E_{i,j} = a_j + \beta_{1j} S_{i,j} + \beta_{2j} D_{i,j} + \beta_{3j} I_{S_{i,j}} A_{S_{i,j}} + \beta_{4j} I_{D_{i,j}} A_{D_{i,j}} + \beta_{5j} I_{L_{i,j}} A_{L_{i,j}} + \beta_{6j} O_{i,j} + \mu_j \quad (4.1)$$

where the subscripts $i \in (1, m)$ and $j \in (1, n)$ denote the tourism demand series to be forecast and the invited forecaster in this study, respectively; m is the number of tourism demand series to be forecast in the experiment; n is the number of forecasters participating in the experiment. Seven components located on the right side of the equation denote the decomposed forecast errors that are driven by statistical forecast bias, desire bias, three types of anchoring bias, and overconfidence bias.

Based on Equation (4.1), forecast errors generated by aided and unaided adjustments using TDFSS need to be collected in a two-round experiment according to the following steps:

Round I:

Step 1. Desired Outcome

Step 2. Judgmental Forecasting

Step 3. Confidence Level in Forecasting

Round II:

Step 1. Desired Outcome

Step 2. Correction of Desired Outcome

Step 3. Judgmental Forecasting

Step 4. Correction of Judgmental Forecasting

Step 5. Confidence Level in Forecasting

Step 6. Correction of Forecasting Regarding Confidence Level

In the first round of forecasting, three sets of data regarding forecasters' judgmental forecasts are collected from three steps. First, forecasters' desired outcome without any decision-making aids are collected as the first set of data. In the second step, forecasters are asked to make their adjustment according to the one-step-ahead statistical forecasts and the real outcomes of tourist arrivals in the history; then the adjustments are collected as the second set of data. Finally, forecasters rate their confidence in their adjustments, which information is collected as the third set of data. According to the data collected from the first round of forecasting, Equation (4.1) can be estimated and the leading indicators of desire, anchoring, and overconfidence biases can be identified. PF and suggestive guidance regarding these leading indicators are available for forecasters' reference in the second round of forecasting.

In the second round of forecasting, the first, third, and fifth steps of adjustment are conducted following the same process in the first round of forecasting; the difference between these two rounds of adjustment is that the follow up adjustments are conducted with PF provided after each of these three steps of adjustment. There are thus six steps to judgmentally forecasting tourism demand in the second round; ultimately, six sets of data about forecasters' adjustment are collected.

4.1 Participants

In order to collect sufficient data for hypotheses testing, it is anticipated to invite at least 60 professional forecasters to participate in the experiment (n>60), who meet any of the following requirements:

- Postgraduate students or students in the same level who are studying tourism and hospitality;
- Currently working in supply chain management, marketing, or business planning in the tourism and hospitality industry for at least three months;
- Familiar with tourism demand forecasting and its characteristics;
- Currently using or have experience in using FSS or software in tourism demand forecasting; and
- Possessing expertise in quantitative forecasting methods.

Before the experiment, participants are informed of the confidentiality of their responses, and that all of their forecasts and adjustments are reported anonymously. Each participant is informed that he/she is free to withdraw from the study at any time.

4.2 Empirical data

In the experiment, real data about international tourist arrivals are used as the indicator of tourism demand. The level of forecasting is specific D-O pairs. As shown in Table 4-1, there are 10 D-O pairs of tourist arrivals to be forecast in the experiment. The data of real tourist arrivals for these 10 D-O pair markets are collected from the World Tourism

Organization (UNWTO) database. Since this experiment focuses on the identification and debiasing of three cognitive bias (desire, anchoring, and overconfidence), two strategies are adopted when preparing the empirical data in order to avoid influences from other biases. First, professional forecasters may have a strong memory about the latest real outcomes of tourist arrivals for a specific D-O pair. If the forecast period covers the recent years of tourist arrivals, forecasters' judgmental forecasting behavior could be seriously biased by their memory. Therefore, the years 2006 and 2007 are defined as forecasting periods for all D-O pairs in this experiment. Second, previous studies revealed that when an event occurred in certain circumstances, the likelihood of such an event happening in the same circumstances will be overestimated; this is called similarity bias (Arnott, 2006; Joram & Read, 1996). In tourism demand forecasting, it is reasonable to believe that people tend to overestimate the similarity of the features of different tourism demand series; however, similarity bias is not one of the target cognitive biases investigated in this study. To avoid the potential influence of similarity bias in the experiment, the time period of historical data and the forecast period are different in each of the tourism demand series. For example, the historical data on tourist arrivals from Chinese Taipei to the Hong Kong SAR were collected for the period 1995–2005, while tourist arrivals from China to Japan were collected for 1997–2005. As a result, the upper limit of parameter i in Equation (4.1) is 10 (m=10).

Table 4-1 Historical data

Destination	Top 5 Market Origin	Historical Data
Australia	China	2002-2008
Hong Kong SAR	Chinese Taipei	1995-2009
Hong Kong SAR	Japan	1995-2008
Hong Kong SAR	Macau SAR	1995-2005
Japan	Chinese Taipei	1997-2008
Japan	China	1997-2007
Japan	Hong Kong SAR	1997-2005
Chinese Taipei	Japan	1997-2009
Chinese Taipei	USA	1997-2008
Chinese Taipei	Korea (ROK)	1997-2005

4.3 Statistical forecasting

According to the forecasting process followed in this study, statistical forecasts are provided in the first place, which mainly reflect the long-term trends, seasonality, and other regular patterns of tourism demand. Therefore, the time series forecasting method was adopted to produce statistical forecasts. One of the most recent innovations in time series forecasting is Hyndman and Khandakar’s (2008) state space exponential smoothing (ETS) approach. As an automatic time series modeling and forecasting approach, ETS is also appropriate for the development of FSS. Considering the advantages and its suitability to being adopted in the design of TDFSS, the ETS approach is adopted in this study to produce statistical forecasts.

4.3.1 Exponential smoothing

According to Hyndman and Khandakar (2008), a time series can be decomposed into two components regarding the characteristics of seasonality and trend. For seasonality, a time series may contain an additive pattern, a multiplicative pattern, or no seasonal pattern; for trend, a time series may contain an additive pattern, a multiplicative pattern, a damped trend of these two patterns, or no trend pattern. Traditional ES methods to forecast time series have 15 variations, the different combinations of a time series' seasonality and trend patterns determining which is used (Table 4-2).

Table 4-2 Traditional exponential smoothing methods

Trend pattern	Seasonal pattern		
	None (N)	Additive (A)	Multiplicative (M)
None (N)	TS(N,N)	TS(N,A)	TS(N,M)
Additive (A)	TS(A,N)	TS(A,A)	TS(A,M)
Multiplicative (M)	TS(M,N)	TS(M,A)	TS(M,M)
Additive damped (Ad)	TS(Ad,N)	TS(Ad,A)	TS(Ad,M)
Multiplicative damped (Md)	TS(Md,N)	TS(Md,A)	TS(Md,M)

Using the ES method for additive trend pattern and additive seasonal pattern (Holt-Winter's additive method) for illustration, the time series forecast using ES can be generated with the following equation:

$$\hat{Y}_{t+h|t} = l_t + b_t h + s_{t-m+h_m^+}$$

where

$$\begin{cases} l_t = a(Y_t - s_{t-m}) + (1-a)(l_{t-1} + b_{t-1}) \\ b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\ s_t = \gamma(Y_t - l_{t-1} - b_{t-1}) + (1-\gamma)s_{t-m} \end{cases}$$

$\hat{Y}_{t+h|t}$ denotes the h-steps ahead forecast; l_t denotes the level of the series; b_t denotes the trend pattern; s_t denotes the seasonal pattern; m indicates the length of seasonality; and $h_m^+ = [(h-1)\ mod\ m] + 1$. Normally, the initial data of l_0, b_0, and s_0, … , s_{1-m} are directly collected from the observed data. A comprehensive description of all variations of ES can be found in Hyndman and Khandakar (2008).

4.3.2 Innovations state space models of exponential smoothing

Based on the traditional exponential methods, Hyndman, Koehler, Ord, and Snyder (2008) argued that forecast error term should be considered the third pattern of the time series in ES besides seasonal and trend patterns. Following this idea, they proposed two possible innovations state space models for each of the methods in Table 4-2. One contains additive errors of time series forecasting and the other contains multiplicative errors. Therefore, the variations of ES have been extended to 30, as shown in Table 4-3.

Table 4-3 Innovations state space exponential smoothing models

		Seasonal pattern		
Error pattern	Trend pattern	N	A	M
A	N	ETS(A,N,N)	ETS(A,N,A)	ETS(A,N,M)
	A	ETS(A,A,N)	ETS(A,A,A)	ETS(A,A,M)
	M	ETS(A,M,N)	ETS(A,M,A)	ETS(A,M,M)
	Ad	ETS(A,Ad,N)	ETS(A,Ad,A)	ETS(A,Ad,M)
	Md	ETS(A,Md,N)	ETS(A,Md,A)	ETS(A,Md,M)
M	N	ETS(M,N,N)	ETS(M,N,A)	ETS(M,N,M)
	A	ETS(M,A,N)	ETS(M,A,A)	ETS(M,A,M)
	M	ETS(M,M,N)	ETS(M,M,A)	ETS(M,M,M)
	Ad	ETS(M,Ad,N)	ETS(M,Ad,A)	ETS(M,Ad,M)
	Md	ETS(M,Md,N)	ETS(M,Md,A)	ETS(M,Md,M)

Note: The triplet ETS(., ., .) refers to: error, trend and seasonality, respectively; N=none, A=Additive, M=Multiplicative, Ad=Additive damped, and Md=Multiplicative damped.

Using the initial letters of error, trend, and seasonality, innovations state space ES is denoted ETS. Thirty variations of ETS can be drawn in the following equation:

$$\begin{cases} y_t = w(x_{t-1}) + r(x_{t-1})\varepsilon_t \\ x_t = f(x_{t-1}) + g(x_{t-1})\varepsilon_t \end{cases} \tag{4.2}$$

where $x_t = (l_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})'$, ε_t is a Gaussian white noise process, $\varepsilon_t \in N(0, \sigma^2)$. Let $\mu_t = w(x_{t-1})$, $r(x_{t-1})$ for the models with additive errors be equal to 1; so $y_t = \mu_t + \varepsilon_t$, while the multiplicative error models has $r(x_{t-1}) = \mu_t$, so $y_t = \mu_t(1 + \varepsilon_t)$. Therefore, $\varepsilon_t = (y_t - \mu_t)/\mu_t$ indicates the relative error for multiplicative models.

In order to use the state space models for forecasting, the values of x_0 and the parameters α, β, and γ need to be estimated in the first place. According to Ord, Koehler, and Snyder (1997), maximum likelihood estimates of the state space model (4.2) can be obtained as below:

$$L(\theta, x_0) = nlog\left(\sum_{t=1}^{n} \varepsilon_t^2\right) + 2\sum_{t=1}^{n} \log|r(x_{t-1})|$$

where $\theta = (a, \beta, \gamma, \varphi)'$. These parameters and initial states of x_0 can be estimated by minimizing L. After the initial states of x_0 have been estimated, the seasonal indices are added to zero for additive seasonality ($s_t + s_{t-1} + \cdots + s_{t-m+1} = 0$) and to m for multiplicative seasonality ($s_t + s_{t-1} + \cdots + s_{t-m+1} = m$).

Furthermore, the various ES models can be interpreted as weighted averages in traditional approaches, thus:

$$\begin{cases} 0 < \alpha < 1, \\ 0 < \beta < \alpha, \\ 0 < \gamma < 1 - \alpha, \\ 0 < \varphi < 1, \end{cases}$$

According to Hyndman, Akram, and Archibald's (2008) study, however, the above restrictions are not necessary in ETS models.

4.3.3 Model selection and automatic forecasting

For a specific time series of data, the general principle is to select the most appropriate form of ETS models from Table 4-3, which contains two steps. The first step is to model the series using all 30 forms of ES; the second step is to compare the measure of forecast accuracy of all 30 models and select the model with minimum forecast error. In order to improve the reliability of model selection, Akaike's Information Criterion (AIC) is used to measure the forecast error of ETS models:

$$AIC = L(\hat{\theta}, \hat{x}_0) + 2q$$

where $\hat{\theta}$ and $\hat{x}_0$ denote the estimates of θ and x_0, respectively, and q is the number of parameters in θ plus the number of free states in x_0. The specific ETS model with the smallest AIC value is considered the most appropriate for the specific time series data.

Based on the model selection criteria, an automatic and robust forecasting algorithm can be developed in four steps:

1. Optimization of θ and x_0 of all ETSs on the time series to be forecast;
2. Selection of the best ETS model according to the AIC indicator;
3. Production of forecasts in the forecast horizon using the best ETS model.

The statistical forecasts are listed in Table 4-4 along with corresponding forecast errors.

Table 4-4 Statistical forecasts using ETS modeling approaches

Year	Source Market	Destination	Q	SF	PE(%)
2006	Australia	China	308,452	273,623	11.29
2006	Hong Kong SAR	Chinese Taipei	2,177,232	2,140,643	1.68
2006	Hong Kong SAR	Japan	1,311,111	1,229,066	6.26
2006	Hong Kong SAR	Macau SAR	577,792	492,344	14.79
2006	Japan	Chinese Taipei	1,309,121	1,497,451	-14.39
2006	Japan	China	811,675	732,598	9.74
2006	Japan	Hong Kong SAR	352,265	296,101	15.94
2006	Chinese Taipei	Japan	1,161,489	1,200,861	-3.39
2006	Chinese Taipei	USA	394,802	373,461	5.41
2006	Chinese Taipei	Korea ROK	196,260	191,558	2.40
2007	Australia	China	357,524	325,397	8.99
2007	Hong Kong SAR	Chinese Taipei	2,238,731	2,165,315	3.28
2007	Hong Kong SAR	Japan	1,324,336	1,341,103	-1.27
2007	Hong Kong SAR	Macau SAR	626,103	561,641	10.30
2007	Japan	Chinese Taipei	1,385,255	1,374,354	0.79
2007	Japan	China	942,439	891,250	5.43
2007	Japan	Hong Kong SAR	432,042	340,301	21.23
2007	Chinese Taipei	Japan	1,166,380	1,243,145	-6.58
2007	Chinese Taipei	USA	397,965	381,534	4.13
2007	Chinese Taipei	Korea ROK	225,814	200,878	11.04

4.4 Experimental procedure

Based on the statistical forecasts produced by ETS, judgmental forecasts by forecasters are made through two rounds of adjustment. The whole process of judgmental forecasting is conducted using a set of demo webpages that simulate the use of a real TDFSS. The demo pages are designed based on HKTDFS, and the data storage also relies on the HKTDFS database. The empirical data of tourism demand to be forecast are first stored in the HKTDFS database. Forecasters are then invited to conduct two-round online judgmental forecasting using the webpages. Finally, forecasters' adjustments are stored in the HKTDFS database for further analysis. In order to avoid any judgmental bias caused by confusion, instructions in both English and Chinese on how to produce a judgmental forecast in the experiment are provided to participants (see Appendix 1 and 2).

4.4.1 The first round of judgmental forecasting

The first round of judgmental forecasting aims to collect forecasters' judgmental forecast without any PF. The forecasting process in this round involves three steps and the corresponding adjustments made by forecasters are collected at the end of each step.

4.4.1.1 Step 1: desired outcome

In the first step, forecasters are asked to provide their desired outcome, if they have one, for the forecasting of all tourism demand series in question. To make sure that forecasters have a correct understanding of desire outcome, the definition of desired outcome is shown in the instruction of experiment (Appendix 1 and 2). In addition, some examples of how to identify and measure desired outcome with related examples are given during the experiment and also when related questions are pointed out by forecasters. Using the same labeling as in Equation (4.1), the desired outcomes of forecaster *j* for the one-step-ahead forecast of tourism demand series *i* in Step 1 of the first round adjustment is labeled $DO_{i,j,h+1}$. Then the percentage error of the desired outcome is calculated by comparing the one-step-ahead forecast with the real outcome:

$$D_{i,j,h+1} = \frac{Desire\ Error}{Real\ Outcome} = \frac{Q_{i,h+1} - DO_{i,j,h+1}}{Q_{i,h+1}}$$

where h denotes the length of the historical data of tourism demand series I, and $Q_{i,h+1}$ indicates the real outcome of the first point on the one-step-ahead forecast (h+1) for tourism demand series *i*.

4.4.1.2 Step 2: judgmental forecasting

In the second step, forecasters need to provide their judgmental adjustment based on the statistical forecasts, the desired outcome, and the historical data for all tourism demand series. The statistical forecast error can then be measured by the difference between the statistical forecast and the real outcome. The statistical forecast error, which is labeled $S_{i,t+1}$ in this step, can thus be measured by the percentage error of the one-step-ahead statistical forecast with the real outcome:

$$S_{i,h+1} = \frac{Q_{i,h+1} - SF_{i,h+1}}{Q_{i,h+1}} \tag{4.3}$$

where $SF_{i,h+1}$ indicates the one-step-ahead statistical forecast of tourism demand series i.

Judgmental forecast in this step is labeled $F_{i,j,h+1}$. According to Equation (3.2), percentage change on the statistical forecast, the desired outcome, and the latest observation are calculated as:

$$\begin{cases} A_{S_{i,j,h+1}} = \frac{F_{i,j,h+1} - SF_{i,h+1}}{SF_{i,t+1}} \\ A_{D_{i,j,h+1}} = \frac{F_{i,j,h+1} - DO_{i,j,h+1}}{DO_{i,j,h+1}} \\ A_{L_{i,j,h+1}} = \frac{F_{i,j,h+1} - LO_{i,h+1}}{LO_{i,h+1}} \end{cases} \tag{4.4}$$

Furthermore, three dummy variables (I_S, I_D, and I_L) were developed to identify the occurrence of different anchoring biases in Equation (3.3). In the experiment, these dummy variables are calculated as:

$$I_{S_{i,j,h+1}} = \begin{cases} 1, if\ min(|F_{i,j,h+1} - SF_{i,h+1}|, |F_{i,j,h+1} - DO_{i,j,h+1}|, |F_{i,j,h+1} - LO_{i,h+1}|) = |F_{i,j,h+1} - SF_{i,h+1}| \\ 0, otherwise \end{cases}$$

$$I_{D_{i,j,h+1}} = \begin{cases} 1, if\ min\left(|F_{i,j,h+1} - SF_{i,h+1}|, |F_{i,j,h+1} - DO_{i,j,h+1}|, |F_{i,j,h+1} - LO_{i,h+1}|\right) = |F_{i,j,h+1} - DO_{i,j,h+1}| \\ 0, otherwise \end{cases}$$

$$I_{L_{i,j,h+1}} = \begin{cases} 1, if\ min\left(|F_{i,j,h+1} - SF_{i,h+1}|, |F_{i,j,h+1} - DO_{i,j,h+1}|, |F_{i,j,h+1} - LO_{i,h+1}|\right) = |F_{i,j,h+1} - LO_{i,h+1}| \\ 0, otherwise \end{cases}$$

4.4.1.3 Step 3: confidence level of forecasting

After the judgmental forecasts in this round are made, forecasters need to rate their confidence in the forecasts. Forecasters' confidence level in this step is measured on a 5-point Likert scale and is labeled $Conf_{i,j,h+1}$. Then the confidence level is transformed into a percentage and the overconfidence error is calculated by comparing the one-step-ahead forecast confidence and the percentage of the correct adjustment:

$$O_{i,j,h+1} = Conf_{i,j,h+1} - \frac{\sum_{i=1}^{m} I_{i,j,h+1}}{m} \tag{4.5}$$

where $I_{i,j,h+1}$ is a dummy variable that denotes the occurrence of correct adjustments of the one-step-ahead forecast of tourism demand series *i* given by forecaster *j*. This is achieved by comparing the absolute percentage error (APE) of adjustment ($\left|pe_{a_{i,j,h+1}}\right|$) and the APE of statistical forecast ($|pe_{s_{i,j,h+1}}|$):

$$I_{i,j,h+1} = \begin{cases} 1, if \left|pe_{a_{i,j,h+1}}\right| < |pe_{s_{i,j,h+1}}| \\ 0, if \ \left|pe_{a_{i,j,h+1}}\right| \geq \left|pe_{s_{i,j,h+1}}\right| \end{cases},$$

$$where \begin{cases} pe_{s_{i,j,h+1}} = \dfrac{Q_{i,h+1} - SF_{i,h+1}}{Q_{i,h+1}} \\ pe_{a_{i,j,h+1}} = \dfrac{Q_{i,h+1} - ReF_{i,j,h+1}}{Q_{i,h+1}} \end{cases}$$

4.4.1.4 Hypotheses testing and leading indicators

According to the data collected from the first round of forecasting, Equation (4.1) can be further specified as:

$$E_{i,j,h+1} = a_j + \beta_{1\ j} S_{i,j,h+1} + \beta_{2\ j} D_{i,j,h+1} + \beta_{3\ j} I_{S_{i,j,h+1}} A_{S\ i,j,h+1} + \beta_{4\ j} I_{D\ i,j,h+1} A_{D\ i,j,h+1}$$

$$+\beta_{5\ j} I_{L\ i,j,h+1} A_{L\ i,j,h+1} + \beta_{6\ j} O_{i,j,h+1} + \mu_{j,h+1} \qquad (4.6)$$

where h+1 denotes the point on the forecast horizon for each tourism demand series. The seven components from Equation (4.6) can be estimated using the ordinary least squares (OLS) method.

For each forecaster j, the mean value of $D_{i,j,h+1}$ and its significance among all of his/her adjustments reflects whether there is desire bias in the judgmental forecasting process. Therefore, the mean values of $D_{i,j,h+1}$ among all forecasters and all of their forecasts reflect a general conclusion on whether there is desire bias in the judgmental forecasting of tourism demand (H2a). Furthermore, the mean value of $\beta_{2\ j} D_{i,j,h+1} * Q_{i,h+1}$ reflects the general influence of desire bias on final forecast error. $\beta_{2\ j} D_{i,j,h+1} * Q_{i,h+1}$ is therefore considered the leading indicator of a forecaster's desire bias. According to this leading indicator, PF and suggestive guidance can be provided in the next round of forecasting. Testing the significance of $\beta_{2\ j} D_{i,j,h+1} * Q_{i,h+1}$ among all forecasters and their judgmental forecasts by one-sample t-test provides evidence about the contribution of desire bias to final forecast error. Therefore, H2b is supported if such a leading indicator is significantly unequal to zero among the forecasters who expressed desired outcomes; otherwise this hypothesis is rejected.

Following the same methods of hypotheses testing and leading indicator identification, the mean values of $I_{S_{i,j,h+1}}A_{S\ i,j,h+1}$, $I_{D\ i,j,h+1}A_{D\ i,j,h+1}$, $I_{L\ i,j,h+1}A_{L\ i,j,h+1}$, and $O_{i,j,h+1}$, as well as their significance, among all forecasters and all of their forecasts, reflect whether overconfidence bias and any of the three anchoring biases exist in judgmental forecasting of tourism demand (H3a, H3b, H3c, and H4a). For forecaster j, $\beta_{1\ j}S_{i,j,h+1} * Q_{i,h+1}$ $\beta_{3\ j}I_{S_{i,j,h+1}}A_{S\ i,j,h+1} * Q_{i,h+1}$, $\beta_{4\ j}I_{D\ i,j,h+1}A_{D\ i,j,h+1} * Q_{i,h+1}$, $\beta_{5\ j}I_{L\ i,j,h+1}A_{L\ i,j,h+1} * Q_{i,h+1}$, and $\beta_{6\ j}O_{i,j,h+1} * Q_{i,h+1}$ in Equation (4.6) are considered the leading indicators of his/her overconfidence bias, anchoring bias, and statistical forecast bias, respectively; which constructs the PF and the suggestive guidance that can be provided in the future. The significance of these leading indicators reflects the contribution of statistical forecast bias, overconfidence bias, and anchoring bias to final forecast error (H1, H3d, H3e, H3f, and H4b).

4.4.2 The second round of judgmental forecasting

The second round of judgmental forecasting is conducted by completing the following three tasks:

- First, collecting data about forecasters' unaided adjustments on all tourism demand series for $h+2$, the process for which is the same as in the first round of forecasting.
- Second, collecting data about forecasters' revised adjustment when PF is given.
- Third, collecting data about the suggested adjustments on all tourism demand series for h+2.

In order to accomplish these three tasks, the second round of forecasting is designed in six steps. Steps 1, 3, and 5 proceed in the same way as in the first round of forecasting; the data collected from these three steps accomplish the first task. In Steps 2, 4, and 6, the system provides PF about desire bias, anchoring bias, and overconfidence bias in each step in order to help forecasters correct their adjustments. At the end of these three steps, a set of forecasters' revised adjustments, and a set of system-suggested adjustments, are collected in order to accomplish the second and the third tasks.

4.4.2.1 Step 1: desired outcome

Following the same process as in Step 1 of the first round of forecasting, the desired outcomes given by forecaster j on the one-step-ahead forecast of tourism demand series i are labeled $DO_{i,j,h+2}$. After forecasters' desired outcomes are collected, the percentage error of desired outcome is calculated as:

$$D_{i,j,h+2} = \frac{Desire\ Error}{Real\ Outcome} = \frac{Q_{i,h+2} - DO_{i,j,h+2}}{Q_{i,h+2}}$$

4.4.2.2 Step 2: correction of desired outcome

In Step 2, the mean forecasting error, associated with desire bias of each forecaster in the first round of forecasting, is provided as PF after desired outcome ($DO_{i,j,h+2}$) is provided. According to Equation (3.5), the mean desire error of $DO_{i,j,h+1}$ for forecaster j is calculated as:

$$DES_j = \frac{1}{m}\sum_{i=1}^{m} \beta_{2\ j} D_{i,j,h+1} * Q_{i,j,h+1}$$

Based on the PF of the mean desire error, forecasters' revision of their desired outcome ($DO^{p}_{i,j,h+2}$, where p indicates correction of adjustment based on PF) is collected.

In addition, another set of forecasts that needs to be collected at the end of the second step is that of system-suggested adjustments regarding forecasters' desire error in the previous round of forecasting. This set of forecasts does not need forecasters' input; it can be calculated by the system itself. According to Equation (3.6), the suggested forecasts can be calculated by the following equation:

$$DO^{s}_{i,j,h+2} = DO_{i,j,h+2} + \frac{1}{m}\sum_{i=1}^{m} \beta_{2_j} D_{i,j,h+1} * Q_{i,h+1}$$

where $DO^{s}_{i,j,h+2}$ indicates the suggested forecasts of tourism demand series *i* for forecaster *j* in this step of forecasting.

4.4.2.3 Step 3: judgmental forecasting

In this step, both the statistical forecasts and forecasters' unaided forecasts for point h+2 on each of the tourism demand series are collected following the same process as in Step 2 of the first round adjustment. Then the percentage error of statistical forecast in the second round is calculated as:

$$S_{i,h+2} = \frac{Statistical\ Forecast\ Error}{Real\ Outcome} = \frac{Q_{i,h+2} - SF_{i,h+2}}{Q_{i,h+2}}$$

Based on the statistical forecasts, desired outcomes, and the latest real outcomes of tourist arrivals, forecasters' one-step-ahead forecast for each tourism demand series in this step is labeled $F_{i,j,h+2}$. According to Equation (3.2), three types of anchor measured by percentage change are calculated as:

$$\begin{cases} A_{S_{i,j,h+2}} = \dfrac{F_{i,j,h+2} - SF_{i,h+2}}{SF_{i,t+2}} \\ A_{D_{i,j,h+2}} = \dfrac{F_{i,j,h+2} - DO_{i,j,h+2}}{DO_{i,j,h+2}} \\ A_{L_{i,j,h+2}} = \dfrac{F_{i,j,h+2} - LO_{i,h+2}}{LO_{i,h+2}} \end{cases}$$

In order to specify three types of anchoring error for each forecaster, the dummy variables $I_{S_{i,j,h+2}}$, $I_{D_{i,j,h+2}}$, and $I_{L_{i,j,h+2}}$ need to be identified following the same approach as in the first round of forecasting.

4.4.2.4 Step 4: correction of judgmental forecasting

In this step, the mean anchoring error that forecasters made in their first round of forecasting is provided as PF to help them further correct their judgmental forecasts. According to Equation (3.7), the mean anchoring error for forecaster *j* is calculated as:

$$\begin{cases} A\_STAT_j = \dfrac{1}{N_j(I_{S_{i,j,h+1}} = 1)} \displaystyle\sum_{i=1}^{N_j(I_{S_{i,j,h+1}}=1)} \beta_{3_j} I_{S_{i,j,h+1}} A_{S_{i,j,h+1}} * Q_{i,h+1} \\ A\_DES_j = \dfrac{1}{N_j(I_{D_{i,j,h+1}} = 1)} \displaystyle\sum_{i=1}^{N_j(I_{D_{i,j,h+1}}=1)} \beta_{4_j} I_{D_{i,j,h+1}} A_{D_{i,j,h+1}} * Q_{i,h+1} \\ A\_LAS_j = \dfrac{1}{N_j(I_{L_{i,j,h+1}} = 1)} \displaystyle\sum_{i=1}^{N_j(I_{L_{i,j,h+1}}=1)} \beta_{5_j} I_{L_{i,j,h+1}} A_{L_{i,j,h+1}} * Q_{i,h+1} \end{cases}$$

For the forecast of tourism demand series *i* at point *h*+2, A_STAT_j will be provided if forecaster *j*'s adjustment is close to the statistical forecast ($I_{S_{i,j,h+2}} = 1$). Similarly,

A_DES_j or A_LAS_j will be provided if forecaster j's adjustment is close to the desired outcome or the latest observation, respectively. Based on the PF of anchoring bias, forecasters' revised adjustments ($F\_A^{p}_{i,j,h+2}$) are collected.

System-suggested forecasts regarding forecasters' anchoring error in the previous round of forecasting also need to be collected at the end of this step. According to Equation (3.11), system-suggested forecasts regarding the influence of anchoring bias on statistical forecast, desired outcome, and the latest observation in the first round of forecasting are calculated as:

$$F\_A^{S}_{i,j,h+2} = \begin{cases} \dfrac{F_{i,j,h+2}}{1 - \beta_{3_j}\dfrac{F_{i,j,h+2} - SF_{i,h+2}}{SF_{i,h+2}}}, if\ I_{S_{i,j,h+2}} = 1 \\ \dfrac{F_{i,j,h+2}}{1 - \beta_{4_j}\dfrac{F_{i,j,h+2} - DO_{i,j,h+2}}{DO_{i,h+2}}}, if\ I_{D_{i,j,h+2}} = 1 \\ \dfrac{F_{i,j,h+2}}{1 - \beta_{5_j}\dfrac{F_{i,j,h+2} - LO_{i,h+2}}{LO_{i,h+2}}}, if\ I_{L_{i,j,h+2}} = 1 \end{cases}$$

4.4.2.5 Step 5: confidence level of forecasting

In Step 4, the impact of forecasters' overconfidence bias on the forecast of point h+2 on the forecast horizon for all tourism demand series ($Conf_{i,j,h+2}$) is identified following the same process as in the first round of forecasting. Following Equation (4.5), forecasters' confidence in their adjustment is transformed into a percentage and the overconfidence error is calculated as:

$$O_{i,j,h+2} = Conf_{i,j,h+2} - \frac{\sum_{i=1}^{m} I_{i,j,h+2}}{m}$$

where

$$I_{i,j,h+2} = \begin{cases} 1, if \ \left|pe_{a_{i,j,h+2}}\right| < |pe_{s_{i,j,h+2}}| \\ 0, if \ \left|pe_{a_{i,j,h+2}}\right| \geq \left|pe_{s_{i,j,h+2}}\right| \end{cases},$$

$$where \begin{cases} pe_{s_{i,j,h+2}} = \dfrac{Q_{i,h+2} - SF_{i,h+2}}{Q_{i,h+2}} \\ pe_{a_{i,j,h+2}} = \dfrac{Q_{i,h+2} - CF_{i,j,h+2}}{Q_{i,h+2}} \end{cases}$$

The new collected data about forecasters' overconfidence bias can be used to further test hypotheses H1a and H1b in order to enhance the reliability of the results.

4.4.2.6 Step 6: correction of forecasting regarding confidence level

In order to correct forecaster's overconfidence bias, the mean overconfidence error of $O_{i,j,h+1}$ that each forecaster made in the first round of forecasting is provided as PF to help him/her further revise his/her judgmental forecasts. According to Equation (3.12), the mean overconfidence error is calculated as:

$$OVE_j = \frac{1}{m}\sum_{i=1}^{m} \beta_{6_j} O_{i,j,h+1} * Q_{i,h+1}$$

Based on the PF of overconfidence bias, forecasters' revised adjustments ($F\_O_{i,j,h+2}^{p}$) are collected.

After forecasters' revised forecasts are collected, the last set of forecasts to be collected in this step is the system-suggested forecasts regarding forecasters' overconfidence error. According to Equation (3.17), $F\_O_{i,j,h+2}^{S}$ is calculated as:

$$F\_O^{S}_{i,j,h+2} = \frac{CF_{i,j,h+2}}{1 - \beta_{6_j}(Conf_{i,j,h+2} - \frac{1}{m}\sum_{i=1}^{m} I_{i,j,h+1})}$$

4.4.2.7 Hypotheses testing and leading indicators

In order to test hypotheses H1–4b, Equation (4.1) can be further specified as:

$$E_{i,j,h+2} = \hat{a}_j + \hat{\beta}_{1\ j} S_{i,j,h+2} + \hat{\beta}_{2\ j} D_{i,j,h+2} + \hat{\beta}_{3\ j} I_{S_{i,j,h+2}} A_{S\ i,j,h+2}$$
$$+ \hat{\beta}_{4\ j} I_{D\ i,j,h+2} A_{D\ i,j,h+2}$$
$$+ \hat{\beta}_{5\ j} I_{L\ i,j,h+2} A_{L\ i,j,h+2} + \hat{\beta}_{6\ j} O_{i,j,h+2} + \varepsilon_{j,h+2} \tag{4.7}$$

According to the assumption that forecasters' cognitive behavior presents no significant change between two close forecasting seasons, coefficients $\hat{a}_j$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$, $\hat{\beta}_5$, and $\hat{\beta}_6$ in Equation (4.7) are estimated in the same way as in Equation (4.6). The mean values of $I_{S_{i,j,h+2}} A_{S\ i,j,h+2}$, $I_{D\ i,j,h+2} A_{D\ i,j,h+2}$, $I_{L\ i,j,h+2} A_{L\ i,j,h+2}$, and $O_{i,j,h+2}$, as well as their significance, among all forecasters and all of their forecasts reflect general conclusions regarding whether desire bias, any of the three types of anchoring bias, and the overconfidence bias exist in the judgmental forecasting of tourism demand (H3a, H3b, H3c, and H4a).

Paired-sample *t*-test also needs to be conducted to evaluate the effectiveness of PF. First, comparison of the APE between unaided forecasts ($\left|\frac{Q_{i,h+2} - D_{i,j,h+2}}{Q_{i,h+2}}\right|$) and revised forecasts ($\left|\frac{Q_{i,h+2} - D^{p}_{i,j,h+2}}{Q_{i,h+2}}\right|$) provides evidence of whether PF of forecasters' desire bias from the previous forecasting season can improve judgmental forecast accuracy. H5a would be supported if the forecast error of revised desired outcome were significantly smaller than the unrevised desired outcome; otherwise H5b would be rejected. In the same way, comparison of $|\frac{Q_{i,h+2}-F_{i,j,h+2}}{Q_{i,h+2}}|$ and $|\frac{Q_{i,h+2}-F\_A^{p}_{i,j,h+2}}{Q_{i,h+2}}|$ in three conditions ($I_{S_{i,j,h+2}} = 1$, $I_{D_{i,j,h+2}} = 1$, and $I_{L_{i,j,h+2}} = 1$) provides evidence of whether PF of forecasters' anchoring bias from the previous forecasting season can improve judgmental forecast accuracy (H6a, H6b, and H6c); comparison of $|\frac{Q_{i,h+2}-ReF_{i,j,h+2}}{Q_{i,h+2}}|$ and $|\frac{Q_{i,h+2}-F\_O^{p}_{i,j,h+2}}{Q_{i,h+2}}|$ provides evidence of whether PF of forecasters' overconfidence bias from the previous forecasting season can improve judgmental forecast accuracy (H7a).

To evaluate the effectiveness of suggestive guidance, a series analysis of variance (ANOVA) needs to be conducted among the APEs of forecasters' unaided judgmental forecast, revised forecast regarding PF, and system-suggested forecasts. First of all, the distribution of three APE series, as well as the assumption of equal variance among them, is tested. If the distribution of three series is normally distributed and no significant difference in variance has been found among them, then parametric ANOVA is conducted to investigate the difference of forecast error between these three forecasting methods. Otherwise, comparison of forecast error between the three methods is conducted by a non-parametric approach if any of the assumptions about normal distribution and the assumption of equal variance is not valid. Since the APEs of participants' unaided judgmental forecasts and the revised forecasts based on the corresponding PF are given by the same group of people, and the calculation of system-suggested forecasts are also based on the unaided judgmental forecasts, these three APEs are considered to be related series. Therefore, the Friedman ANOVA test is appropriate for testing the differences between these three types of APE. Wilcoxon signed-rank tests are also conducted after the Friedman ANOVA test in order to compare the differences between the three APEs. Additional attention should be paid to the significance level. In most of the hypotheses testing approaches in this study, the level of significance is defined as $\alpha = 0.05$. According to the principle of Bonferroni correction (Field, 2013), the level of significance changes to 0.05/the number of comparisons in Wilcoxon signed-rank tests, which is $0.05/3 \approx 0.017$.

Considering the PF and system-suggested forecast regarding desire bias, the APE of (i) forecasters' unrevised desired outcome, (ii) their revised desired outcome regarding the PF of desire error in the previous round of forecasting, and (iii) the system-suggested revision of desired outcome are compared through parametric ANOVA or Friedman ANOVA. The selection of the ANOVA method depends on the distribution and variance of these three variables. If the APE of system-suggested desired outcome is significantly smaller than other two forecast errors, it is supposed that the system-suggested forecast regarding forecasters' desire bias performs better in reducing cognitive bias in judgmental forecasting of tourism demand (H5b). Similarly, another three sets of ANOVA or Friedman ANOVA tests need to be conducted in order to examine the accuracy of system-suggested adjustments regarding forecasters' anchoring bias. If the suggested forecasts regarding the three types of anchoring bias are generally more accurate than forecasters' unrevised forecasts and their revised forecasts regarding the corresponding PF, it is proven that the system-suggested forecasts regarding forecasters' anchoring bias perform better in reducing forecasters' cognitive bias in the judgmental forecasting of tourism demand (H6d, H6e, and H6f). Finally, the accuracy of system-suggested adjustments regarding forecasters' overconfidence bias, forecasters' unrevised forecasts, and their revised adjustment regarding the PF of overconfidence error in the previous round of forecasting need to be compared. If the APE of system-suggested adjustments is significantly smaller than the other two APEs, it is supposed that the system-suggested forecasts regarding forecasters' overconfidence bias perform better in reducing forecasters' cognitive bias in judgmental forecasting of tourism demand (H7b). Details of the (Friedman) ANOVA analysis are listed in Table 4-4, above.

Table 4-5 The effectiveness of suggestive guidance – (Friedman) ANOVA analysis

(Friedman) ANOVA Analysis	Required Variables	Conditions	Hypothesis to be Tested
1	$DO_{i,j,h+2}, D^{p}_{i,j,h+2}, D^{s}_{i,j,h+2}$	$i \in$ (1,m) $j \in$ (1,n)	H5b
2	$F_{i,j,h+2}, F\_A^{p}_{i,j,h+2}, F\_A^{s}_{i,j,h+2}$	$I_{S_{i,j,h+2}} = 1$	H6d
3	$F_{i,j,h+2}, F\_A^{p}_{i,j,h+2}, F\_A^{s}_{i,j,h+2}$	$I_{D_{i,j,h+2}} = 1$	H6e
4	$F_{i,j,h+2}, F\_A^{p}_{i,j,h+2}, F\_A^{s}_{i,j,h+2}$	$I_{L_{i,j,h+2}} = 1$	H6f
5	$CF_{i,j,h+2}, F\_O^{p}_{i,j,h+2}, F\_O^{S}_{i,j,h+2}$	$i \in$ (1,m) $j \in$ (1,n)	H7b

4.5 Design of judgmental forecasting process

The two-round judgmental forecasting process is implemented with an online forecasting survey. At the beginning of the survey, an attitude statement is given, as below, to explain the environment that the experiment is going to simulate:

1) The online survey includes two sessions simulating the **two consecutive years, 2006 and 2007**. In the first session, imagine that you are a manager in a company's marketing department at the beginning of 2006. Based on the historical data of tourism demand and your own judgment, you need to predict the tourism demand for specific D-O pairs for the coming year, 2006 (one-step-ahead forecast). In this survey, tourism demand is measured by total tourist arrivals, and overall there are **10 D-O pairs** for which the tourism demand needs to be forecast. For each D-O pair, the historical data of tourist arrivals is collected from the World Tourism Organization (UNWTO). The length of the historical data varies among D-O pairs, dependent upon the availability of data. For example, historical data on tourist arrivals from Chinese Taipei to the Hong Kong SAR were collected for the period 1995–2005, while historical data on tourist arrivals from China to Japan were collected for 1997–2005. The process of judgmental forecasting in the first session is conducted in **three steps**. The **historical data of tourist arrivals** for each D-O pair and **statistical forecasts** are the only references provided for your information.

2) In the second session, imagine that the year 2006 has passed and it is the beginning of 2007. The real outcome of tourist arrivals in 2006 for each D-O pair is now available and becomes the latest observation of the historical series of tourist arrivals. You are asked to judgmentally forecast the tourist arrivals of the same 10 D-O pairs for 2007 (one-step-ahead forecast). In the second session, the judgmental forecast is conducted in **six steps** with three types of information provided: the **updated historical data** on tourist arrivals for each D-O pair, the

statistical forecast, and **suggestive information** based on your forecasting performance from the previous session.

For the first round of the experiment, three webpages (p1_1, p1_2, and p1_3) are designed to process the three steps of judgmental forecasting for the 10 D-O pairs of tourist arrivals. Participants are asked to provide their desired outcomes in p1_1 (Figure 4-1); to provide their judgmental adjustments based on the historical data and statistical forecasts in p1_2 (Figure 4-2); and to rate their confidence in their forecasts in p1_3 (Figure 4-3).

Destination	Origin	Forecast Year		Desired Outcome
Australia	China	2006	☑	10,000
Chinese Taipei	Japan	2006	☑	20,000
	Korea ROK	2006	☑	30,000
	USA	2006	☑	40,000
Hong Kong SAR	Chinese Taipei	2006	☐	
	Japan	2006	☐	
	Macau SAR	2006	☑	50,000
Japan	China	2006	☐	
	Chinese Taipei	2006	☑	60,000
	Hong Kong SAR	2006	☑	70,000

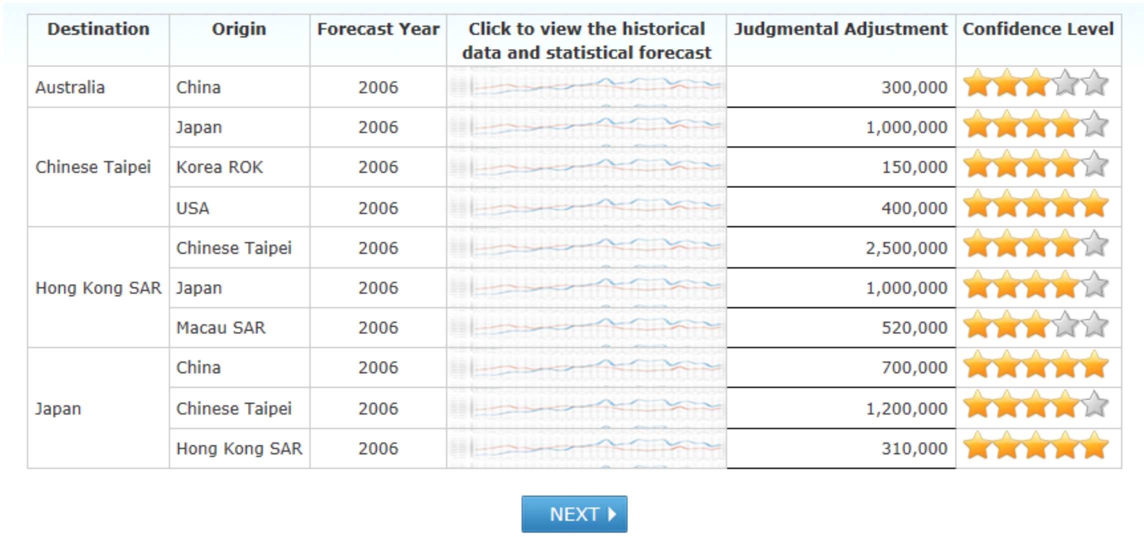
NEXT

Figure 4-1 Participants' desired outcomes (p1_1)

Destination	Origin	Forecast Year	Click to view the historical data and statistical forecast	Judgmental Adjustment
Australia	China	2006		300,000
Chinese Taipei	Japan	2006		1,000,000
	Korea ROK	2006		
	USA	2006		
Hong Kong SAR	Chinese Taipei	2006		
	Japan	2006		
	Macau SAR	2006		
Japan	China	2006		
	Chinese Taipei	2006		
	Hong Kong SAR	2006		

Year	Historical Data (1997-2005) & Statistical Forecast (2006)	Judgmental Adjustment
1997	905,527	
1998	826,632	
1999	826,222	
2000	916,301	
2001	971,190	
2002	998,497	
2003	657,053	
2004	887,311	
2005	1,124,334	
2006	1,200,861	1,000,000

Figure 4-2 Participants' judgmental forecasting (p1_2)

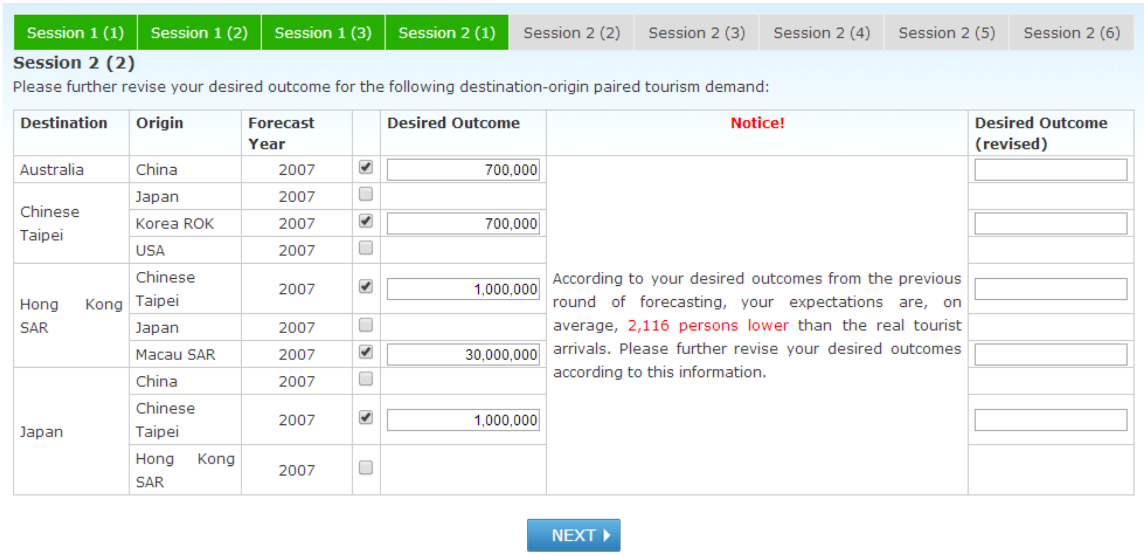


Destination	Origin	Forecast Year	Click to view the historical data and statistical forecast	Judgmental Adjustment	Confidence Level
Australia	China	2006		300,000	★★★☆☆
Chinese Taipei	Japan	2006		1,000,000	★★★★☆
	Korea ROK	2006		150,000	★★★★☆
	USA	2006		400,000	★★★★★
Hong Kong SAR	Chinese Taipei	2006		2,500,000	★★★★☆
	Japan	2006		1,000,000	★★★★☆
	Macau SAR	2006		520,000	★★★☆☆
Japan	China	2006		700,000	★★★★★
	Chinese Taipei	2006		1,200,000	★★★★☆
	Hong Kong SAR	2006		310,000	★★★★★

NEXT

Figure 4-3 Participants' confidence in their forecasts (p1_3)

These three pages are also used in the second round of experiment for the same purposes. Another three pages are also designed to provide performance feedback, and to capture forecasters' revisions of their forecasts. Figure 4-4 is presented after participants provide their desired outcomes. In this case, the system shows that the participants' desired outcomes in the previous round of forecasting were, on average, 2,116 persons lower than the real tourist arrivals. The participants can further revise their desired outcomes of forecast in the second round based on this feedback.


Session 1 (1) | Session 1 (2) | Session 1 (3) | Session 2 (1) | Session 2 (2) | Session 2 (3) | Session 2 (4) | Session 2 (5) | Session 2 (6)

Session 2 (2)

Please further revise your desired outcome for the following destination-origin paired tourism demand:

Destination	Origin	Forecast Year		Desired Outcome	Notice!	Desired Outcome (revised)
Australia	China	2007	☑	700,000	According to your desired outcomes from the previous round of forecasting, your expectations are, on average, 2,116 persons lower than the real tourist arrivals. Please further revise your desired outcomes according to this information.	
Chinese Taipei	Japan	2007	☐			
	Korea ROK	2007	☑	700,000		
	USA	2007	☐			
Hong Kong SAR	Chinese Taipei	2007	☑	1,000,000		
	Japan	2007	☐			
	Macau SAR	2007	☑	30,000,000		
Japan	China	2007	☐			
	Chinese Taipei	2007	☑	1,000,000		
	Hong Kong SAR	2007	☐			

NEXT

Figure 4-4 Performance feedback and the revision of desired outcomes (p2_2)

Similarly, another page is designed to provide performance feedback after participants provide their judgmental forecasts (Figure 4-5). In this page, forecasters' adjustments are first grouped by the detected anchoring bias; then the system provides feedback of forecast errors associated with the three types of anchoring bias in the previous round of forecasting. The participants can then further revise their judgmental forecasts accordingly. The performance feedback on the detected anchoring bias is displayed in two columns. The mean forecast errors, associated with different anchoring biases that are identified from judgmental forecasts in the first forecasting season, are displayed in the first column; the type of potential anchoring biases detected from participants' second forecasting season are listed in the second column. In the second column, SF indicates the statistical forecast anchor; DO indicates the desire anchor; and LO indicates the latest observation anchor. If no anchoring bias is detected from the forecast of a specific D-O pair, the corresponding cell in the second column is blank (Figure 4-5).

Destination	Origin	Forecast Year	Click to view the historical data and statistical forecast	Judgmental Adjustment	Notice! Mean Anchoring Error	Anchor on:	Judgmental Adjustment (revised)
Australia	China	2007		350,000	According to your judgmental adjustment in the previous round of forecasting, the mean anchoring errors of your adjustment are listed as below: 1. Anchor on Statistical Forecast (SF):-200,067 2. Anchoring on the latest outcome (LO): 51,485 The potential anchoring bias of your adjustments in this round are listed in next column, please further revise your adjustment for this round of forecasting.	SF	200,000
Chinese Taipei	Japan	2007		1,000,000		LO	1,050,000
	Korea ROK	2007		230,000		SF	150,000
	USA	2007		100,000			
Hong Kong SAR	Chinese Taipei	2007		2,500,000		LO	3,000,000
	Japan	2007		1,100,000		LO	1,150,000
	Macau SAR	2007		600,000		LO	650,000
Japan	China	2007		900,000		SF	700,000
	Chinese Taipei	2007		1,500,000		SF	1,300,000
	Hong Kong SAR	2007		320,000		SF	150,000

NEXT

Figure 4-5 Performance feedback and the revision of judgmental forecasts (p2_4)

Finally, the system provides feedback about the forecast error associated with participants' overconfidence/lack of confidence in the first round of the experiment; it also asks participants to revise their forecast in response to the informed forecast error (Figure 4-6).

Session 1 (1) | Session 1 (2) | Session 1 (3) | Session 2 (1) | Session 2 (2) | Session 2 (3) | Session 2 (4) | Session 2 (5) | Session 2 (6)

Session 2 (6)

Please further revise your adjustment according to your confidence error:

Destination	Origin	Forecast Year	Click to view the historical data and statistical forecast	Judgmental Adjustment	Confidence Level	Notice!	Judgmental Adjustment (revised)
Australia	China	2007		360,000	★★★☆☆	According to your judgmental adjustment from the previous round of forecasting, your levels of confidence have resulted your judgmental forecasts, on average, 30,489 persons lower than the real tourist arrivals. Please further revise your judgmental forecasts according to this information.	
Chinese Taipei	Japan	2007		1,500,000	★★★★☆		
	Korea ROK	2007		703,000	★★★★★		
	USA	2007		430,000	★★★★☆		
Hong Kong SAR	Chinese Taipei	2007		2,500,000	★★★☆☆		
	Japan	2007		1,500,000	★★★☆☆		
	Macau SAR	2007		30,002,100	★★★★★		
Japan	China	2007		930,000	★★★★☆		
	Chinese Taipei	2007		1,500,000	★★★★☆		
	Hong Kong SAR	2007		370,000	★★★★★		

NEXT

Figure 4-6 Performance feedback and the revision of forecasts regarding confidence (p2_6)

4.6 A preliminary test

To assess the clarity and content validity of the online survey, the online webpages were tested before the main experiment. Eight research students from Master's and Doctoral degrees in the School of Hotel and Tourism Management, Hong Kong Polytechnic University, were invited to participate in this test. All of these students were studying in the research area of tourism and hospitality management. They were first required to complete the survey questionnaire and then to assess whether the whole survey process was easily understood, and whether the information presented on the webpages was correctly worded.

Invitations were emailed to participants. Two students responded with concerns about their contribution in the preliminary test since they were not familiar with tourism demand forecasting and might bias the test result. These two students were not involved in the test. Another student only finished the first session of the experiment due to a failure to save the forecasting results in one step of the first session. As a result, no feedback could be produced in the second session and the student's survey was terminated. Ultimately, five students completed the survey and provided comments and suggestions on the webpages. The test results and the participants' comments revealed some problems with the survey, summarized in Table 4-6.

First, performance feedback could not be produced if nonstandard inputs were given by participants. One student mentioned that "the forecast error given by the system in the second session of the experiment are all zero, so the notices are actually meaningless for me in the revision of my judgmental forecast." According to the records of his forecasts stored in the database, it was found that this student had committed a typographic error: the letter “t” was inputted after one of his judgmental forecasts in the first session of experiment and rendered the forecast unrecognizable by the system. As a result, seven coefficients in Equation (4.6) could not be produced, and so the system’s feedback provided “0” as the forecast error in the second session of the experiment. The same problem could also occur when participants inputted other non-integer letters, symbols, or blank spaces as part of their judgmental forecasts. Furthermore, the system could produce inaccurate feedback if no input was given by a participant. For example, one student did not want to judgmentally adjust the forecasts based on the historical data and the statistical forecasts in the second step of the first forecasting season, so she left the cells of judgmental adjustment blank for all D-O pairs. In that case, “null” values for this set of data were delivered to the system and “null” values of coefficients β_4^1, β_5^1, and β_6^1 in Equation (4.6) were produced; and finally the performance feedback of anchoring bias in the second forecasting season was zero. However, when a participant provides nothing in judgmental adjustment, it does not mean his/her adjustments are equal to zero; it means that he/she does not want to make any change based on the statistical forecasts. Therefore, statistical forecasts had to be considered for further analysis in the same way as judgmental adjustments.

In order to solve this problem, two sets of input checking functions were developed for the purpose of helping forecasters correct typos and produce correct performance feedback information. These two sets of checking function include holistic input checking and specific input checking:

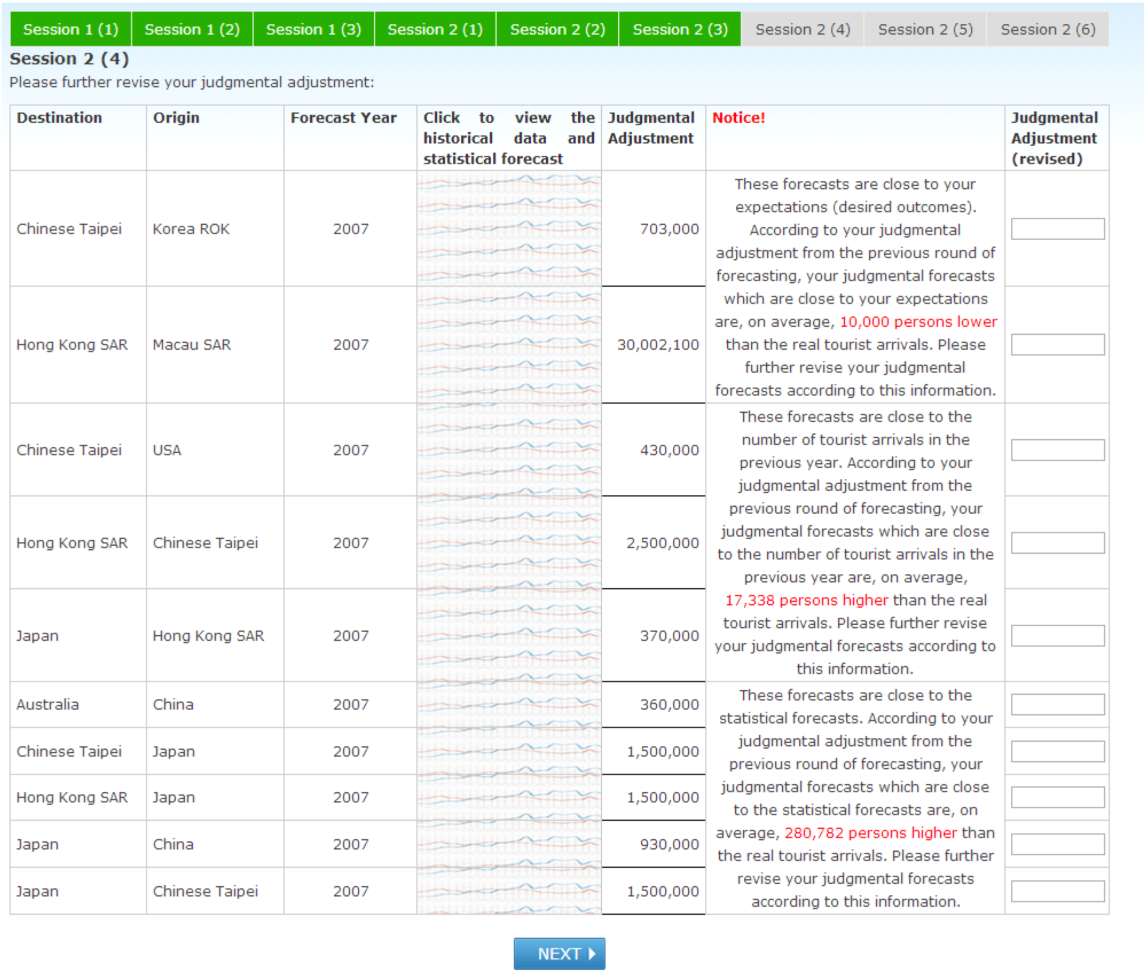
(i) Holistic input checking: all inputs had to be integers only. No letter, symbol, or blank space could be accepted as forecasts and no decimal was required for input. This set of input checking was applied to all pages of the online survey. When participants finished inputting one page, all inputs were checked by this function before moving to the next page. The system would alert participant with detailed information if any judgmental forecast did not pass the input checking.

(ii) Specific input checking: for p1_2, statistical forecast would be considered as judgmental forecasts if no adjustment had been made by participants for any D-O pair. For p1_3, it was compulsory to evaluate participants' confidence level in forecasts for all D-O pairs. For p2_2, if participants did not revise his/her desired outcome for a specific D-O pair market in response to the performance feedback, his/her desired outcomes provided in p2_1 would be used as the final desired outcome; the same input checking principle was applied in p2_4 and p2_6.

Table 4-6 Results of preliminary test

	Problem	Solution
1	When a participant move from one page to another without inputting any values on judgmental adjustment or confidence level, the forecast error associated with anchoring and overconfidence bias cannot be generated.	Input checking functions were designed, including: (i) Holistic input checking; and (ii) Specific input checking.
2	The classification of the three types of anchoring bias in p2-4 was difficult to be understood.	The table in p2-4 was redesigned with all D-O pairs grouped by the identified anchoring bias.
3	Some participants were not clear about what they were going to do in the experiment, even an attitude statement and an instruction were given before the survey.	A verbal guidance would be delivered in the main experiment.

Second, the performance feedback shown in the table of p2_4, regarding the three types of anchoring bias, was difficult to understand and some participants were confused about which forecasts should be revised in this step. One student was confused about the types of potential anchoring bias, especially when there were blank cells. He said that "there seem to be three types of anchoring error identified, but only two were observed. It might be caused by space limitation." Indeed, the omitted type of anchoring bias in this case was the one that had not been detected from his forecasts in the first forecasting season; the blank cell indicated that no anchoring bias had been detected from the forecast of the corresponding D-O pair in the second forecasting season. Furthermore, in some cases, the participants struggled to revise their judgmental forecasts according to the detected anchoring bias. One student mentioned in his comments that "It's a bit difficult to revise the judgmental forecasts with identified anchoring bias since people have to match the potential anchoring bias with the corresponding forecast error one by one. In that case, about 30 (three types of anchoring bias * 10 D-O pairs) matches would need to be checked in this process!"



Session 1 (1) | Session 1 (2) | Session 1 (3) | Session 2 (1) | Session 2 (2) | Session 2 (3) | Session 2 (4) | Session 2 (5) | Session 2 (6)

Session 2 (4)

Please further revise your judgmental adjustment:

Destination	Origin	Forecast Year	Click to view the historical data and statistical forecast	Judgmental Adjustment	Notice!	Judgmental Adjustment (revised)
Chinese Taipei	Korea ROK	2007		703,000	These forecasts are close to your expectations (desired outcomes). According to your judgmental adjustment from the previous round of forecasting, your judgmental forecasts which are close to your expectations are, on average, 10,000 persons lower than the real tourist arrivals. Please further revise your judgmental forecasts according to this information.	
Hong Kong SAR	Macau SAR	2007		30,002,100		
Chinese Taipei	USA	2007		430,000	These forecasts are close to the number of tourist arrivals in the previous year. According to your judgmental adjustment from the previous round of forecasting, your judgmental forecasts which are close to the number of tourist arrivals in the previous year are, on average, 17,338 persons higher than the real tourist arrivals. Please further revise your judgmental forecasts according to this information.	
Hong Kong SAR	Chinese Taipei	2007		2,500,000		
Japan	Hong Kong SAR	2007		370,000		
Australia	China	2007		360,000	These forecasts are close to the statistical forecasts. According to your judgmental adjustment from the previous round of forecasting, your judgmental forecasts which are close to the statistical forecasts are, on average, 280,782 persons higher than the real tourist arrivals. Please further revise your judgmental forecasts according to this information.	
Chinese Taipei	Japan	2007		1,500,000		
Hong Kong SAR	Japan	2007		1,500,000		
Japan	China	2007		930,000		
Japan	Chinese Taipei	2007		1,500,000		

NEXT

Figure 4-7 Redesigned Page 2_4

In order to solve these problems, the table on p2_4 was redesigned (Figure 4-7) with the following features:

(i) For each participant, the 10 D-O pairs of tourism market were grouped according to the potential anchoring biases detected in his/her judgmental forecasts on p2_3. For each group of forecasts, the system provided feedback about the mean forecast error associated with the corresponding anchoring bias detected from the first forecasting season. Therefore, only one column of performance feedback was provided and participants did not have to match the potential anchoring bias with the corresponding forecast error.

(ii) For the forecasts of D-O pairs with no anchoring bias detected, no revision should be made in this step, so they were automatically hidden in the redesigned table. Therefore, blank cells in the column of system notice would not be displayed and participants would be clearer and focused on which forecasts should be revised in this step.

One participant commented that "This is the first time I have done this kind of judgmental forecasting and I am not sure whether my understanding of the experiment is correct." In order to make sure that participants fully understood the experiment, verbal guidance was provided as a supplementary instruction and an attitude statement was made during the main experiment.

4.7 Main experiment design

The main experiment was conducted in the period August 13–September 20, 2014. To meet the requirements for participants, three groups of part-time master (MSc) and doctoral (DHTM) students in the School of Hotel and Tourism Management (SHTM), Hong Kong Polytechnic University, were invited to participate. All of these students had a background in hotel and tourism education. The majority were working at the time for a variety of stakeholders in the tourism and hospitality industry, like travel, hotel, catering, retail, and other tourism-related sectors; planning and marketing agencies/departments of various tourism sectors; and education and research establishments focusing on tourism and hospitality. Their work or research area was to some extent related to the field of tourism demand forecasting, so they were qualified to participate in the main experiment.

Verbal guidance during the experiment was required; however, it was impossible to deliver verbal guidance individually to each participant. All the participants were part-time students working in different locations; the majority were located in Mainland China, Hong Kong, Chinese Taipei, Singapore, and other Asia Pacific countries/regions. Therefore, it was necessary to find a time to call participants together for the experiment and to offer the verbal guidance. Due to this condition, the experiment was conducted at the beginning or the end of three MSc/DHTM classes in different locations at different times (Table 4-7).

Table 4-7 Three sessions of the experiment for data collection

No.	Date	Location	No. of participants	Student Level	Length of experiments
1	13-Aug, 2014	Hong Kong	22	Doctoral	45 minutes
2	16-Sep, 2014	Hangzhou	37	Master	55 minutes
3	20-Sep, 2014	Hangzhou	31	Master	50 minutes

4.7.1 The first session of the experiment

The first session was conducted in Hong Kong on August 13, 2014. At that time, 22 DHTM students from different Asia Pacific areas were taking a course in the School of Hotel and Tourism Management. These students were invited based on the principle of voluntary participation. In order to encourage their participation, all students were told that participation would be anonymous. Also, a lucky draw session to win a portable cell phone charger was held at the end of the experiment as an incentive to complete the experiment. As a result, all 22 DHTM students were willing to participate in this session of the experiment.

Each participant needed a computer with the most up-to-date version of Internet Explorer (IE) installed, so this session of the experiment was held in a well-established computer library in the School of Hotel and Tourism Management, with laptops provided to all participants. To ensure that the experiment would not be interrupted by any technical issues, all the laptops were checked before the experiment and confirmed to be in good condition to use. In addition, the experiment was arranged before classes started in order to ensure that participants' judgmental forecasts would not be influenced by any unexpected interruptions.

In the first five minutes of the experiment, the attitude statement was delivered in both hard copy and verbally. Participants were then asked to visit the webpages for the experiment through IE and individually complete the two-round judgmental forecasting process following the researcher's verbal guidance. Instructions for the experiment were also provided in hard copy to each participant for reference. During the experiment, participants were free to raise their hands and ask questions in any step of the forecasting process if there was any unclear statement or misunderstanding. The researcher would respond in detail to solve participants' problems, allowing them to move on to the next step. When a participant completed the two-round judgmental forecasting process, the webpage for the lucky draw opened automatically on his/her web browser. The participant could then choose to participate in the online lucky draw.

After each step of the forecasting, participants' input was stored in a HKTDFS database when they clicked "NEXT" on each page. When a participant completed the three steps in the first round of forecasting, the system automatically calculated the leading indicators of his/her cognitive bias, which were then used as PF in the second round of forecasting. When participants completed the second (p2_2), the fourth (p2_4), and the sixth step (p2_6) of the second round of forecasting, the system produced system-suggested forecasts and stored them in the database.

The whole experiment lasted 45 minutes; two participants gave up in the middle of the experiment because of personal issues. The other 20 participants completed the whole experiment and moved on to the lucky draw session. Ultimately, 20 sets of data were stored in the database after the first session of the experiment. Valid answers would be further analyzed by standard data screening processes after all experimental data were collected.

4.7.2 The second session of the experiment

The second session of the experiment was conducted in Hangzhou on September 16, 2014. Forty-five MSc students from different areas of Mainland China taking the course Managing Human Resources in the Hotel and Tourism Industry in the joint education center of Zhejiang University were invited to be voluntary participants. Again, they were told that their participation would be anonymous, and a lucky draw for portable chargers as incentives was arranged to encourage participation in the experiment. Since public computers were not available in the education center, the students who did not bring their personal computer or other electronic devices to the class were not able to process judgmental forecasting through web browsers. As a result, only 37 MSc students in that class were able to join the experiment.

Unlike the first session of experiment, this session was held at the end of the class. In order to ensure good condition for the participants to conduct judgmental forecasting, a 15-minute tea break between the class and the experiment was held. After the tea break, all eligible participants expressed that they were refreshed and ready for the experiment.

Following the same process as in the first session, a five minutes attitude statement was delivered before the two-round judgmental forecasting process. Verbal guidance was provided by the researcher throughout the whole experiment and help was provided to participants who had any problem regarding the forecasting process. The process of storing forecasters' input, as well as system-suggested forecasts, was the same as in the first session.

The judgmental forecasting process took longer than in the first session because more students asked questions regarding forecasting techniques. For example, many were not clear about the process by which statistical forecasts were generated and the influencing factors that had been considered in generating the statistical forecasts, even though they were clearly stated in the instructions. The researcher was also asked to explain the theory of cognitive bias (especially anchoring bias) before the participants conducted the corresponding steps of judgmental forecasting. The whole experiment therefore lasted about 55 minutes.

During the experiment, the majority of the 37 participants used their personal laptops; some tried to complete the judgmental forecasting process using tablets. The webpages for this experiment were designed specifically for computer use; thus, some of the tables/figures were not properly displayed on other electronic devices. Because of this, some participants using tablets were not able to complete the forecasting process. Ultimately, 29 participants successfully completed the two-round judgmental forecasting process and stored their inputs in the database.

4.7.3 The third session of the experiment

The third session of the experiment was conducted on September 20, 2014 at the same place as the second session; the students were second-year part-time MSc students. There were 45 students in the class; 14 of them had no personal computer or other electronic device with them. Therefore, the number of participants in this session was 31.

Similar to the second session, this session was held at the end of the class and a tea break was organized in order to refresh participants before the experiment. The experimental process was the same as that followed in the second session; it was completed within 50 minutes.

Device support was again an issue: five participants who browsed the experiment webpages using tablets were not able to proceed beyond the second step of the first round of forecasting. Therefore, their inputs were not considered valid data. Besides these, 26 complete data sets were stored in the database for further data screening and analysis.

4.8 Summary

This chapter has elaborated the research methodology. It first discussed the appropriateness of testing the hypotheses using a judgmental forecasting experiment, followed by the qualification of participants to be invited, as well as the empirical data to be used. To support the experiment, statistical forecasts were produced by a state space ES approach. The method of automatic model selection for ES was also explained in detail. Specifically, the experiment involved two-round judgmental forecasting. The first round of forecasting aimed to detect forecasters' cognitive bias; the second aimed to help forecasters correct their cognitive bias using two debiasing strategies: PF and suggestive guidance. Methods of hypotheses testing were further discussed based on the information collected from each step of the experiment.

This chapter also describes the design of the preliminary test and the main experiment. The preliminary test was conducted with six research students in SHTM. According to the results of the preliminary test, further revisions were made to the webpages designed for the experiment and the instruction. The data of the main experiment were collected from three forecasting sessions at different times and in more than one location. Participants in the main experiment were all part-time MSc and DHTM students.

5 HYPOTHESES TEST RESULTS

5.1 Data screening

Table 5-1 Main series to describe the data set

Series		Description
Raw Data	Real_1	Real outcome of tourist arrivals for a specific D-O pair in 2006
	SF_1	Statistical forecast of tourist arrivals for a specific D-O pair in 2006
	Des_1	Desired outcome in the first round of forecasting
	Adj_1	Judgmental forecast in the first round of forecasting
	Conf_1	Confidence of forecast in the first round of forecasting
	Real_2	Real outcome of tourist arrivals for a specific D-O pair in 2007
	SF_2	Statistical forecast of tourist arrivals for a specific D-O pair in 2007
	Des_2	Desired outcome in the second round of forecasting
	Des_2_P	Revised desired outcome based on the performance feedback of desire bias
	Adj_2	Judgmental forecast in the second round of forecasting
	Adj_2_p	Revised judgmental forecast based on the performance feedback of anchoring bias
	Conf_2	Confidence of forecast in the second round of forecasting
	Conf_2_p	Revised judgmental forecast based on the performance feedback of overconfidence bias
Coefficients	β_1	statistical forecast error
	β_2	desire error
	β_3	anchoring error on statistical forecast
	β_4	anchoring error on desired outcome
	β_5	anchoring error on the latest observation
	β_6	overconfidence error
Leading Indicator	DES	potential forecast error associated with desire bias
	A_STAT	potential forecast error associated with anchoring bias on statistical forecast
	A_DES	potential forecast error associated with anchoring bias on desired outcome
	A_LAS	potential forecast error associated with anchoring bias on the latest observation
	OVE	potential forecast error associated with overconfidence bias

Seventy-five participants' forecasting data (750 cases) were successfully collected from the main experiment. Table 5-1 lists the main variables to describe the whole dataset, which can be further classified into three groups: items of raw data, coefficients, and leading indicators. For data screening, missing data and outliers were identified in the raw data.

5.1.1 Missing data

The handling of missing data is important because it is the main source of bias in statistical results (Hair, Tatham, Anderson, & Black, 2006). A variety of methods are proposed in the literature to deal with missing data: direct maximum likelihood, multiple imputation, maximum likelihood listwise deletion, arithmetic mean imputation, stochastic regression imputation, pairwise deletion, and similar response pattern imputation (Enders, 2006; Olinsky, Chen, & Harlow, 2003). Among all these well-developed methods, listwise deletion is the most commonly used (Enders, 2006; Gilley & Leone, 1991). This method is based on the assumptions that the sample is large enough for reliable statistical analysis and that missing values represent less than 10% of the whole dataset. When the assumptions are valid, listwise deletion is more robust than other sophisticated methods (Allison, 2001). In the current study, 75 participants completed their judgmental adjustments in the experiment, so the sample for hypotheses testing is 750 (10 D-O pair forecasts for each participant), which is large enough to carry out statistical tests. Therefore, the listwise deletion method is appropriate for use in this study, which means removing cases with missing values before statistical analysis.

Identification of missing values using the listwise deletion method is conducted by comparing participants' inputs and the system-generated leading indicators. When checking participants' inputs, blank inputs are considered missing values, with one exception: if a participant did not provide the desired outcome for a particular D-O pair, it indicates that he/she did not have any expected tourism demand in certain markets. Such cases are considered as normal inputs rather than missing values. Furthermore, the leading indicators automatically calculated by the system may be equal to zero for two possible reasons. First, the coefficients (α, β_i) estimated by the OLS method may be zero if the corresponding independent variable is not significant in the cognitive bias detection model. Second, for anchoring bias, leading indicator(s) regarding one or two types of anchor are not detected from a participant's judgmental forecasting in the second round of forecasting. In the first case, it is possible that all the coefficients estimated by OLS are equal to zero, which means that the independent variables in the cognitive bias detection model are all insignificant. This happens when participants provide extreme values of judgmental forecasts, probably caused by typos or mistakes regarding the unit of forecasts. This either means that the participant's judgmental forecasts in the second round were not guided by any PF information and there would thus be no difference between the system-suggested forecasts and the participant's unaided judgmental forecasts; or it can be assumed that a debiasing strategy has been applied to the participant's decision-making. Therefore, such cases are also considered as missing values. Details about the participants and valid answers are listed in Table 5-2.

In the first session of the experiment, 20 participants successfully submitted their judgmental forecasts and no missing values were detected from the first round of forecasting; two participants did not complete their second round of forecasting. Because no extreme scenario was discovered in leading indicators production, 18 valid answers were given by participants in the first session of the experiment, which indicates that 180 valid cases were collected. In the second session, 29 participants completed the first round of forecasting and successfully stored their judgmental forecasts in the database. One participant did not complete the second round of forecasting, so 28 valid answers were obtained from the two rounds of forecasting in the second session. However, there were some participants whose leading indicators were all equaled to zero. The cases given by certain participants are considered as missing values even they have completed two rounds of forecasting. Ultimately, 27 valid answers (270 valid cases) were collected from the second session. Twenty-six participants provided valid answers in the first round of forecasting in the third session, but three participants did not complete the second round; and no extreme scenarios were observed in the leading indicators. Therefore, 23 valid answers (230 valid cases) were collected from the final session of the experiment. After the missing data was extracted, valid cases amounted to 680 and the missing values only represented 9.3% of the whole dataset.

Table 5-2 Participants and valid answers

	1st session (13 Aug)	2nd session (16 Sep)	3rd session (20 Sep)	Total
Participants in Round_1 (Valid)	20 (20)	29 (29)	26 (26)	75 (75)
Participants in Round_2 (Valid)	20 (18)	29 (28)	24 (23)	73 (69)
Total Participants (Valid)	20 (18)	29 (27)	26 (23)	75 (68)

5.1.2 Outliers

Outliers are defined as observations substantially different from the majority of observations regarding one or more characteristic (Hair et al., 2006). Like missing values, outliers also bias the mean of the data characteristics and influence the data distribution (Field & Hole, 2003). According to the literature, boxplot is the most widely adopted method to identify outliers (Hoaglin & Iglewicz, 1987; Hoaglin, Iglewicz, & Tukey, 1986; Tukey, 1977). A boxplot derives its name from the rectangular box with two whiskers located at the top and the bottom of the box. For a particular batch of data, a traditional boxplot displays four features:

- The center of the boxplot is usually the sample median, which is identified as a line across the box;
- The upper and bottom line of the box is constructed by the sample quartiles or the fourths, which locate the middle half of the data;
- From the top and bottom of the box, dashed lines extend outward to the two adjacent values, the whiskers of the boxplot, which covers the outermost observations that are not extreme enough to be flagged as outliers by an exploratory rule of thumb.
- Extreme values, which are labeled as the stars beyond the ends of the whiskers, are considered as potential outliers and need to be further examined.

The quartiles are the basis of the fences for flagging real outliers. Let Q_1 and Q_3 indicate the lower and upper quartile, respectively; the fences therefore lie at $Q_1 - k(Q_3 - Q_1)$ and $Q_3 + k(Q_3 - Q_1)$. A change in the value of k can have a great impact on the identification of outliers. In previous studies, several values of k have been recommended for the use of flagging observations as outliers, which include $k = 1.0$ (McNeil, 1977), $k = 1.5$ and 3.0 (Tukey, 1977), and $k = 2.0$ (Ingelfinger, Mosteller, Thibodeau, & Ware, 1987). Hoaglin et al. (1986) examined the performance of these values by Gaussian samples and revealed that "the main resistant rule" ($k = 1.5$) has a number of advantages when the sample size is small (less than 20) and $k = 2.0$ performs better for larger samples. The rule based on $k = 3.0$ is extremely conservative. Based on these findings, another study by Hoaglin and Iglewicz (1987) recommended that $k \approx 2.2$ is a robust criterion to identify fences and outliers. Therefore, the outliers of the raw data collected from the experiment are identified using the boxplot method with the rule $k = 2.2$. The results of outlier identification are shown in Table 5-2.

Table 5-3 Outliers identified using boxplot (*k*=2.2)

	Q1	Median	Q3	Fence	Outliers
Des_1	100000	220369	900000	1760000	44
Adj_1	296101	500000	1229066	2052523	2
Conf_1	3	4	5	4.4	0
Des_2	200000	400000	1000000	1760000	28
Des_2_P	196227.5	350000	1000000	1768299.5	27
Adj_2	315000	587000	1299250	2165350	13
Adj_2_p	340301	580000	1330000	2177337.8	4
Conf_2	3	4	5	4.4	0
Conf_2_p	340000	564731.5	1330000	2178000	8

5.2 Profile of participants

Table 5-4 presents the demographic characteristics of the participants in the main experiment. Among the 68 participants, 55.41% were male and 45.59% female. The age group 36–45 provided the largest numbers, followed by the group aged 26–35. These two groups represented the majority of participants (82.36%). No participant was over 55. In terms of education background, 35.29% participants were college/university graduates or held a degree of the same level; 64.71% were postgraduates. No participant had any other educational background, which indicates that all were relatively highly educated. Furthermore, three-quarters were in tourism and hospitality-related industrial professions, and academic researchers in this area only represented 25%. This sample distribution shows that the majority of the participants had rich working experience in the industry and that their judgmental forecasting would have shared a variety of cognitive features. Also because of this sample distribution, the results of the current study say more about the real judgmental forecasting features of tourism demand in the industry. The debiasing strategies proposed in this study can be easily applied in tourism-related sectors.

Table 5-4 Profile of participants in the main experiment (N=68)

Characteristics	Valid N	Percentage
Gender		
Male	37	54.41%
Female	31	45.59%
Age		
16-25	2	2.94%
26-35	26	38.24%
36-45	30	44.12%
46-55	10	14.71%
Education		
College/university	24	35.29%
Postgraduate	44	64.71%
Field		
Tourism and hospitality industrial profession	51	75.00%
Tourism and hospitality academic researcher	17	25.00%

5.3 Hypotheses testing

5.3.1 Descriptive analysis of cognitive biases

The features of five cognitive bias identified in participants' unaided judgmental forecasting process (the first round of experiment) were analyzed before testing the hypotheses. First, the number of each cognitive bias observed in the experiment (N) were counted, and the proportions of them in the entire judgmental forecasts (%) were calculated. Furthermore, the measures of each cognitive bias were produced according to the methods developed in Chapter 3; and the mean, minimum, and maximum values of each measure were also analyzed. Regarding participants' demographical characteristics, such as gender and age group, the features of cognitive biases were further categorized as shown in Table 5-5.

Regarding the desire bias, participants performed too optimistic in their expectation because the desired outcomes were generally higher than the real outcome in a large scale (MPE=-0.58). Furthermore, male participants' desire bias was almost twice as female participants'. Regarding the mean percentage change and standard deviation of three anchoring bias, the percentage changes on statistical forecast anchor is relatively smaller than the percentage changes on desire anchor and latest observation anchor. Specifically, female participants performed more conservatively than male participants; and so did the younger participants comparing with older participants. Different from the measures of previous four cognitive biases, overconfidence bias was measured by the difference between participants' own confidence evaluation and their real performance – the proportion of correct adjustment from the baseline forecasts. As shown in Table 5-5, participants were generally overconfident on their judgmental forecasts (Mean=0.24), and such phenomenon was consistent between male and female participants, as well as among the participants in different age groups.

Table 5-5 Indicators of cognitive biases

		N	%	Mea	Min.	Max.	Sd.
Desire Bias							
	Total	401	58.97	-0.58	-1.00	1.00	0.41
	Gender						
	Male	241	35.44	-0.44	-0.79	1.00	0.73
	Female	160	23.53	-0.22	-1.00	1.00	0.76
	Age						
	16-25	1	0.15	0.87	0.87	0.87	
	26-35	132	19.41	-0.13	-0.79	1.00	0.36
	36-45	189	27.79	-0.61	-1.00	1.00	0.86
	46-55	79	11.62	0.81	-0.43	1.00	0.38
Statistical Forecast Anchor	Total	461	67.79	0.04	-0.28	0.41	0.02
	Gender						
	Male	246	36.18	0.07	-0.28	0.41	0.61
	Female	215	31.62	0.01	-0.22	0.25	0.05
	Age						
	16-25	16	2.35	0.01	-0.20	0.15	0.07
	26-35	175	25.74	0.01	-0.22	0.25	0.05
	36-45	204	30	0.03	-0.27	0.64	0.10
	46-55	66	9.71	0.16	-0.28	0.41	0.16
Desire Anchor	Total	57	8.38	0.57	-0.90	1.03	0.44
	Gender						
	Male	27	3.97	0.37	-0.89	0.95	0.76
	Female	30	4.41	0.75	-0.90	1.03	0.85
	Age						
	16-25	0	0				
	26-35	18	2.65	0.72	-0.79	0.72	0.34
	36-45	31	4.56	-0.01	-0.90	0.98	0.51
	46-55	8	1.18	0.54	-0.40	1.03	0.48
Latest Observation Anchor	Total	154	22.65	0.12	-0.45	0.41	0.06
	Gender						
	Male	92	13.53	0.19	-0.45	0.41	0.45
	Female	62	9.12	0.01	-0.29	0.28	0.07
	Age						
	16-25	4	0.59	0.02	0.00	0.05	0.02
	26-35	61	8.97	0.05	-0.29	0.31	0.34
	36-45	63	9.26	0.23	-0.07	0.41	0.09
	46-55	26	3.82	0.00	-0.45	0.28	0.15
Overconfidence Bias	Total	680	100	0.24	-0.50	0.90	0.01
	Gender						
	Male	370	54.41	0.26	-0.50	0.90	0.29
	Female	310	45.59	0.22	-0.50	0.90	0.25
	Age						
	16-25	20	2.94	0.05	-0.20	0.40	0.16
	26-35	260	38.24	0.26	-0.50	0.90	0.27
	36-45	300	44.12	0.22	-0.50	0.80	0.28
	46-55	100	14.71	0.32	-0.30	0.90	0.24

Table 5-6 Cognitive Errors

		N	%	Mean	Min.	Max.	Sd.
Desire Bias	Total	401	58.97	-1470190	-147603780	1957232	554876
	Gender						
	Male	241	35.44	-2472375	-147603780	394415	14254279
	Female	160	23.53	39352	-1726377	1957232	298236
	Age						
	16-25	1	0.15	73461	73461	73461	
	26-35	132	19.41	9286	-474603	394415	84619
	36-45	189	27.79	-3133603	-147603780	1957232	16044231
	46-55	79	11.62	17793	-316445	343473	79046
Statistical Forecast Anchor	Total	287	42.21	-88215	-11294791	130340	40058
	Gender						
	Male	171	25.15	-114988	-11294791	130340	866332
	Female	116	17.06	-48749	-1819208	128591	182183
	Age						
	16-25	6	0.88	-44315	-88797	-2064	36346
	26-35	105	15.44	-28982	-370618	128591	69890
	36-45	130	19.12	-73635	-1819208	130340	191202
	46-55	46	6.76	-270355	-11294791	89232	1663638
Desire Anchor	Total	33	4.85	389350	-1151996	1915556	128929
	Gender						
	Male	23	3.38	370804	-176789	1863225	618491
	Female	10	1.47	432005	-1151996	1915556	1006186
	Age						
	16-25	0	0				
	26-35	9	1.32	336735	-113	1863225	690378
	36-45	18	2.65	289092	-1151996	1915556	741883
	46-55	6	0.88	769048	-19275	1872847	814003
Latest Observation Anchor	Total	153	22.5	-24581	-4314209	457206	29587
	Gender						
	Male	92	13.53	-53208	-4314209	457089	465683
	Female	61	8.97	18593	-139996	457206	85180
	Age						
	16-25	4	0.59	-16088	-27409	-2375	10417
	26-35	61	8.97	16387	-364553	457206	96438
	36-45	62	9.12	-99024	-4314209	143243	552008
	46-55	26	3.82	55511	-98438	457089	154904
Overconfidence Bias	Total	533	78.38	74594	-307896	3593477	14797
	Gender						
	Male	287	42.21	53622	-307896	3098867	296043
	Female	246	36.18	99061	-184491	3593477	387275
	Age						
	16-25	18	2.65	-274	-5280	3180	1924
	26-35	194	28.53	167769	-236278	3593477	534765
	36-45	236	34.71	16585	-307896	596728	88135
	46-55	85	12.5	38853	-200212	1191850	168264

Table 5-6 summarizes the forecast errors associated with different cognitive biases. Generally, desire bias led to negative forecast error among male participants but such bias had opposite effect on female participants. Desire bias generated huge negative forecast error from the participants aged between 36 and 45, but such effect was relatively smaller and positive among the participants in other age groups. Regarding the three variations of anchoring bias, statistical forecast anchor led to negative forecast errors and desire anchor led to positive forecast errors. Such effect was consistent among all gender and age groups. Latest observation anchor led to negative forecast errors for male participants but positive forecast errors for female participants. Except participants in the age group of 16-25, overconfidence bias led to positive forecast errors.

5.3.2 Detection of statistical forecast bias (H1)

The influence of statistical forecast bias on final forecast error is not the main focus of this study; however, the validity of the cognitive bias detection model greatly depends on the significance of the decomposed forecast error associated with statistical forecast bias. Therefore, it is necessary to test whether statistical forecast bias contributes significantly to forecast error after judgmental forecasting. According to the hypothesis testing method in Section 4.4.1.4, the series used to test the significance of statistical forecast bias were the real outcome, statistical forecast of tourist arrivals for all D-O pair markets in 2006, and the system-generated coefficient $\beta_{1,j}$. Using these three series, the decomposed forecast error associated with statistical forecast bias in participants' first round of forecasting ($STAT_{2006}$) is calculated with the following equation:

$$STAT_{i,j,2006} = \beta_{1_j} * \frac{Q_{i,2006} - SF_{i,2006}}{Q_{i,2006}} * Q_{i,2006}$$

The result of one-sample *t*-test on $STAT_{i,j}$ is shown in Table 5-7.

Table 5-7 Result of hypothesis testing (H1)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
$STAT_{2006}$	680	9148	96297	679	2.477	0.013

Forecast errors associated with statistical forecast bias for all participants are statistically significant (*M*=9148, *SD*=96297, *t*(679)=2.477, *p*=0.013). Therefore, **H1 is rejected**: statistical forecast bias has a significant influence on final forecast error after judgmental forecasting. This result provides evidence that, when developing the cognitive bias detection model, it is valid to incorporate the component of statistical forecast bias in order to achieve a better fit with the real data, even if this component is not a part of cognitive bias.

5.3.3 Detection of desire bias

5.3.3.1 Desire bias (H2a)

To test the existence of desire bias, two series of data must be further analyzed, including the real outcome and participants' desired outcome of tourist arrivals for all D-O pair markets in 2006. The percentage error of desired outcome (PE_DO_{2006}) is calculated as:

$$PE\_DO_{i,j,2006} = \frac{Q_{i,2006} - DO_{i,j,2006}}{Q_{i,2006}}$$

Some participants did not have a desired outcome for all D-O pair markets in the experiment; desire bias cannot be reflected in such cases. Therefore, the cases with no desired outcome were excluded from this part of the analysis. Furthermore, the variable PE_DO also contains some outliers that may bias statistical analysis. Following the same approach to detecting outliers in the raw data, 34 outliers were also identified from PE_DO. As a result, $PE\_DO_{i,j}$ was developed with 327 valid cases. The result of one-sample *t*-test on PE_DO is shown in Table 5-8.

Table 5-8 Result of hypothesis testing (H2a)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
PE_DO_{2006}	327	0.566	0.481	326	21.268	<0.001

Among the participants who provided desired outcomes, desire error is significant (M=0.566, SD=0.481, $t(326)$=21.268, p<0.001). Therefore, **H2a is supported**. A positive value of mean percentage error (MPE=0.566) reveals that, when participants had a desired outcome of tourist arrivals for a particular D-O pair market, such expectations were generally lower than the real outcome. In other words, participants showed conservatism (or pessimism) bias in their desired outcomes.

5.3.3.2 The contribution of desire bias to final forecast error (H2b)

Besides the real outcome and participants' desired outcome of tourist arrivals for all D-O pair markets in 2006, the system-generated coefficient β_2 is also required to test the contribution of desire bias to final forecast error. The decomposed forecast error associated with desire bias in participants' first round of forecasting (DES_{2006}), which is also defined as the leading indicator of desire bias, is calculated with the following equation:

$$DES_{i,j,2006} = \beta_{2j} * \frac{Q_{i,2006} - DO_{i,2006}}{Q_{i,2006}} * Q_{i,2006}$$

Following the same method of extracting irrelevant data from the variable $PE\_DO_{i,j}$, the cases with no desired outcome were excluded from this part of the analysis. The outliers of PE_DO were also detected and removed from the series using the boxplot method. As a result, $DES_{i,j}$ with 397 valid cases was developed. The result of one-sample *t*-test on $DES_{i,j}$ is shown in Table 5-9.

Table 5-9 Result of hypothesis testing (H2b)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
DES_{2006}	397	-467473	402151	396	-2.316	0.021

Forecast errors associated with desire bias among all participants were statistically significant (*M*=-467473, *SD*=402151, *t*(396)=-2.316, *p*=0.021). Therefore, **H2b is supported**; desire bias significantly contributes to final forecast error.

5.3.4 Detection of anchoring bias

The way to test the existence of anchoring bias in people's judgmental forecasting differs from the methods in previous hypotheses testing. Considering the fact that the anchoring effect exists when judgmental forecasting begins with an initial position and gives forecasts around it, anchoring bias is observed if the judgmental forecasts are lacking adjustment. Therefore, judgmental forecasts with anchoring bias should be normally distributed with a mean value equal to the value of the anchor (e.g., statistical forecast, desired outcome, or latest observation). With a concern that the judgmental forecasts and the real outcomes of tourist arrivals for 10 D-O pairs were in different scales (e.g., the tourist arrivals from Mainland China to Hong Kong in 2006 were up to 2,177,232; while those from Korea ROK to Chinese Taipei were only 196,260), standardized distance between judgmental forecast and the anchoring would be appropriate to measure anchoring bias. Therefore, percentage change on detected anchors was adopted as the indicator of anchoring bias in this study.

5.3.4.1 Anchoring bias on statistical forecast (H3a)

To test the existence of anchoring bias on statistical forecast, four series of data were extracted from the dataset, including participants' judgmental forecasts and the statistical forecasts of tourist arrivals for all D-O pair markets in 2006 and 2007. When participants' judgmental forecast anchored on statistical forecast, the percentage change was calculated as:

$$\begin{cases} A_{S_{i,j,2006}} = \dfrac{F_{i,j,2006} - SF_{i,2006}}{SF_{i,2006}} \\ A_{S_{i,j,2007}} = \dfrac{F_{i,j,2007} - SF_{i,2007}}{SF_{i,2007}} \end{cases}$$

Considering the valid cases of $A_{S_{i,j,2006}}$ and $A_{S_{i,j,2007}}$ in this part of the analysis, only cases whose statistical forecast anchor was detected in the two rounds of forecasting were collected as valid cases ($I_{S_{i,j,2006}} = 1$ and $I_{S_{i,j,2007}} = 1$). Outlier checking was also conducted on variables $A_{S_{i,j,2006}}$ and $A_{S_{i,j,2007}}$. Finally, these two variables were developed with 461 and 293 valid cases, respectively. The results of one-sample *t*-tests on $A_{S_{i,j,2006}}$ and $A_{S_{i,j,2007}}$, as well as the combination of these two series, are shown in Table 5-10.

Table 5-10 Result of hypothesis testing (H3a)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
$A_{S_{2006}}$	461	0.043	0.445	460	2.090	0.037
$A_{S_{2007}}$	293	0.081	0.736	292	1.892	0.060
$A_{S_{all}}$	754	0.058	0.575	753	2.771	0.006

The percentage changes on statistical forecast in the first round of forecasting were significantly larger than zero (*M*=0.043, *SD*=0.481, *t*(460)=2.090, *p*=0.037), but not significant in the second round of forecasting (*M*=0.081, *SD*=0.736, *t*(292)=1.892, *p*=0.060). Further investigation of all judgmental forecasts with detected statistical forecast anchors reveals that the percentage change was significantly larger than zero (*M*=0.058, *SD*=0.575, *t*(753)=2.771, *p*=0.006). Therefore, **H3a is rejected**, indicating that anchoring bias in statistical forecast is not significant when participants' forecasts are close to the statistical forecasts. Positive values of mean percentage change reveal that, in this scenario, participants' forecasts were always larger than the statistical forecasts, which shows optimism bias in judgmental forecasting.

Fildes et al. (2009) investigated people's optimism bias in two judgmental forecasting situations: judgmental adjustment in the right direction and in the wrong direction. They identified significant differences of optimism bias in these two situations. Their study inspired the current study to investigate whether significant differences in anchoring bias can be identified in these two situations. Specifically, right direction of adjustment in the condition of the statistical forecast anchor detected means that participants' judgmental forecast and the real outcome of tourist arrivals are both either larger or smaller than the statistical forecast; wrong direction of adjustment in the same condition means that the statistical forecast is located between participants' judgmental forecast and the real outcome. Regarding the direction of adjustment, the cases of judgmental forecast with the statistical forecast anchor detected are further classified and Table 5-11 shows the valid number of cases in each group:

Table 5-11 Sample size grouped by forecasting round and direction of adjustment

	As_{2006}	As_{2007}	As_{all}
Right direction (R)	136	97	233
Wrong direction (W)	325	196	521

Further investigation of the significance of $As_{i,j,2006}$ amd $As_{i,j,2007}$ was conducted by a series of one-sample *t*-tests. AS shown in Table 5-12, the percentage change on statistical forecast was significantly larger than zero in both rounds of forecasting when adjustment was in the right direction; the same indicator was not significantly different from zero when participants made adjustment in the wrong direction. As a result, **H3a is partially supported**. Anchoring bias on the statistical forecast is significant in the judgmental forecasting of tourism demand when participants made adjustment in the wrong direction. However, such anchoring bias was not significant when participants made adjustment in the right direction. In this situation, the anchoring effect is mainly accompanied by optimism bias since the mean percentage change is positive, which may be the main cause of such a conclusion.

Table 5-12 Result of hypothesis testing (H3a_additional)

	Test Value = 0					
Variable	N	Mean	SD	*df*	*t*	Sig.
$A^R_{S\ 2006}$	136	0.074	0.111	135	7.855	<0.001
$A^R_{S\ 2007}$	97	0.411	0.779	96	5.194	<0.001
$A^R_{S\ all}$	233	0.214	0.535	232	6.124	<0.001
$A^W_{S\ 2006}$	325	0.030	0.524	324	1.040	0.299
$A^W_{S\ 2007}$	196	-0.082	0.656	195	-1.743	0.083
$A^W_{S\ all}$	521	-0.012	0.580	520	-0.468	0.640

** A^R_S indicates adjustments in the right direction; A^W_S indicates adjustments in the wrong direction.*

5.3.4.2 Anchoring bias on desired outcome (H3b)

Participants' judgmental forecasts and their desired outcomes of tourist arrivals for all D-O pair markets in 2006 and 2007 were extracted from the data set to test the existence of anchoring bias on the desired outcome. When participants' judgmental forecasts were anchored on desired outcome, the percentage change is calculated as:

$$\begin{cases} A_{D\,i,j,2006} = \dfrac{F_{i,j,2006} - DO_{i,j,2006}}{DO_{i,2006}} \\ A_{D\,i,j,2007} = \dfrac{F_{i,j,2007} - DO_{i,j,2007}}{DO_{i,j,2007}} \end{cases}$$

The valid cases for $A_{D\,i,j,2006}$ and $A_{D\,i,j,2007}$ in this part of the analysis were extracted in three steps. In the first step, the cases with desired outcome provided in both rounds of forecasting were selected. Then the cases in which the desire anchor was detected were further collected as valid cases ($I_{D\,i,j,2006} = 1$ and $I_{D\,i,j,2007} = 1$). In the last step, outlier checking was conducted to finalize the valid cases for these two variables. Through these three steps, $A_{D\,i,j,2006}$ and $A_{D\,i,j,2007}$ were developed with 57 and 87 valid cases, respectively. The results of one-sample *t*-tests on these two variables, as well as their combination, are shown in Table 5-13.

Table 5-13 Result of hypothesis testing (H3b)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
$A_{D_{2006}}$	57	0.571	3.352	56	1.285	0.204
$A_{D_{2007}}$	87	-1.493	7.920	86	-1.758	0.082
$A_{D_{all}}$	144	-0.676	6.569	143	-1.235	0.219

The percentage changes on desired outcome in the experiment were not significantly different from zero (M=-0.676, SD=6.569, $t(143)$=-1.235, p=0.219). This conclusion is consistent among the subsamples of the first round of forecasting (M=0.571, SD=3.352, $t(56)$=1.285, p=0.204) and the second round of forecasting (M=-1.493, SD=7.920, $t(87)$=-1.758, p=0.082). Therefore, **H3b is supported**: anchoring bias on desired outcome exists in participants' judgmental forecasting. The value of mean percentage change and the standard deviation reveals that participants made dramatic changes to their desired outcomes. According to the PE of desired outcome (Table 5-8), the error of participants' desired outcome is also significantly large. As a result, participants' desired outcome usually performed very poorly and they made larger changes to it when new and reliable information was provided.

5.3.4.3 Anchoring bias on the latest observation (H3c)

Similar to the means of testing anchoring biases on statistical forecast and desired outcome, four series were required to test the existence of anchoring bias on the latest observation, including participants' judgmental forecasts and the latest observation for all D-O pair markets in 2006 and 2007. In the first round of forecasting, participants were asked to forecast tourist arrivals in 2006; the latest observation in the first round of forecast was thus the real outcome of tourist arrivals in 2005; in the second round, the target year was 2007, so the tourist arrivals in 2006 became the latest observation. Therefore, the percentage change in participants' judgmental forecast anchoring on the latest observation is calculated according to the following equations:

$$\begin{cases} A_{L_{i,j,2006}} = \dfrac{F_{i,j,2006} - Q_{i,2005}}{Q_{i,2005}} \\ A_{L_{i,j,2007}} = \dfrac{F_{i,j,2007} - Q_{i,2006}}{Q_{i,2006}} \end{cases}$$

The valid cases for $A_{L_{i,j,2006}}$ and $A_{L_{i,j,2007}}$ in this part of the analysis were extracted in the same way as in the construction of $A_{S_{i,j,2006}}$ and $A_{S_{i,j,2007}}$: only the cases with the latest observation anchor detected in both rounds of forecasting were collected as valid cases ($I_{L_{i,j,2006}} = 1$ and $I_{L_{i,j,2007}} = 1$). After outlier checking, 154 and 300 cases were extracted for the two developed variables $A_{S_{i,j,2006}}$ and $A_{S_{i,j,2007}}$, respectively. The results of one-sample *t*-tests on $A_{L_{i,j,2006}}$ and $A_{L_{i,j,2007}}$, as well as the combination of these two series, are shown in Table 5-14.

Table 5-14 Result of hypothesis testing (H3c)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
$A_{L_{2006}}$	154	0.115	0.736	153	1.946	0.053
$A_{L_{2007}}$	300	-0.357	4.480	299	-1.381	0.168
$A_{L_{all}}$	454	-0.197	3.672	453	-1.142	0.254

According to the testing result, the percentage changes on the latest observation in the experiment are not significantly different from zero (*M*=-0.197, *SD*=3.672, *t*(453)=-1.142, *p*=0.254). This conclusion is consistent with the subsamples of the first round of forecasting (*M*=0.115, *SD*=0.736, *t*(153)=1.946, *p*=0.053) and the second round of forecasting (*M*=-0.357, *SD*=4.080, *t*(299)=-1.381, *p*=0.168). Therefore, **H3c is supported**, anchoring bias on the latest observation exists in participants' judgmental forecasting. The negative value of mean percentage change (-0.197) further reveals that, when participants anchored their judgmental forecast on the latest observation of tourist arrivals, the adjustments were generally lower than the latest observations. In other words, participants expressed conservative (or pessimism bias) in their adjustment in this situation.

5.3.4.4 The contribution of the statistical forecast anchor to final forecast error (H3d)

Regarding the statistical forecast anchor, the contribution of anchoring bias to the final forecast error is measured by the leading indicator A_STAT_{2006}, which is the decomposed forecast error associated with the statistical forecast anchor. Based on the percentage change on statistical forecast (A_S) and the system-generated coefficient β_3, A_STAT_{2006} can be calculated as:

$$A\_STAT_{i,j,2006} = \beta_{3_j} * A_{S_{i,j,2006}} * Q_{i,2006}$$

Considering the valid cases of $A\_STAT_{i,j,2006}$, only the cases with the statistical forecast anchor detected in the first round of forecasting are extracted ($I_{S_{i,j,2006}} = 1$). In addition, some participants' β_{3_j} were equal to zero, which indicates that the relationships between these participants' statistical anchor and the final forecast error were not significant. Therefore, the cases with $\beta_{3_j} = 0$ were excluded from the valid cases of $A\_STAT_{i,j,2006}$. Outlier checking was also conducted to remove the cases with extreme values in the series.

Finally, $STAT_{i,j,2006}$ was developed with 287 valid cases. The result of one-sample *t*-test on $STAT_{i,j}$ is shown in Table 5-15.

Table 5-15 Result of hypothesis testing (H3d)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
A_STAT_{2006}	287	-88215	678621	286	-2.202	0.028

The decomposed forecast errors associated with the statistical forecast anchor among all participants were statistically significant (*M*=-88215, *SD*=678621, *t*(286)=-2.202, *p*=0.028). Thus, **H3d is supported**, anchoring bias on statistical forecast significantly contributes to the final forecast error.

5.3.4.5 *The contribution of the desire anchor to final forecast error (H3e)*

Regarding the desire anchor, the contribution of anchoring bias to the final forecast error is measured by the leading indicator A_DES_{2006}, which is the decomposed forecast error associated with the desire anchor. Using the percentage change on desired outcome (A_D) and the system-generated coefficient β_4, A_DES_{2006} can be calculated according to the following equation:

$$A\_DES_{i,j,2006} = \beta_{4_j} * A_{D_{i,j,2006}} * Q_{i,2006}$$

Following the same method to extract irrelevant data of variable $A\_STAT_{i,j,2006}$, the cases with no anchoring bias on desired outcome in the first round of forecasting were not considered as valid cases and deleted from the series. In addition, β_{4_j} for some participants was equal to zero, which indicates that the relationship between the statistical anchor and the final forecast error were not significant for these participants. Therefore, the cases with $\beta_{4_j} = 0$ were also excluded from the series. After the outlier checking, 33 valid cases were collected for variable $A\_DES_{i,j,2006}$. The result of one-sample *t*-test on $A\_DES_{i,j}$ is shown in Table 5-16.

Table 5-16 Result of hypothesis testing (H3e)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
A_DES_{2006}	33	389350	740639	32	3.020	0.005

The decomposed forecast errors associated with the desire anchor among all participants were statistically significant (*M*=-389350, *SD*=740639, *t*(32)=3.020, *p*=0.005). Thus, **H3e is supported**, anchoring bias on desired outcome significantly contributes to the final forecast error. However, it is noted that the sample used to test H3e is relatively small; only 33 valid cases were collected from the data set. The first reason for limited number of available cases is that many participants did not have any expectations for specific D-O pair markets in the first round of forecasting, so they did not provide desired outcomes of tourist arrivals for these markets. As a result, 41% of the cases did not contain desired outcome data. Second, among the 59% cases in which participants provided desired outcomes, only 14% are identified with the desire anchor. As a result, less than 5% of the cases were kept after outlier checking.

5.3.4.6 *The contribution of the latest observation anchor to final forecast error (H3f)*

Regarding the anchor of latest real outcome of tourist arrivals, the contribution of anchoring bias to the final forecast error is measured by the leading indicator A_LAS_{2006}, which is the decomposed forecast error associated with the latest observation anchor. Using the percentage change on the latest observation (A_L) and the system-generated coefficient β_5, A_LAS_{2006} can be calculated as:

$$A\_LAS_{i,j,2006} = \beta_{5_j} * A_{L_{i,j,2006}} * Q_{i,2006}$$

Considering the valid cases of $A\_LAS_{i,j,2006}$, only the cases with the latest observation anchor detected in the first round of forecasting were extracted ($I_{L_{i,j,2006}} = 1$). For a particular participant j, $\beta_{5_j} = 0$ indicates that the relationship between the participant's latest observation anchor and the final forecast error is not significant; cases with $\beta_{5_j} = 0$ were thus considered invalid and removed from the series. Outlier checking was also conducted to extract cases with extreme values. Finally, $A\_LAS_{i,j,2006}$ was developed with 153 valid cases. The result of one-sample *t*-test on $A\_LAS_{i,j,2006}$ is shown in Table 5-17.

Table 5-17 Result of hypothesis testing (H3f)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
A_LAS_{2006}	153	-24581	365977	152	-0.831	0.407

The decomposed forecast errors associated with the latest observation anchor among all participants were not statistically significant (*M*=-24581, *SD*=365977, *t*(152)=-0.831, *p*=0.407). Thus, **H3f is rejected**: the contribution of anchoring bias to final forecast error was not significant regarding the anchor of the latest observation.

5.3.5 Detection of overconfidence bias

Unlike forecast-related variables, the method used to measure participants' confidence in their judgmental forecast was a 5-point Likert scale. Overconfidence indicates that a forecaster's confidence in a specific forecast exceeds the average level of correct forecasts. Correct forecast means that the judgmental forecast is closer to the real outcome than the baseline (statistical) forecast. Overconfidence is measured by the difference between confidence in a certain forecast and the percentage of correct forecasts. Lack of confidence is identified when confidence in a forecast is lower than the percentage of correct forecasts.

5.3.5.1 Overconfidence bias (H4a)

To test the existence of overconfidence bias, participants' judgmental forecasts and their evaluations of confidence in those forecasts, as well as the real outcome of tourist arrivals for all D-O pair markets in 2006 and 2007, were selected. First, the series of participants' confidence was transformed into a standardized variable on a 0–1 scale by dividing by the maximum evaluation of confidence (5 points). Second, the judgmental forecast error and the baseline forecast error were compared and cases with smaller judgmental forecast error were labeled as correct forecasts. For each participant, the number of correct forecasts was counted and divided by the total number of his/her forecasts (10), which gives a percentage of correct forecasts. Therefore,

$$\begin{cases} O_{i,j,2006} = Confidence_{i,j,2006} - \frac{\sum_{i=1}^{10} I_{i,j,2006}}{10}, \quad I_{i,j,2006} = \begin{cases} 1, if \left|pe(F_{i,j,2006})\right| < \left|pe(SF_{i,j,2006})\right| \\ 0, if \left|pe(F_{i,j,2006})\right| > \left|pe(SF_{i,j,2006})\right| \end{cases} \\ O_{i,j,2007} = Confidence_{i,j,2007} - \frac{\sum_{i=1}^{10} I_{i,j,2007}}{10}, \quad I_{i,j,2007} = \begin{cases} 1, if \left|pe(F_{i,j,2007})\right| < \left|pe(SF_{i,j,2007})\right| \\ 0, if \left|pe(F_{i,j,2007})\right| > \left|pe(SF_{i,j,2007})\right| \end{cases} \end{cases}$$

Since all participants provided their confidence in the judgmental forecasts and no outlier was identified using the boxplot method, all cases were valid for use in this part of the analysis. The results of one-sample *t*-test on $O_{i,j,2006}$ and $O_{i,j,2007}$, as well as their combination, are shown in Table 5-18.

Table 5-18 Result of hypothesis testing (H4a)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
O_{2006}	680	0.244	0.271	679	23.478	<0.001
O_{2007}	680	0.074	0.381	679	5.034	<0.001
O_{all}	1360	0.159	0.341	1359	17.168	<0.001

Participants' expressed overconfidence was significant (M=0.159, SD=0.341, $t(1359)$=17.168, p<0.001). Such a conclusion is consistent within the subsamples of the first round of forecasting (M=0.244, SD=0.271, $t(679)$=23.478, p<0.001) and the second round of forecasting (M=0.074, SD=0.381, $t(679)$=5.034, p<0.001). Therefore, **H4a is supported**: overconfidence bias was significant in participants' judgmental forecasting. According to Table 5-18, the mean value of $O_{i,j,2006}$ was 0.244. This means that participants' confidence in their judgmental forecasts was, on average, 24.4% higher than their real performance, which further indicates a high level of overconfidence observed in the first round of forecasting. In the second round, the mean value of $O_{i,j,2007}$ was smaller than the mean value of $O_{i,j,2006}$. A paired-sample *t*-test on overconfidence in the first and second rounds of forecasting (Table 5-19) also revealed that participants' overconfidence bias was significantly reduced in the second round (M=0.171, SD=0.359, $t(679)$=12.410, p<0.001).

Table 5-19 Mean difference of overconfidence in two rounds of forecasting

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
*Diff_O**	680	0.171	0.359	679	12.410	<0.001

** Diff_O indicates the difference of overconfidence between the 1st and the 2nd round of forecasting.*

5.3.5.2 The contribution of overconfidence bias to final forecast error (H4b)

The contribution of overconfidence bias to final forecast error is measured by the leading indicator OVE_{2006}, which is actually the decomposed forecast error on overconfidence. This part of the analysis uses three variables, including the real outcome of tourist arrivals for all D-O pair markets in 2006, the system-generated coefficient β_6, and the newly generated measurement of overconfidence (O_{2006}). OVE_{2006} can be calculated with the following equation:

$$OVE_{i,j,2006} = \beta_{6_j} * O_{i,j,2006} * Q_{i,2006}$$

Some participants' β_{6_j} was equal to zero, which indicates that the relationships between these participants' overconfidence bias and the final forecast error were not significant. Considering the valid cases of $OVE_{i,j,2006}$, such cases with $\beta_{6_j} = 0$ were excluded. Also, because no outlier was detected in $O_{i,j,2006}$, there were 560 valid cases for $OVE_{i,j}$. The result of one-sample *t*-test on $OVE_{i,j,2006}$ is shown in Table 5-20.

Table 5-20 Result of hypothesis testing (H4b)

Variable	Test Value = 0					
	N	Mean	SD	*df*	*t*	Sig.
OVE_{2006}	560	70998	333646	559	5.036	<0.001

The decomposed forecast errors associated with overconfidence bias among all participants were statistically significant (*M*=70988, *SD*=333646, *t*(559)=5.036, *p*<0.001). Therefore, **H4b is supported**: overconfidence bias contributes significantly to the final forecast error.

5.3.6 Debiasing of desire bias

5.3.6.1 Performance feedback of desire bias (H5a)

To test whether PF, which is the leading indicator of desire bias in the first round of forecasting, can effectively improve judgmental forecasting in the second round of forecasting, the APEs of unaided desired outcomes and the PF-based revisions in the second round of forecasting had to be compared. Three series from the dataset had to be used, including the real outcome of tourist arrivals, participants' unaided desired outcome, and participants' revised desired outcome based on the PF of their desire bias for all D-O pair markets in 2007. Let $D_{i,j,2007}$ and $D_{i,j,2007}^{P}$ denote the APEs of unaided desired outcome and the PF-based revision, respectively, calculated as:

$$\begin{cases} D_{i,j,2007} = \left| \dfrac{Q_{i,2007} - DO_{i,j,2007}}{Q_{i,2007}} \right| \\ D_{i,j,2007}^{P} = \left| \dfrac{Q_{i,2007} - DO_{i,j,2007}^{p}}{Q_{i,2007}} \right| \end{cases}$$

The valid cases for these two developed variables depend on the availability of desired outcome and outlier checking. Since some participants did not have desired outcomes for some D-O pair markets in the second round of forecasting, the cases with no desired outcome were excluded. Through outlier checking, 274 valid cases were finally selected for $D_{i,j,2007}$ and $D^{P}_{i,j,2007}$. Both unaided desired outcome and PF-based revision were given by the same group of participants; therefore, a paired-sample *t*-test was conducted on these two variables to test whether forecast accuracy differed significantly. The result is shown in Table 5-21.

Table 5-21 Result of hypothesis testing (H5a)

Variable	N	Mean	SD	*df*	*t*	Sig.
$D_{i,j,2007}$	274	0.442	0.306			
$D^{P}_{i,j,2007}$	274	0.424	0.279			
$D_{i,j,2007} - D^{P}_{i,j,2007}$	274	0.018	0.149	273	2.005	0.046

There was significant difference in the APEs of the unaided desired outcome (*M*=0.442, *SD*=0.306) and the PF-based revisions (*M*=0.424, *SD*=0.279); *t*(273)=2.005, *p*=0.046) of the participants who provided a desired outcome in the second round of forecasting. A positive value of mean difference (0.018) between $D_{i,j,2007|}$ and $D^{P}_{i,j,2007}$ reveals that the APE of PF-based revision of desired outcome was smaller than unaided desired outcome. Therefore, **H5a is supported**. This suggests that PF of a participant's desire bias in the previous forecasting season significantly reduced the desire bias in his/her following forecasting season. However, the extent of improvement in forecast accuracy was relatively small: on average, APE of 1.8% was observed. There was still a large forecast error in participants' desired outcome, which would seriously bias subsequent judgmental forecasting if participants were anchoring on their desired outcome.

5.3.6.2 System-suggested forecasts regarding desire bias (H5b)

As an example of suggestive guidance, system-suggested forecasts regarding desire bias is not achieved from participants' input, but is automatically calculated by TDFSS with

Equation (3.6). Therefore, five series are required: participants' unaided desired outcome, PF-based revision, system-suggested desired outcome, and the real outcome of tourist arrivals for all D-O pair markets in 2007, as well as the system-generated coefficient $\beta_{2,j}$ for all participants. Let $DO^{s}_{i,j,2007}$ denote the system-suggested desired outcome, which is calculated with the following equation:

$$DO^{s}_{i,j,2007} = DO_{i,j,2007} + \frac{1}{10}\sum_{i=1}^{10} \beta_{2_j} D_{i,j,2006} * Q_{i,2006}$$

Then the APE of $DO^{s}_{i,j,2007}$ can be calculated following the same method used to calculate $D_{i,j,2007}$ and $D^{P}_{i,j,2007}$. Let $D^{s}_{i,j,2007}$ denote the APE of $DO^{s}_{i,j,2007}$; further analysis will focus on the three APEs in order to test whether $D^{s}_{i,j,2007}$ is significantly smaller than the other two.

Considering the valid cases of these three APEs, the cases with desired outcome provided by forecasters in the second round of forecasting are selected. Since $\beta_{2,j}$ indicates the relationship between participant's desire bias and the final forecast error, when $\beta_{2,j}$ is equal to zero that relationship is not significant; the cases of $D^{s}_{i,j,2007}$ with $\beta_{2,j} = 0$ are not valid. Therefore, the cases with $\beta_{2,j} = 0$ are considered invalid. Based on these two criteria, as well as the outlier checking of the three APEs, 183 valid cases for $D_{i,j,2007}$, $D^{P}_{i,j,2007}$ and $D^{s}_{i,j,2007}$ were finally collected. Table 5-22 shows the result of the normality test of the three APEs; and Figure 5-1 offers a visual description of the series distribution, the Normal Q-Q plots of three APEs.

Table 5-22 Normality tests of $D_{i,j,2007}$, $D^{P}_{i,j,2007}$, and $D^{s}_{i,j,2007}$

Variable	N	Shapiro-Wilk		
		Statistic	*df*	Sig.
$D_{i,j,2007}$	182	0.875	182	<0.001
$D^{P}_{i,j,2007}$	182	0.941	182	<0.001
$D^{s}_{i,j,2007}$	182	0.854	182	<0.001

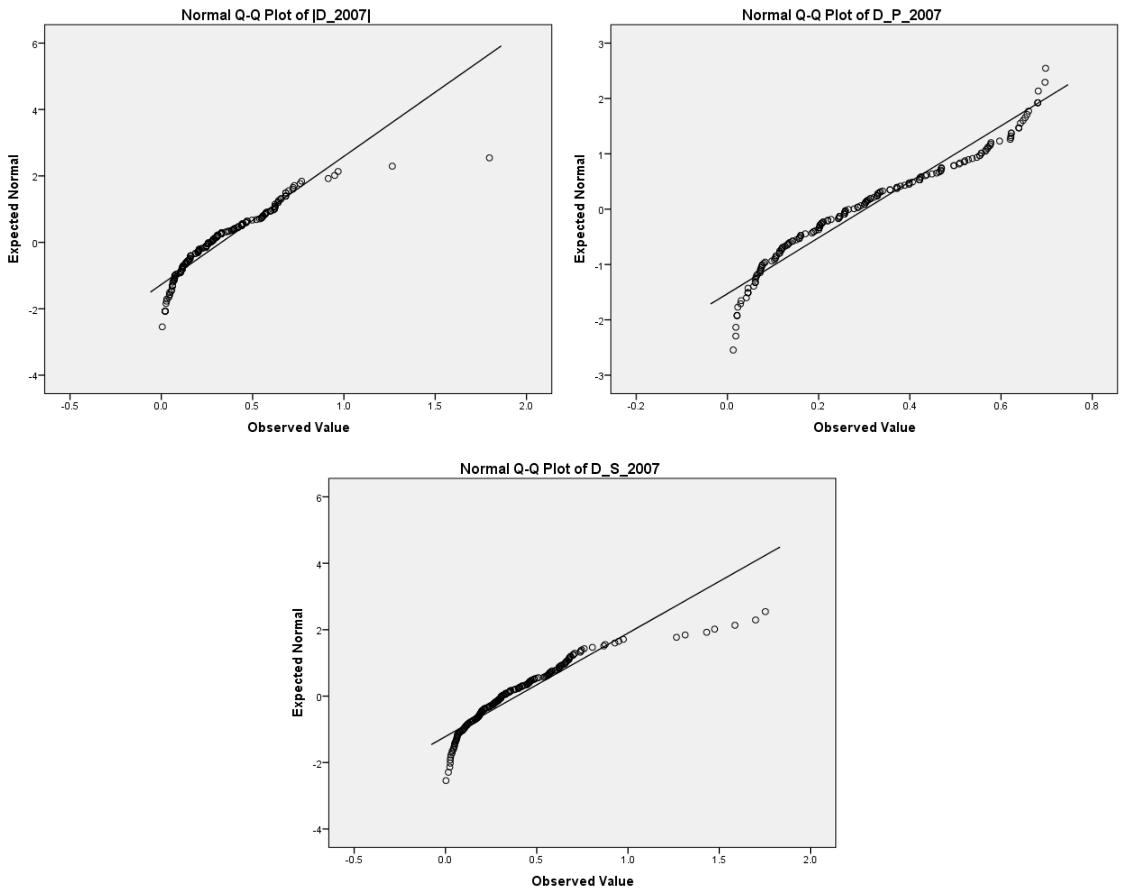


Figure 5-1 Normal Q-Q Plot of $D_{i,j,2007}$, $D^{P}_{i,j,2007}$, and $D^{s}_{i,j,2007}$

According to the results, all three APE series were significantly non-normal. Therefore, non-parametric tests are appropriate for the comparison of $D_{i,j,2007}$, $D^{P}_{i,j,2007}$, and $D^{s}_{i,j,2007}$.

Considering that $DO_{i,j,2007}$ and $DO^{P}_{i,j,2007}$ were given by the same group of people and that $DO^{s}_{i,j,2007}$ is calculated based on $DO_{i,j,2007}$, these three series of desired outcomes are related; so the APEs calculated based on the three sets of forecasts are also related. Therefore, the Friedman ANOVA test is the appropriate non-parametric method to test the differences between the three APEs. Table 5-23 shows the result of the Friedman ANOVA test on $D_{i,j,2007}$, $D^{P}_{i,j,2007}$, and $D^{s}_{i,j,2007}$.

Table 5-23 Result of hypothesis testing (H5b)

Variable	N	Mean	SD	Mean Rank	Friedman Test	
$D_{i,j,2007}$	182	0.329	0.259	1.978	Chi-Square	5.150
$D^{P}_{i,j,2007}$	182	0.302	0.198	1.901	*df*	2
$D^{s}_{i,j,2007}$	182	0.391	0.321	2.121	Asymp. Sig.	0.076

Among the three sets of desired outcome, PF-based revision of desired outcome had the lowest MAPE, and the lowest standard deviation (*M*=0.302, *SD*=0.198), followed by participants' unaided desired outcome (*M*=0.329, *SD*=0.259). System-suggested forecasts had the highest MAPE and the biggest standard deviation (*M*=0.391, *SD*=0.321). However, the differences between the three APEs are not statistically significant ($x^2(2) = 5.150, p = 0.076$). Therefore, **H5b is rejected**: system-suggested forecast with desire error correction did not outperform participants' unaided desired outcome and PF-based revisions. Wilcoxon tests were also used to follow up this result. According to the principle of Bonferroni correction, all effects were reported at a significance level of 0.017. It appears that $D^{s}_{i,j,2007}$ was significantly larger than $D^{P}_{i,j,2007}$ (*Z*=-3.282, *p*<0.001), likewise the difference between $D^{s}_{i,j,2007}$ and $D_{i,j,2007}$ (*Z*=-2.847, *p*=0.004); while, the difference between $D^P_{i,j,2007}$ and $D_{i,j,2007}$ was not significant (*Z*=-0.892, *p*=0.372). We can conclude that, in terms of desire bias, PF-based revision performs significantly more accurately than system-suggested desired outcome. PF-based revision also outperforms participants' unaided desired outcome but is not statistically significant.

5.3.7 Debiasing of anchoring bias

5.3.7.1 Performance feedback of anchoring bias on statistical forecast (H6a)

To test whether the PF of a forecaster's anchoring bias in the statistical forecast in the first round of forecasting can effectively improve his/her judgmental forecasting in the following round, the APEs of unaided judgmental forecasts and the PF-based revisions must be compared. Therefore, (i) the real outcome of tourist arrivals, (ii) participants' unaided judgmental forecast, and (iii) participants' revised forecasts based on the PF of their statistical forecast anchor for all D-O pair markets in 2007 need to be used. Let $AS_{i,j,2007}$ and $AS^P_{i,j,2007}$ denote the APEs of unaided judgmental forecast and the PF-based revision, respectively, when the statistical forecast anchor is detected; they are calculated as:

$$\begin{cases} AS_{i,j,2007} = |\dfrac{Q_{i,2007} - F_{i,j,2007}}{Q_{i,2007}}| \\ AS^P_{i,j,2007} = |\dfrac{Q_{i,2007} - F_{A^{\,p}_{i,j,2007}}}{Q_{i,2007}}| \end{cases}$$

The valid cases for these two APEs were extracted in three steps. First, cases with the statistical forecast anchor detected in the second round of forecasting were selected ($I_{S_{i,j,2007}} = 1$). Then the cases with $\beta_{3,j}$ equal to zero were removed because they indicate that the relationship between the statistical forecast anchor and the final forecast error is not significant. Finally, 12 outliers were identified and removed from the series. Following these three steps, 201 valid cases were collected for $AS_{i,j,2007}$ and $AS^{P}_{i,j,2007}$. Since unaided judgmental forecast and PF-based revision were given by the same group of participants, a paired-sample *t*-test was conducted on these two variables to test whether their effect on forecast accuracy was significantly different. Table 5-24 shows the result of the hypothesis testing.

Table 5-24 Result of hypothesis testing (H6a)

Variable	N	Mean	SD	*df*	*t*	Sig.
$AS_{i,j,2007}$	201	0.306	0.387			
$AS^{P}_{i,j,2007}$	201	0.093	0.127			
$AS_{i,j,2007} - AS^{P}_{i,j,2007}$	201	0.213	0.375	200	8.049	<0.001

Among the forecasts with a statistical forecast anchor detected from the second round of forecasting, there was a significant difference in the APEs of unaided judgmental forecasts (*M*=0.306, *SD*=0.387) and PF-based revisions (*M*=0.093, *SD*=0.127); *t*(200)=8.049, *p*<0.001. A positive value of mean difference (0.213) between $AS_{i,j,2007}$ and $AS^{P}_{i,j,2007}$ reveals that the APE of PF-based revision of anchoring bias was smaller than unaided judgmental forecast when statistical forecast anchor was detected. Thus, **H6a is supported**, suggesting that the PF of participants' anchoring bias in statistical forecast in the previous forecasting season can significantly reduce such anchoring bias in their following forecasting season. In addition, the mean difference of two APEs was 21.3% and the MAPE of PF-based revision was less than 10%, which indicates that PF produces a dramatic improvement in forecast accuracy.

5.3.7.2 *Performance feedback of anchoring bias on desired outcome (H6b)*

To test whether the PF of one's anchoring bias on desired outcome in the first round of forecasting can effectively improve judgmental forecasting in the second round, the APEs of unaided desired outcomes and the PF-based revisions in the second round of forecasting need to be compared. Three series need to be used: the real outcome of tourist arrivals, participants' unaided judgmental forecast, and participants' revised forecasts based on the PF of their anchoring bias for all D-O pair markets in 2007. Let $AD_{i,j,2007}$ and $AD^{P}_{i,j,2007}$ denote the APEs of unaided judgmental forecast and PF-based revision, respectively; these two variables are calculated in the same way as $AS_{i,j,2007}$ and $AS^{P}_{i,j,2007}$. The difference is that these two groups of variables contain different cases extracted from the dataset.

First of all, the cases with the desire anchor detected in the second round of forecasting were selected ($I_{D_{i,j,2007}} = 1$). Then the cases with $\beta_{4,j} = 0$ need to be removed; however, the result shows that no case with $\beta_{4,j}$ equal to zero was identified among the cases with the desire anchor detected. After that, outlier checking was conducted among the remaining cases and 33 valid cases were collected for $AD_{i,j,2007}$ and $AD^{P}_{i,j,2007}$. Following the same method used to test H5a and H6a, a paired-sample *t*-test was conducted on these two variables in order to test whether they produced a significant difference in forecast accuracy. The result is shown in Table 5-25.

Table 5-25 Result of hypothesis testing (H6b)

Variable	N	Mean	SD	*df*	*t*	Sig.
$AD_{i,j,2007}$	33	1.246	1.771			
$AD^{P}_{i,j,2007}$	33	0.429	0.496			
$AD_{i,j,2007} - AD^{P}_{i,j,2007}$	33	0.817	1.884	32	2.490	0.018

Among the forecasts with the desire anchor detected from the second round of forecasting, there was a significant difference in the APEs of unaided judgmental forecasts (M=1.246, SD=1.771) and PF-based revisions (M=0.429, SD=0.496); $t(32)$=2.490, p=0.018. A positive value of mean difference (0.817) between $|AD_{i,j,2007}|$ and $|AD^{P}_{i,j,2007}|$ reveals that the APE of PF-based revision of anchoring bias was smaller than unaided judgmental forecast when the desire anchor was detected. Therefore, **H6b is supported**. This suggests that the PF of participants' anchoring bias on the desired outcome in the previous forecasting season can significantly reduce such anchoring bias in the following forecasting season. Furthermore, it appears that the APE of participants' forecast improved by 81.7% with the help of PF.

5.3.7.3 Performance feedback of anchoring bias on the latest observation (H6c)

To test whether the PE of one's anchoring bias on the latest observation in the first round of forecasting can effectively improve judgmental forecasting in the following round, the APEs of unaided judgmental forecasts and the PF-based revisions need to be compared. Therefore, (i) the real outcome of tourist arrivals, (ii) participants' unaided judgmental forecasts, and (iii) participants' revised forecasts based on the PF of their latest observation anchor for all D-O pair markets in 2007 need to be used. Let $AL_{i,j,2007}$ and $AL^{P}_{i,j,2007}$ denote the APEs of unaided judgmental forecast and the PF-based revision when the latest observation anchor is detected, respectively; these two variables are calculated in the same way as $AD_{i,j,2007}$ and $AD^{P}_{i,j,2007}$; the difference here is the extraction of valid cases.

First, the cases with the latest observation anchor detected in the second round of forecasting were selected ($I_{L_{i,j,2007}} = 1$). Then the cases with $\beta_{5,j}$ equal to zero were identified and removed from the series because such cases indicate that the relationship between the latest observation anchor and the final forecast error is not significant. After that, outlier checking of the remaining cases was conducted using the boxplot method. As a result, 141 valid cases were collected for $AL_{i,j,2007}$ and $AL^{P}_{i,j,2007}$. The result of paired-sample *t*-test is shown in Table 5-26.

Table 5-26 Result of hypothesis testing (H6c)

Variable	N	Mean	SD	*df*	*t*	Sig.
$AL_{i,j,2007}$	141	0.186	0.170			
$AL^{P}_{i,j,2007}$	141	0.084	0.069			
$AL_{i,j,2007} - AL^{P}_{i,j,2007}$	141	0.102	0.174	140	6.927	<0.001

Among the forecasts with the latest observation anchor detected from the second round of forecasting, there is a significant difference in the APEs of unaided judgmental forecasts (*M*=0.186, *SD*=0.170) and PF-based revisions (*M*=0.084, *SD*=0.069); *t*(140)=6.927, *p*<0.001. The positive value of the mean difference (0.102) between the two APEs reveals that the APE of PF-based revision of anchoring bias is smaller than that of unaided judgmental forecast when the latest observation anchor is detected. Thus, **H6c is supported**: the PF of participants' anchoring bias on the latest observation in the previous forecasting season can significantly reduce such anchoring bias in the following forecasting season. Moreover, the mean difference between $AD_{i,j,2007}$ and $AD^{P}_{i,j,2007}$ was 10.2% and the MAPE of PF-based revision was less than 10%, which indicates that PF dramatically improves forecast accuracy.

5.3.7.4 System-suggested forecasts regarding the statistical forecast anchor (H6d)

Unlike participants' unaided judgmental forecast and PF-based revisions, system-suggested forecasts regarding anchoring bias were not given by participants, but calculated by TDFSS with Equation (3.11). Five series were used to test the performance of system-suggested forecasts: participants' unaided judgmental forecast, PF-based revision, system-suggested forecast, the real outcome of tourist arrivals for all D-O pair markets in 2007, and the system-generated coefficient $\beta_{3,j}$ for all participants. Let $F\_A^{S}_{i,j,2007}$ denote the system-suggested forecast regarding the anchoring bias on statistical forecast, which is calculated as:

$$F\_AS^{S}_{i,j,2007} = \frac{F_{i,j,2007}}{1 - \beta_{3_j}\dfrac{F_{i,j,2007} - SF_{i,2007}}{SF_{i,2007}}}$$

Let $AS^{s}_{i,j,2007}$ denote the APE of $F\_AS^{S}_{i,j,2007}$; the calculation of $AS^{s}_{i,j,2007}$ follows the same method as the calculation of $AS_{i,j,2007}$ and $AS^{P}_{i,j,2007}$. Further analysis focuses on the comparison of these three APEs in order to see whether $AS^{s}_{i,j,2007}$ is significantly smaller than the other two.

The first criterion for selecting valid cases of $AS^{s}_{i,j,2007}$ was whether anchoring bias on statistical forecast was detected in the second round of forecasting ($I_{S_{i,j,2007}} = 1$). Since

$\beta_{3,j}$ indicates the relationship between a participant's anchoring bias (statistical forecast anchor) and final forecast error, $\beta_{3,j} = 0$ indicates that this relationship is not significant. Therefore, cases with $\beta_{3,j}$ equal to zero were not valid. Fourteen outliers were identified and removed from the series and finally $|AS^{s}_{i,j,2007}|$ was developed with 199 valid cases. Table 5-27 shows the result of the normality test on the three APEs, and the Normal Q-Q plots of the three APEs' distribution are shown in Figure 5-27.

Table 5-27 Normality tests of $AS_{i,j,2007}$, $AS^{P}_{i,j,2007}$, and $AS^{s}_{i,j,2007}$

Variable	N	Shapiro-Wilk		
		Statistic	*df*	Sig.
$AS_{i,j,2007}$	199	0.714	199	<0.001
$AS^{P}_{i,j,2007}$	199	0.608	199	<0.001
$AS^{s}_{i,j,2007}$	199	0.728	199	<0.001

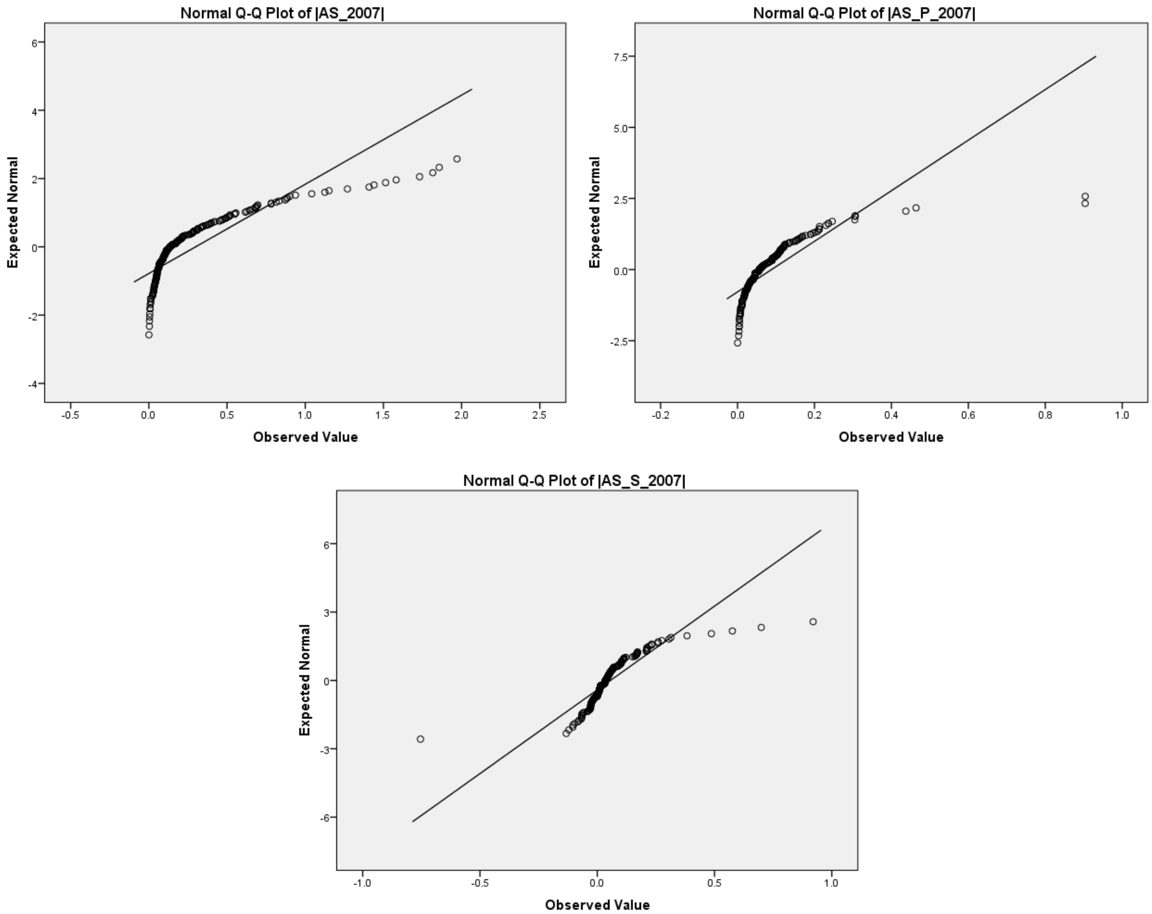


Figure 5-2 Normal Q-Q Plot of $AS_{i,j,2007}$, $AS^{P}_{i,j,2007}$, and $AS^{s}_{i,j,2007}$

According to the results, $AS_{i,j,2007}$, $AS^{P}_{i,j,2007}$, and $AS^{s}_{i,j,2007}$ were significantly non-normal. Thus, the parametric ANOVA test was not appropriate in this part of the analysis. Instead, the Friedman ANOVA test was appropriate for comparing the three APEs. Table 5-28 shows the result of the Friedman ANOVA test on $AS_{i,j,2007}$, $AS^{P}_{i,j,2007}$, and $AS^{s}_{i,j,2007}$.

Table 5-28 Result of hypothesis testing (H6d)

	N	Mean	SD	Mean Rank	Friedman Test	
$AS_{i,j,2007}$	199	0.299	0.383	2.563	Chi-Square	128.681
$AS^{P}_{i,j,2007}$	199	0.088	0.112	1.942	*df*	2
$AS^{s}_{i,j,2007}$	199	0.056	0.136	1.495	Asymp. Sig.	<0.001

Among the three sets of forecast errors, system-suggested forecast had the lowest MAPE (*M*=0.056, *SD*=0.136), followed by the PF-based revision of anchoring bias (*M*=0.088, *SD*=0.112); participants' unaided judgmental forecast had the highest MAPE and the largest standard deviation (*M*=0.299, *SD*=0.383). The result of the Friedman test reveals that the differences between the three APEs were statistically significant ($x^2(2) = 128.618, p < 0.001$). Therefore, **H6d is supported**: system-suggested forecast with statistical forecast anchor correction is significantly more accurate than unaided judgmental forecast and revision based on PF. Therefore, three Wilcoxon tests were conducted as post-hoc tests to investigate the differences between the three APEs. Using a significance level of 0.017, the result shows that the three APEs were significantly different. Specifically, $AS_{i,j,2007}^{s}$ was significantly smaller than $AS_{i,j,2007}^{P}$ (*Z*=-5.697, *p*<0.001) and significantly smaller than $AS_{i,j,2007}$ (*Z*=-9.925, *p*<0.001). According to the MAPE of three sets of forecasts, the forecast accuracy has been improved 21% with the help of PF, and it would be improved a further 3.2% if system-suggested forecast was adopted as the final forecast. Therefore, we can conclude that, regarding anchoring bias on statistical forecast, both PF-based revision and system-suggested forecast significantly outperform unaided judgmental forecast, while system-suggested forecast gives the best performance in this scenario.

5.3.7.5 System-suggested forecasts regarding the desire anchor (H6e)

System-suggested forecasts regarding the desire anchor were generated with Equation (3.11). In order to compare participants' unaided judgmental forecasts, PF-based revisions and system-suggested forecasts when the desire anchor is detected in the second round of forecasting, five series from the dataset are required. Four series of raw data were selected as in the test of H6d; the fifth series used in this part of the analysis was the system-generated coefficient $\beta_{4,j}$ for all participants. Let $F\_AD^{S}_{i,j,2007}$ denote the system-suggested forecast concerning the anchoring bias on desired outcome; $F\_AD^{S}_{i,j,2007}$ was calculated using the equation below:

$$F\_AD^{S}_{i,j,2007} = \frac{F_{i,j,2007}}{1 - \beta_{4_j} \frac{F_{i,j,2007} - DO_{i,2007}}{DO_{i,2007}}}$$

Let $AD^{s}_{i,j,2007}$ denote the APE of $F\_AD^{S}_{i,j,2007}$; $AD^{s}_{i,j,2007}$ is calculated in the same way as $AD_{i,j,2007}$ and $AD^{P}_{i,j,2007}$. Further analysis focused on a comparison of $AD_{i,j,2007}$, $AD^{P}_{i,j,2007}$, and $AD^{s}_{i,j,2007}$ and identifying whether $AD^{s}_{i,j,2007}$ was the smallest, as well as its significance.

Considering the valid cases of these three APEs, the cases with the desire anchor detected in the second round of forecasting were first selected ($I_{D\,i,j,2007} = 1$). Since β_{4_j} reflects the relationship between participants' anchoring bias (desire anchor) and the final forecast error, β_{4_j} equal to zero indicates that this relationship is not significant. Therefore, the cases with $\beta_{4_j} = 0$ were considered invalid and removed from the series. Although no outlier was identified among the remaining cases, there were only 29 valid cases for the three APEs, which is almost the minimum requirement for sample size. Table 5-29 shows the result of the normality test on these three APEs and the normal Q-Q plots of the three series' distribution are shown in Figure 5-3.

Table 5-29 Normality tests of $AD_{i,j,2007}$, $AD^{P}_{i,j,2007}$, and $AD^{s}_{i,j,2007}$

Variable	N	Shapiro-Wilk		
		Statistic	*df*	Sig.
$AD_{i,j,2007}$	29	0.549	29	<0.001
$AD^{P}_{i,j,2007}$	29	0.771	29	<0.001
$AD^{s}_{i,j,2007}$	29	0.704	29	<0.001

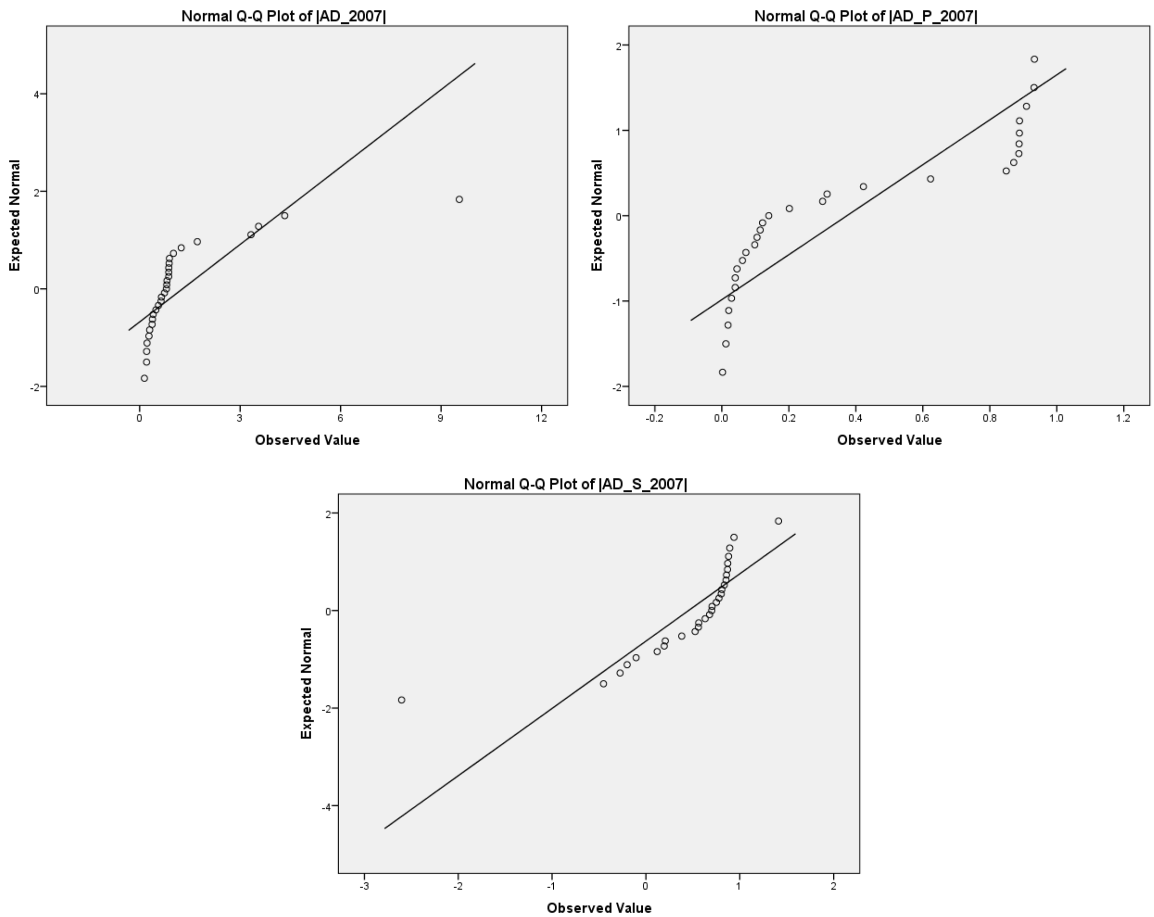


Figure 5-3 Normal Q-Q Plot of $AD_{i,j,2007}$, $AD^{P}_{i,j,2007}$, and $AD^{s}_{i,j,2007}$

According to the results, all three APEs were significantly non-normal. Therefore, the Friedman ANOVA test was appropriate to conduct the comparison of $AD_{i,j,2007}$, $AD^{P}_{i,j,2007}$, and $AD^{s}_{i,j,2007}$.

Table 5-30 Result of hypothesis testing (H6e)

	N	Mean	SD	Mean Rank	Friedman Test	
$AD_{i,j,2007}$	29	1.280	1.890	2.345	Chi-Square	6.276
$AD^{P}_{i,j,2007}$	29	0.373	0.380	1.966	*df*	2
$AD^{s}_{i,j,2007}$	29	0.454	0.725	1.690	Asymp. Sig.	0.043

According to Table 5-30, PF-based revision of anchoring bias had the lowest MAPE, and the lowest standard deviation (*M*=0.373, *SD*=0.380), followed by system-suggested forecasts (*M*=0.454, *SD*=0.725); participants' unaided judgmental forecast had the highest MAPE and the biggest standard deviation (*M*=1.280, *SD*=1.890). The result of the Friedman test shows that the differences between $AD_{i,j,2007}$, $AD^{P}_{i,j,2007}$, and $AD^{s}_{i,j,2007}$ were statistically significant ($x^2(2) = 6.276, p = 0.043$). Therefore, **H6e is rejected**. Although the accuracy of three types of forecast were significantly different, system-suggested forecast with the desire anchor correction was not the most accurate; instead, PF-based revision outperforms participant's unaided judgmental forecast and system-suggested forecast in this situation. Moreover, Wilcoxon tests were conducted to further investigate the differences between three types of forecast. With a significance level of 0.017, it appears that both $AD^{P}_{i,j,2007}$ (*Z*=-2.908, *p*=0.004) and $AD^{s}_{i,j,2007}$ (*Z*=-2.670, *p*=0.008) were significantly smaller than $AD_{i,j,2007}$. The mean value of $AD^{P}_{i,j,2007}$ was slightly smaller than $AD^{S}_{i,j,2007}$; however, the difference between them was not statistically significant (*Z*=-0.897, *p*=0.370). Therefore, we can conclude that, regarding anchoring bias on desired outcome, both PF-based revision and system-suggested forecast significantly outperform unaided judgmental forecast. It seems that PF-based revision performs the best, but the difference between the latter and system-suggested forecast was not statistically significant.

5.3.7.6 System-suggested forecasts regarding the latest observation anchor (H6f)

System-suggested forecasts regarding the latest observation anchor can be generated following the same method as the suggested forecasts covering other two types of anchoring bias. To compare participants' unaided judgmental forecasts, PF-based revisions, and system-suggested forecasting when latest outcome anchor was detected in the second round of forecasting, five series from the dataset were needed: four of them were the raw data, the same data used in the tests for H6d and H6e; the fifth series was the system-generated coefficient $\beta_{5,j}$ for all participants. Let $F\_AL^S_{i,j,2007}$ denote the system-suggested forecast regarding the anchoring bias on the latest observation; $F\_AL^S_{i,j,2007}$ is calculated as below:

$$F\_AL^S_{i,j,2007} = \frac{F_{i,j,2007}}{1 - \beta_{5_j}\dfrac{F_{i,j,2007} - LO_{i,2007}}{LO_{i,2007}}}$$

Let $AL^s_{i,j,2007}$ denotes the APE of $F\_AL^S_{i,j,2007}$; $AL^s_{i,j,2007}$ was calculated in the same way as $AL_{i,j,2007}$ and $AL^P_{i,j,2007}$. Further analysis focused on comparison of these APEs and identifying whether $AL^s_{i,j,2007}$ could outperform the other two.

Considering the valid cases of $AL_{i,j,2007}$, $AL^P_{i,j,2007}$, and $AL^s_{i,j,2007}$, the cases with the latest observation anchor detected in the second round of forecasting were selected ($I_{L_{i,j,2007}} = 1$). Since the value of β_{5_j} reflects the relationship between a participant's anchoring bias (latest observation anchor) and the final forecast error, $\beta_{5_j} = 0$ indicates that this relationship is not significant. Therefore, the cases with $\beta_{5_j} = 0$ were considered invalid. Twenty-three outliers were identified as invalid cases and removed from the series; ultimately, 141 cases were collected for $AL_{i,j,2007}$, $AL^{P}_{i,j,2007}$, and $AL^{s}_{i,j,2007}$. Table 5-31 shows the normality test results on these three APEs; their Normal Q-Q plots are shown in Figure 5-4.

Table 5-31 Normality tests of $AL_{i,j,2007}$, $AL^{P}_{i,j,2007}$,and $AL^{s}_{i,j,2007}$

Variable	N	Shapiro-Wilk		
		Statistic	*df*	Sig.
$AL_{i,j,2007}$	141	0.887	141	<0.001
$AL^{P}_{i,j,2007}$	141	0.893	141	<0.001
$AL^{s}_{i,j,2007}$	141	0.316	141	<0.001

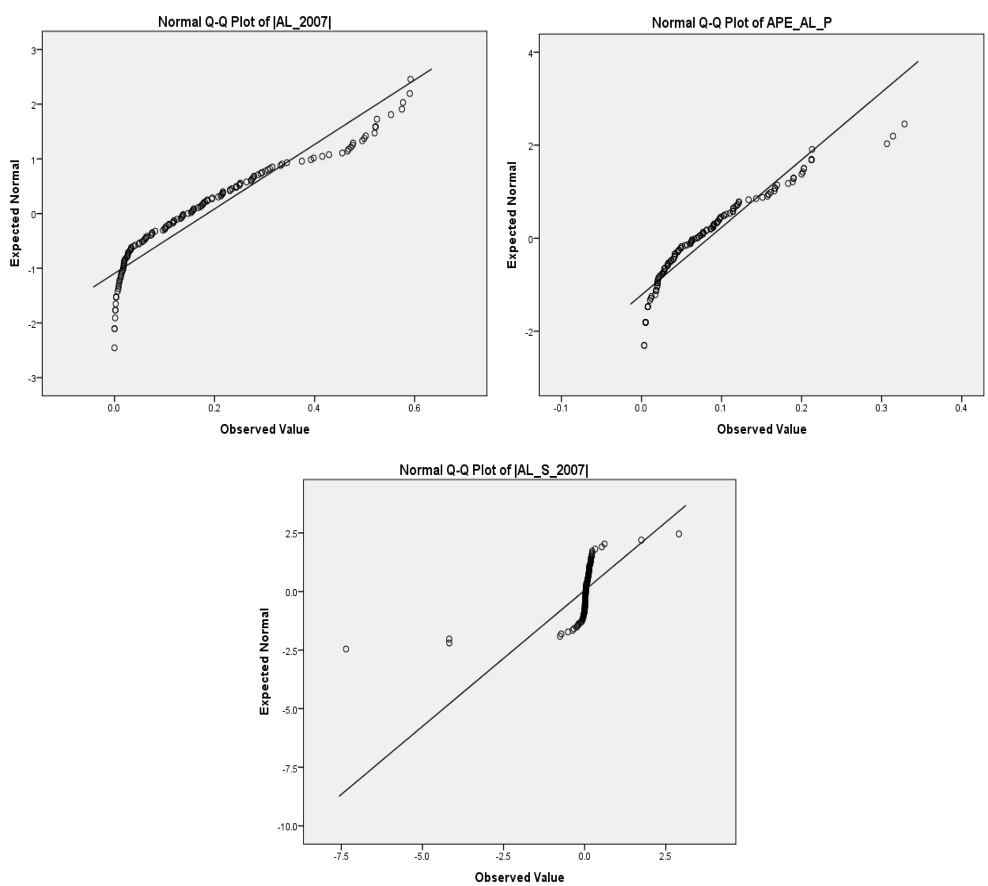


Figure 5-4 Normal Q-Q Plot of $AL_{i,j,2007}$, $AL^{P}_{i,j,2007}$, and $AL^{s}_{i,j,2007}$

According to the results, the three APEs were significantly non-normal. Thus, the Friedman ANOVA test was appropriate for the comparison of the three APEs. Table 5-32 shows the result of the Friedman ANOVA test.

Table 5-32 Result of hypothesis testing (H6f)

	N	Mean	SD	Mean Rank	Friedman Test	
$AL_{i,j,2007}$	141	0.186	0.170	2.333	Chi-Square	32.128
$AL_{i,j,2007}^{P}$	141	0.084	0.069	2.007	*df*	2
$AL_{i,j,2007}^{S}$	141	0.042	0.861	1.660	Asymp. Sig.	<0.001

Among the three sets of forecasts, system-suggested forecast had the lowest MAPE (*M*=0.042, *SD*=0.861), followed by the PF-based revision of anchoring bias (*M*=0.084, *SD*=0.069); participants' unaided judgmental forecast had the highest MAPE (*M*=0.186, *SD*=0.170). The result of the Friedman test shows that the differences between the three APEs were statistically significant ($x^2(2) = 32.128, p < 0.001$). Therefore, **H6f is supported**: system-suggested forecast with the latest observation anchor correction is significantly the most accurate forecast method. Based on the Friedman ANOVA test result, three Wilcoxon tests were conducted as post-hoc tests to further identify the differences between the three APEs. With a significance level of 0.017, the three APEs were significantly different. In particular, $AL_{i,j,2007}^{S}$ was significantly smaller than both $AL_{i,j,2007}^{P}$ (*Z*=-3.901, *p*<0.001) and $AL_{i,j,2007}$ (*Z*=-6.017, *p*<0.001). According to the MAPE of the three sets of forecasts, PF helped participants improve their forecast accuracy by 10.2%, and system-suggested forecast performed 14.4% more accurately than participants' unaided judgmental forecast. Therefore, it can be concluded that, regarding anchoring bias on the latest observation, both PF-based revision and system-suggested forecast are significantly more accurate than participants' unaided judgmental forecast, and system-suggested forecast also significantly outperforms PE-based revision.

5.3.8 Debiasing of overconfidence bias

5.3.8.1 Performance feedback of overconfidence bias (H7a)

To test whether the PF of participants' overconfidence bias in the first round of forecasting can effectively improve the accuracy of their judgmental forecast in the second round of forecasting, the APEs of corrected judgmental forecast (CF) and the PF-based revision regarding overconfidence bias in the second round of forecasting had to be compared. Here, CF is the one forecast method with desire bias and anchoring bias removed, but still contains overconfidence bias. Therefore, CF is similar to the PF-based revision with anchoring bias eliminated. Three series from the dataset had to be used for this part of the analysis: (i) the real outcome of tourist arrivals, (ii) participants' CF, and (iii) participants' revised forecast based on the PF of their overconfidence bias for all D-O pair markets in 2007. Let $OV_{i,j,2007}$ and $OV^{P}_{i,j,2007}$ denote the APEs of CF and the PF-based revision when overconfidence bias is detected; the calculation of these two variables are:

$$\begin{cases} OV_{i,j,2007} = \left| \dfrac{Q_{i,2007} - CF_{i,j,2007}}{Q_{i,2007}} \right| \\ OV^{P}_{i,j,2007} = \left| \dfrac{Q_{i,2007} - F{O}^{p}_{i,j,2007}}{Q_{i,2007}} \right| \end{cases}$$

Since all participants were required to rate their confidence in the CF of each D-O pair markets, the whole sample was used as the basic sample for this part of the analysis.

Thirteen outliers were identified and removed from the series. Thus, the two APEs were developed with 667 valid cases. A paired-sample *t*-test was conducted on these two variables and the result is shown in Table 5-33.

Table 5-33 Result of hypothesis testing (H7a)

Variable	N	Mean	SD	*df*	*t*	Sig.
$OV_{i,j,2007}$	667	0.112	0.172			
$OV_{i,j,2007}^{P}$	667	0.135	0.198			
$OV_{i,j,2007} - OV_{i,j,2007}^{P}$	667	-0.023	0.113	666	-5.244	<0.001

According to the result, the difference between $OV_{i,j,2007}$ and $OV_{i,j,2007}^{P}$ was statistically significant (M=-0.023, SD=0.113; $t(666)$=-5.244, p<0.001), and the APE of participants' CF (M=0.112, SD=0.172) was slightly smaller than the PF-based revisions (M=0.135, SD=0.198). Therefore, **H7a is rejected**: the PF of participants' overconfidence bias in the previous forecasting season does not help participants to reduce their overconfidence bias in the following forecasting season.

5.3.8.2 System-suggested forecasts regarding overconfidence bias (H7b)

System-suggested forecasts regarding overconfidence bias were produced with Equation (3.17). Five series from the dataset were required to test whether system-suggested forecast performs best in terms of forecast accuracy: participants' CF, PF-based revision, system-suggested forecast, the real outcome of tourist arrivals for all D-O pair markets in 2007, and the system-generated coefficient β_{6_j} for all participants. Let $F\_O_{i,j,2007}^{S}$ denote the system-suggested forecast regarding the overconfidence bias; it is calculated using the following equation:

$$F\_O^S_{i,j,2007} = \frac{CF_{i,j,2007}}{1 - \beta_6(Conf_{i,j,2007} - \frac{1}{m}\sum_{i=1}^{m} I_{i,j,2006})}$$

Let $OV^s_{i,j,2007}$ denote the APE of $F\_O^S_{i,j,2007}$; $OV^s_{i,j,2007}$ is calculated the same way as $OV_{i,j,2007}$ and $OV^P_{i,j,2007}$. Further analysis focused on comparison of the three APEs in order to see whether $OV^s_{i,j,2007}$ was significantly smaller than the other two, and identify its significance.

Considering the valid cases of these three APEs, cases with $\beta_{6_j} = 0$ were eliminated as invalid because an insignificant relationship between overconfidence bias and the final forecast error was identified for these cases and thus system-suggested forecasts cannot be properly calculated. According to this criterion, and following outlier checking, 421 valid cases were collected for $OV_{i,j,2007}$, $OV^P_{i,j,2007}$, and $OV^s_{i,j,2007}$. Table 5-34 shows the results of the normality test on these three APEs, and their normal Q-Q plots are shown in Figure 5-5.

Table 5-34 Normality tests of $OV_{i,j,2007}$, $OV^P_{i,j,2007}$, and $OV^s_{i,j,2007}$

Variable	N	Shapiro-Wilk		
		Statistic	*df*	Sig.
$OV_{i,j,2007}$	421	0.403	421	<0.001
$OV^P_{i,j,2007}$	421	0.504	421	<0.001
$OV^s_{i,j,2007}$	421	0.504	421	<0.001

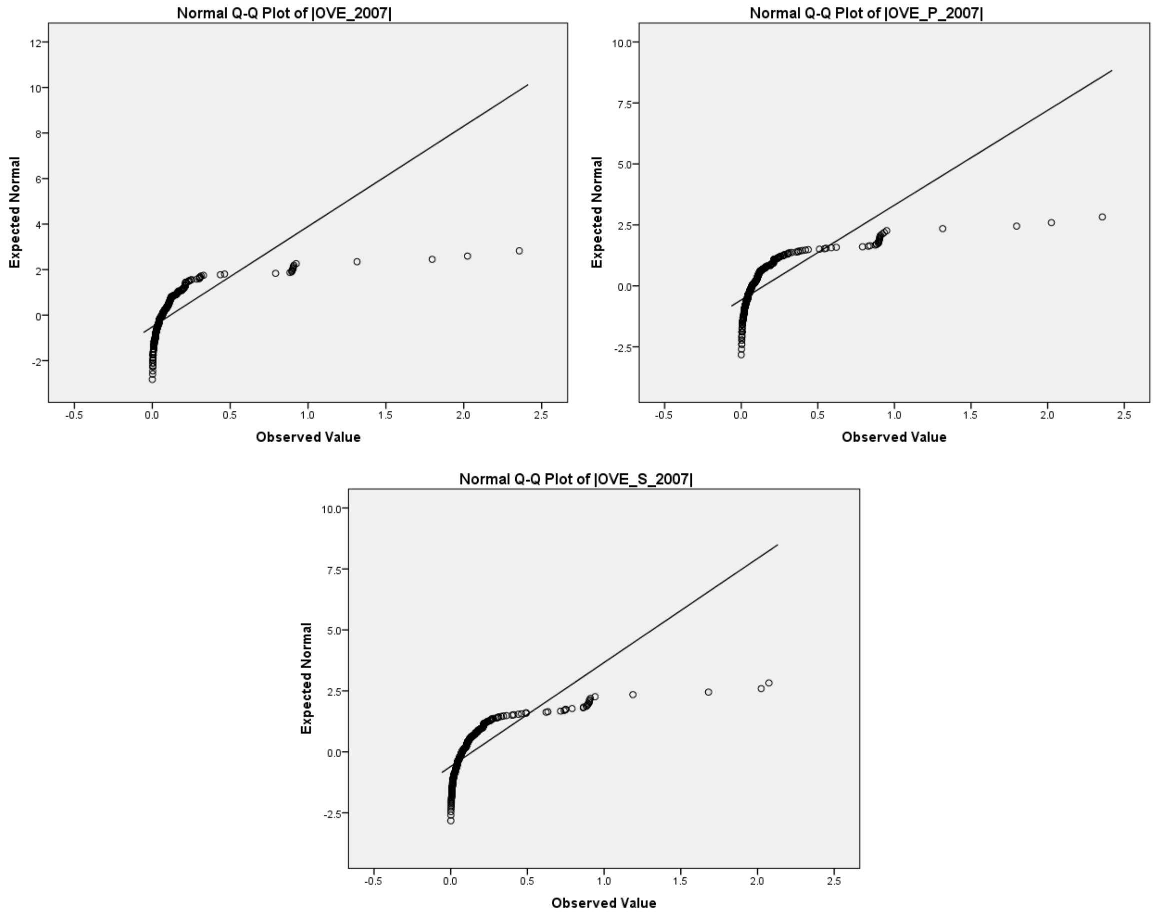


Figure 5-5 Normal Q-Q Plot of $OV_{i,j,2007}$, $OV^{P}_{i,j,2007}$, and $OV^{s}_{i,j,2007}$

All three APEs are significantly non-normal. Therefore, the Friedman ANOVA test was appropriate for the comparison of $OV_{i,j,2007}$, $OV^{P}_{i,j,2007}$, and $OV^{s}_{i,j,2007}$. Table 5-35 shows the result of the Friedman ANOVA test.

Table 5-35 Result of hypothesis testing (H7b)

	N	Mean	SD	Mean Rank	Friedman Test	
$OV_{i,j,2007}$	421	0.116	0.227	1.885	Chi-Square	10.802
$OV^{P}_{i,j,2007}$	421	0.149	0.257	2.070	*df*	2
$OV^{s}_{i,j,2007}$	421	0.141	0.235	2.045	Asymp. Sig.	0.005

Among the three sets of forecasts, participants' CF had the lowest MAPE and the lowest standard deviation (*M*=0.116, *SD*=0.227), followed by system-suggested forecasts (*M*=0.141, *SD*=0.235); PF-based revision had the highest MAPE and the biggest standard deviation (*M*=0.149, *SD*=0.257). The result of the Friedman test shows that the differences between $OV_{i,j,2007}$, $OV^{P}_{i,j,2007}$, and $OV^{s}_{i,j,2007}$ were statistically significant ($x^2(2) = 10.802, p = 0.005$). Therefore, **H7b is rejected**. Although the accuracy of the three types of forecast was significantly different, system-suggested forecast with overconfidence bias correction cannot outperform participants' forecast. Wilcoxon tests were also conducted to further identify the differences between them. With a significance level of 0.017, $OV_{i,j,2007}$ was significantly smaller than $OV^{P}_{i,j,2007}$ (*Z*=-5.635, *p*<0.001) and $OV^{s}_{i,j,2007}$ (*Z*=-3.657, *p*<0.001), while the difference between $OV^{P}_{i,j,2007}$ and $OV^{s}_{i,j,2007}$ was not statistically significant (*Z*=-0.159, *p*=0.874). Therefore, it is concluded that, regarding overconfidence bias, there is no significant difference in forecast accuracy between PF-based revision and system suggest forecast; neither can outperform forecasters' own judgment.

5.4 Summary

This chapter has presented the results of the main experiment, including data screening, profile of participants, and hypotheses testing. It started with checks for missing data and outliers in the raw data. As a result of these steps, 680 valid cases were retained. Then the profile of the participants was discussed. In the hypotheses testing, both parametric and non-parametric statistics were adopted for statistical analysis; data normality tests were also conducted to determine whether non-parametric approaches should be conducted in the hypothesis testing.

According to the results of the hypotheses tests, 14 of 21 hypotheses are supported; the hypothesis regarding the existence of a statistical anchor is partially supported when participants' judgmental forecast falls in the wrong direction of adjustment of the baseline forecast (statistical forecast); six hypotheses were rejected. Two rejected hypotheses related to desired outcome (desire bias and desire anchor), showing that PF is more effective than system-suggested forecast in participants' judgmental forecasting; another two rejected hypotheses related to overconfidence bias, revealing that neither PF nor system-suggested forecast is effective in helping participants reduce their forecast error.

6 FINDINGS AND DISCUSSION

According to the results of the data analysis, 14 of 21 hypotheses are supported by the data; six are rejected; and the hypothesis regarding whether there is anchoring bias on statistical forecast is partially supported when participants' adjustments were in the wrong direction. A summary of the hypotheses test results is shown in Table 6-1.

Table 6-1 Summary of hypotheses test results

Hypotheses	N	Indicator	STAT	Result
H1. Statistical forecast bias has no influence on the final forecast error after judgmental forecasting.	670	$M(STAT_{2006}) = 9148$	t=2.477*	Reject
H2a. Desire bias exists in forecasters' judgmental forecasting.	327	$M(PE\_DO_{2006}) = 0.566$	t=21.268**	Support
H2b. Desire bias contributes to the final forecast error.	397	$M\ (DES_{2006}) = -467473$	t=-2.316*	Support
H3a. Anchoring bias in statistical forecast is unavoidable in forecasters' judgmental forecasting.	461	$M(A_{S_{2006}}) = 0.043$	t=2.090*	Partially Support
	293	$M(A_{S_{2007}}) = 0.081$	t=1.892*	
	754	$M(A_{S_{all}}) = 0.058$	t=2.771**	
Judgmental adjustment in the right direction	136	$M(A^{R}_{S_{2006}}) = 0.074$	t=7.855**	(Reject)
	97	$M(A^{R}_{S_{2007}}) = 0.411$	t=5.194**	
	233	$M(A^{R}_{S_{all}}) = 0.214$	t=6.124**	
Judgmental adjustment in the wrong direction	325	$M(A^{W}_{S_{2006}}) = 0.030$	t=1.040	(Support)
	196	$M(A^{W}_{S_{2007}}) = -0.082$	t=-1.743	
	521	$M(A^{W}_{S_{all}}) = -0.012$	t=-0.468	
H3b. Anchoring bias in desired outcome is unavoidable in forecasters' judgmental forecasting.	57	$M(A_{D_{2006}}) = 0.571$	t=1.285	Support
	87	$M(A_{D_{2007}}) = -1.493$	t=-1.758	
	144	$M(A_{D_{all}}) = -0.676$	t=-1.235	

Hypotheses	N	Indicator	STAT	Result
H3c. Anchoring bias in the latest observation is unavoidable in forecasters' judgmental forecasting.	154	$M(A_{L_{2006}})$ = 0.115	t=1.946	Support
	300	$M(A_{L_{2007}})$ = -0.357	t=-1.381	
	454	$M(A_{L_{all}})$ = -0.197	t=-1.142	
H3d. Anchoring bias in statistical forecast contributes to the final forecast error.	287	$M(A\_STAT_{2006})$ =-88215	t=-2.202*	Support
H3e. Anchoring bias in desired outcome contributes to the final forecast error.	33	$M(A\_DES_{2006})$ = 389350	t=3.020**	Support
H3f. Anchoring bias in the latest observation contributes to the final forecast error.	153	$M(A\_LAS_{2006})$ = -24581	t=-0.831	Reject
H4a. Overconfidence is unavoidable in forecasters' judgmental forecasting.	680	$M(O_{2006})$ = 0.244	t=23.478**	Support
	680	$M(O_{2007})$ = -0.074	t=5.034**	
	1360	$M(O_{all})$ = 0.159	t=17.168**	
H4b. Overconfidence bias contributes to the final forecast error.	560	$M(OVE_{2006})$ = 70998	t=5.036**	Support
H5a. Feedback of the mean desire error in a forecaster's previous forecasting season reduces desire bias in the following forecasting season.	274	MD = 0.018	t=2.005*	Support
H5b. Suggested forecast with desire error correction is the most accurate adjustment, better than a forecaster's unaided adjustment and the adjustment based on the corresponding PF.	182	$M(\lvert D_{i,j,2007}\rvert)$ = 0.329	x^2=5.150	Reject
		$M(\lvert \mathrm{D}^{\mathrm{P}}_{\mathrm{i,j,2007}}\rvert)$ = 0.302		
		$M(\lvert \mathrm{D}^{\mathrm{s}}_{\mathrm{i,j,2007}}\rvert)$ = 0.391		

Hypotheses	N	Indicator	STAT	Result
H6a. Feedback of the mean anchoring error in the statistical forecast in the previous forecasting season reduces anchoring bias in the following forecasting season.	201	MD = 0.213	t=8.049**	Support
H6b. Feedback of the mean anchoring error in the desired outcome in the previous forecasting season reduces anchoring bias in the following forecasting season.	33	MD = 0.817	t=2.490*	Support
H6c. Feedback of the mean anchoring error in the latest observation in the previous forecasting season reduces anchoring bias in the following forecasting season.	141	MD = 0.102	t=6.927**	Support
H6d. System-suggested forecast with statistical forecast anchor correction is the most accurate forecast, better than unaided forecast and adjustment based on the corresponding PF.	199	$M(\lvert AS_{i,j,2007}\rvert)$=0.299 $M(\lvert AS^{P}_{i,j,2007}\rvert)$=0.088 $M(\lvert AS^{s}_{i,j,2007}\rvert)$=0.056	x^2=128.681**	Support
H6e. System-suggested forecast with desire anchor correction is the most accurate forecast, better than unaided forecast and adjustment based on the corresponding PF.	29	$M(\lvert \mathrm{AD}_{\mathrm{i,j},2007}\rvert)$=1.280 $M(\lvert \mathrm{AD}^{\mathrm{P}}_{\mathrm{i,j},2007}\rvert)$=0.373 $M(\lvert \mathrm{AD}^{\mathrm{s}}_{\mathrm{i,j},2007}\rvert)$=0.454	x^2=6.276	Reject
H6f. System-suggested forecast with latest observation anchor correction is the most accurate forecast, better	141	$M(\lvert \mathrm{AL}_{\mathrm{i,j},2007}\rvert)$=0.186 $M(\lvert \mathrm{AL}^{\mathrm{P}}_{\mathrm{i,j},2007}\rvert)$=0.084 $M(\lvert \mathrm{AL}^{\mathrm{s}}_{\mathrm{i,j},2007}\rvert)$=0.042	x^2=32.128**	Support

Hypotheses	**N**	**Indicator**	**STAT**	**Result**
than unaided forecast and adjustment based on the corresponding PF.				
H7a. Feedback of the mean overconfidence error in the previous forecasting season reduces overconfidence bias in the following forecasting season.	667	*MD* = -0.023	*t*=-5.244**	Reject
H7b. System-suggested forecast with overconfidence error corrected is the most accurate forecast, better than comparing the statistical forecast, unaided forecast, and the adjustment based on the PF of the mean overconfidence error in the previous forecasting season.	421	$M(\lvert OV_{i,j,2007}\rvert)$=0.116 $M(\lvert OV^{P}_{i,j,2007}\rvert)$=0.149 $M(\lvert OV^{s}_{i,j,2007}\rvert)$=0.141	x^2=10.802**	Reject

*M: mean value; MD: mean difference; *: 5% significance level; **: 1% significance level;*

Unfortunately, the debiasing strategies related to overconfidence bias are all rejected, which indicates that the two debiasing strategies (PF and system-suggested forecast) proposed in this study are not effective in reducing overconfidence bias in judgmental forecasting. Therefore, these two strategies should not be applied in the design of TDFSS. Further discussion of the hypotheses test results are presented in the following sections.

6.1 Modeling cognitive bias with statistical forecast error

As mentioned in the reviewed literature of time series forecasting, the tourism demand of a specific D-O pair market can be decomposed into two components named regular and irregular patterns (Fildes et al., 2006). Regular patterns (e.g., trend, seasonality, and relationships with the main influencing factors) can be well captured by statistical modeling techniques; therefore, this component of tourism demand can be well predicted by statistical forecasting methods. However, tourism demand has features that are highly sensitive to special events and high uncertainty, which means that a large proportion of tourism demand presents irregular patterns (Zhang, Song, & Huang, 2009). Therefore, the application of statistical methods to tourism demand forecasting, which focuses on the accurate prediction of regular patterns, is far from sufficient to generate accurate forecasts. Effective prediction of irregular patterns in tourism demand is as important as effective prediction of regular patterns (Song & Li, 2008; Witt & Witt, 1995).

Professional forecasters in this area can effectively estimate the influence of special events on tourism demand and accurately predict the irregular patterns based on their domain knowledge, expertise, and working experience. It is reasonable that judgmental forecasting performs better than statistical techniques in estimating the influence of irregular interruptions on tourism demand. In fact, many studies have revealed that a combination of statistical forecast and judgmental forecast performs extremely well and better than the application of any individual method in tourism demand forecasting (Oh & Morzuch, 2005; Song & Li, 2008). Statistical forecast error in tourism demand forecasting is considered to be the bias caused by the irregular patterns. Judgmental adjustment of statistical forecast is expected to be more accurate with the elimination statistical forecast bias; hence, the results of judgmental adjustments should ideally be free of such bias. In other words, in an ideal situation, the forecast error of judgmental forecast based on statistical forecast should have no significant relationship with statistical forecast error.

Since it has been the focus of many studies in tourism demand forecasting (e.g., Tideswell et al., 2001; Wong, Song, Witt, & Wu, 2007), the existence of statistical forecast error and whether it biases the forecast accuracy is not repeatedly examined in this study. Regarding the expected power of judgmental forecasting, this study focused on testing whether judgmental adjustment of statistical forecast plays an ideal role as discussed. Investigation of statistical forecast error reveals that the mean percentage error of statistical forecast ($\frac{Q_{i,2006}-SF_{i,2006}}{Q_{i,2006}}$) is significantly positive (*M*=0.050, *SD*=0.086, *t*(679)=14.918, *p*<0.001), which indicates that tourism demand in the D-O markets is generally underestimated by the ES method. Specifically, 5% of forecast error is caused by such conservatism biasing of statistical forecasts. This result is consistent with some studies of tourism demand forecasting computation. For example, Coshall's (2009) study about international tourists to the United Kingdom revealed that ES forecasts are generally lower than the real outcomes in all three forecast horizons. Another forecast computation conducted by Athanasopoulos, Hyndman, Song, and Wu (2011) provided a more general conclusion on this phenomenon. They found that, based on over 1,300 series of tourism demand, the ES and Forecast Pro methods generally underestimated both point forecasts and interval forecasts of tourist arrivals. As a result, conservatism is widely observed in the statistical forecasting of tourism demand.

The result of the test of H1 provides further evidence supporting this conclusion: the decomposed error of forecasters' unaided forecasts associated with statistical forecast error is significantly different from zero (*M*=9148, *SD*=96297, *t*(679)=2.477, *p*=0.013), which reveals that the existence of statistical forecast bias is significant even when judgmental adjustment is conducted. Therefore, modeling the relationships between forecasters' unaided judgmental forecast and cognitive bias should incorporate statistical forecast error in order to accurately estimate the influence of cognitive bias on forecast error. A review of the coefficients estimated by OLS for all participants after the first round of forecasting provides evidence of this conclusion: 67 out of 68 β_{1_j} were statistically significant. A positive mean value of β_{1_j} (*M*=0.538, *SD*=1.098, *t*(68)=12.688, *p*<0.001) indicates that the relationship between statistical forecast error and the error of forecasters' unaided judgmental forecast is generally positive: the more seriously tourism demand is underestimated by statistical forecasting techniques, the greater the forecast errors of forecasters' judgmental forecast obtained.

6.2 Cognitive bias in judgmental forecasting of tourism demand

According to the hypotheses test results, all the hypotheses about the existence of cognitive bias in judgmental forecasting are supported (H2a, H3b, H3c, and H4a) or partially supported (H3a). These results show that the desire bias, overconfidence bias, and three types of anchoring bias studied are commonly found in the judgmental forecasting of tourism demand. Furthermore, the results of the tests of hypotheses regarding the contribution of cognitive bias to judgmental forecast error reveal that cognitive bias caused by desired outcome, statistical forecast anchor, desire anchor, and overconfidence contribute significantly to the error of judgmental forecast (H2b, H3d, H3e, and H4b), while the contribution of anchoring bias on the latest observation anchor to judgmental forecast error is relatively small and insignificant (H3f).

6.2.1 Desire bias

In Arnott's (2006) study, desire bias is considered one of the main cognitive biases; it is also the foundation of anchoring bias on desired outcome. In this study, participants provided their desired outcome for tourist arrivals for 58.97% of all D-O pair markets in the first round of forecasting, based on which significant desire bias was observed (M=0.566, SD=0.481, $t(326)$=21.268, p<0.001). This conclusion further confirms that desire bias is one of the main cognitive biases and is commonly found in the judgmental forecasting of tourism demand. Furthermore, a positive MPE of participants' desired outcome indicates that forecasters' desired outcomes are generally conservative. This conclusion differs from previous studies in other research areas. For example, Mathews and Diamantopoulous (1990) provided evidence that managers of a health products company usually provided judgmental forecasts with an optimism bias. Fildes et al. (2009) conducted a broader investigation of the features of judgmental forecasting based on forecasters from four supply-chain companies in different industries and over 60,000 demand forecasts. One of their findings was a general bias towards optimism. From the perspective of cognitive bias, these studies provide evidence that desire bias always occurs in tandem with optimism bias. Two studies from the perspective of predictive psychology have provided some clues to explain this phenomenon. First, Armor and Taylor’s (2002) explanation is that people tend to “infer the likelihood of different outcomes on the basis of case-specific plans or scenarios about how the future will unfold” and “the very processes of constructing and considering these scenarios tend to render people prone to bias” (Armor & Taylor, 2002, p. 342). However, this does not explain why desire bias comes with optimism bias, rather than with conservatism. Massey, Simmons, and Armor (2011) further explained that people’s expectations are often over-optimistic because “they frequently believe that their preferred outcomes are more likely than is merited” (Massey et al., 2011, p. 274). Massey and colleagues’ study on professional football predictions also provided evidence that optimism is driven by desire. On the other hand, a few studies argue that there are links between desire bias and conservatism (Eroglu, 2006; Van Den Boom, 2009). Incentive concavity theory may give a reasonable explanation of the joint cognitive bias of desire and conservatism: the rewards of making bold but probably accurate forecasts are smaller than the penalties of making bold but probably inaccurate forecasts (Lamont, 2002). In forecasting decision-making, forecasters have to balance the probability of a forecast being more accurate and its rewards/penalties. When the forecasting circumstances are uncertain and complex, conservative forecast is the optimal choice following the risk-averse principle. In this case, positive influence on forecast would be decreased and negative influence would be increased, which leads to a lower level of desired outcome. Therefore, the uncertainty and complexity of a forecasting task are the key factors leading to the joint cognitive bias of desire and conservatism. As mentioned earlier, high uncertainty is a typical feature of tourism demand, which encourages forecasters to follow the incentive concavity theory in their cognitive behavior.

In addition, the contribution of desire bias to final forecast error is measured by the decomposed forecast error associated with desire bias. The testing of H2b indicates that desire bias causes forecast error of over 460,000 units on average. However, the mean forecast error of participants' unaided judgmental forecast was only 44,275. It is surprising that the forecast errors associated with desire bias are much larger than the final forecast errors, and that the errors are in different directions (-467,343 vs 44,275). The only explanation for this is that some participants had very poor desired outcomes, which were revised on an even larger scale after additional information (statistical forecasts and historical data of tourist arrivals) was available. Further evidence of this conclusion is collected from the estimated coefficient ${\beta_2}_j$. It is shown that 22 of 58 participants' ${\beta_2}_j$ were negative (${\beta_2}_j = 0$ for 10 participants and they are not counted in this part of the discussion), which indicates that when large desire errors were made by these participants, they tended to make even larger revisions in the opposite direction. According to the literature, such revision of judgmental forecast may lead to two different results. Some studies have revealed that large revisions cause overreaction most of the time (Fildes et al., 2006; Goodwin & Fildes, 2001; Lawrence et al., 2002). When forecasters recognized the shortcomings of their forecasting ability, they attempted to make up for them by making large judgmental adjustments; certainly, these adjustments were poor substitutes. On the other hand, some studies have revealed that large revisions usually depend on reliable information, which leads to necessary adjustments; in contrast, small revisions are usually based on unreliable information and the adjustments based on unreliable information are usually unnecessary and damage forecast accuracy (Fildes et al., 2009). Therefore, they argued that large-scale revision of judgmental forecasts always produces better performance than small amendments. A comparison of the final forecast error and the error associated with desire bias reveals that final forecast error is significantly smaller (M=-423198, SD=4083739, $t(396)$=-2.497, p=0.013), which supports the second argument.

6.2.2 Anchoring bias

The anchoring effect is one of the most robust cognitive heuristics in the judgmental forecasting literature (Clements & Hendry, 2008; Furnham & Boo, 2011). Anchoring bias in this study has been further categorized according to different anchors, including the statistical forecast anchor, the desire anchor, and the latest observation anchor. According to the judgmental forecasts collected in the two-round experiment, anchoring bias was detected in 1,352 of 1,360 forecasts. Furthermore, 55.77% of these were anchored on the statistical forecasts; 33.58% were anchored on the latest observations; forecasts anchored on desired outcome were the least frequently identified, representing 10.65% of cases.

6.2.2.1 Statistical forecast anchor

With a powerful prediction of trend, seasonality, the long-term relationship with its influence factors, and other regular patterns of timer series, statistical forecast is considered one of the main references in judgmental forecasting; its importance has been widely examined (Eroglu, 2006; Goodwin, 2005; Goodwin & Fildes, 1999; Sanders & Ritzman, 2001; Song, Gao, & Lin, 2012). However, the shortcomings of statistical forecast in predicting the irregular patterns of timer series are usually neglected by forecasters.

The test of H3a shows that anchoring bias on statistical forecast is significantly right skewed. A follow-up examination reveals that this phenomenon is significant when forecasters' judgmental forecast is in the right direction; judgmental forecasts in the wrong direction were given with a center equal to the statistical forecast. As a result, anchoring bias on statistical forecast was significant when forecasters made adjustments in the wrong direction; such anchoring bias is insignificant when forecasters' adjustments are in the right direction. The distribution of percentage change when adjustment is in the right direction shows that optimism bias occurred in this situation. A joint effect of optimism bias and anchoring bias on forecast error is why participants' adjustments were not distributed with a mean value equal to the statistical forecast anchor when forecasts were adjusted in the right direction. When using mean PC to indicate the general magnitude of PC on the statistical forecast, the mean PC of right-direction adjustments (0.214) is relatively larger than the mean PC of wrong-direction adjustments (-0.012).

Furthermore, some studies in the judgmental forecasting literature have provided evidence that judgmental forecasts perform differently depending on the direction of adjustment. For example, Trapero, Fildes, and Davydenko (2011) and Fildes et al. (2009) concluded that forecast accuracy is significantly different when forecasters make right/wrong direction of adjustment. Both studies also concluded that adjustments in the right direction perform more accurately than adjustments in the wrong direction since optimism bias was frequently observed within the wrong-side adjustments. In this study, judgmental forecasts with anchoring bias on statistical forecast are further grouped according to the direction of adjustment. A comparison of the APE of these two groups of forecasts reveals that adjustment in the right direction is significantly more accurate than adjustment in the wrong direction (*M*=0.068, M^-=0.105, *SD*=0.08, *p*<0.001). Generally, participants' right-direction adjustments with the statistical forecast anchor are 3.71% more accurate than wrong-direction adjustments with the same anchor. Also, it has been found that adjustment in the right direction is relatively larger than adjustment in the wrong direction, which also indicates that larger adjustments are relatively more accurate than smaller adjustments. Therefore, this finding further supports the conclusions in Trapero et al. (2011) and Fildes et al. (2009) that large-scale adjustments are more accurate than small adjustments, and that wrong-side adjustments should be avoided.

In addition, the test of H3d shows a significant negative forecast error associated with the statistical forecast anchor (*M*=-88215, *SD*=678621, *t*(286)=-2.202, *p*=0.028). In other words, participants' judgmental forecasts are always too high when the statistical forecast anchor is detected.

6.2.2.2 Desire anchor

Desire bias in judgmental forecasting has been widely studied (Arnott, 2006; Fildes et al., 2009; Goldfarb et al., 2005; Mathews & Diamantopoulos, 1989); however, the anchoring effect on desired outcome is not widely explored, though there are a few exceptions. Blackley and DeBoer's (1993) study of the revenue and budget of the United States government revealed that the forecasters of outlays and the deficit budget in the Republican administration's Office of Management and Budget (OMB) continuously proposed biased proposals of total outlays and the budget deficit. Since the budget is used as a tactical negotiating tool, it seems that the forecasters in the Republican OMB always anchor their predictions on expectations of lower total outlays with lower domestic spending and higher defense spending. In Shavelson, Cadwell, and Izu's (1977) study of teachers' pedagogical decision-making, teachers' prior information (experience) about students guides their expectations of students, which may serve to anchor teachers' subsequent estimates of students' aptitudes. Erroneous estimates would be made by teachers if such prior information contained errors or was misrepresented.

In the current study, 807 desired outcomes were collected from two rounds of judgmental forecasting; however, only 144 forecasts with anchoring bias on the desired outcome were detected in the judgmental adjustments. This shows that participants had desired outcomes for tourist arrivals in 59.34% of the D-O pair markets, but that only 17.84% of judgmental forecasts "stuck" on the desired outcome after judgmental forecasting. This result further reveals that participants made dramatic changes to their decisions when they received new and reliable information (statistical forecasts and historical data of tourist arrivals); only a few forecasts still relied on the desired outcome, and the desire anchor was detected. According to the test of H3b, significant anchoring bias on desired outcome was observed (M=0.571, SD=3.352, $t(56)$=1.285, p=0.204). Although the cases with desire anchor detected were relatively few (144 cases), they were normally distributed with a mean value equal to the desired outcome. This result supports the conclusion of the few previous studies that considered desired outcome to be an anchor biasing forecasters' judgmental forecasting.

Furthermore, since no significant evidence has been found that adjustments with the desire anchor are skewed from the desired outcome, the standard deviation of PC is used to describe the magnitude of changes on desired outcomes. The standard deviation of changes in the second round of forecasting (7.920) was twice that of the changes in the first round (3.352). This result indicates that, when using the desire anchor in judgmental forecasts, participants made larger adjustments to their desired outcomes in the second round of forecasting. A possible reason for this phenomenon is that participants learned from the first round that the reliability of their desired outcome was low and increased the uncertainty of their forecast, leading to larger variations on their desired outcome in the second round, even though they still relied on their expectations of tourism demand.

In addition, the test of H3e reveals that the forecast error associated with the desire anchor is larger than the components of forecast errors associated with the statistical forecast anchor and the latest observation anchor (M=-389350, SD=740639, $t(32)$=3.020, p=0.008). Regarding this finding and previous findings related to desired outcomes (H2a, H3b), it is concluded that participants' desired outcome, as well as further adjustments based on it, make the most inaccurate judgmental forecasts.

6.2.2.3 Latest observation anchor

The latest observation anchor is the last type of anchoring bias examined in this study. As stated by Harvey and Bolger (1996), people sometimes put more weight on the latest data point of the forecasting series than trends, even if the series is significantly trended. Therefore, the anchoring bias on the latest observation of a forecasting series would be as significant as the anchoring bias on statistical forecast or desired outcome. According to the test of H3c, anchoring bias on the latest observation is observed with a mean value equal to the latest observations of each D-O pair markets (M=-0.197, SD=3.672, $t(453)$=-

1.142, *p*=0.254). Like statistical forecast and desired outcome, the latest observation of a forecasting series is also considered a significant anchor in judgmental forecasting. In cases where latest observation anchor was detected, participants attempted to make adjustments around the latest observations of the forecasting series and no significant optimism or pessimism adjustment was identified based on that.

Considering the different number of forecasts with anchoring biases detected in the two rounds of forecasting, forecasts with the statistical forecast anchor decreased by 36.44%; forecasts with the desire anchor increased by 52.63%; and the number of forecasts with the latest observation anchor detected in the second round of forecasting (300) was almost double that in the first round (154), with the highest increase of 94.81%. It seems that, when forecasting experience increased, participants preferred to anchor their judgmental forecasts on the latest observation rather than on other anchors like statistical forecast and desired outcome. In other words, participants put more weight on the latest observations as they became more experienced in the forecasting tasks. This strategy of judgmental forecasting looks similar to the Naïve I method, which simply uses the latest observation of the time series as the forecast (Brodie & De Kluyver, 1987). Some review studies in tourism demand forecasting concluded that Naïve forecasts are more accurate than other forecasting techniques for annual forecasts (Athanasopoulos et al., 2011; Witt & Witt, 1995). Therefore, it is expected that, as forecasting experience increases, participants' forecasts anchoring on the latest observation become more accurate than forecasts anchoring on statistical forecast and desired outcome. However, a comparison of forecasts with the three anchors in the second round of forecasting reveals that forecasts with the statistical forecast anchor were significantly more accurate than the forecasts with the two other types of anchor (M_1=0.320, M_2=1.448, M_3=0.536, W(2,271)=4.798, p=0.009).

Furthermore, the test of the MAPE of forecasts with the latest observation anchor in the first round reveals significant forecast error (M=0.167, SD=0.626, t(153)=3.326, p=0.001). However, the test of H3f indicates that the errors of forecasts associated with anchoring bias on the latest observation were smaller than the errors of forecasts associated with the two other types of anchoring bias; and they were not statistically significant (M=-24581, SD=365977, t(152)=-0.831, p=0.407). These findings indicate that, although significant forecast error (in absolute value) using the latest observation anchor was observed in the experiment, the contribution of the latest observation anchor to the judgmental forecast error was insignificant when the direction (positive or negative) of adjustment is considered.

6.2.3 Overconfidence bias

Unlike desire bias and anchoring bias, overconfidence bias cannot be measured by PE or PC. In this study, it is measured by the percentage of inefficient adjustment based on the baseline forecasts ($\left|pe(F_{i,j,2006})\right| < |pe(SF_{i,j,2006})|$). Therefore, the evaluations of overconfidence bias for a participant are the same for his/her forecasts of all D-O pair markets but differ from one participant to the next. In two rounds of forecasting, participants' mean confidence was about 3.8 out of 5, which indicates that participants generally felt confident in their judgmental forecasts.

Previous studies have indicated that overconfidence bias is frequently observed among non-experts but is not significant among experienced forecasters (Aukutsionek & Belianin, 2001; Bolger & Önkal-Atay, 2004; Eroglu & Croxton, 2010). However, the test of H4a shows that significant overconfidence bias was observed among the experienced forecasters in this experiment (*M*=0.159, *SD*=0.341, $t(1359)=17.168$, $p<0.001$). This shows that participants' confidence in their judgmental forecasts was, on average, 15.9% higher than the forecasts' real performance. As a result, overconfidence bias is not a characteristic limited to non-expert forecasters but is also widely observed among forecasters experienced in tourism demand forecasting.

Further investigation of two rounds of forecasting revealed that participants' overconfidence bias in the first round was significantly greater than in the second round (*MD*=0.178, *SD*=0.017, $t(1227)=9.527$, $p<0.001$). To be specific, participants' confidence in the accuracy of their judgmental forecasts was generally 24.4% higher than their real performance in the first round of forecasting, but this overconfidence bias dropped significantly to 7.35% in the second round. This result indicates that participants are able to better understand their forecasting ability as their forecasting experience increases. As a result, the findings related to overconfidence bias in this study partially support the conclusions of previous studies: although overconfidence bias is significant among professional tourism demand forecasters, the seriousness of overconfidence bias is negatively correlated to forecasters' experience.

In addition, the test of H4b shows significant forecast errors associated with overconfidence bias in participants' judgmental forecasts (*M*=70988, *SD*=333646, $t(559)=5.036$, $p<0.001$). The mean forecast error associated with overconfidence bias reveals that, besides the insignificant forecast error associated with the latest observation anchor, forecast error associated with overconfidence bias is smaller than the other cognitive biases. A correlation test on the APEs of judgmental forecasts and participants' overconfidence also confirms that forecast error is positively correlated to overconfidence bias; however, this correlation is quite weak (r=0.095, p<0.001). Therefore, the contribution of overconfidence bias to the final forecast error is the lowest but statistically significant.

6.3 Debiasing strategies in judgmental forecasting of tourism demand

In this study, two debiasing strategies are proposed with the aim of reducing forecasters' cognitive bias in judgmental forecasting. One strategy is PF: giving forecasters information about their cognitive biases detected in the forecasting process, as well as the forecast errors associated with such cognitive biases detected in the previous forecasting season. According to the literature, PF is widely recommended as a type of informative guidance (Balzer et al., 1989; Benson & Önkal, 1992; Fildes et al., 2006). The other strategy is system-suggested forecasts: a kind of suggestive guidance that not only provides information but also directly suggests courses of action to forecasters. As mentioned in Fildes et al. (2006), system-suggested forecast can also be used to challenge forecasters' positions and assumptions in judgmental forecasting. Based on the hypotheses test results, the effectiveness of these two strategies in judgmental forecasting is further discussed below.

6.3.1 Effectiveness of performance feedback

Previous studies revealed that informative guidance is effective when applied to decision-making that is unrelated to forecasting (Fildes et al., 2006; Montazemi et al., 1996; Singh, 1998). This study provides supplements to the research of informative guidance in judgmental forecasting. PF as one type of informative guidance was provided to the participants in the experiment with expectations of reducing their cognitive bias and improving forecast accuracy. However, PF only takes the role of supportive information and no specific action was suggested to forecasters, who had to make the forecast decisions by themselves. Therefore, the effectiveness of PF was measured by the improvement in forecast errors before and after forecasters received PF; specifically, it was measured by the improvement of MAPE. The results support four out of five hypotheses (H5a, H6a–c), but not the hypothesis about participants' overconfidence bias (H7a). The effectiveness of PF regarding five types of cognitive bias is summarized in Table 6-2.

Table 6-2 Effectiveness of performance feedback

	Cognitive Bias	Reduced Forecast Error
Effective		
	Desire Bias	1.8% *
	Anchoring Bias on Statistical Forecast	21.3%**
	Anchoring Bias on Desired Outcome	81.7% *
	Anchoring Bias on Latest Observation	10.2%**
Ineffective		
	Overconfidence Bias	-2.3%**

**: 5% significance level; **: 1% significance level.*

The differences of forecast error before and after PF was provided were all significant, including the ineffective PF regarding overconfidence bias. The most significant improvement was achieved with PF regarding anchoring bias on desired outcome, which improved by 81.7% on the mean forecast accuracy. As examined in previous hypotheses, desired outcome produces less accurate forecasts than statistical forecast and judgmental adjustments. The second and third most significant improvements in forecast accuracy were also produced by PF of anchoring bias, 21.3% and 10.2% improvement regarding the statistical forecast anchor and the latest observation anchor, respectively. Although the effectiveness of PF regarding desire bias is relatively small, with only 1.8% improvement in the mean forecast accuracy, this improvement is statistically significant. PF has been proved to be ineffective only when it reports participants' overconfidence bias; a decrease of 2.3% in the mean forecast accuracy was observed after the relevant PF was provided. In summary, PF is an effective debiasing strategy to reduce desire bias and anchoring bias, especially the three types of anchoring bias; however, it is proved to be ineffective in reducing overconfidence bias.

These three findings are also of interest based on the above summary. First, both desire bias and desire anchor are based on participants' desired outcome, but the effectiveness of PF on these two desired outcome-related cognitive biases differs. When participants have a pre-established expectation of the tourism demand, it seems difficult to change their mind even if reliable information about their desire bias is provided. However, a 1.8% improvement of forecast accuracy is better than nothing. It is still recommended to provide PF during forecasters' judgmental forecasting when they have a desired outcome for tourism demand. On the other hand, forecast accuracy is significantly improved when anchoring bias is detected based on the desired outcome. Considering these two applications of PF related to participants' desired outcome, the difference between these two scenarios is whether or not PF is provided as a single reference to support decision-making. This shows that when performance feedback is provided as a single support to forecasters' decision-making, the effectiveness of PF is quite restricted. This effectiveness would be dramatically improved if PF were provided to forecasters accompanied by other kinds of reliable information, such as statistical forecasts and historical data of tourism demand.

Second, judgmental forecasts anchoring on statistical forecast had 30.6% forecast error (MAPE), and PF effectively reduced this to 9.3%. However, the statistical forecasts in this situation are still hard to beat. As shown in Table 6-3, the MAPE of statistical forecasts was 2% more accurate than PF-based revisions and this difference is statistically significant. The standard deviations of these two types of forecast also show that the deviation of statistical forecasting is only half that of the deviation generated by PF-based forecasts. As a result, statistical forecasts are more accurate than participants' judgmental adjustments when the statistical forecast anchor is detected, no matter whether PF is provided. In this situation, suggestive guidance would be more effective than informative guidance, which is to suggest that forecasters keep the baseline forecasts unchanged if they anchor their adjustments on PF.

Table 6-3 Comparison of PF-based revisions and statistical forecasts

Variable	N	Mean	SD	*df*	*t*	Sig.
$AS_{i,j,2007}^{P}$	201	0.093	0.127			
$SF_{i,2007}$	201	0.073	0.060			
$AS_{i,j,2007}^{P} - SF_{i,2007}$	201	0.020	0.010	286	2.043	0.042

Third, judgmental forecasts with the latest observation anchor had 18.6% forecast error (MAPE), and PF effectively reduced this to 8.4%. Further comparison of MAPE between PF-based revisions and the forecasts equal to the latest observations (Naïve I forecast) reveals that although the MAPE of Naïve I forecasts was smaller than PF-based revisions, the difference between them was not statistically significant (Table 6-4).

Table 6-4 Comparison of PF-based revisions and Naïve I forecasts

Variable	N	Mean	SD	*df*	*t*	Sig.
$AL^P_{i,j,2007}$	141	0.084	0.069			
$NI_{i,2007}$	141	0.073	0.064			
$AL^P_{i,j,2007} - NI_{i,2007}$	141	0.011	0.079	280	1.392	0.165

NI indicates Naïve I forecasts.

According to the findings regarding anchoring bias on the latest observation, two debiasing strategies can be developed:

(i) From the perspective of informative guidance, PF about forecasters' anchoring error on the latest observation is a good strategy to reduce their anchoring bias;

(ii) From the perspective of suggestive guidance, recommending forecasters to replace their judgmental adjustments with Naïve I forecasts (the latest observation) is also a good strategy, similar to the PF.

6.3.2 Effectiveness of system-suggested forecast

The second type of debiasing strategy in this study, suggestive guidance directly provides optimal forecasts based on forecasters' judgmental forecasts and the detected cognitive biases. Such forecasts can be used as the final forecasts when an FSS is designed with a high level of restrictiveness; or they can be used as supportive information to further aid forecasters' decision-making. Since the information of system-suggested forecasts contains clear suggestions of action, it is not classified as informative guidance but suggestive guidance when used in systems with low restrictiveness (Fildes et al., 2006).

Similar to the effectiveness test of PF, the effectiveness of system-suggested forecast is also measured by the improvement of MAPE. According to the test results, three out of five hypotheses regarding system-suggested forecasts are rejected (H5b, H6e, and H7b) and the two supported in this part of the analysis are both about debiasing anchoring bias (H6d and H6f). Table 6-5 summarizes the effectiveness of system-suggested forecasts in reducing participants' cognitive bias.

Table 6-5 Effectiveness of system-suggested forecast

	Cognitive Bias	Reduced Forecast Error	
		vs. UF	vs. PF
Effective			
	Anchoring Bias on Statistical Forecast	24.3%**	3.0%**
	Anchoring Bias on Latest Observation	14.4%**	4.2%**
Ineffective			
	Desire Bias	-6.2%	-8.9% *
	Anchoring Bias on Desired Outcome	82.6% *	0.9%
	Overconfidence Bias	-2.5% *	0.8%

*UF denotes unaided judgmental forecast; PF denotes PF-based revision; *: 5% significance level; **: 1% significance level.*

Generally speaking, system-suggested forecasts outperform participants' unaided judgmental forecasts when three types of anchoring biases are detected. Similar to the effectiveness of PF, the most significant improvement in forecast accuracy is achieved when participants anchor their judgmental forecasts on desired outcome, which reduces MAPE by 82.6%. System-suggested forecasts perform better when the desire anchor is detected than when statistical forecast anchor and a latest observation anchor are detected; but the improvements are still significant, reducing MAPE by 24.3% and 14.4%, respectively. It seems that system-suggested forecasts are harmful to forecast accuracy when desire bias and overconfidence bias are detected; this harmful effect is only statistically significant among the judgmental forecasts in which overconfidence bias detected.

System-suggested forecasts are more accurate than participants' PF-based revision when forecasters anchor on statistical forecast or the latest observation. Specifically, system-suggested forecasts are 3.0% more accurate than PF-based revisions when anchoring bias on statistical forecast is detected; this improvement in forecast accuracy increases to 4.2% when forecasters anchor on the latest observations of the forecasting series. When overconfidence bias is detected, system-suggested forecasts perform slightly better than PF-based revisions, reducing MAPE by 0.8%; this improvement is not statistically significant. The results also show that system-suggested forecasts are not as accurate as PF when cognitive bias related to desired outcome is detected: an 8.9% increase in MAPE is observed among cases with desire bias and this effect is statistically significant; a 0.9% increase in MAPE is observed among cases in which the desire anchor detected, but it is not statistically significant. As a result, system-suggested forecasts are most effective in improving forecast accuracy when participants anchor their judgmental forecasts on statistical forecast or the latest observation. System-suggested forecasts are not as good as PF at dealing with the cognitive biases related to participants' desired outcome.

The above findings further revised and supplemented the findings on the effectiveness of PF. First of all, they show that the MAPE of statistical forecast is 2% more accurate than the MAPE of performance-based revisions when the statistical forecast anchor is detected; and they imply that forecasters should be advised to maintain the baseline (statistical) forecasts unchanged in this situation. However, a 3.2% improvement in MAPE is produced by system-suggested forecasts compared with PF-based revisions, which shows that system-suggested forecasts are more accurate than statistical forecasts. Comparison of the MAPE of system-suggested forecasts and PF-based revisions confirms this finding, indicating that system-suggested forecast is, on average, 1.1% more accurate than statistical forecast (Table 6-6). Therefore, system-suggested forecast, with the error of anchoring bias on statistical forecast eliminated, is preferred. However, according to Equation (3.3) and Equation (3.11), system-suggested forecasts following the method proposed in this study may not be available if coefficient $\beta_{3,j}$ for a forecaster is not significant or is equal to zero. In such a case, keeping the baseline forecast unchanged would be the second best choice.

Table 6-6 Comparison of system-suggested forecasts and statistical forecasts

Variable	N	Mean	SD	*df*	$\chi 2$	Sig.
$AS_{i,j,2007}^{S}$	199	0.056	0.136			
$SF_{i,2007}$	199	0.071	0.059			
$AS_{i,j,2007}^{S} - SF_{i,2007}$	199	-0.015		1	32.514	<0.001

Furthermore, system-suggested forecast, with the error of the latest observation anchor eliminated, is more accurate than PF-based revisions in the same scenario. Previous discussion reveals that the accuracy of PF-based revisions is similar to the accuracy of Naïve I forecasts and that both outperform participants' unaided judgmental forecasts when they anchor on the latest observation of a forecasting series. Comparing system-suggested forecast and Naïve I forecast in this scenario reveals that system-suggested forecast is significantly more accurate, with a 3.1% improvement in MAPE (Table 6-7). Therefore, Naïve I forecasts are as good as PF-based revisions, but not as good as system-suggested forecasts in this scenario. However, according to Equation (3.3) and Equation (3.11), system-suggested forecasts following the method used in this study may not be available if coefficient $\beta_{5,j}$ for a forecaster is not significant or is equal to zero. If this is the case, it is necessary to decide whether to provide PF as informative guidance or Na ïve I forecasts as suggestive guidance. According to Table 6-4, the MAPE of Na ïve I forecasts is 7.3% and that of PF-based revisions is 8.4%; the standard deviation of Na ïve I forecasts (0.064) is also slightly smaller than the standard deviation of PF-based revisions (0.069). Therefore, though these two MAPEs are not significantly different, Na ïve I forecasts have priority over PF-based revision. As a result, system-suggested forecast, with the error of anchoring bias on the latest observation eliminated, is the best choice when participants' judgmental forecasts anchor on the latest observations of a forecasting series. If the system forecasts are not available, the second best option is to advise forecasters to replace their judgmental forecasts with Na ïve I forecasts.

Table 6-7 Comparison of system-suggested forecasts and Na ïve I forecasts

Variable	N	Mean	SD	*df*	x^2	Sig.
$AL_{i,j,2007}^{S}$	141	0.042	0.861			
$NI_{i,2007}$	141	0.073	0.064			
$AL_{i,j,2007}^{S} - NI_{i,2007}$	141	-0.031		1	5.965	0.015

6.4 A guidance-based debiasing model in the design of TDFSS

According to the above discussions, the conceptual debiasing framework proposed in this study is further revised as shown in Figure 6-1.

Considering the detection of cognitive bias, statistical forecast error is significant in judgmental forecast and should be involved in the model to estimate the leading indicators of cognitive bias. All types of cognitive bias researched in this study significantly exist in judgmental forecasting of tourism demand, with the exception of anchoring bias on statistical forecast when judgmental adjustment is in the right direction; all types of cognitive bias researched in this study contribute significantly to the forecast error of tourism demand, with the exception of anchoring bias on the latest observation.

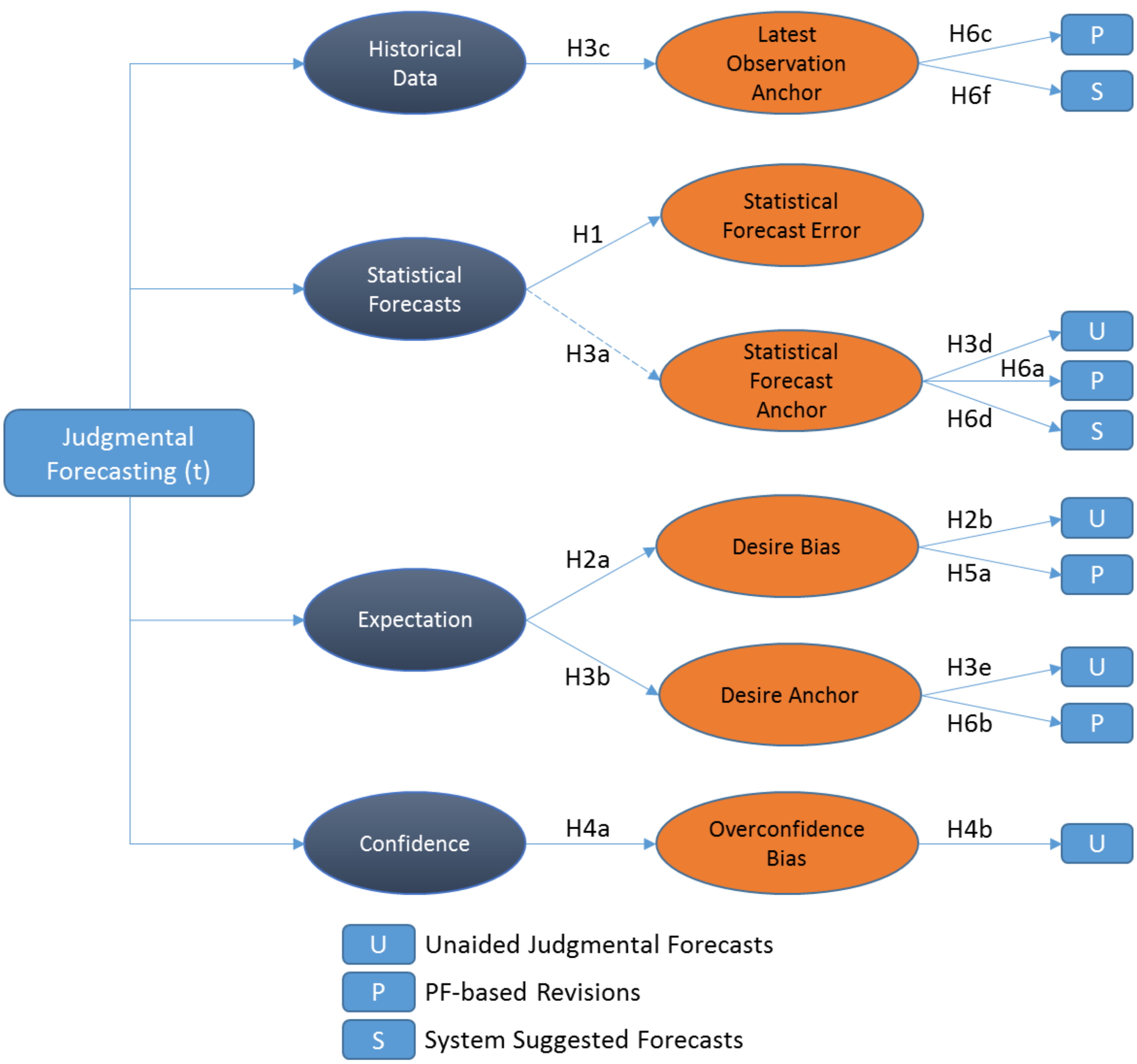


Figure 6-1 Confirmed conceptual debiasing framework in the design of TDFSS

Considering the debiasing strategies, PF is effective in reducing desire bias and three types of anchoring bias; however, it is harmful to forecast accuracy when dealing with overconfidence bias. System-suggested forecasts are effective in reducing two types of anchoring bias, the statistical forecast anchor and the latest observation anchor, and performs better than PF in these two situations. However, it is ineffective in dealing with cognitive biases related to forecasters' desired outcome (desire bias and anchoring bias on desired outcome) and overconfidence bias.

The final task for this study is to apply the confirmed conceptual framework to the design of TDFSS. An application model of debiasing in the module of judgmental forecasting in TDFSS is proposed, with two components of cognitive bias detection and debiasing.

6.4.1 Implementation of cognitive bias detection

Cognitive bias detection serves the effective debiasing strategy; so a paired detection-debiasing approach should be established for each type of cognitive bias. As discussed in previous sections, all five types of cognitive bias (including three types of anchoring bias) are identified as significant in the judgmental forecasting of tourism demand. However, both PF and system-suggested forecasts with overconfidence error eliminated have been proven ineffective in reducing such cognitive bias. In other words, cognitive bias detection of overconfidence bias cannot be paired with effective debiasing strategy using the two methods (PF and system-suggested forecast) proposed in this study. Therefore, overconfidence bias is not considered a target in the implementation of cognitive bias detection.

The implementation of cognitive bias detection is shown in Figure 6-2. At the beginning of a forecasting season (t), the presence of desired outcome should be identified by asking the forecaster's expectations for tourism demand of D-O pair markets. Then judgmental forecasting is processed, with historical data and statistical forecasts provided as supportive information for the forecaster's reference.

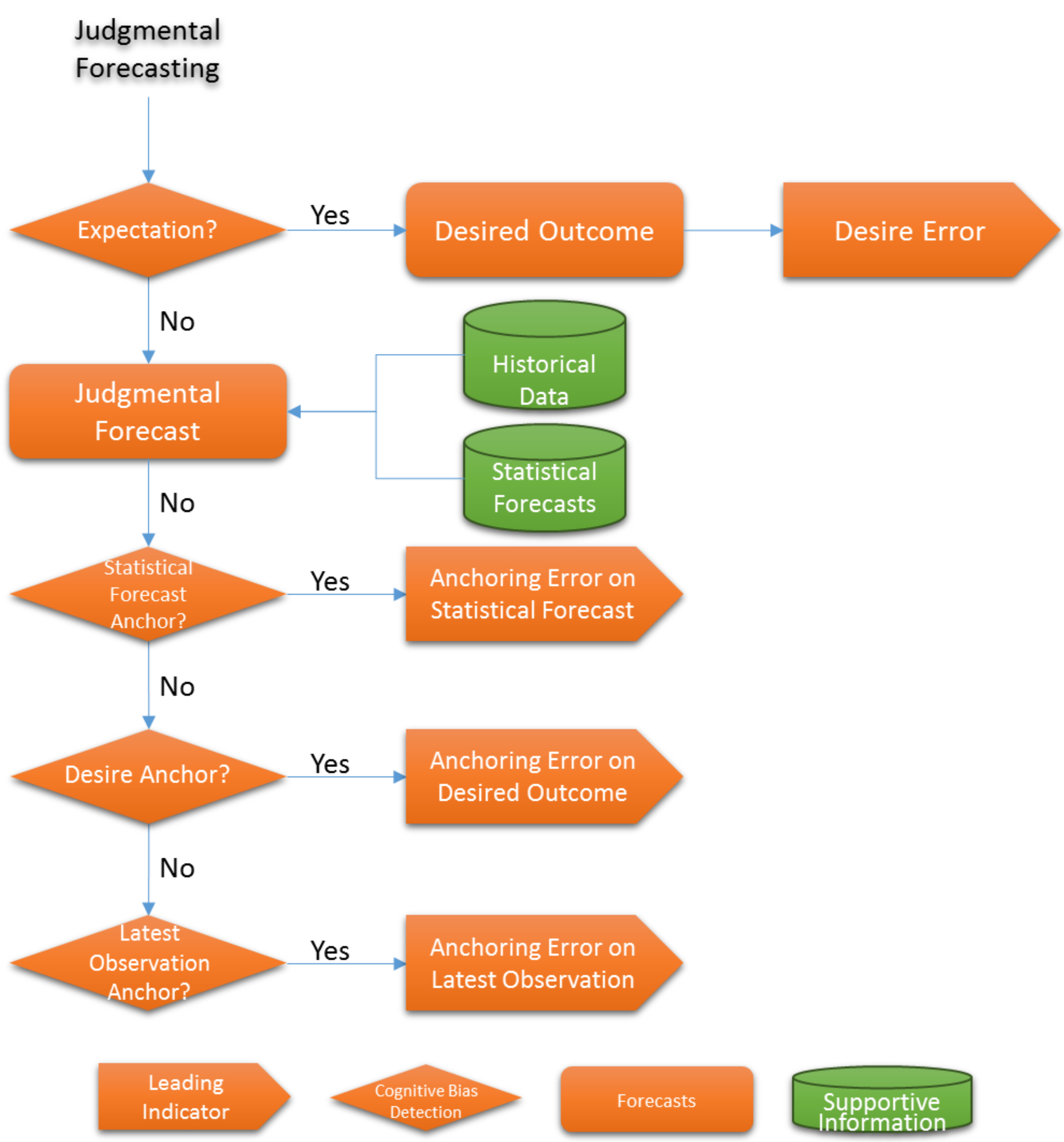


Figure 6-2 Implementation of cognitive bias detection

When the real outcomes of tourism demand at *t* are available, the percentage error of the judgmental forecast is regressed by (i) the percentage error of the desired outcome, (ii) the percentage changes regarding the three anchors of statistical forecast, desired outcome, and the latest observation, and (iii) the level of overconfidence (confidence level subject to the level of correct adjustments). According to the estimated regression coefficients ($\beta_2 \sim \beta_6$), forecast error can be further decomposed and each component associated with an identified cognitive bias, named desire error, anchoring error on statistical forecast, anchoring error on desired outcome, and anchoring error on the latest observation. Such decomposed forecast errors are used as the leading indicators of forecasters' cognitive bias.

6.4.2 Implementation of guidance-based debiasing

As discussed in previous sections, the two debiasing strategies proposed in this study perform differently in reducing cognitive bias in judgmental forecasting. Generally, PF is the most effective strategy for reducing the cognitive biases related to desired outcomes (desire bias and anchoring bias on desired outcome); system-suggested forecasts perform best in reducing the anchoring biases on statistical forecast and the latest observation.

Based on the discussion of debiasing effectiveness, seven principles of debiasing are proposed for implementation in the design of TDFSS:

1. Provide PF about forecasters' desire error when they have desired outcomes for tourism demand.
2. Provide PF about forecasters' error of the desire anchor when they anchor their judgmental forecasts on desired outcome.

3. Providing other reliable information (e.g., statistical forecasts and historical data of tourism demand) with PF would be more helpful to improve the accuracy of judgmental forecasting.
4. Advise forecasters to replace their judgmental forecasts with system-suggested forecasts (with anchoring error extracted) when they anchor their judgmental forecasts on statistical forecasts.
5. Advise forecasters to keep statistical forecast unchanged when they anchor their judgmental forecasts on statistical forecast and no system-suggested forecast is available.
6. Advise forecasters to replace their judgmental forecasts with system-suggested forecasts (with anchoring error extracted) when they anchor their judgmental forecasts on the latest observation of a forecasting series.
7. Recommend Naïve I forecasts to forecasters when they anchor their judgmental forecasts on the latest observation of a forecasting series and neither system-suggested forecast nor PF is available.

Using these seven principles, the implementation of debiasing in the design of TDFSS is shown in Figure 6-3.

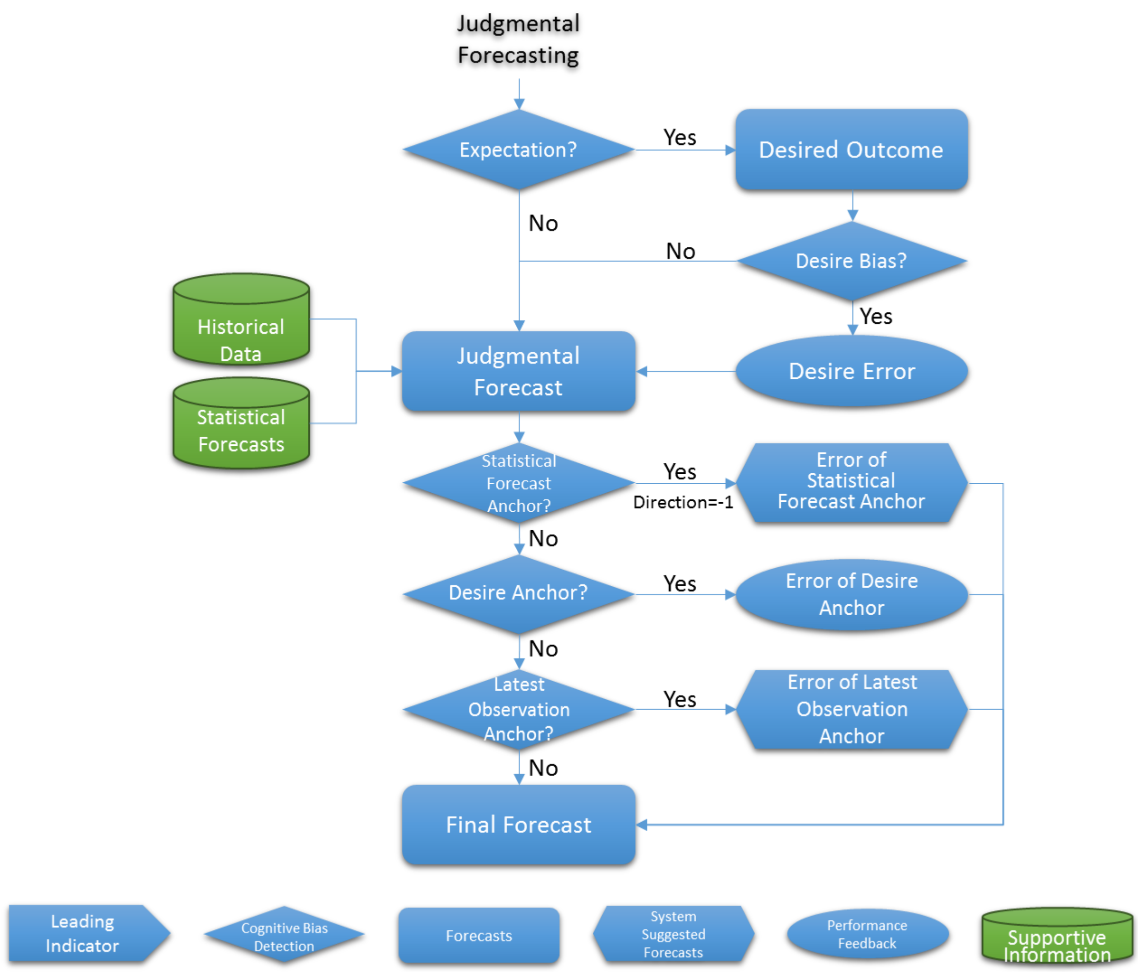


Figure 6-3 Implementation of guidance-based debiasing

6.4.3 The debiasing model

An application model of debiasing in the design of TDFSS is proposed by combing the above two implementations (Figure 6-4). In this model, four leading indicators of cognitive bias are estimated in the forecasting of tourism demand at time *t*, which are used in the forms of PF or system-suggested forecasts in the following forecasting season (*t*+1).

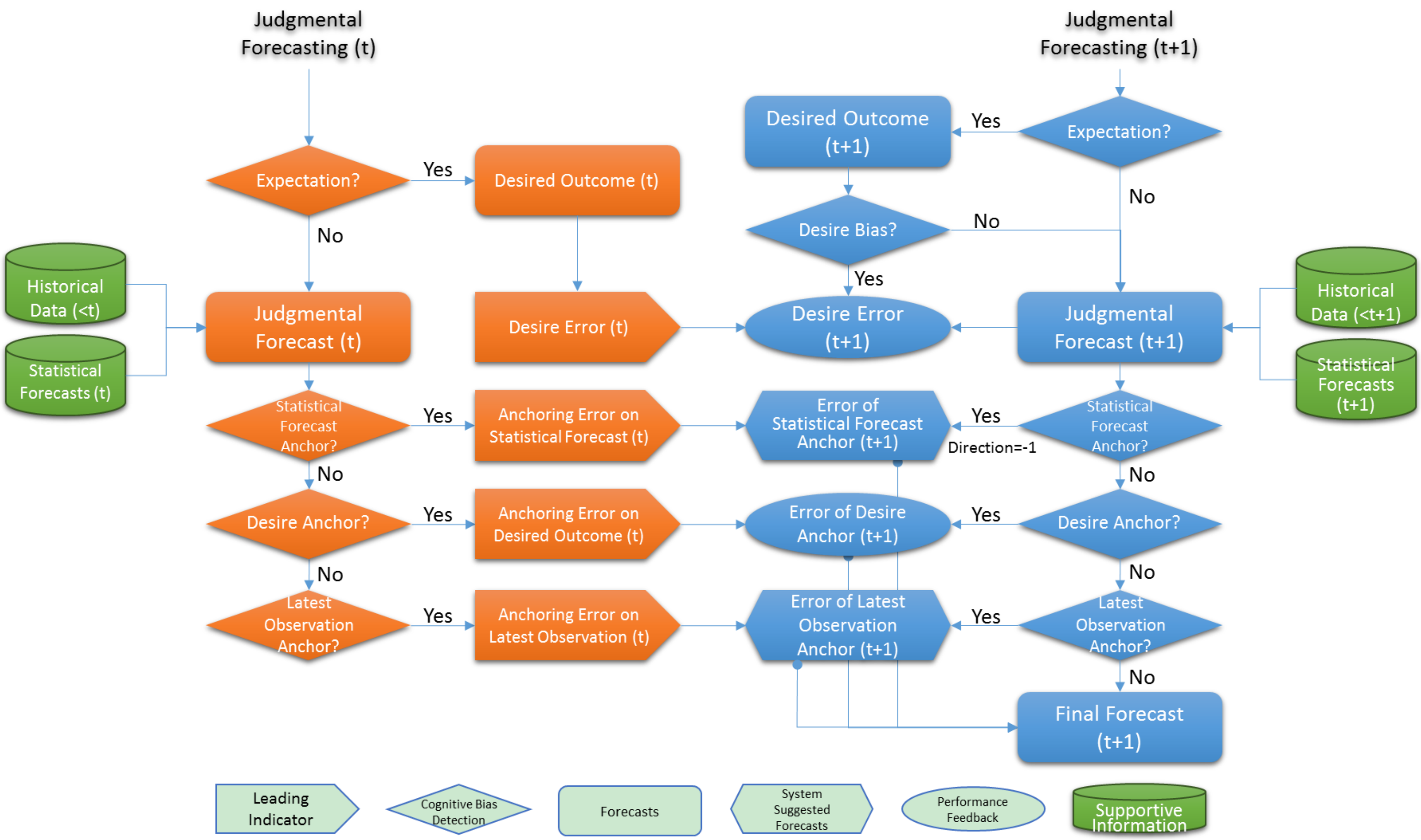


Figure 6-4 The debiasing model in the design of TDFSS

6.5 Summary

A summary of the hypotheses test results was presented at the beginning of this chapter and further discussion of these results followed. First, it has been proven that statistical forecast error is still significant even when judgmental forecasting is processed. Therefore, it should be involved in modeling forecast error and cognitive biases. Regarding the cognitive bias detected in judgmental forecasting of tourism demand, it is found that participants' desired outcome and further adjustments based on it produce the most inaccurate forecasts. Furthermore, desire bias occurs jointly with conservatism bias and optimism bias occurs jointly with anchoring bias on statistical forecast. It has also been proven that large-scale adjustments are more accurate than small adjustments and wrong-side adjustment should be avoided. Regarding the debiasing strategies, PF is the most effective way to reduce cognitive biases related to participants' desired outcome, and system-suggested forecast perform best when participants anchor their forecasts on statistical forecasts or the latest observations of a forecasting series.

7 CONCLUSION

This chapter reviews the entire study, including the research objectives, the conceptual framework and hypotheses, methodology, hypotheses test results, findings, and discussion, and brings it to a conclusion. In addition, the significance of this study is addressed in terms of theoretical contribution and practical implementation in the design of TDFSS. The chapter closes with concluding remarks.

7.1 Overview of the study

As mentioned at the beginning of this study, previous studies in the areas of FSS and TDFSS development mainly focused on the adoption of advanced statistical forecasting techniques in order to improve forecast accuracy, or the combination of statistical forecasting and judgmental forecasting techniques in order to incorporate the advantage of forecasters' domain knowledge and experience into the forecasting process. However, the negative effect of forecasters' judgment, their cognitive bias, on forecast accuracy and the strategy to eliminate or reduce such influence has largely been left unexplored. Therefore, the main purposes of this study are to improve the design of TDFSS by further enhancing forecasters' ability to find appropriate times to conduct judgmental forecast, and to apply accurate judgmental interventions when appropriate.

The research objectives of this study are threefold: first of all, it aims to identify the cognitive bias that occurs when forecasters use tourism demand FSS, and to measure the contribution of cognitive bias to final forecast error; secondly, this study focuses on the development of two debiasing strategies and their effectiveness in reducing cognitive bias; finally, the third purpose of this study is to improve the design of TDFSS by proposing a guidance-based debiasing model using the findings obtained by fulfilling the first two research purposes.

Relevant studies that examine the key concepts used in this study were reviewed in the literature review chapter. In Arnott's (2006) taxonomy of cognitive biases, seven are commonly identified in studies of tourism demand forecasting, FSS development, and other research areas involving judgmental forecasting: herding bias, conservatism bias, desire bias, overconfidence bias, and three types of anchoring bias. Five cognitive biases are ultimately considered in the debiasing tasks in this study.

Empirical studies have also revealed that informative and suggestive guidance are two broad ways to reduce cognitive bias in judgmental forecasting. Specifically, PF and system-suggested forecasts are the two recommended debiasing strategies, representing informative guidance and suggestive guidance, respectively. The current study then proposed two algorithms to generate PF and system-suggested forecasts. Then a conceptual framework with 21 hypotheses regarding the influence of cognitive bias on forecast error and the effectiveness of the two debiasing strategies were proposed in the conceptual framework chapter.

Given the nature of this study, a prototype cognitive bias detection and debiasing function was designed based on HKTDFS and the data used for the experiment were real outcomes of tourist arrivals in different D-O pair markets. A two-round experiment was designed for data collection and a preliminary test was conducted before the main experiment. Following convenience sampling, 75 experts in tourism demand forecasting were invited to participate in this experiment. Each of them produced 20 forecasts for each D-O pair market over two rounds of judgmental forecasting. After data screening and outlier checking, 1,360 cases were kept for further analysis.

Both parametric and non-parametric statistical methods were adopted to test the hypotheses developed and the results of these hypotheses tests were reported. Fifteen of the 21 hypotheses were fully or partially supported; four of the six rejected hypotheses concerned the effectiveness of two debiasing strategies. Both PF and system-suggested forecasts with overconfidence error eliminated were found to be ineffective in improving forecast accuracy. Further discussion based on the test results was presented in the following chapter.

The results show some interesting findings. First of all, statistical forecast error is significant after judgmental forecasting is processed, which indicates that the component of statistical forecast error should be involved in the regression model of forecast error. Moreover, joint efforts are observed in some situations: conservatism bias and desire bias jointly affect forecasters' judgmental forecasting, while optimism bias always occurs with the statistical forecast anchor. Although there is significant anchoring bias on the latest observation, its contribution to the forecast error is the smallest and most insignificant of all types of cognitive bias. Besides, overconfidence is not exclusively a feature of non-experienced forecasters; it is also commonly observed among experts of tourism demand forecasting; however, the seriousness of overconfidence bias in judgmental forecasting decreases as forecasting experience increases. Considering the contribution of cognitive bias on forecast error, desire bias usually leads to excessively high forecasts and generates the largest forecast error of all types of forecast; on the other hand, the forecast error associated with overconfidence bias is the smallest.

Considering the two debiasing strategies proposed in this study, PF performs best in reducing cognitive bias related to forecasters' desired outcome, and other supportive information like statistical forecasts and historical data of tourism demand are helpful. System-suggested forecast is the best option when forecasters' anchor on statistical forecast and the latest observation. However, the second best choice varies depending on the type of cognitive bias if PF or system-suggested forecast are not available. Keeping the baseline forecast unchanged is the best choice if anchoring bias on statistical forecast is detected but no relevant system-suggested forecast is available; when the latest observation anchor is detected, Naïve I forecasts would have similar benefits as system-suggested forecasts on forecast accuracy.

The conceptual framework proposed in this study is further revised according to the findings and discussion. In order to guide the design of TDFSS using the conceptual framework, the implementation of cognitive bias detection and debiasing were further discussed. Finally, a guidance-based debiasing model was proposed at the end of Chapter 6.

7.2 Theoretical contribution

7.2.1 Exploration of judgmental forecasting of tourism demand

At present, research into cognitive bias in judgmental forecasting and the design of FSS is limited and studies in the field of tourism demand forecasting are extremely rare. Due to the unique characteristics of tourism demand, many research gaps relating to forecasters' cognitive bias in judgmental forecasting of tourism demand are left unexplored. This study focuses on the identification of cognitive bias in judgmental forecasting of tourism demand, as well as the development of effective debiasing strategies to reduce the influence of cognitive bias on forecast accuracy. Some previous studies proposed different frameworks and taxonomies to classify and characterize cognitive bias, but the majority had problems of overlapping and causal relationships between cognitive biases. Up to now, Arnott's (2006) taxonomy of cognitive bias is considered the most exhaustive and mutually exclusive one with which to classify and categorize cognitive biases. However, no review study in this area using Arnott's taxonomy has been found in the recent literature. This study presents the first comprehensive review of empirical studies in the area of judgmental forecasting using such a taxonomy; its results can be used for other researcher's reference in future studies.

7.2.2 The conceptual framework

The literature review identified some major limitations in previous studies and a conceptual framework with 21 hypotheses of cognitive bias detection and debiasing were developed. The framework contains two parts with different concepts: cognitive bias detection and debiasing; the former concepts offer foundation to the latter. With the hypotheses developed, the conceptual framework provides theoretical guidance on how to incorporate cognitive bias detection and debiasing into the design of TDFSS. Using the data collected in the experiment, the hypotheses were tested and the conceptual framework further revised according to the results. Specifically, the majority of the concepts were found to be significant, and were further confirmed to be kept in the framework. However, some of the concepts were proven to be insignificant. For example, there is significant anchoring bias on the latest observation but its contribution to forecast error is not significant; system-suggested forecast cannot outperform PF-based revisions when forecasters' cognitive bias is based on their desired outcomes; and both system-suggested forecasts and PF are insignificant in reducing forecasters' cognitive bias. As a result, the concepts with insignificant test results were removed from the conceptual framework and the revised conceptual framework is considered to be the most reliable framework with significant and important concepts.

Regarding the components of cognitive bias detection in the framework, five cognitive biases – desire bias, overconfidence, and three types of anchoring bias – were selected as the debiasing tasks. According to the literature, these cognitive biases are the most commonly identified in studies of judgmental forecasting in non-tourism research areas; however, their significance in tourism demand forecasting is still unknown. Using the real data of tourism demand in 10 D-O pair markets, the existence of these five cognitive biases, as well as their contribution to forecast error in judgmental forecasting of tourism demand was examined in a two-round experiment. Some findings of this study provide evidence to support some previous studies. Consistent with the conclusions in Coshall (2009) and Athanasopoulos and colleagues' (2011) studies, for example, tourism demand is found to be generally underestimated by ES forecasts. Large-scale revision of the desired outcome always performs better than small revisions, and wrong-direction adjustment should be avoided when the statistical forecast anchor is detected (Fildes et al., 2009; Trapero et al., 2011). On the other hand, some unique conclusions of this study contradict the findings of previous studies. For example, forecasters' desired outcome for tourism demand performs too conservatively in this study; however, optimism bias is frequently identified as being accompanied by forecasters' expectation in previous studies (Fildes et al., 2009; Mathews & Diamantopoulous, 1990). It is believed that overconfidence is significant among non-experienced forecasters but not among experienced ones (Bolger & Önkal-Atay, 2004; Eroglu & Croxton, 2010), but the result in this study shows that such cognitive bias is also significant among tourism demand forecasting experts. For the findings that are consistent in previous studies, this study provides more evidence to support conclusions in the area of tourism demand forecasting. For the findings contradicting previous studies, this study provides new viewpoints and corresponding evidence, which helps to further increase the diversity of research and to stimulate new research related to these inconsistent findings.

Regarding the components of debiasing in the framework, the current study proposed two debiasing strategies, PF and system-suggested forecasts, as well as the algorithms to generate the information required for these two strategies. PF contains information about forecasters' mean forecast error associated with a specific cognitive bias if it is identified; system-suggested forecasts are the revisions with the error of corresponding cognitive bias eliminated. The hypotheses test results regarding these two strategies provide strong evidence that the strategies are effective in reducing forecast error in different scenarios. For example, PF is the most effective strategy for reducing forecasters' cognitive bias based on their desired outcomes; system-suggested forecasts, with errors of cognitive bias eliminated, perform best in improving forecast accuracy when forecasts are anchored on statistical forecast or the latest observation. Additionally, other supportive guidance is also provided as supplementary options if system-suggested forecasts or PF are not calculable with the proposed algorithms. Therefore, a set of comprehensive debiasing strategies is proposed in this study, contributing to research into TDFSS development and debiasing in other areas.

7.2.3 Generalization of the study approach

The approach of this research can be easily generalized to other research areas in four ways. First of all, a preliminary review of the literature regarding cognitive bias reveals two limitations in the field: (i) a lack of reviews of cognitive bias in the research areas of judgmental forecasting, tourism demand forecasting, and the design of FSS, and (ii) cognitive bias can be expressed in different forms and each form is sourced by a different cognitive heuristic. The categorization or taxonomy of cognitive bias also varies in previous studies. Therefore, the fundamental aim of this study was to identify an appropriate taxonomy of cognitive bias. After a comparison of existing studies of cognitive bias categorization, one of the most exhaustive and mutually exclusive taxonomies (Arnott, 2006) was selected to guide the literature review. Then a comprehensive review of empirical studies on cognitive bias detection and debiasing was conducted, based on which, a conceptual framework with related hypotheses was developed. Therefore, the establishment of the conceptual framework of this study was conducted following a taxonomy – the literature review – conceptual framework (TLC) approach. Such an approach can be used to establish the conceptual/theoretical framework of cognitive bias and debiasing strategies regarding other types of cognitive biases in judgmental forecasting of tourism demand that are not investigated in this study. Also, such a TLC approach can be applied in other research areas with conditions similar to this study, that is, areas lacking a comprehensive review study and a widely accepted taxonomy of research objective(s).

A two-round laboratory experiment was designed for data collection in this study. With the support of web-based FSS, such a method is good for large-scale data collection of judgmental forecasts among mass participants. Also, the data collected from this approach include different types of forecast (desired outcomes, judgmental forecasts, and statistical forecasts) as well as the historical data of real outcomes of tourism demand. Once the real outcomes of tourism demand at the forecasting point are available, it is easy to calculate different measures of forecast accuracy (e.g., PE, MPE, APE, and MAPE) for each type of forecast. More important, forecasts collected from two rounds of forecasting contain different information. Considering the first round of forecasting as the control group and the second round as the experiment group, two dimensions of information are reflected from the collected forecasts: one is different types of forecast like desired outcomes, statistical forecasts, and judgmental forecasts in different conditions; the other dimension is whether new influencing factors are involved in judgmental forecasting, such as PF and system-suggested forecasts, as identified in this study. Two-dimensional comparison of forecast accuracy could thus be conducted with this data collection approach. Because of the diversity of comparison capability, such an approach can be adopted to examine the significance of other types of cognitive bias, testing the effectiveness of other debiasing strategies, and other studies related to forecast accuracy competition.

Moreover, it is not difficult to test for the existence of cognitive bias, or to test the contribution of cognitive bias to forecast error. Statistical techniques of mean comparison, such as independent-sample *t*-test, paired-sample *t*-test, ANOVA, and non-parametric ANOVA, were adopted as the data analysis methods for these two purposes in this study. Such methods are the simplest and most efficient methods for forecast accuracy comparison. The selection of a particular method depends on various conditions like whether two or more groups of forecast accuracy must be compared, whether they are independent, and whether the forecast accuracy in different groups follows the assumption of equal variance. Detailed criteria of method selection can be found on the ForPrin website (http://www.forecastingprinciples.com/). Because of their efficiency and ease of control, these methods can be adopted to test the existence of cognitive bias and its influence in other fields of judgmental forecasting; they can also be used to test the influence of other types of cognitive bias on judgmental forecasting of tourism demand.

Finally, this study examined the effectiveness of debiasing strategy by checking the improvement of forecast accuracy measured by MAPE. Since non-equal variances were widely observed in each comparison, non-parametric ANOVA was adopted to test effectiveness; however, parametric ANOVA should be used if there is equal variance assumption among comparison groups. These methods and the criteria by which to select the appropriate one are easy to follow in order to test the effectiveness of other debiasing strategies or other algorithms.

7.3 Practical contribution

7.3.1 The guidance-based debiasing model

According to the revised conceptual framework, a guidance-based debiasing model was proposed at the end of this study. This model is an application of the conceptual framework in practice, which can be used as a blueprint to improve the design of TDFSS. The advantages of this model in TDFSS design are twofold and related to the two components. The component of cognitive bias detection (Figure 6-2) can help the system successfully detect a user's cognitive bias in a forecasting process. The leading indicators of corresponding cognitive bias can also be automatically calculated when the real outcome at the forecasting point is available. The other component of the model is debiasing (Figure 6-3), which provides appropriate debiasing strategies to the system user in order to reduce his/her cognitive bias in the next forecasting season. According to the results of this study, incorporating this model into the design of the judgmental forecasting module in TDFSS could improve the forecast accuracy, and can benefit both TDFSS developers and system users.

7.3.2 Support for TDFSS developers

For TDFSS developers, the guidance-based debiasing model clearly identifies the steps of judgmental forecasting, as well as the key supportive materials and corresponding products in each step. Three key issues need to be highlighted regarding the contribution of the application model. First, both desire bias and desire anchor are based on forecasters' desired outcome for the tourism demand of D-O pair markets. Therefore, the model proposes an additional step before judgmental forecasting for system users to provide their desired outcome. Since desired outcome should ideally be provided based on system users' domain knowledge and expertise, no supportive information should be provided in this step.

The application model also focuses on debiasing four cognitive biases, including desire bias and three types of anchoring bias. With the exception of overconfidence bias, the most commonly identified cognitive biases in the judgmental forecasting literature have all been addressed in this model. Since a variety of cognitive bias detection and debiasing is involved, the application model is presented in the form of a flow chart for TDFSS developers to follow easily.

The two debiasing strategies of PF and system-suggested forecasts should be used in different scenarios. As shown at the end of Section 6.4.2, seven principles of debiasing are proposed with the application model. Following these principles, the system developer is able to design different debiasing strategies for different cognitive biases. When a specific cognitive bias is detected, the corresponding debiasing strategy can be activated. Therefore, an optimal use of these two debiasing strategies can be achieved in practice.

Besides the application model, this study also provides detailed algorithms for cognitive bias detection and debiasing. As shown in the equations in Chapter 3, the first imported function is to estimate the coefficients $\beta_{1,j}$~$\beta_{6,j}$ using OLS methods; then clear description of how to calculate leading indicators for each cognitive bias is presented; finally, the way to produce system-suggested forecasts, with corresponding cognitive bias eliminated, is also described in detail. It is easy for system developers to incorporate such algorithms into TDFSS using different programming languages (e.g., java, c^{++}, and R).

7.3.3 Support for tourism demand forecasters

Besides the benefits to TDFSS developers, the application model also benefits system users. First of all, the judgmental forecasting process using TDFSS is redesigned. A new step for collecting the system user's desired outcome is added to the beginning of the forecasting process. Relevant information to support the system user's decision-making is also provided with specific purposes during the forecasting process. The forecasting process proposed in the application model can help system users to better recognize which potential cognitive biases are involved in their judgmental forecasting of tourism demand and the features of such biases (e.g., always optimistic or pessimistic).

Secondly, PF and system-suggested forecasts provide quantified information about system users' cognitive bias as detected in their forecasting process in the previous forecasting season. With other supportive information, such as baseline forecasts and historical data of tourism demand, system users are able to critically evaluate the reliability of their expectations, domain knowledge, experience, as well as other information they receive for judgmental forecasting. As a result, critical thinking helps system users make more reliable decisions in judgmental forecasting. For example, to make necessary changes to the baseline forecast, or to keep the baseline forecast unchanged if no reliable information is available. Also, small changes to baseline forecasts generally reflect forecasters' uncertainty, which should be avoided after critical thinking.

Furthermore, after the use of TDFSS with the application model of debiasing involved, forecasters are able to recognize that adjustments to baseline forecasts regarding the PF or directly accepting the system-suggested forecast are actually wise choices when they are offered by the system. Specifically, the current study has proved that revision based on PF produces the most accurate forecast when desire bias or desire anchor is involved in judgmental forecasting; system-suggested forecasts with anchoring bias eliminated should be directly accepted when system users anchor their forecast on statistical forecasts or the latest observations of a forecasting series.

Regarding the ease of use, the redesigned judgmental forecasting process does not have many differences comparing with the process using traditional FSS. Additional inputs required from forecasters are also limited, which dependents on the specific debiasing strategies proceeded for debiasing. If performance feedback is used as the debiasing strategy, only one additional value of forecast (PF-based revision) in each forecasting exercise requires forecasters' input; while, if system suggested forecast is proceeded as the debiasing strategy, no additional input is required from forecasters and they can finish the forecasting process in the same way as using the traditional FSSs. Therefore, the ease of use of redesigned FSS will not be significantly influenced by the adoption of the debiasing model.

7.4 Limitations of this study

In the process of accomplishing the research objective and answering the research questions, several limitations of this study have been identified. First, this study focuses on the debiasing of five cognitive biases: desire bias, overconfidence bias, and anchoring bias on statistical forecast, desired outcome, and the latest observation. These cognitive biases were targeted because they are commonly identified in empirical studies on judgmental forecasting. However, the empirical studies found in the literature are few:

only 55 articles have focused on, examined, or mentioned cognitive bias, and only seven were conducted in the field of tourism demand forecasting. Therefore, there is a risk that the five cognitive biases studied in this research may not be the most commonly identified in judgmental forecasting of tourism demand.

Furthermore, one of the findings in this study is that desire bias occurs jointly with conservatism bias, which contradicts previous literature. The current study explains the possible reason for such a result, but another possible reason for this finding is that the sample of cases with desire bias detected is relatively small. The valid cases for hypotheses testing regarding desire bias represent less than half of the whole sample (680 cases). Moreover, the problem of small sample size commonly arose in the testing of the hypotheses related to desired outcome, which may also be the reason why H6e was rejected.

Another finding generated from this study is that overconfidence is negatively correlated with forecasters' experience. The data for hypotheses testing were collected from a two-round laboratory experiment, which means that the evidence to support this conclusion comes from only two rounds of forecast. Because of this, the dynamic of the relationship between overconfidence and forecasters' experience is unexplored. For example, it is not clear whether this negative relationship is the strongest in the short term or whether the relationship would become positive in the long term. A long-term observation covering several rounds of forecasting would be better to further test such a relationship.

The statistical forecasts produced in this study follow the ETS approach. One of the conclusions about anchoring bias on statistical forecast is that anchoring on statistical forecast performs as poorly as anchoring on other targets (the latest observations or desired outcomes). However, such a finding may not be valid if other statistical forecasting methods are adopted to produce the baseline forecast of tourism demand. As concluded in Song and Li's (2008) review study, no statistical forecasting method can outperform others in all scenarios in tourism demand forecasting. Therefore, it can be assumed that the findings generated from this study are valid when the baseline forecast is produced by the ETS approach; however, the findings would probably change if other methods were adopted to produce the baseline forecast.

Finally, overconfidence bias, as one of the commonly identified cognitive biases in the literature, was treated as one of the debiasing tasks in this study; both PF and system-suggested forecasts were applied to reduce the influence of overconfidence bias on forecast accuracy. The findings reveal that both debiasing strategies failed to help forecasters improve their forecast accuracy. However, it is difficult to conclude definitively that these two debiasing strategies are ineffective in reducing overconfidence bias in judgmental forecasting of tourism demand because the current study only examined the effectiveness of one algorithm for each debiasing strategy; other ways to produce PF or system-suggested forecasts need to be tested before a reliable conclusion can be reached.

7.5 Future research

More review studies in the area of cognitive bias in judgmental forecasting and FSS development are urgently needed, for two reasons. First, extended literature review in this area is necessary in order to confirm whether the five cognitive biases studied in this study are really the most commonly identified in judgmental forecasting of tourism demand. If new cognitive biases are identified, or other cognitive biases in Arnott's (2006) taxonomy are identified as commonly occurring in the field of tourism demand forecasting, further research on the debiasing of the newly identified cognitive biases can be conducted following the same approach as proposed in this study. In order to further improve the conceptual framework and the application model, an extended literature review of cognitive bias from the perspectives of DSS development and user experience is also required.

In this study, cognitive bias is classified according to Arnott's (2006) taxonomy. Previous studies also reveal that cognitive bias is sourced by different cognitive heuristics, such as availability, anchoring effect, and representativeness (Arnott, 2006; Hogarth, 1987; Remus & Kottemann, 1986; Tversky & Kahneman, 1974). Therefore, the way in which forecasters process cognitive heuristics may influence their cognitive bias in judgmental forecasting. However, cognitive bias detection and debiasing remain unexplored from this perspective. More efforts in the future should be made to identify and measure the relationship between cognitive bias and cognitive heuristics in judgmental forecasting and how to improve the design of TDFSS with better support for users' cognitive heuristic approach. Based on the results of the current study, another interesting topic in future study will be how to incorporate forecasters' cognitive heuristics into the conceptual framework and the application model.

Regarding the data collected from the two-round experiment, the sample with desire bias or desire anchor detected is relatively small, which is considered to threaten conclusion validity. Only 29 valid cases were identified to test hypothesis H6e, which would be the main reason why this hypothesis was rejected. In the future, more studies should focus on the identification and debiasing of cognitive bias related to forecasters' desired outcome.

As mentioned in Section 7.4, one of the limitations in this study is the adoption of a single statistical forecasting approach. The statistical forecasts, as the baseline forecasts in this study, were generated by the ETS approach. Compared with other statistical forecasting methods, the ETS approach has its advantages in terms of automatic modeling and forecasting, easy understanding, and strong forecasting of trends and seasonality. However, as discussed by two review studies in tourism demand forecasting, no single method can be the best in all scenarios (Song & Li, 2008; Witt & Witt, 1995). The conceptual framework and the application model proposed in this study can be re-examined in the future using different statistical forecasting methods to produce baseline forecasts.

The results of this study reveal that both PF and system-suggested forecasts are ineffective in reducing forecasters' overconfidence bias in judgmental forecasting. Considering these two debiasing strategies, the mean forecast error associated with a particular cognitive bias is used as the information for PF, and an algorithm to calculate system-suggested forecasts is also proposed. However, the information used in PF is not unique, nor is the algorithm used to calculate system-suggested forecasts. For example, the mean forecast error that a forecaster makes regarding one specific cognitive bias can also be used in PF. Besides PF and system-suggested forecasts, other debiasing strategies may also be effective in reducing forecasters' cognitive bias. Therefore, further research should focus on new PF content, new algorithms to produce system-suggested forecasts, and new debiasing strategies to deal with forecasters' cognitive bias, especially overconfidence bias.

7.6 Summary

This chapter concludes the current study. It began with an overview summarizing the key aspects of each chapter. Research gaps discovered in the literature review are now filled by the exploration of judgmental forecasting of tourism demand from the perspective of cognitive bias and the innovative conceptual framework. The generalizability of the research approach is also a theoretical contribution of this study. Both TDFSS developers and system users will benefit from this study in different ways. However, no research can be perfect, lacking any limitation for further improvement. The last two parts of this chapter summarized the limitations of this study and proposed directions for further study based on the current research. It is hoped that studying cognitive bias and debiasing in the design of TDFSS will continuously contribute to both theory and industry practice.

Appendix 1 Instruction of the experiment (English)

Thank you for agreeing to participate in this study on tourism demand judgmental forecasting. Please consider the following points when/before completing the online survey:

1. The attitude statement

1) The online survey includes two sessions simulating **two consecutive years of 2006 and 2007**. In the first session, assuming that you are a manager from a company's marketing department at the beginning of 2006. Based on the historical data of tourism demand and your judgment, you need to predict the tourism demand for specific Destination-Origin (D-O) pairs for the coming year 2006 (one-step ahead forecast). In this survey, tourism demand is measured by total tourist arrivals and overall there are **10 destination-origin (D-O) pairs** for which the demand of tourism need to be forecasted. For each D-O pair, the historical data of tourist arrivals is collected from the World Tourism Organization (UNWTO). The length of the historical data varies among different D-O pairs, which is dependent upon the availability of the data. For example, the historical data on tourist arrivals from Chinese Taipei to Hong Kong SAR were collected for the period 1995-2005; while tourist arrivals from China to Japan were collected for 1997-2005. The process of judgmental forecasting in the first session is conducted in **three steps**. The **historical data of tourist arrivals** for each D-O pair and **statistical forecasts** are the only references provided for your information.

2) In the second session, assuming that year 2006 was passed and you are at the beginning of 2007. The real outcome of tourist arrivals in 2006 for each D-O pair is available now and becomes the latest observation of the historical series of tourist arrivals. You are asked to judgmentally forecast the tourist arrivals of the 10 D-O pairs for 2007 (one-step ahead forecast). In the second session, the judgmental forecast is conducted in **six steps** with three types of information provided: the **updated historical data** on tourist arrivals for each D-O pair, the **statistical forecast** and the **suggestive information** based on your forecasting performance from the previous session.

2. Preparation

1) Please visit http://www.tourismforecasting.net/hktdfs/onlinesurvey/language.jsp using the most updated versions of Internet Explorer, Mozilla Firefox, or Google Chrome.

2) Before the survey starts, please select the language that you prefer and input your email address in the next page.

3) Apart from the references provided by the system, please do not refer to any other information but only rely on your own judgment.

3. Forecasting Process (Session 1)

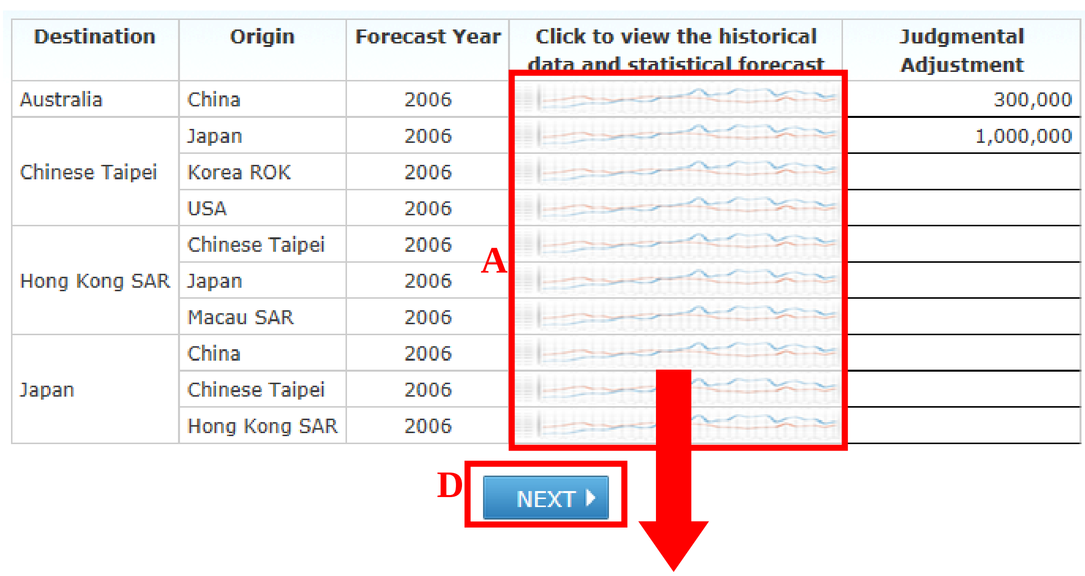
After you have selected the language and inputted your invitation code, the first session of judgmental forecasting (session 1) will start and please provide your judgmental forecasts following the below steps:

1) In the first step of Session 1 (Figure 1), please provide your desired outcome of tourist arrivals in 2006 for any specific D-O pair.

Figure 5. The first step of Session 1

Desired outcome means the number of tourist arrivals that you think would be achieved for a specific D-O pair. First, you need to select the D-O pair which you would like to input the desired outcomes by clicking on column (**A**) of the table; then input the desired outcomes in the corresponding cell in column (**B**). For any of the D-O pairs that you selected in column (**A**), it means that you would like to input a desired outcome (tourist arrivals in 2006). **Please input integers only**. For the D-O pairs that you do not want to input the desired outcomes, please do not select the corresponding cells in column (**A**). After you have provided all desired outcomes, please move to the second step by clicking the "NEXT" button (**C**).

2) In the second step (Figure 2), please provide your forecasts based on the historical data and the statistical forecasts of tourist arrivals for all D-O paired markets.

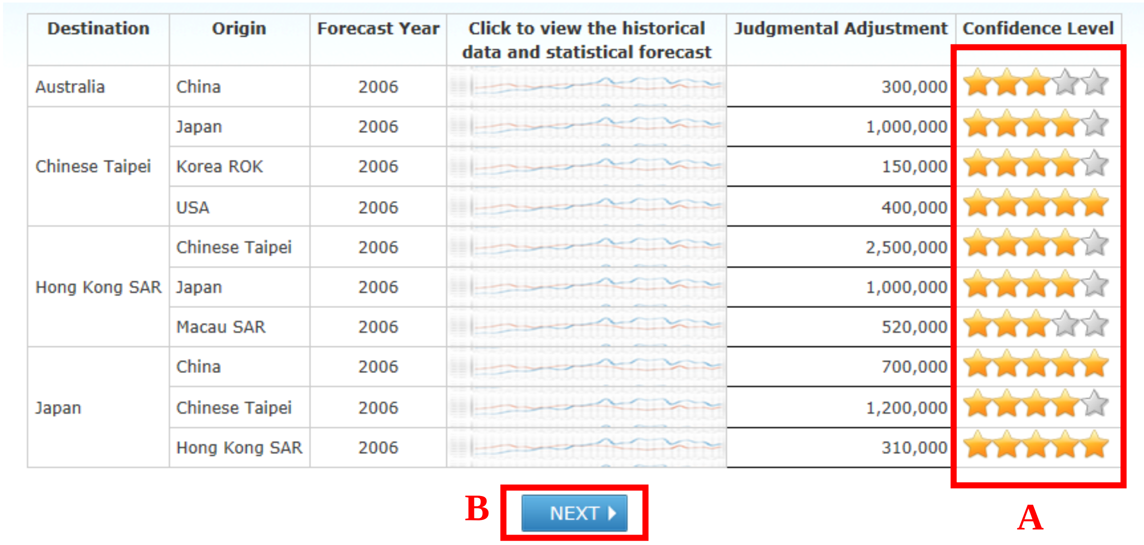


Destination	Origin	Forecast Year	Click to view the historical data and statistical forecast	Judgmental Adjustment
Australia	China	2006		300,000
Chinese Taipei	Japan	2006		1,000,000
	Korea ROK	2006		
	USA	2006		
Hong Kong SAR	Chinese Taipei	2006		
	Japan	2006		
	Macau SAR	2006		
Japan	China	2006		
	Chinese Taipei	2006		
	Hong Kong SAR	2006		

Figure 6. The second step of Session 1

First, you need to click on the images in column (**A**) and view the historical data of tourist arrivals for a specific D-O pair. In the pop up window, both line graph (on the left) and a table of the historical tourist arrivals (on the right) for the chosen D-O pair are provided. The one-step ahead statistical forecast for a certain market is provided below the historical data on the right hand side of the table and is highlighted in red. Please input your judgmental adjustment (**integer only**) based on the statistical forecast in cell (**B**) of the table. After you have provided the judgmental forecast, your input will also appear on the line graph. Please return to the main page by clicking the "DONE" button (**C**) and continue your jugmetnal adjustments for the rest D-O pairs following the same process. You are required to provide judgmental adjsutments for all D-O paired markets and, after that, please move to the third step by clicking the "NEXT" button (**D**).

3) In the third step (Figure 3), please provide the confidence level of your forecasts.



Destination	Origin	Forecast Year	Click to view the historical data and statistical forecast	Judgmental Adjustment	Confidence Level
Australia	China	2006		300,000	★★★☆☆
Chinese Taipei	Japan	2006		1,000,000	★★★★☆
	Korea ROK	2006		150,000	★★★★☆
	USA	2006		400,000	★★★★★
Hong Kong SAR	Chinese Taipei	2006		2,500,000	★★★★☆
	Japan	2006		1,000,000	★★★★☆
	Macau SAR	2006		520,000	★★★☆☆
Japan	China	2006		700,000	★★★★★
	Chinese Taipei	2006		1,200,000	★★★★☆
	Hong Kong SAR	2006		310,000	★★★★★

Figure 7. The third step of Session 1

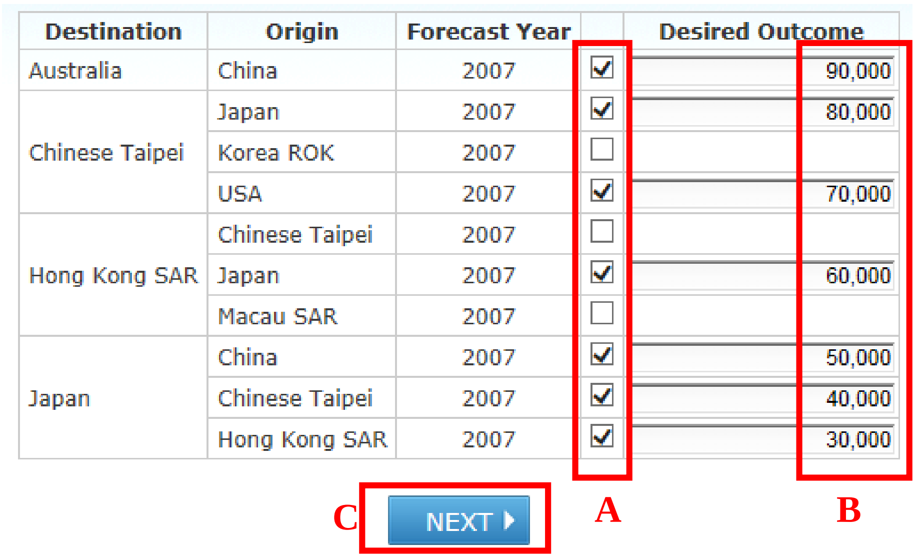
Confidence level of your forecasts in this step is measured by five-point Likert scale: one star in the last column of the above table indicates **Very unconfident** and five star indicates **Very confident** (**A**). You are required to click on the stars to identify your confidence levels with all of the forecasts. After you have provided the confidence levels for all judgmental adjustments, please move to the second session of the judgmental forecast by clicking the "NEXT" button (**B**).

4. Forecasting Process (Session 2)

The second session starts with the following page (Figure 4) and please click the "START" button (**A**) to carry out the judgmental forecasts in six steps.

Figure 8. Start of Session 2

1) In the first step of Session 2 (Figure 5), please provide your desired outcome of tourist arrivals in 2007 for any specific D-O pairs.



Destination	Origin	Forecast Year		Desired Outcome
Australia	China	2007	☑	90,000
Chinese Taipei	Japan	2007	☑	80,000
	Korea ROK	2007	☐	
	USA	2007	☑	70,000
Hong Kong SAR	Chinese Taipei	2007	☐	
	Japan	2007	☑	60,000
	Macau SAR	2007	☐	
Japan	China	2007	☑	50,000
	Chinese Taipei	2007	☑	40,000
	Hong Kong SAR	2007	☑	30,000

Figure 9. The first step of Session 2

The first step of Session 2 is the same as the first step of Session 1, you need to select the D-O pairs that you would like to input desired outcomes by clicking on the cells in column (**A**) of the table and input the desired outcomes in the corresponding cell in collumn (**B**). After you have provided all the desired outcomes, please move to the second step by clicking the "NEXT" button (**C**).

2) In the second step (Figure 6), please provide your adjustments on the desired outcomes based on the system generated statement.

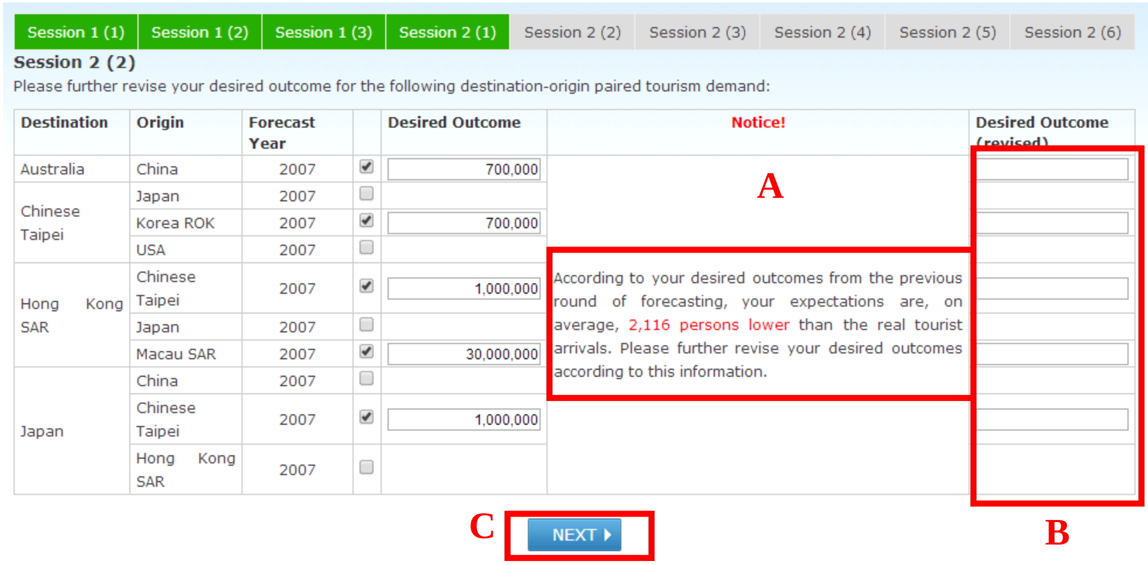
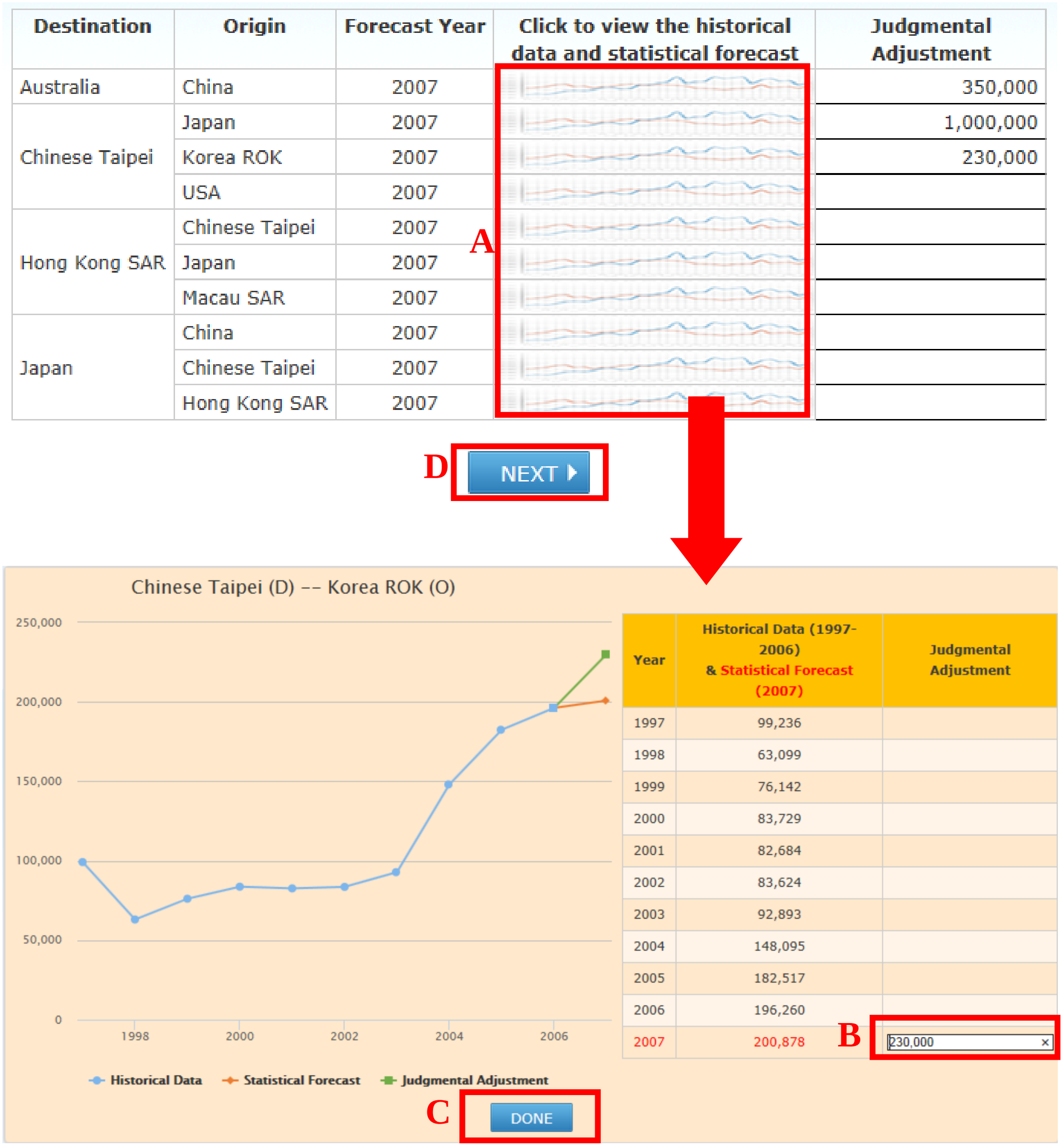


Figure 10. The second step of Session 2

In column (**A**) of the above table, the system provides information about your performance in the previous round of forecasting. For example, as shown in Figure 6, the statement suggests that your desired outcomes in 2006 for the 10 D-O pairs are, on average, **2,116 persons lower than the real tourist arrivals**.

According to the desired outcomes of tourist arrivals in 2007 and the system generated statement, please revise the desired outcomes in column (**B**). For any D-O pair that you do not want to change the desired outcome, you can **keep the corresponding cell in blank**. After you have revised the desired outcomes, please move to the next step by clicking the "NEXT" button (**C**).

3) In the third step (Figure 7), please provide your forecasts based on the historical data and the statistical forecasts of tourist arrivals for all D-O pairs.



Destination	Origin	Forecast Year	Click to view the historical data and statistical forecast	Judgmental Adjustment
Australia	China	2007		350,000
Chinese Taipei	Japan	2007		1,000,000
	Korea ROK	2007		230,000
	USA	2007		
Hong Kong SAR	Chinese Taipei	2007		
	Japan	2007		
	Macau SAR	2007		
Japan	China	2007		
	Chinese Taipei	2007		
	Hong Kong SAR	2007		

Year	Historical Data (1997-2006) & Statistical Forecast (2007)	Judgmental Adjustment
1997	99,236	
1998	63,099	
1999	76,142	
2000	83,729	
2001	82,684	
2002	83,624	
2003	92,893	
2004	148,095	
2005	182,517	
2006	196,260	
2007	200,878	230,000

Figure 11. The third step of Session 2

The third step of Session 2 is the same as the second step of Session 1. Based on the historical data and the statistical forecasts of tourist arrivals in 2007, you are required to provide forecasts for all D-O pairs by clicking on the images in the fourth column of the table (**A**) and input the adjustments in the pop up cell (**B**). Please click the "DONE" button in the pop up window to finish the adjustment for each of the D-O pairs (**C**) and click the "NEXT" button (**D**) to move to the next step.

4) In the fourth step (Figure 8), please further revise your forecasts based on the system generated statement.

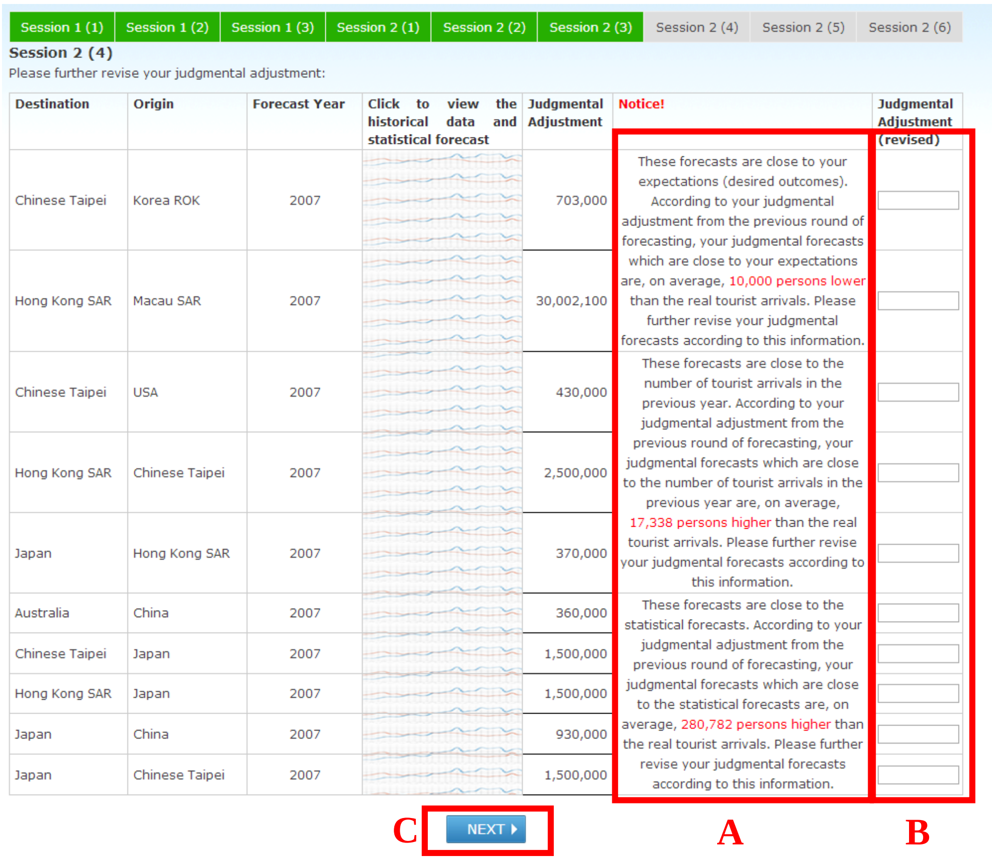


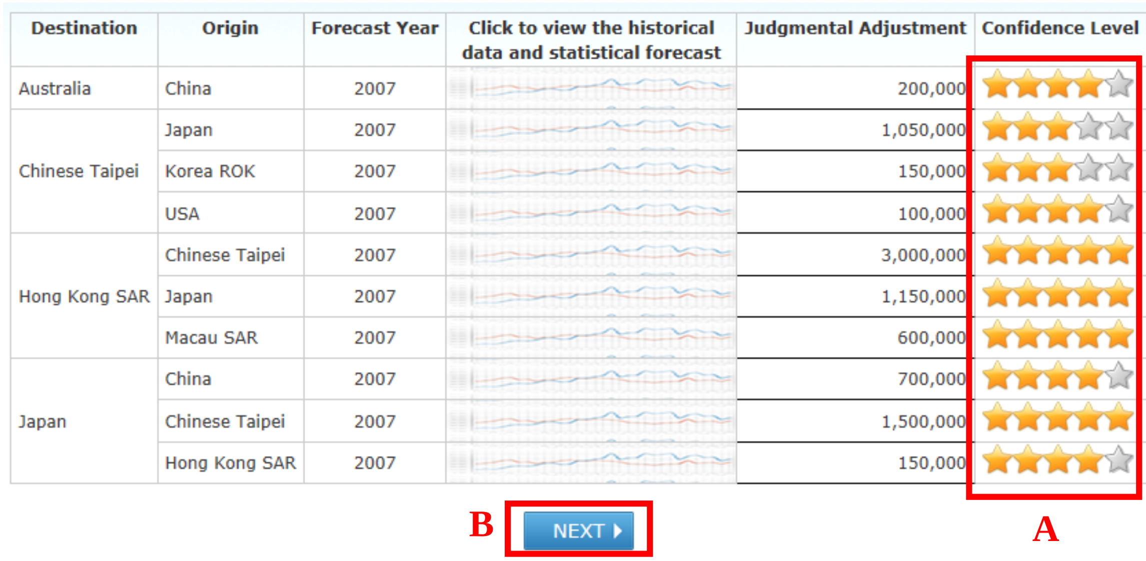
Figure 12. The fourth step of Session 2

In this step, the system provides the potential anchoring bias involved in your forecasts. Anchoring bias indicates that the adjustments from an initial position (anchor) is usually insufficient. In that case, anchoring bias mainly contributes to the forecast error when forecaster anchors on an initial point of forecast. The initial point of forecast can be the statistical forecast of tourist arrivals, the real outcome of tourist arrivals in last year, or your desired forecast, which are named as **statistical forecast anchor**, **latest observation**

anchor, and **desire anchor**, respectively. In this study, these three types of anchors are detected by the system according to your forecasts in 2006; and your judgmental forecasts are categorized in different groups according to the potential anchoring bias detected. For each group, the system provides information about your performance in the previous round of forecasting when the corresponding anchoring bias occurred (**A**). For example, as shown in Figure 8, the first two judgmental forecasts are categorized in the group with desire anchor detected. The statement suggests that your forecasts with desire anchor in the previous round of forecasting are, on average, **10,000 persons lower than the real tourist arrivals**.

According to your judgmental forecasts of tourist arrivals in 2007 and the system generated statements, please further revise your forecasts in the last column of the table (**B**). If you do not want to revise your forecast for any D-O pairs, please **keep the corresponding cell blank**. After you have finished the revision of your forecasts, please move to the next step by clicking the “NEXT” button (**C**).

5) In the fifth step (Figure 9), please provide the confidence level of your forecasts.



Destination	Origin	Forecast Year	Click to view the historical data and statistical forecast	Judgmental Adjustment	Confidence Level
Australia	China	2007		200,000	
Chinese Taipei	Japan	2007		1,050,000	
	Korea ROK	2007		150,000	
	USA	2007		100,000	
Hong Kong SAR	Chinese Taipei	2007		3,000,000	
	Japan	2007		1,150,000	
	Macau SAR	2007		600,000	
Japan	China	2007		700,000	
	Chinese Taipei	2007		1,500,000	
	Hong Kong SAR	2007		150,000	

Figure 13. The fifth step of Session 2

Following the same process in the third step of Session 1, you are required to identify your confidence levels for all of your forecasts (**A**). After that, please move to the next step by clicking the "NEXT" button (**B**).

6) In the sixth step (Figure 10), please further revise your forecasts according to the system generated statement.

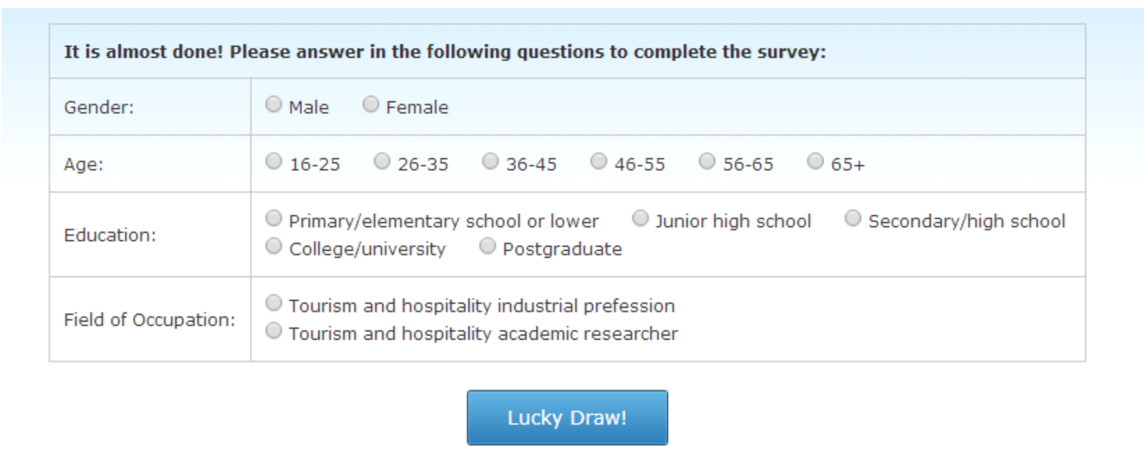
Destination	Origin	Forecast Year	Click to view the historical data and statistical forecast	Judgmental Adjustment	Confidence Level	Notice!	Judgmental Adjustment (revised)
Australia	China	2007		360,000		According to your judgmental adjustment from the previous round of forecasting, your levels of confidence have resulted your judgmental forecasts, on average, 30,489 persons lower than the real tourist arrivals. Please further revise your judgmental forecasts according to this information.	
Chinese Taipei	Japan	2007		1,500,000			
	Korea ROK	2007		703,000			
	USA	2007		430,000			
Hong Kong SAR	Chinese Taipei	2007		2,500,000			
	Japan	2007		1,500,000			
	Macau SAR	2007		30,002,100			
Japan	China	2007		930,000			
	Chinese Taipei	2007		1,500,000			
	Hong Kong SAR	2007		370,000			

Figure 14. The sixth step of Session 2

Column (**A**) of the table displays the mean forecast error reflected by the confidence level of your forecasts of tourist arrivals in 2006. For example, as shown in Figure 10, the forecaster's level of confidence have resulted the judgmental forecasts, on average, **30,489 persons lower than the real tourist arrivals**.

According to the mean forecast error of over-adjustment or under-adjustment, please further revise your forecasts in the last column (**B**) of the table. If you do not want to revise the forecast for any of the D-O pairs, please **keep the corresponding cell blank**. After you revised your forecasts, please move to the final step by clicking the "NEXT" button (**C**).

7) In the final step, please answer the following questions and complete the survey by clicking the "FINISH" button.



It is almost done! Please answer in the following questions to complete the survey:	
Gender:	Male Female
Age:	16-25 26-35 36-45 46-55 56-65 65+
Education:	Primary/elementary school or lower Junior high school Secondary/high school College/university Postgraduate
Field of Occupation:	Tourism and hospitality industrial prefession Tourism and hospitality academic researcher

Lucky Draw!

Figure 15. The final step of the online survey

Thank you for your participation! Please be assured that the information you have provided will be used for academic research only and will be kept strictly confidential.

Appendix 2 Instruction of the experiment (Chinese)

非常感谢您的参与以及对本次问卷调查中做出的贡献。本次问卷调查是一项对旅游市场需求进行主观预测的模拟实验，请根据以下说明完成一次以在线问卷调查的形式进行的主观预测。

1. 前提假设

1) 本次问卷调查分为两个阶段模拟现实中的旅游需求预测，第一阶段假设在 2006 年初进行，而第二阶段假设在 2007 年初进行。您是某公司的市场部负责人，面对一个旅游需求预测系统，根据系统提供的信息以及您自己的主观判断分别在这两年对不同旅游目的地和客源地的市场需求量进行连续两年的主观预测。在本次模拟中，旅游需求量以年度游客人数为测量指标，以 10 个特定旅游目的地——客源地 (D-O) 市场为单位，进行连续两年的年度游客人数预测。所使用的游客人数历史数据均来自于联合国世界旅游组织 (UNWTO)。由于该组织提供的历史数据时间分布不同，本次模拟所使用的 D-O 市场游客人数历史数据区间也不尽相同。例如，在本模拟中提供的中国台北到香港特别行政区的年度游客人数历史区间为 1995-2005，而数据提供的中国到日本的年度游客人数历史区间为 1997-2005。

2) 在第一阶段模拟中，主观预测分为三个步骤进行；10 个 D-O 市场的年度历史游客人数、统计预测值是仅有的可供您参考的系统资料。在第二阶段模拟中，主观预测分为六个步骤。2006 年的年度实际产生的游客人数将作为已知的历史数据提供给您作为 2007 年各 D-O 市场年度游客人数预测的参考资料。此外，系统将提供有关每个 D-O 市场的统计预测值以及您在 2006 年度预测的准确度等提示信息，进一步支持您进行 2007 年度游客人数预测。

2. 准备工作

1) 请使用最新版本的 IE 浏览器 或 Google Chrome 浏览器访问如下网页：
http://www.tourismforecasting.net/hktdfs/onlinesurvey/language.jsp

2) 本次模拟提供中文和英文两种备选语言，在模拟开始之前请选择您偏好的语言。在随后的页面中，请填写您的邮箱地址。

3) 除本次模拟中预测系统提供的参考信息以外，请不要参考其他资料或他人的建议，请根据您的主观判断独立完成本次模拟预测。

3. 第一阶段预测过程

在选择语言和提供邀请码之后，点击页面下方的下一步按钮，第一阶段模拟预测随即开始。请根据各页面的要求完成三个步骤的主观预测：

1) 请在第一个页面中（图 1）提供各 D-O 市场 2006 年度游客人数所能达到的期望值。

目的地	客源地	预测年份		期望值
澳大利亚	中国	2006	☑	10,000
中国台湾	日本	2006	☑	20,000
	韩国	2006	☑	30,000
	美国	2006	☑	40,000
香港特别行政区	中国台湾	2006	☐	
	日本	2006	☐	
	澳门特别行政区	2006	☑	50,000
日本	中国	2006	☐	
	中国台湾	2006	☑	60,000
	香港特别行政区	2006	☑	70,000

图 1 第一阶段步骤 1

首先，该期望值指的是根据你对该 D-O 市场的了解，您觉得该市场在 2006 年所应达到的游客人数。请在页面表格中的第四列选择您觉得有确切期望值的 D-O 市场 (**A**)，然后在右侧一列的相应位置输入您的期望值 (**B**)。对于在第四列选出的 D-O 市场，其右侧的期望值为必填项（仅限正整数）；而对于您觉得没有确切期望值的 D-O 市场，请不要在第四列勾选，也无需在右侧相应位置输入任何期望值。当期望值输入完毕后，请点击下一步按钮进入本阶段步骤 2。

2) 根据在第二个页面中（图 2）显示的历史数据和统计预测值，请对所有 D-O 市场提供您的主观预测值。

目的地	客源地	预测年份	点击下图浏览历史数据和统计预测值	主观预测值
澳大利亚	中国	2006		300,000
中国台湾	日本	2006		1,000,000
	韩国	2006		230,000
	美国	2006		350,000
香港特别行政区	中国台湾	2006		2,500,000
	日本	2006		1,000,000
	澳门特别行政区	2006		510,000
日本	中国	2006		600,000
	中国台湾	2006		1,200,000
	香港特别行政区	2006		320,000

年份	历史数据 (1997-2005) & 统计预测值 (2006)	主观预测值
1997	905,527	
1998	826,632	
1999	826,222	
2000	916,301	
2001	971,190	
2002	998,497	
2003	657,053	
2004	887,311	
2005	1,124,334	
2006	1,200,861	1,000,000

图 2 第一阶段步骤 2

首先，依次点击页面表格中第四列的图片 (**A**)，在弹出窗口中查看各 D-O 市场的历史年度游客人数及统计预测值。弹出窗口分左右两部分，左侧显示历史数据和统计预测值的折线图，右侧显示数据表格，其中最下一行显示 2006 年统计预测值。请根据页面显示的历史数据、统计预测值及个人判断，将您的主观预测值填入表格右下方的单元格内 (**B**)，完成后点击完成按钮返回主页面 (**C**)。按同样方法点击主页面中表格第四列的其他图

片，并提供相应的主观预测值。本页所有 D-O 市场的主观预测值均为必填项，完成后请点击下一步按钮进入下一步骤 (D)。

3) 请在第三个页面中（图 3）对您的预测值进行信心评价。

目的地	客源地	预测年份	点击下图浏览历史数据和统计预测值	主观预测值	对主观预测值的信心
澳大利亚	中国	2006		300,000	★★★☆☆
中国台湾	日本	2006		1,000,000	★★★★☆
	韩国	2006		230,000	★★★★☆
	美国	2006		350,000	★★★★★
香港特别行政区	中国台湾	2006		2,500,000	★★★★☆
	日本	2006		1,000,000	★★★★☆
	澳门特别行政区	2006		510,000	★★★☆☆
日本	中国	2006		600,000	★★★★★
	中国台湾	2006		1,200,000	★★★★☆
	香港特别行政区	2006		320,000	★★★★★

图 3 第一阶段步骤 3

本模拟中使用 5 度李克特量表测量对预测值的信心水平。如图 3 所示，表格最后一列显示信心水平：一星代表非常没有信心；五星代表非常有信心 (A)。对于所有 D-O 市场预测值的信心评价均为必填项，完成后请点击下一步按钮进入第二阶段预测 (B)。

4. 第二阶段预测过程

在第二阶段的起始页面（图 4），请点击页面下方的开始按钮 (A) 进行六个步骤的主观预测

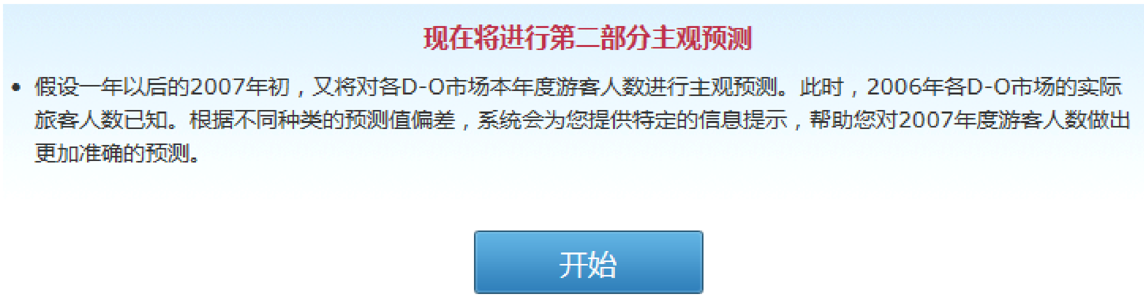


图 4 第二阶段起始页

1) 请在第一个页面中（图 5）提供各 D-O 市场 2007 年度游客人数所能达到的期望值。

目的地	客源地	预测年份		期望值
澳大利亚	中国	2007	☑	90,000
中国台湾	日本	2007	☑	80,000
	韩国	2007	☐	
	美国	2007	☑	70,000
香港特别行政区	中国台湾	2007	☐	
	日本	2007	☑	60,000
	澳门特别行政区	2007	☐	
日本	中国	2007	☑	50,000
	中国台湾	2007	☑	40,000
	香港特别行政区	2007	☑	30,000

图 5 第二阶段步骤 1

与第一阶段步骤 1 相同，请在页面表格的第四列选择您觉得有确切期望值的 D-O 市场 (A)，并在右侧一列的相应位置输入期望值 (B)。对于在第四列选出的 D-O 市场，其右侧的期望值为必填项（仅限正整数）；而对于您觉得没有确切期望值的 D-O 市场，请不要在第四列勾选，也无需在右侧相应位置输入任何期望值。 输入完毕过后请点击下一步按钮进入第二阶段步骤 2。

2) 在步骤 2 中（图 6），请根据系统提示信息进一步修改您的期望值。

图 6 第二阶段步骤 2

在页面表格中(**A**)列，系统提示您在上一次旅游人数预测中期望值的表现。以图 6 为例，系统提示信息显示预测者在 2006 年各 D-O 市场的**期望值平均比实际游客人数低 2,166 人次**。

根据您对各 D-O 市场 2007 年游客人数的期望值和您在 2006 年的期望值表现，请在页面表格中最后一列进一步修改您的期望值 (**B**)。如果您不想进一步调整某 D-O 市场的期望值，可在最后一列对应行的单元格中保持空白即可。期望值调整完毕后请点击下一步按钮进入步骤 3 (**C**)。

3)　根据在第三个页面中（图 7）显示的历史数据和统计预测值，请对所有 D-O 市场提供您的主观预测值。

图 7 第二阶段步骤 3

该步骤与第一阶段步骤 2 相同，请依次点击页面表格中第四列的图片 (**A**)，在弹出窗口中基于历史数据、统计预测值及个人判断输入您的主观预测值 (**B**)，完成后点击完成按钮返回主页面 (**C**)。按同样方法点击主页面中表格第四列的其他图片完成所有 D-O 市场的主观预测后点击下一步按钮进入步骤 4。

4)　根据第四个页面的表格中（图 8）的系统提示，请进一步修改您的主观预测值。

目的地	客源地	预测年份	点击下图浏览历史数据和统计预测值	主观预测值	系统提示！	主观预测值（调整后）
中国台湾	韩国	2007		703,000	这些预测值接近您的期望值。根据以往预测表现，当您的预测值接近期望值的时候，预测值平均比实际旅客人数低10,000人次。请根据此误差进一步调整您的预测值。	
香港特别行政区	澳门特别行政区	2007		30,002,100		
中国台湾	美国	2007		430,000	这些预测值接近上一年旅客人数。根据以往预测表现，当您的预测值接近上一年旅客人数的时候，预测值平均比实际旅客人数高17,338人次。请根据此误差进一步调整您的预测值。	
香港特别行政区	中国台湾	2007		2,500,000		
日本	香港特别行政区	2007		370,000		
澳大利亚	中国	2007		360,000	这些预测值接近统计预测值。根据以往预测表现，当您的预测值接近统计预测值的时候，预测值平均比实际旅客人数高280,782人次。请根据此误差进一步调整您的预测值。	
中国台湾	日本	2007		1,500,000		
香港特别行政区	日本	2007		1,500,000		
日本	中国	2007		930,000		
日本	中国台湾	2007		1,500,000		

图 8 第二阶段步骤 4

该页提供的系统提示信息分为两部分，第一部分的内容为您在 2006 年预测中表现出的锚误差。锚误差是主观预测偏差的一种常见类型，指在某一初始预测值的基础上进行调整时所表现出的调整不足现象。当锚误差出现时，它将是构成最终预测误差的主要因素之一。常见的锚误差有三种情况，分别锚定于统计预测值，上一年历史值，和期望值。三种锚误差分别被命名为**统计锚误差**，**最近实际值锚误差**，和**期望锚误差**。本次模拟中所使用的预测系统会根据您在 2006 年的预测表现自动检查在 2007 年预测中可能出现的锚误差，并将不同类型的锚误差分类并显示于页面表格的第六列 (**A**)。以图 8 为例，预测者在对韩国到中国台湾以及澳门特别行政区到香港特别行政区的两组旅游人数预测中可能出现期望锚误差，并提示用户 2006 年由于期望锚误差导致**主观预测值比实际旅客人数低 10000 人次**。

根据上一年您在主观预测中表现出的不同类型的平均锚误差以及当前主观预测中检测出的潜在锚误差，请进一步调整并在页面表格的最后一列输入您的主观预测值 (**B**)。如果对于某 D-O 市场的主观预测值不想做进一步调整，可在最后一列对应行的单元格中保持空白即可。预测值调整完毕后请点击下一步按钮进入步骤 5 (**C**)。

5)　请在第五个页面中（图 9）对您的预测值进行信心评价。

目的地	客源地	预测年份	点击下图浏览历史数据和统计预测值	主观预测值	对主观预测值的信心
澳大利亚	中国	2007		200,000	★★★★☆
中国台湾	日本	2007		1,050,000	★★★☆☆
	韩国	2007		150,000	★★★☆☆
	美国	2007		60,000	★★★★☆
香港特别行政区	中国台湾	2007		3,000,000	★★★★★
	日本	2007		1,100,000	★★★★★
	澳门特别行政区	2007		580,000	★★★★★
日本	中国	2007		930,000	★★★★☆
	中国台湾	2007		1,300,000	★★★★★
	香港特别行政区	2007		300,000	★★★★☆

图 9 第二阶段步骤 5

与第一阶段步骤 3 相同，请评价您对所有 D-O 市场预测值的信心水平 (**A**)，并在完成后点击下一步按钮进入步骤 6 (**B**)。

6)　请根据第六个页面中的系统提示（图 10）进一步调整主观预测值。

目的地	客源地	预测年份	点击下图浏览历史数据和统计预测值	主观预测值	对主观预测值的信心	系统提示！	主观预测值（调整后）
澳大利亚	中国	2007		360,000		根据以往预测表现，您对预测值的信心水平导致预测值平均比实际旅客人数低30,489人次。请根据此误差进一步调整您的期望值。	
中国台湾	日本	2007		1,500,000			
	韩国	2007		703,000			
	美国	2007		430,000			
香港特别行政区	中国台湾	2007		2,500,000			
	日本	2007		1,500,000			
	澳门特别行政区	2007		30,002,100			
日本	中国	2007		930,000			
	中国台湾	2007		1,500,000			
	香港特别行政区	2007		370,000			

图 10 第二阶段步骤 6

页面表格中第七列显示的系统提示内容是您在 2006 年预测中表现出的过度自信（或缺乏自信）而导致的平均预测误差 (**A**)。以图 10 为例，系统提示信息表明预测者在 2006 年主观预测中表现出的信心水平导致**预测值平均比实际游客人数低 30,489 人次**。

根据您在 2006 年预测中由于信心水平导致的平均预测误差，请在页面表格中最后一列进一步调整主观预测值 (**B**)。如果对于某 D-O 市场的主观预测值不想做进一步调整，可在最后一列对应行的单元格中保持空白即可。预测值调整完毕后请点击下一步按钮进入本次模拟的最后一个环节分类信息采集。

7) 请填写如下分类信息并点击完成按钮结束本次问卷调查。

感谢您的参与，请填写如下信息以完成本次问卷调查：	
性别：	○男 ○女
年龄：	○16-25 ○26-35 ○36-45 ○46-55 ○56-65 ○65+
教育背景：	○小学及以下 ○初中 ○高中 ○大学本科 ○研究生及以上
所属领域：	○旅游及酒店管理行业从业人 ○旅游及酒店管理研究人员

完成

图 11 分类信息

您所提供的所有信息将仅用于科研目的，我们会对您提供的所有信息严格保密。再次感谢您的参与！

REFERENCES

Adya, M., Collopy, F., Armstrong, J. S., & Kennedy, M. (2001). Automatic identification of time series features for rule-based forecasting. *International Journal of Forecasting*, *17*(2), 143–157.

Allison, P. D. (2001). *Missing Data*. Thousand Oaks, CA: SAGE Publications.

Ankomah, P. K., & Crompton, J. L. (1992). Tourism cognitive distance: A set of research propositions. *Annals of Tourism Research*, *19*(2), 323–342.

Archer, B. H. (1987). Demand forecasting and estimation. In J. R. B. Ritchie & C. R. Goeldner, *Travel, tourism and hospitality research* (pp. 77–85). New York: Wiley.

Arkes, H. R. (2001). Overconfidence in judgmental forecasting. In J. S. Armstrong, *Principles of forecasting* (pp. 495–515). Norwell, MA: Kluwer Academic Publishers. Retrieved from http://link.springer.com/chapter/10.1007/978-0-306-47630-3_22

Armor, D. A., & Taylor, S. E. (2002). When Predictions Fail: The Dilemma of Unrealistic Optimism. In T. Gilovich, D. Griffin, & D. Kahneman, *Heuristics and Biases* (pp. 334–347). Cambridge, UK: Cambridge University Press. Retrieved from http://dx.doi.org/10.1017/CBO9780511808098.021

Armstrong, J. S. (2001). *Principles of forecasting: A handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers.

Armstrong, J. S., & Collopy, F. (1994). How serious are methodological issues in surveys? A reexamination of the Clarence Thomas Polls. Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.37.2861&rep=rep1&type=pdf

Arnott, D. (1998). *A taxonomy of decision biases* (Technical Report No. 1/98). Melbourne, Australia: School of Information Management & Systems Monash University.

Arnott, D. (2006). Cognitive biases and decision support systems development: A design science approach. *Information Systems Journal*, *16*(1), 55–78.

Ashton, R. H. (1990). Pressure and performance in accounting decision settings: Paradoxical effects of incentives, feedback, and justification. *Journal of Accounting Research*, *28*, 148–180.

Asimakopoulos, S. (2008). *A human-computer interaction perspective on forecasting systems design* (Ph.D., School of Management Service). Lancaster University, United Kingdom -- Lancaster.

Asimakopoulos, S., & Dix, A. (2013). Forecasting support systems technologies-in-practice: A model of adoption and use for product forecasting. *International Journal of Forecasting*, *29*(2), 322–336.

Athanasopoulos, G., Hyndman, R. J., Song, H., & Wu, D. C. (2011). The tourism forecasting competition. *International Journal of Forecasting*, *27*(3), 822–844. doi:10.1016/j.ijforecast.2010.04.009

Aukutsionek, S. P., & Belianin, A. V. (2001). Quality of forecasts and business performance: A survey study of Russian managers. *Journal of Economic Psychology*, *22*(5), 661–692.

Ayton, P., Pott, A., & Elwakili, N. (2007). Affective forecasting: Why can't people predict their emotions? *Thinking & Reasoning*, *13*(1), 62–80.

Baggio, R., & Corigliano, M. A. (2008). A practical forecasting method for a tourism organization. In *Economic Contribution at Sub-National Levels*. Malaga, Spain. Retrieved from http://www.inroutenetwork.org/attachments/036_4.2.d%20A%20practical%20forecasting%20method%20for%20a%20tourism%20organization.pdf

Balzer, W. K., Doherty, M. E., & O'Connor, R. (1989). Effects of cognitive feedback on performance. *Psychological Bulletin*, *106*(3), 410–433.

Bar-Hillel, M. (1973). On the subjective probability of compound events. *Organizational Behavior and Human Performance*, *9*(3), 396–406.

Batchelor, R. (2007). Bias in macroeconomic forecasts. *International Journal of Forecasting*, *23*(2), 189–203.

Batchelor, R., & Dua, P. (1992). Conservatism and consensus-seeking among economic forecasters. *Journal of Forecasting*, *11*(2), 169–181.

Batson, C. D. (1975). Rational processing or rationalization? The effect of disconfirming information on a stated religious belief. *Journal of Personality and Social Psychology*, *32*(1), 176.

Bazerman, M. H. (1998). *Judgment in managerial decision making* (4th ed.). New York: Wiley.

Bazerman, M. H., & Moore, D. A. (2008). *Judgment in managerial decision making* (7th ed.). New York: Wiley. Retrieved from http://digitalcommons.usu.edu/unf_research/44/

Benbasat, I., & Dexter, A. S. (1985). An experimental evaluation of graphical and color-enhanced information presentation. *Management Science*, *31*(11), 1348–1364.

Bennett, S. K. (1982). Student perceptions of and expectations for male and female instructors: Evidence relating to the question of gender bias in teaching evaluation. *Journal of Educational Psychology*, *74*(2), 170.

Benson, P. G., & Önkal, D. (1992). The effects of feedback and training on the performance of probability forecasters. *International Journal of Forecasting*, *8*(4), 559–573.

Blackley, P. R., & DeBoer, L. (1993). Bias in OMB's economic forecasts and budget proposals. *Public Choice*, *76*(3), 215–232.

Bolger, F., & Önkal-Atay, D. (2004). The effects of feedback on judgmental interval predictions. *International Journal of Forecasting*, *20*(1), 29–39.

Brannon, L. A., Tagler, M. J., & Eagly, A. H. (2007). The moderating role of attitude strength in selective exposure to information. *Journal of Experimental Social Psychology*, *43*(4), 611–617. doi:10.1016/j.jesp.2006.05.001

Briggs, L. K., & Krantz, D. H. (1992). Judging the strength of designated evidence. *Journal of Behavioral Decision Making*, *5*(2), 77–106.

Brodie, R. J., & De Kluyver, C. A. (1987). A comparison of the short term forecasting accuracy of econometric and naive extrapolation models of market share. *International Journal of Forecasting*, *3*(3), 423–437.

Buehler, R., Messervey, D., & Griffin, D. (2005). Collaborative planning and prediction: Does group discussion affect optimistic biases in time estimation? *Organizational Behavior and Human Decision Processes*, *97*(1), 47–63.

Burger, C., Dohnal, M., Kathrada, M., & Law, R. (2001). A practitioners guide to time-series methods for tourism demand forecasting–a case study of Durban, South Africa. *Tourism Management*, *22*(4), 403–409.

Camerer, C., Babcock, L., Loewenstein, G., & Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time. *The Quarterly Journal of Economics*, *112*(2), 407–441.

Camerer, C., & Johnson, E. J. (1997). The process-performance paradox in expert judgment: How can experts know so much and predict so badly? In W. M. Goldstein & M. H. Robin, *Research on judgment and decision making: Currents, connections, and controversies* (pp. 342–364). Cambridge: Cambridge University Press. Retrieved from http://www.google.com/books?hl=en&lr=&id=9Jo5v25Yn3AC&oi=fnd&pg=PA342&dq=expert+less+overconfidence&ots=eT-Zvcvp-k&sig=CCmG-rtT71OEHxyGx4sFuw1KxVQ

Camerer, C., Loewenstein, G., & Rabin, M. (2004). *Advances in Behavioral Economics*. New York: Russell Sage Foundation.

Carter, C. R., Kaufmann, L., & Michel, A. (2007). Behavioral supply management: A taxonomy of judgment and decision-making biases. *International Journal of Physical Distribution & Logistics Management*, *37*(8), 631–669.

Chapman, L. J., & Chapman, J. P. (1967). Genesis of popular but erroneous psychodiagnostic observations. *Journal of Abnormal Psychology*, *72*(3), 193–204.

Clements, M. P., & Hendry, D. F. (2008). *A companion to economic forecasting*. Malden, USA: Blackwell Publishing Ltd.

Cohen, J., Chesnick, E. I., & Haran, D. (1972). A confirmation of the inertial -ψ effect in sequential choice and decision. *British Journal of Psychology*, *63*(1), 41–46.

Collopy, F., & Armstrong, J. S. (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Management Science*, *38*(10), 1394–1414.

Coshall, J. T. (2009). Combining volatility and smoothing forecasts of UK demand for international tourism. *Tourism Management*, *30*(4), 495–511. doi:10.1016/j.tourman.2008.10.010

Croce, V., & Wöber, K. W. (2011). Judgemental forecasting support systems in tourism. *Tourism Economics*, *17*(4), 709–724.

Cunliffe, S. (2002). Forecasting risks in the tourism industry using the Delphi technique. *Tourism*, *50*(1), 31–41.

De Menezes, A. G., & Vieira, J. C. (2008). Willingness to pay for airline services attributes: Evidence from a stated preferences choice game. *European Transport*, *39*(1), 1–13.

DuCharme, W. M. (1970). Response bias explanation of conservative human inference. *Journal of Experimental Psychology*, *85*(1), 66.

Du, N., & Budescu, D. V. (2007). Does past volatility affect investors' price forecasts and confidence judgements? *International Journal of Forecasting*, *23*(3), 497–511.

Easley, D., & Kleinberg, J. (2010). *Networks, crowds, and markets: Reasoning about a highly connected world* (Vol. 8). Cambridge, UK: Cambridge University Press.

Ebert, R. J. (1972). Environmental structure and programmed decision effectiveness. *Management Science*, *19*(4), 435–445.

Egeth, H. (1967). Selective attention. *Psychological Bulletin*, *67*(1), 41.

Eggleton, I. R. (1982). Intuitive time-series extrapolation. *Journal of Accounting Research*, *20*(1), 68–102.

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgment and choice. *Journal of Accounting Research*, *19*(1), 1–31.

Enders, C. K. (2006). A primer on the use of modern missing-data methods in psychosomatic medicine research. *Psychosomatic Medicine*, *68*(3), 427–436.

English, J. M., & Kernan, G. L. (1976). The prediction of air travel and aircraft technology to the year 2000 using the Delphi method. *Transportation Research*, *10*(1), 1–8.

Ernst, B. (2013). *The role of errors and feedback for learning and decision making* (Unpublished Ph.D. Dissertation). Department of Psychology, the University of Konstanz, Konstanz, Germany. Retrieved from http://kops.ub.uni-konstanz.de/handle/urn:nbn:de:bsz:352-243593

Eroglu, C. (2006). *An investigation of accuracy, learning and biases in judgmental adjustments of statistical forecasts* (Unpublished Ph.D. Dissertation). The Graduate School of The Ohio State University, Columbus, OH. Retrieved from http://rave.ohiolink.edu/etdc/view?acc_num=osu1150398313

Eroglu, C., & Croxton, K. L. (2010). Biases in judgmental adjustments of statistical forecasts: The role of individual differences. *International Journal of Forecasting*, *26*(1), 116–133.

Ferguson, N. (2008). *The ascent of money: A financial history of the world*. London: Penguin Press.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). London: SAGE.

Field, A., & Hole, G. (2003). *How to design and report experiments* (7th ed.). London: Sage publications Limited.

Fildes, R., & Beard, C. (1992). Forecasting systems for production and inventory control. *International Journal of Operations & Production Management*, *12*(5), 4–27.

Fildes, R., Goodwin, P., & Lawrence, M. (2006). The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, *42*(1), 351–361.

Fildes, R., Goodwin, P., Lawrence, M., & Nikolopoulos, K. (2009). Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting*, *25*(1), 3–23.

Fischer, G. W. (1982). Scoring-rule feedback and the overconfidence syndrome in subjective probability forecasting. *Organizational Behavior and Human Performance*, *29*(3), 352–369.

Fisher, C., & Malde, S. (2011). Moral imagination or heuristic toolbox? Events and the risk assessment of structured financial products in the financial bubble. *Business Ethics: A European Review*, *20*(2), 148–158.

Frechtling, D. (2012). *Forecasting tourism demand, methods and strategies*. Oxford, UK: Butterworth Heinemann.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, *40*(1), 35–42. doi:10.1016/j.socec.2010.10.008

Garrod, B., & Fyall, A. (2005). Revisiting Delphi: The Delphi technique in tourism research. In B. W. Ritchie, P. M. Burns, & C. A. Palmer, *Tourism research methods: Integrating theory with practice* (pp. 85–98). Wallingford, UK: CABI.

Gelardi, A. (2010). Information quantity and order in students' tax research judgements. *Asian Journal of Finance & Accounting*, *2*(1), 25–46.

Gilley, O. W., & Leone, R. P. (1991). A two-stage imputation procedure for item nonresponse in surveys. *Journal of Business Research*, *22*(4), 281–291.

Goldfarb, R. S., Stekler, H. O., & David, J. (2005). Methodological issues in forecasting: Insights from the egregious business forecast errors of late 1930. *Journal of Economic Methodology*, *12*(4), 517–542.

Gönül, M. S., Önkal, D., & Lawrence, M. (2006). The effects of structural characteristics of explanations on use of a DSS. *Decision Support Systems*, *42*(3), 1481–1493.

Goodwin, P. (2000a). Correct or combine? Mechanically integrating judgmental forecasts with statistical methods. *International Journal of Forecasting*, *16*(2), 261–275.

Goodwin, P. (2000b). Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting*, *16*(1), 85–99.

Goodwin, P. (2002). Integrating management judgment and statistical methods to improve short-term forecasts. *Omega*, *30*(2), 127–135.

Goodwin, P. (2005). How to integrate management judgment with statistical forecasts. *Foresight: The International Journal of Applied Forecasting*, *2005*(1), 8–12.

Goodwin, P., & Fildes, R. (1999). Judgmental forecasts of time series affected by special events: Does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making*, *12*(1), 37–53.

Goodwin, P., & Fildes, R. (2001). Tracing the process of using a forecasting support system. Presented at the Twenty-First International Symposium of Forecasting, Atlanta.

Goodwin, P., Fildes, R., Lawrence, M., & Nikolopoulos, K. (2007). The process of using a forecasting support system. *International Journal of Forecasting*, *23*(3), 391–404.

Goodwin, P., Fildes, R., Lawrence, M., & Stephens, G. (2011). Restrictiveness and guidance in support systems. *Omega*, *39*(3), 242–253.

Goodwin, P., Lee, J. K., Fildes, R., Nikolopoulos, K., & Lawrence, M. (2007). Understanding the use of forecasting systems: An interpretive study in a supply-chain company. *Bath University Management School Working Paper, Working paper*.

Goodwin, P., Sinan Gönül, M., & Önkal, D. (2013). Antecedents and effects of trust in forecasting advice. *International Journal of Forecasting*, *29*(2), 354–366.

Gregor, S., & Benbasat, I. (1999). Explanations from intelligent systems: Theoretical foundations and implications for practice. *MIS Quarterly*, 497–530.

Haag, G., Häggman, J., & Mattsson, J. (2010). *Currency Trading in the FX market: Will spectral analysis improve technical forecasting?*. JÖNKÖPING, Sweden. Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2:325611

Hageman, A. M. (2010). The role of confidence in tax return preparation using tax software. *Advances in Accounting Behavioral Research*, *13*(1), 31–57.

Hair, J. F., Tatham, R. L., Anderson, R. E., & Black, W. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall. Retrieved from http://library.wur.nl/WebQuery/clc/1809603

Hartvigsen, G. (1992). Limitations of knowledge-based systems for financial analysis in banking. *Expert Systems with Applications*, *4*(1), 19–32.

Harvey, N. (2007). Use of heuristics: Insights from forecasting research. *Thinking & Reasoning*, *13*(1), 5–24.

Harvey, N., & Bolger, F. (1996). Graphs versus tables: Effects of data presentation format on judgemental forecasting. *International Journal of Forecasting*, *12*(1), 119–137.
Harvey, N., & Harries, C. (2004). Effects of judges' forecasting on their later combination of forecasts for the same outcomes. *International Journal of Forecasting*, *20*(3), 391–409.
Hevner, A., & Chatterjee, S. (2010). Design science research frameworks. In A. Hevner & S. Chatterjee, *Design Research in Information Systems: Theory and Practice* (Vol. 22, pp. 23–31). New York: Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-5653-8_3
Hoaglin, D. C., & Iglewicz, B. (1987). Fine-tuning some resistant rules for outlier labeling. *Journal of the American Statistical Association*, *82*(400), 1147–1149.
Hoaglin, D. C., Iglewicz, B., & Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, *81*(396), 991–999.
Hogarth, R. M. (1987). *Judgement and choice: The psychology of decision* (2nd ed.). Chichester, UK: John Wiley & Sons.
Hogarth, R. M., & Makridakis, S. (1981). Forecasting and planning: An evaluation. *Management Science*, *27*(2), 115–138.
Hubers, C., & Lyons, G. (2013). Assessing future travel demand: A need to account for non-transport technologies? *Foresight*, *15*(3), 211–227.
Hyndman, R. J., Akram, M., & Archibald, B. C. (2008). The admissible parameter space for exponential smoothing models. *Annals of the Institute of Statistical Mathematics*, *60*(2), 407–426.
Hyndman, R. J., & Billah, B. (2003). Unmasking the Theta method. *International Journal of Forecasting*, *19*(2), 287–290.
Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*, *27*(3), 1–22.
Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer. Retrieved from http://www.google.com/books?hl=en&lr=&id=GSyzox8Lu9YC&oi=fnd&pg=PA2&dq=forecating+with+exponential+smoothing+:+the+state+space+approach&ots=1p2oEHpFq6&sig=IPiMp1pBCnDSG5bIxLG4rYljdT0
Ingelfinger, J. A., Mosteller, F., Thibodeau, L. A., & Ware, J. H. (1987). *Biostatistics in clinical medicine*. New York: Macmillan Publishing Company.
Jacobsen, D. I. (2002). *Vad, hur och varför?: om metodval i företagsekonomi och andra samhällsvetenskapaliga ämnen*. Lund: Studentlitteratur.
Joram, E., & Read, D. (1996). Two faces of representativeness: The effects of response format on beliefs about random sampling. *Journal of Behavioral Decision Making*, *9*(4), 249–264.
Joyce, E. J., & Biddle, G. C. (1981). Are auditors' judgments sufficiently regressive? *Journal of Accounting Research*, *19*(2), 323–349.
Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.
Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. *Cognitive Psychology*, *3*(3), 430–454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237–251.
Kasper, G. M. (1996). A theory of decision support system design for user calibration. *Information Systems Research*, *7*(2), 215–232.
Keen, P. G. W., & Morton, M. S. S. (1978). *Decision support systems: an organizational perspective*. Addison-Wesley Reading, MA.
Keil, M., Depledge, G., & Rai, A. (2007). Escalation: The role of problem recognition and cognitive bias. *Decision Sciences, 38*(3), 391–421.
Keren, G. (1990). Cognitive aids and debiasing methods: can cognitive pills cure cognitive ills?
Keren, G. (1997). On the calibration of probability judgments: Some critical comments and alternative perspectives. *Journal of Behavioral Decision Making*, *10*(3), 269–278.
Kibedi, G. (1981). Future trends in international tourism. *Tourism Review*, *36*(1), 3–6.
Kirs, P. J., Pflughoeft, K., & Kroeck, G. (2001). A process model cognitive biasing effects in information systems development and usage. *Information & Management*, *38*(3), 153–165.
Koester, R., & Luthans, F. (1979). The impact of the computer on the choice activity of decision makers: A replication with actual users of computerized MIS. *Academy of Management Journal*, *22*(2), 416–422.
Kottemann, J. E., Davis, F. D., & Remus, W. E. (1994). Computer-assisted decision making: Performance, beliefs, and the illusion of control. *Organizational Behavior and Human Decision Processes*, *57*(1), 26–37.
Krueger, J. I. (2003). Return of the ego–Self-referent information as a filter for social prediction: Comment on Karniol (2003). Retrieved from http://psycnet.apa.org/journals/rev/110/3/585/
Kydd, C. T. (1989). Cognitive biases in the use of computer-based decision support systems. *Omega*, *17*(4), 335–344.
Lamont, O. A. (2002). Macroeconomic forecasts and microeconomic forecasters. *Journal of Economic Behavior & Organization*, *48*(3), 265–280.
Landeta, J. (2006). Current validity of the Delphi method in social sciences. *Technological Forecasting and Social Change*, *73*(5), 467–482. doi:10.1016/j.techfore.2005.09.002
Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology*, *32*(2), 311.
Langer, E. J. (1977). The Psychology of Chance. *Journal for the Theory of Social Behaviour*, *7*(2), 185–203.
Lawrence, M. (1983). An exploration of some practical issues in the use of quantitative forecasting models. *Journal of Forecasting*, *2*(2), 169–179.
Lawrence, M., Edmundson, R., & O'Connor, M. (1985). An examination of the accuracy of judgmental extrapolation of time series. *International Journal of Forecasting*, *1*(1), 25–35.
Lawrence, M., Goodwin, P., & Fildes, R. (2002). Influence of user participation on DSS use and decision accuracy. *Omega*, *30*(5), 381–392.
Lee, C.-F., & King, B. (2009). A determination of destination competitiveness for Taiwan's hot springs tourism sector using the Delphi technique. *Journal of Vacation Marketing*, *15*(3), 243–257.

Lee, J. K., & Yum, C. S. (1998). Judgmental adjustment in time series forecasting using neural networks. *Decision Support Systems*, *22*(2), 135–154.

Lee, W. Y., Goodwin, P., Fildes, R., Nikolopoulos, K., & Lawrence, M. (2007). Providing support for the use of analogies in demand forecasting tasks. *International Journal of Forecasting*, *23*(3), 377–390.

Levy, M., & Tasoff, J. (2012). Exponential-growth bias and lifecycle consumption. *Working Paper,* Department of Economics, London School of Economics & Political Science.

Li, G., Song, H., & Witt, S. F. (2005). Recent developments in econometric modeling and forecasting. *Journal of Travel Research, 44*(1), 82.

Lim, J. S., & O'Connor, M. (1994). Systems to support judgemental adjustment of initial forecasts. In *System Sciences* (Vol. 3, pp. 263–271).

Lim, J. S., & O'Connor, M. (1995). Judgemental adjustment of initial forecasts: its effectiveness and biases. *Journal of Behavioral Decision Making*, *8*(3), 149–168.

Liu, J. C. (1988). Hawaii tourism to the year 2000: A Delphi forecast. *Tourism Management*, *9*(4), 279–290.

Lopes, L. L., & Oden, G. C. (1987). Distinguishing between random and nonrandom events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *13*(3), 392–400. doi:http://dx.doi.org/10.1037/0278-7393.13.3.392

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476.

Makridakis, S., Hibon, M., & Moser, C. (1979). Accuracy of forecasting: An empirical investigation. *Journal of the Royal Statistical Society. Series A (General)*, *142*(2), 97–145.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems, 15*(4), 251–266.

Massey, C., Simmons, J. P., & Armor, D. A. (2011). Hope over experience desirability and the persistence of optimism. *Psychological Science, 22*(2), 274–281.

Mathews, B. P., & Diamantopoulos, A. (1989). Judgemental revision of sales forecasts: A longitudinal extension. *Journal of Forecasting*, *8*(2), 129–140. doi:10.1002/for.3980080206

Mathews, B. P., & Diamantopoulous, A. (1990). Judgemental revision of sales forecasts: Effectiveness of forecast selection. *Journal of Forecasting*, *9*(4), 407–415.

Mazanec, J. A. (1986). A decision support system for optimizing advertising policy of a national tourist office model outline and case study. *International Journal of Research in Marketing*, *3*(2), 63–77.

McKercher, B. (1995). The destination-market matrix: A tourism market portfolio analysis model. *Journal of Travel & Tourism Marketing*, *4*(2), 23–40.

McNees, S. K. (1992). The uses and abuses of 'consensus' forecasts. *Journal of Forecasting*, *11*(8), 703–710.

McNeil, B. J., Pauker, S. G., Sox Jr, H. C., & Tversky, A. (1982). On the elicitation of preferences for alternative therapies. *The New England Journal of Medicine*, *306*(21), 1259–1262.

McNeil, D. R. (1977). *Interactive data analysis*. New York: John Wiley. Retrieved from http://cds.cern.ch/record/233214

Montazemi, A. R., Wang, F., Khalid Nainar, S. M., & Bart, C. K. (1996). On the effectiveness of decisional guidance. *Decision Support Systems*, *18*(2), 181–198.
Moskowitz, H., Schaefer, R. E., & Borcherding, K. (1976). 'Irrationality'of managerial judgments: Implications for information systems. *Omega*, *4*(2), 125–140.
Mules, T., & McDonald, S. (1994). The economic impact of special events: The use of forecasts. *Festival Management and Event Tourism*, *2*(1), 45–53.
Murphy, A. H., & Daan, H. (1984). Impacts of feedback and experience on the quality of subjective probability forecasts: Comparison of results from the first and second years of the Zierikzee experiment. *Monthly Weather Review*, *112*(3), 413–423.
Nettelhorst, S. C., Brannon, L. A., & Trey Hill, W. (2013). Examining the impact of consumer feedback on internet product evaluation: Comparing base-rate and case history information. *Computers in Human Behavior*, *29*(4), 1290–1294.
Nikolopoulos, K., & Assimakopoulos, V. (2003). Theta intelligent forecasting information system. *Industrial Management & Data Systems*, *103*(9), 711–726.
O'Connor, M., Remus, W., & Griggs, K. (1993). Judgemental forecasting in times of change. *International Journal of Forecasting*, *9*(2), 163–172.
O'Connor, M., Remus, W., & Griggs, K. (1997). Going up–going down: How good are people at forecasting trends and changes in trends? *Journal of Forecasting*, *16*(3), 165–176.
Ofir, C., & Mazursky, D. (1997). Does a surprising outcome reinforce or reverse the hindsight bias? *Organizational Behavior and Human Decision Processes*, *69*(1), 50–57.
Oh, C., & Morzuch, B. J. (2005). Evaluating Time-Series Models to Forecast the Demand for Tourism in Singapore Comparing Within-Sample And Postsample Results. *Journal of Travel Research*, *43*(4), 404–413.
Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*, *151*(1), 53–79.
Olsen, R. A. (1997). Desirability bias among professional investment managers: Some evidence from experts. *Journal of Behavioral Decision Making*, *10*(1), 65–72.
Önkal, D., Yates, J. F., Simga-Mugan, C., & Öztin, Ş. (2003). Professional vs. amateur judgment accuracy: The case of foreign exchange rates. *Organizational Behavior and Human Decision Processes*, *91*(2), 169–185.
Ord, J. K., Koehler, A. B., & Snyder, R. D. (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of the American Statistical Association*, *92*(440), 1621–1629.
Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science*, *11*(4), 404–428.
Parikh, M., Fazlollahi, B., & Verma, S. (2001). The effectiveness of decisional guidance: an empirical evaluation. *Decision Sciences*, *32*(2), 303–332.
Patel, R., & Davidson, B. (2003). *Forskningsmetodikens grunder. Att planera, genomföra och rapportera en undersökning.* Lund: Studentlitteratur. Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2:259327
Pearce, S. L. (1995). *Comparing an expert system vs traditional approach to forecasting item demand in a distribution inventory environment* (PhD Dissertation). Texas A&M University, College Station. Retrieved from http://dl.acm.org/citation.cfm?id=922516

Perrin, B. M., Barnett, B. J., & Walrath, L. C. (1993). Decision making bias in complex task environments. In *the Human Factors and Ergonomics Society Annual Meeting* (Vol. 37, pp. 1117–1121). Retrieved from http://pro.sagepub.com/content/37/16/1117.short

Perrin, B. M., Barnett, B. J., Walrath, L., & Grossman, J. D. (2001). Information order and outcome framing: An assessment of judgment bias in a naturalistic decision-making context. *The Journal of the Human Factors and Ergonomics Society*, *43*(2), 227–238.

Petropoulos, C., Patelis, A., Metaxiotis, K., Nikolopoulos, K., & Assimakopoulos, V. (2003). SFTIS: a decision support system for tourism demand analysis and forecasting. *Journal of Computer Information Systems*, *44*(1), 21–32.

Petter, S. (2007). A Process to Reuse Experiences via Written Narratives among Software Project Managers: A Design Science Research Proposal. In V. K. Vaishnavi & W. Kuechler Jr, *Design Science Research Methods and Patterns: Innovating Information and Communication Technology* (pp. 41–54). New York: Auerbach Publications. Retrieved from http://www.citeulike.org/group/13195/article/7001219

Poulton, E. C. (1994). *Behavioral decision theory: A new approach*. Cambridge: Cambridge University Press. Retrieved from http://www.google.com/books?hl=en&lr=&id=kdag6FwF0CEC&oi=fnd&pg=PP1&dq=Behavioral+decision+theory:+A+new+approach.&ots=89SevHZIks&sig=McPOhJ8Tpbzvx7brAMbBH5yrHLA

Prideaux, B., Laws, E., & Faulkner, B. (2003). Events in Indonesia: Exploring the limits to formal tourism trends forecasting methods in complex crisis situations. *Tourism Management*, *24*(4), 475–487.

Rabin, M. (2002). Inference by believers in the law of small numbers. *The Quarterly Journal of Economics*, *117*(3), 775–816.

Ralph, P. (2011). Toward a theory of debiasing software development. In S. Wrycza, *Research in Systems Analysis and Design: Models and methods* (pp. 92–105). Berlin/Heidelberg, Germany: Springer.

Reimers, S., & Harvey, N. (2011). Sensitivity to autocorrelation in judgmental time series forecasting. *International Journal of Forecasting*, *27*(4), 1196–1214.

Remus, W. E. (1977). Bias and variance in Bowman's managerial coefficient theory. *Omega*, *5*(3), 349–351.

Remus, W. E. (1984). An empirical investigation of the impact of graphical and tabular data presentations on decision making. *Management Science*, *30*(5), 533–542.

Remus, W. E., & Kottemann, J. (1986). Toward intelligent decision support systems: An artificially intelligent statistician. *MIS Quarterly*, *10*(4), 403–418.

Russo, J. E., & Schoemaker, P. J. (1992). Managing overconfidence. *Sloan Management Review*, *33*(2), 7–17.

Samuelson, W., & Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of Risk and Uncertainty*, *1*(1), 7–59.

Sanders, N. R., & Ritzman, L. P. (2001). Judgmental adjustment of statistical forecasts. In J. S. Armstrong, *Principles of Forecasting: A Handbook for Researchers and Practitioners* (pp. 405–416). Boston: Kluwer Academic Publishers.

Schnaars, S. P., & Topol, M. T. (1987). The use of multiple scenarios in sales forecasting: An empirical test. *International Journal of Forecasting*, *3*(3–4), 405–419.

Schulster, J. R. (2004). Time-critical decision-making in spaceflight operations: the mars express experience. Retrieved from http://arc.aiaa.org/doi/pdf/10.2514/6.2004-631-427

Schwenk, C. H. (1986). Information, cognitive biases, and commitment to a course of action. *Academy of Management Review*, *11*(2), 298–310.

Sedlmeier, P., & Gigerenzer, G. (1997). Intuitions about sample size: The empirical law of large numbers. *Journal of Behavioral Decision Making*, *10*(1), 33–51.

Shavelson, R. J., Cadwell, J., & Izu, T. (1977). Teachers' sensitivity to the reliability of information in making pedagogical decisions. *American Educational Research Journal*, *14*(2), 83–97.

Sheldon, P. J., & Var, T. (1985). Tourism forecasting: A review of empirical research. *Journal of Forecasting*, *4*(2), 183–195.

Silver, M. S. (1991). Decisional guidance for computer-based decision support. *MIS Quarterly*, *15*(1), 105–122.

Singh, T. D. (1998). Incorporating cognitive aids into decision support systems: the case of the strategy execution process. *Decision Support Systems*, *24*(2), 145–163.

Sjöberg, L. (1982). Aided and unaided decision making: Improving intuitive judgement. *Journal of Forecasting*, *1*(4), 349–363.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance*, *6*(6), 649–744.

Soll, J. B., & Mannes, A. E. (2011). Judgmental aggregation strategies depend on whether the self is involved. *International Journal of Forecasting*, *27*(1), 81–102.

Song, H., Gao, B. Z., & Lin, V. S. (2013). Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system. *International Journal of Forecasting*, *29*(2), 295–310.

Song, H., Gao, Z., & Lin, S. (2012). Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system. *International Journal of Forecasting*, In press. doi:10.1016/j.ijforecast.2011.12.003

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting – A review of recent research. *Tourism Management*, *29*(2), 203–220.

Song, H., & Witt, S. F. (2000). *Tourism demand modelling and forecasting: modern econometric approaches*. Oxford: Pergamon Press.

Song, H., Witt, S. F., & Zhang, X. (2008). Developing a web-based tourism demand forecasting system. *Tourism Economics*, *14*(3), 445–468.

Spenceley, A. (2008). Requirements for sustainable nature-based tourism in transfrontier conservation areas: A southern African Delphi consultation. *Tourism Geographies*, *10*(3), 285–311.

Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychologica*, *47*(2), 143–148.

Takeda, H., Veerkamp, P., & Yoshikawa, H. (1990). Modeling design process. *AI Magazine*, *11*(4), 37–48.

Taylor, S. E., & Thompson, S. C. (1982). Stalking the elusive 'vividness' effect. *Psychological Review*, *89*(2), 155–181.

Tideswell, C., Mules, T., & Faulkner, B. (2001). An integrative approach to tourism forecasting: A glance in the rearview mirror. *Journal of Travel Research*, *40*(2), 162–171.

Timmers, H., & Wagenaar, W. A. (1977). Inverse statistics and misperception of exponential growth. *Perception & Psychophysics*, *21*(6), 558–562.

Todd, P. M., & Gigerenzer, G. (2007). Environments that make us smart: Ecological rationality. *Current Directions in Psychological Science*, *16*(3), 167–171.

Topolinski, S., & Reber, R. (2010). Gaining insight into the 'aha' experience. *Current Directions in Psychological Science*, *19*(6), 402–405.

Trapero, J. R., Fildes, R., & Davydenko, A. (2011). Nonlinear identification of judgmental forecasts effects at SKU level. *Journal of Forecasting*, *30*(5), 490–508.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley. Retrieved from http://xa.yimg.com/kq/groups/16412409/1159714453/name/exploratorydataanalysis.pdf

Tullis, T. S. (1981). An evaluation of alphanumeric, graphic, and color information displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *23*(5), 541–550.

Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, *76*(2), 105–110.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions. *Science*, *211*(4481), 453–458.

Tversky, A., & Kahneman, D. (1982). Evidential impact of base rates. In D. Kahneman, P. Slovic, & A. Tversky, *Judgment under uncertainty: Heuristics and biases* (pp. 153–160). New York: Cambridge University Press.

Vaishnavi, V., & Kuechler, W. (2004). Design research in information systems. Retrieved from http://www.citeulike.org/group/4795/article/6505471

Vaishnavi, V., & Kuechler, W. (2007). *Design science research methods and patterns: innovating information and communication technology*. Boca Raton: Auerbach Publications. Retrieved from http://www.google.com/books?hl=en&lr=&id=sI2Y9Jh8tq8C&oi=fnd&pg=PP1&dq=design+science+research+methods+and+patterns:+innovating+information+and+communication+technology&ots=k6aEX6PrZJ&sig=Xnj6qiQbWd8kjTyDerFNxoqel7k

Van Den Boom, A. (2009). *Failing of mega projects* (Master of Science in Finance and International Business). Aarhus School of Business, University of Aarhus, Aarhus, Denmark. Retrieved from http://pure.au.dk/portal-asb-student/files/7043/Thesis_May_2009_-_Arne_VAN_DEN_BOOM__Aarhus_School_of_Business.pdf

VandenBos, G. R. (2007). *APA Dictionary of Psychology*. California: American Psychological Association.

Vokurka, R. J., Flores, B. E., & Pearce, S. L. (1996). Automatic feature identification and graphical support in rule-based forecasting: a comparison. *International Journal of Forecasting*, *12*(4), 495–512.

Wagenaar, W. A. (1988). *Paradoxes of gambling behaviour*. East Sussex: Lawrence Erlbaum Associates, Inc. Retrieved from http://psycnet.apa.org/psycinfo/1989-98042-000

Wagenaar, W. A., & Timmers, H. (1978). Extrapolation of exponential time series is not enhanced by having more data points. *Attention, Perception, & Psychophysics*, *24*(2), 182–184.

Whitley, B. E. (1996). Behavioral Science: Theory, Research, and Application. In B. E. Whitley Jr & M. E. Kite, *Principles of Research in Behavioral Science* (pp. 1–30). California: Mayfield Publishing.

Witt, S. F., & Witt, C. A. (1995). Forecasting tourism demand: A review of empirical research. *International Journal of Forecasting*, *11*(3), 447–475.

Wöber, K. W. (2003). Information supply in tourism management by marketing decision support systems. *Tourism Management*, *24*(3), 241–255.

Wong, K. K. F., Song, H., Witt, S. F., & Wu, D. C. (2007). Tourism forecasting: To combine or not to combine? *Tourism Management*, *28*(4), 1068–1078.

Yaniv, I. (2004). Receiving other people’s advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, *93*(1), 1–13.

Yates, J. F., Lee, J.-W., & Shinotsuka, H. (1996). Beliefs about overconfidence, including its cross-national variation. *Organizational Behavior and Human Decision Processes*, *65*(2), 138–147.

Yeoman, I., & McMahon-Beattie, U. (2005). Developing a scenario planning process using a blank piece of paper. *Tourism and Hospitality Research*, *5*(3), 273–285.

Yong, Y. W., Keng, K. A., & Leng, T. L. (1989). A Delphi forecast for the Singapore tourism industry: Future scenario and marketing implications. *International Marketing Review*, *6*(3), 35–46.

Yüksel, S. (2007). An integrated forecasting approach to hotel demand. *Mathematical and Computer Modelling*, *46*(7–8), 1063–1070. doi:10.1016/j.mcm.2007.03.008

Zhang, X., Song, H., & Huang, G. Q. (2009). Tourism supply chain management: A new research agenda. *Tourism Management*, *30*(3), 345–358.