



Universiteit
Leiden
The Netherlands

Judgmental adjustments in revenue management : exploring the impact of user overrides on the accuracy of system-generated occupancy forecasts at multiple forecasting horizons

Koupriouchina, L.A.

Citation

Koupriouchina, L. A. (2019, November 12). *Judgmental adjustments in revenue management : exploring the impact of user overrides on the accuracy of system-generated occupancy forecasts at multiple forecasting horizons*. Retrieved from <https://hdl.handle.net/1887/80328>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/80328>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/80328> holds various files of this Leiden University dissertation.

Author: Koupriouchina, L.A.

Title: Judgmental adjustments in revenue management : exploring the impact of user overrides on the accuracy of system-generated occupancy forecasts at multiple forecasting horizons

Issue Date: 2019-11-12

JUDGMENTAL ADJUSTMENTS IN REVENUE MANAGEMENT

Exploring the Impact of User Overrides
on the Accuracy of System-Generated Occupancy Forecasts
at Multiple Forecasting Horizons

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag 12 november 2019
klokke 11.15 uur

door

Larissa Alekseevna Koupriouchina
geboren te Semipalatinsk, Kazakhstan
in 1979

Promotor: prof. dr. J.I. van der Rest

Co-promotoren: dr. H. Pluut
dr. P.W. van der Zwan

Promotiecommissie: prof. dr. J.A.A. Adriaanse
dr. C.Y. Heo (École hôtelière de Lausanne, Switzerland)
prof. dr. P. Legohérel (University of Angers, France)
prof. dr. R.D. van der Mei (Vrije Universiteit Amsterdam)



This research was supported by the Netherlands Organisation for Scientific Research (NWO) under project number: 023.002.090.

Cover design: Guus Gijben (Proefschrift-AIO)
Lay-out: Guus Gijben (Proefschrift-AIO)
ISBN: 978-94-93184-01-5

© 2019 L.A. Koupriouchina

Behoudens de in of krachtens de Auteurswet van 1912 gestelde uitzonderingen mag niets uit deze uitgave worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.

Voor zover het maken van reprografische verveelvoudiging uit deze uitgave is toegestaan op grond van artikel 16h Auteurswet 1912 dient men de daarvoor wettelijk verschuldigde vergoedingen te voldoen aan de Stichting Reprorecht (Postbus 3051, 2130 KB Hoofddorp, www.reprorecht.nl). Voor het overnemen van (een) gedeelte(n) uit deze uitgave in bloemlezingen, readers en andere compilatiewerken (art. 16 Auteurswet 1912) kan men zich wenden tot de Stichting PRO (Stichting Publicatie- en Reproductierechten Organisatie, Postbus 3060, 2130 KB Hoofddorp, www.cedar.nl/pro).

No part of this book may be reproduced in any form, by print, photo print, microfilm or any other means without written permission from the publisher.

I dedicate this dissertation to my beloved family:

To my husband Alexandre, who taught me that the main product of the PhD is not the dissertation, but the researcher.

To my daughter Giulia, whose desire to write books motivated me to finish mine.

To my parents Tamara and Kairzhan, who gave me everything they could.

PREFACE

I submit this dissertation for the degree of Doctor of Philosophy at Leiden University. My interest in this research field stems from my passion for technology. Ever since receiving my master's degree, I have gravitated towards any project related to technology, such as being part of the technology risk audit group and IT special audit services at various international companies. Still fascinated by technology and data, I found a job in higher hospitality management education and started searching for interesting topics to challenge myself further. My ambition was to find a topic that would combine my work and interests. It would have to be future-looking, important from an academic and business perspective, interesting, and intriguing. I decided to use technology to find my path. I followed what Revenue Management professionals were discussing online and analyzed the literature as well as all the available content of specialized discussion groups on LinkedIn. In this way, I discovered that forecasting was a highly-discussed issue.

In the hospitality industry, computerized systems provide an estimate of future demand and the Revenue Manager has to decide whether he or she agrees with these estimates. As currently, increasing numbers of algorithms are used to support human decisions in a variety of fields, there will be an even greater need to successfully combine computer output with human knowledge and intuition. Can humans improve on decisions suggested by these data-intensive and sophisticated algorithms?

Even though I have been working on this project for several years, I am still intrigued by this topic. I see applications of these broader questions to other areas of human life. Every day we see examples unfolding. Computerized diagnostic software help doctors with diagnosis, however it is still up to the doctor to decide whether to agree with it or to trust their own knowledge and intuition. And, with the proliferation of technology in our every-day lives, these examples multiply: the way we get credit from the bank, potential partners suggested by a dating site, judicial decision making, and (much hyped in the media) self-driving cars. All these traditionally human areas have been 'invaded' by the growing power of computer algorithms. They come with fancy names and high price tags, but should we blindly follow them? How do we know if they are right? How can we evaluate the influence of our own interventions? I am fascinated by all these questions, unfortunately they cannot be answered in one dissertation. These are big, important questions which I believe we as humans should keep asking ourselves.

The dissertation should not only be of interest to scholars of revenue management and operations research and forecasting, but also to revenue management system providers and Hotel Revenue Managers in practice.

ACKNOWLEDGEMENTS

This work would not have been possible without the support of my supervisors, co-supervisors, coaches, colleagues, friends and family. Moreover, financial support granted by the Netherlands Organisation for Scientific Research (NWO) was paramount to the completion of this work.

My deepest gratitude and appreciation goes to my supervisor, Prof. Jean-Pierre van der Rest, for encouraging me to start on my PhD journey, suggesting to apply for an NWO research grant, and for encouraging and helping me to re-apply and succeed with funding on a second attempt. Jean-Pierre played a crucial role in my development as a researcher, guided me through all stages of this project, and shared precious knowledge and experience. Jean-Pierre's patience, wisdom, positivity, insight, constant expression of support, and faith in my abilities sustained me enormously through many stages of this complex undertaking.

I am also grateful to Dr. Helen Pluut and Dr. Peter van der Zwan who have been instrumental in helping me develop my research methodology, knowledge, and skills. From the moment that they were involved in the project, I have been incredibly impressed by their knowledge of methods and professional supervision.

This project would have been impossible without the knowledge, feedback, and participation of Prof. Zvi Schwartz. His academic work has been one of my sources of inspiration, and it has been an honor to collaborate with him on a number of journal articles, a book chapter, and in the revenue management education meetings (RevME) launched on his initiative. I am very grateful that he presented my research ideas to the revenue management system company (IDeaS), and for supporting my research proposal, participating in the numerous project conference calls, and providing feedback on project deliverables.

The professional and friendly environment of the Leiden University Dual PhD Centre (DPC) helped to create the atmosphere necessary for me to complete this dissertation. I extend my gratitude to Dr. Mark Dechesne who revised several versions of my early papers and provided his input on the unfolding research. I am grateful to the creator of the DPC Dr. Adriaan in 't Groen, to the current director Dr. Johannes Magliano-Tromp, and to the Program Support Officer Ms. Petra Evers and my coaches Ms. Charlotte de Roon and Dr. Yvonne Kleistra.

I am deeply grateful to my colleagues at the Hotelschool The Hague, especially colleagues from the Research Center, IT Helpdesk, Mediacenter, my fellow colleagues with whom I teach Revenue Management, the Director of the Research Center Dr. Jeroen Oskam, faculty managers Ms. Maryna Jooste and Mr. Ajay Kapur and the administrator of the Research Centre, Ms. Ludmila Snapkovskaya.

This research was made possible due to the collaboration with IDeaS and I am thankful for having been granted access to the unique data, for letting me complete company online training courses, and for sharing technical and business knowledge.

I received enormous support from my family and friends. My close friends Nuriya, Serena and Lara listened to me and cheered for my success from Kazakhstan, Italy and The Netherlands. At home I was lucky to experience endless patience, positivity, humor and care from my husband Alexandre and my daughter Giulia. Alexandre, thank you for listening to me compassionately, for sharing your experiences as a researcher, for leaving me space to make my own decisions, and for supporting me and my choices without any judgement. Giulia, thank you for inspiring me with your achievements and determination. I am deeply grateful to my parents, Kairzhan and Tamara. Although thousands of kilometers have separated us all these years, I have always felt your continuous support, encouragement, and love.

TABLE OF CONTENTS

Preface	V
Acknowledgements	VII
Table of contents	IX
List of figures	XI
List of tables	XII
List of abbreviations	XIV
Chapter 1. Introduction	15
1.1 Background and Scope of the Research	15
1.2 Aim and Objectives	20
1.3 Outline of the Methodology	21
1.4 Delimitations	23
1.5 Practical Significance	23
1.6 Structure of the Thesis	24
Chapter 2. Hotel Occupancy Forecasting Accuracy Measures	27
2.1 Introduction	28
2.2 Literature Review	29
2.3 Method	31
2.3.1 Data	31
2.3.2 Accuracy Measures	33
2.3.3 Computational Problems	33
2.3.4 Dealing With Multiple Forecasting Horizons	35
2.4 Illustration	37
2.5 Discussion and Future Research	41
Chapter 3. Time Dimension of Hotel Forecast Accuracy Evaluation	43
3.1 Introduction	43
3.2 Literature Review	43
3.3 Method	49
3.3.1 Data	49
3.3.2 Dependent Variables	52
3.3.3 Independent Variables	55
3.3.4 Analytical Approach	56
3.4 Results	63
3.4.1 Descriptive Statistics	63
3.4.2 Multilevel Analysis of the Scale-Dependent Measure	66
3.4.3 Multilevel Analysis of Scale-Independent Accuracy Measures	71
3.5 Discussion and Future Research	76
	IX

Chapter 4. Judgmental Adjustments and Forecast Accuracy	79
4.1 Introduction	79
4.2 Literature Review	80
4.3 Method	85
4.3.1 Data	85
4.3.2 Dependent Variables	86
4.3.3 Independent Variables	88
4.3.4 Analytical Approach	90
4.4 Results	93
4.4.1 Descriptive Statistics	93
4.4.2 Multilevel Analysis	95
4.4.3 Additional Analysis	109
4.5 Discussion and Future Research	122
 Chapter 5. Conclusion	125
5.1 General Contribution	125
5.2 Limitations of the Research	128
5.3 Directions for Future Research	130
5.4 Implications	133
 References	135
Summary	141
Samenvatting (Dutch Summary)	143
Curriculum Vitae	147
Appendices	149
Appendix A: Scale-Dependent Forecasting Accuracy Measures	149
Appendix B: Scale-Independent Forecasting Accuracy Measures	150
Appendix C: Literature Overview of Forecasting in Hotel RM	153
Appendix D: Data Management Plan	155
Appendix E: Dataset One Data Definition	162
Appendix F: Dataset Two Data Definition	164
Appendix G: Detailed Additional Analysis from Chapter 4	168

LIST OF FIGURES

Figure 1:	The relationship between aim, objectives, and research questions	22
Figure 2:	Structure of the thesis	24
Figure 3:	Mean Average Error (MAE) June 2011, all segments	36
Figure 4:	Mean Average Error (MAE) June 2011, at a segment level	36
Figure 5:	Data structure of Dataset one (partial)	50
Figure 6:	Relationship between mean forecast accuracy and DTA	63
Figure 7:	Data structure of Dataset two (partial)	86
Figure 8:	Judgmental adjustments per DTA category	95
Figure 9:	Mean improvement in forecast accuracy per DTA category	95
Figure 10:	Regression lines for improvement in sAPE by DTA, for business type, direction, override type and special event based on Model 6.1.6	113
Figure 11:	Regression lines for improvement in sAPE by DTA, for override size based on Model 6.1.6	113
Figure 12:	Regression lines for improvement in sAPE by Business Type, for override direction	115
Figure 13:	Regression lines for improvement in sAPE by judgmental adjustment size, for business type	118

LIST OF TABLES

Table 1:	MAPE of system and manual forecasts at segment level (1-month out)	32	Table 26:	Models without and with level-1 interaction terms for additional analysis of relationship between business type and improvement in accuracy	116
Table 2:	Forecasting accuracy measures for “one month plus” segment-generated forecast, total hotel	39	Table 27:	Models without and with level-1 interaction terms for additional analysis of relationship between business type and improvement in accuracy	117
Table 3:	Forecasting accuracy measures for 1-month out system-generated vs. manual forecast, total hotel	40	Table 28:	Models without and with level-1 interaction terms for additional analysis of relationship between size of adjustment and improvements in accuracy	119
Table 4:	Source data summary for dataset one	50	Table 29:	Models without and with level-1 interaction terms for additional analysis of relationship between size of adjustment and improvements in accuracy	120
Table 5:	Number of hotels by region and type	51	Table 30:	Summary results for main hypotheses and additional research questions	121
Table 6:	Hotel size by hotel type	51	Table 31:	Data definition of the dataset one variables selected for analysis	162
Table 7:	Illustrative comparison of AE, APE, sAPE and LnQ	55	Table 32:	Data definition of the dataset two variables selected for analysis	164
Table 8:	Means and standard deviations of AE, APE, sAPE and LnQ by DTA	64	Table 33:	Models with level-1 interaction terms for additional analysis of relationship between DTA and improvement in accuracy	168
Table 9:	Forecast errors by hotel type	65	Table 34:	Models with level-1 interaction terms for additional analysis of relationship between DTA and improvement in accuracy	170
Table 10:	Results of the multilevel regression models for Absolute Error	67	Table 35:	Models with level-1 interaction terms for additional analysis of relationship between DTA and improvement in accuracy	172
Table 11:	Results of the multilevel regression models for Absolute Percentage Error	72	Table 36:	Models without and with level-1 interaction terms for additional analysis of relationship between size of adjustment and improvement in accuracy	174
Table 12:	Results of the multilevel regression models for Symmetric Absolute Percentage Error	73	Table 37:	Models without and with level-1 interaction terms for additional analysis of relationship between size of adjustment and improvement in accuracy	176
Table 13:	Results of the multilevel regression models for LnQ	74	Table 38:	Models without and with level-1 interaction terms for additional analysis of relationship between size of adjustment and improvement in accuracy	178
Table 14:	Summary results hypotheses 1-3	76	Table 39:	Models without and with level-1 interaction terms for additional analysis of relationship between business type and improvement in accuracy	180
Table 15:	Source data summary for dataset two	85	Table 40:	Models without and with level-1 interaction terms for additional analysis of relationship between business type and improvement in accuracy	181
Table 16:	Number of judgmental adjustments per DTA per override type for Transient and Group	96	Table 41:	Models without and with level-1 interaction terms for additional analysis of relationship between business type and improvement in accuracy	182
Table 17:	Forecast error per adjustment direction	96			
Table 18:	Improvement in forecast errors by hotel type	96			
Table 19:	Fixed effects estimates and variance-covariance estimates for improvement in sAPE	97			
Table 20:	Fixed effects estimates and variance-covariance estimates for models of the predictors of improvement in LnQ	102			
Table 21:	Fixed effects estimates and variance-covariance estimates for models of the predictors of improvement in APE	104			
Table 22:	Fixed effects estimates and variance-covariance estimates for models of the predictors of improvement in AE	106			
Table 23:	Summary results hypotheses 1-4	109			
Table 24:	Models with level-1 interaction terms to further explore the relationship between DTA and sAPE _{imp}	110			
Table 25:	Models with level-1 interaction terms for additional analysis of relationship between DTA and improvement in accuracy	114			

LIST OF ABBREVIATIONS

ADR	Average Daily Rate
AE	Absolute Error
AIC	Akaike Information Criterion
APE	Absolute Percentage Error
ARI	Average Daily Rate Index
BIC	Bayesian Information Criterion
CBT	Critical booking threshold
DBA	Days before arrival
DL	Days left to departure
DTA	Days to arrival
GOPPAR	Gross Operating Profit Per Available Room
LCC	Low-Cost Carrier
MPI	Market Penetration Index
OTA	Online Travel Agent
RevPAR	Revenue Per Available Room
RGI	Revenue Generation Index
RM	Revenue Manager
RMA	Revenue Management Analyst
RMS	Revenue Management System
SKU	Stock-keeping unit
TRevPAR	Total Revenue Per Available Room

Chapter 1

Introduction

Prediction is very difficult, especially if it's about the future.
- Niels Bohr -

1 .1 Background and Scope of the Research

Revenue Management (RM) is a systematic approach to optimizing revenues by setting prices and managing product availability based on patterns of demand and customer willingness to pay (Talluri & Van Ryzin, 2004). It was initially introduced in the airline industry, where it became “an integral part of commercial operations for virtually all airlines” (Cross, Higbie, & Cross, 2011, p. 10). Boyd (1998) analyzed the impact of RM at US Airlines and Delta Airlines and credited it with US\$500 million and US\$300 million of increased revenue, respectively. After the enormous success of RM in the airline industry, it was adopted by the hotel industry. Revenue management brought US\$100 million of additional annual revenue to Marriott Hotel (Cross, 1997b).

Today, many other industries seek benefits from RM techniques and approaches, among these are cruise lines, restaurants, golf courses, the sports and entertainment industries, television networks, transportation firms, and natural gas storage and transmission (Talluri & Van Ryzin, 2004). RM famously helped National Car Rental to avert bankruptcy thereby safeguarding 7500 jobs (Geraghty & Johnson, 1997). According to FedEx Chairman and CEO Frederick Smith, “a significant portion of their 10 per cent revenue increase and 33 per cent growth in profits” can be attributed to RM and the pricing approach (Cross et al., 2011, p. 13).

RM includes a revenue optimization methodology that uses data and analytics to anticipate customer behavior and optimize price and product availability. It helps companies in certain industries to drive their revenues and profits. Essential

characteristics that these service industries all share, are constrained capacity, perishable inventory, reservations made in advance, and a segmentable market (Kimes, 1989, 2003).¹

The birth of RM in the airline industry is commonly attributed to the following three key turning points: (1) the creation of the Littlewood rule² in 1972; (2) the introduction of the US Domestic Deregulation Act of 1978; and (3) the doctoral thesis by Peter Belobaba (*Air Travel Demand and Airline Seat Management*) which introduced the concept of Expected Marginal Seat Revenue (EMSR) in 1987 (Yeoman, 2011; Yeoman & McMahon-Beattie, 2017). While these turning points were undoubtedly important, technology played a crucial role in shaping modern RM and was the main enabler of the vast majority of these turning points. Indeed, the Littlewood rule was significant as it established the first static single resource-based RM model. However, successful implementation of the Littlewood algorithm was feasible due to a favorable pre-condition. Prior to 1972, British Overseas Airways Corporation introduced a sophisticated computer reservation system (Yeoman & McMahon-Beattie, 2017). Similarly, although the US Domestic Deregulation Act changed the face of the airline industry forever by allowing entrance of Low-Cost Carriers (LCCs), the resulting changes were possible due to the advanced state of technology. In 1985, American Airlines responded to increased competitive pressure from a highly competitive LCC - PEOPLExpress by using advanced analytics and introducing the Ultimate Super Saver. Due to the newly developed computer systems, it was able to control the availability of deeply discounted fares. This important advantage helped American Airlines to drive PEOPLExpress out of business (Yeoman & McMahon-Beattie, 2017). Moreover, EMSR had a significant impact on businesses because it became feasible to implement and build its logic into the foundation of many revenue management systems. In other words, technology played an important role in all three turning points, and it continues to play an important role in current RM development.

Technology has been paramount not only for the development of airline RM, but also for hotel RM, and its importance is still growing. Marriott is often associated with pioneering the RM concepts in the hotel industry. For example, it first experimented with rate fences and length of stay controls (Hanks, Cross, & Noland, 2002; Kimes, 2016). In the mid-1980s, Marriott developed its own automated RMS,

which generated daily demand forecasts and inventory recommendations for each of its 160,000 rooms at its Marriott, Residence Inn and Courtyard brands (Marriott & Cross, 2000). Hilton, Holiday Inn and Sheraton and other major hotel chains in North America followed Marriott's example (Cross et al., 2011). In the late 1980s, many hotels with sophisticated electronic property management systems introduced yield management systems (Kimes, 1989). Since the 1990s, the hotel industry has experienced many drastic changes and pressures: 1) price transparency leading to the shift of balance towards consumer power; 2) the rise of Online Travel Agents (OTAs) resulting in increased distribution costs; 3) changing ownership structures leading to the need to generate revenues from all departments; 4) the tragic events of 9/11 which resulted in a hotel occupancy decrease of 15-20% and a decrease in profits for US hotels by more than \$600 million. During these shifts and turbulent times, evolving technological capabilities represented an important competitive advantage. In the late 1990s, Cross (1997a) highlighted the role of technology in RM, emphasizing the importance of forecasting and optimization components of a revenue management system. At the same time, there was growing awareness that an RMS would become one of the main weapons to win this battle. Kimes and Wagner (2001) emphasized the importance of revenue management systems and called for protecting it as a trade secret. Overall, a shift can be observed towards the increasing power of science, and a dominance of technology and complex algorithms.

Although technology plays an increasingly important role, it cannot operate without human input. In 1992, Jones and Hamilton (1992, p. 91) warned about an excessive emphasis and reliance on yield management systems: "no computerized system will ever be successful without a wide range of skilled personnel who are involved in the process." Similarly, a decade later, Kimes (2003, p. 138) stated that the "mere possession of a revenue management system does not guarantee success." Despite increased technological developments, this view has been supported in more recent studies. Bobb and Veral (2008, p. 299) warned that, "revenue management applications and software cannot, and should not, be seen as only a computerized solution but rather as a tool to aid the human decision-making process. In fact, a hybrid system may prove to be the most effective economically as well as managerially." For example, in the context of automated room rate pricing decision-making, Van der Rest and Roper (2013) and Van der Rest, Roper, and Wang (2018) stressed the importance of human and social capabilities. In the context of hotel occupancy forecasting, human judgement plays an important role as many algorithms are based on historical data, which do not include present and future changes and may not predict a discontinuous change in the business environment (Ghalia & Wang, 2000). Revenue managers often possess important additional knowledge, exogenous to the RMS (Schwartz, Uysal, Webb, & Altin, 2016). For these

¹ Segmentation is a key characteristic of revenue management. It allows the strategy to be tailored to match the needs of various customers and to increase bookings, revenue and profit. Segmentation can be done in various ways and with various degrees of data aggregation. A first step and common approach is to distinguish between transient and group business. Group business is "typically defined as 10 or more rooms per night, sold pursuant to a signed agreement" (STR, 2019). The transient segment are "guests who book individually rather than with a group" (IDeaS, 2019).

² The first static single resource quantity based RM model developed by Ken Littlewood.

reasons, RM has been referred to as the 'art and science' of forecasting demand, while simultaneously adjusting price and availability of inventory to meet demand (Cross, Higbie, & Cross, 2009; Erdem & Jiang, 2016; Josephi, Stierand, & Van Mourik, 2015).

There is general consensus on the importance and need for more work on the role of human judgment in technology-supported RM decision-making. For example, in a review of academic journal articles, books and monographs published in the period 2002–2012, Ivanov and Zhechev (2012) concluded that RM technology is important, but far too little attention was paid to its impact on RM decisions. Also, after systematically reviewing 163 studies (from 2004–2013) using three major online hospitality and tourism research databases, Denizci Guillet and Mohammed (2015, p. 546) argued that, "there are still unanswered questions about the interaction of users with technology regarding the impact of user overrides." Similarly, Josephi et al. (2015, pp. 4-5) who provided a brief historic overview of RM, pointed out that, "future avenues for research should explore the role of automation and human judgment in RM decision making." Despite its recognized importance, there is very little research on human judgment in technology-supported RM decision-making.

Several studies in the airline forecasting literature reveal that human judgement is valuable (Bach, 1999; Loew, 2000; Mukhopadhyay, Samaddar, & Colville, 2007; Weatherford, 2016; Zeni, 2003). Yet, judgmental adjustments have rarely been studied in the context of hotel revenue management forecasting. Kimes (1999) conducted a study on forecast data of approximately 90 hotels of a large North American hotel chain and concluded that more accurate group forecasts were achieved in larger hotels with a higher dependence on group business and with more frequent forecast updates in the month before arrival. However, these results were limited to group business and are therefore not representative of total hotel occupancy, which also includes transient business. In their analysis of 57 experienced revenue managers, Schwartz and Cohen (2004a) found that the nature of the user interface influenced the way revenue managers adjusted computer forecasts. In their subsequent study conducted on the same data, Schwartz and Cohen (2004b) demonstrated that subjective estimates of forecast uncertainty depend on an individual's industry experience and gender. However, the generalizability of these studies is limited. Albeit hotel and airline revenue management share key characteristics, there are important conceptual differences such as that the length of stay of hotels has a network structure (i.e., the displacement effect), and that arrival demands for multi-night stays and the length of stay are stochastic (Chiang, Chen, & Xu, 2007; Lai & Ng, 2005). Also, these two hotel studies were conducted with a relatively small number of participants using an experimental design and on simulated data. Moreover, the studies pertained to a limited timeframe; four months and seven weeks of data respectively. Additionally, computing power has increased drastically in recent years,

enabling the utilization of more complex RM models and computer algorithms. Therefore, there is a need for research on the current role of human judgment in technology-supported RM decision-making.

Revenue management systems generally consist of multiple dynamically connected modules, where the forecasting module is seen as a starting point of the process and its output has a crucial impact on other modules (Ghalia & Wang, 2000; Talluri & Van Ryzin, 2004). Accurate forecasts are important and estimates suggest that a 20% reduction of forecast error could translate into a 1% increase in revenue generated from the RMS (Talluri & Van Ryzin, 2004).³ Most of the literature on forecasting in RM largely focuses on mathematical design and implementation of more accurate forecasting models, such as time-series forecasting methods and Bayesian forecasting methods (Talluri & Van Ryzin, 2004). Also, most research attention is given to the airline and tourism industries, thereby ignoring the unique features of hotel occupancy forecasting (e.g., pick-up forecasting, length of stay).

The academic empirical work on how the hotel RMS forecasts are used and modified by revenue managers is scarce in hospitality research literature. In the generic operations research literature, forecasting accuracy has been greatly improved by the use of automated forecasting support systems. However, Goodwin, Fildes, Lawrence, and Nikolopoulos (2007, p. 403) observe that "while modern commercial forecasting systems provide access to powerful statistical methods and the ability to link easily to large databases, their design has largely neglected the role that the judgment of users plays in the forecasting process." This human interaction with the automated forecasting system typically happens when the forecasting module generates forecasts and before these forecasts are fed into the optimization module to generate proper recommendations. At this stage, most RMSs provide a possibility to override the forecasts. These overrides are also referred to in the literature as judgmental adjustments. The research findings on the value of human expertise, however, are contradictory. According to some studies, a combination of statistical forecasting with judgmental forecasting should lead to improved forecasting accuracy (Donihue, 1993; Fildes et al., 2009; McNees, 1990; D. S. Turner, 1990), whereas other work demonstrates that forecasters often make unnecessary adjustments (Goodwin & Wright, 1994; Lawrence, Goodwin, O'Connor, & Onkal, 2006; J. S. Lim & Oconnor, 1995). These conflicting findings indicate a need to understand deeper the role of judgment in the hotel forecasting context.

3 The forecasting module generates constrained and unconstrained demand forecasts. While there are other types of revenue management forecasts (e.g., strategic forecasts, revenue forecasts, and operational forecasts), demand forecasts are generally considered as key (Cullen, 2015, p. 47).

To summarize, hotel revenue managers increasingly rely on their RMSs, in which the forecasting module plays a crucial role as it influences all other modules and critical decisions, including optimization and pricing. Although human judgement is necessary for a successful forecasting process, only few empirical studies have been carried out to examine judgmental adjustments in a hotel RM forecasting context. Thus, a better understanding of the complex phenomenon of user-system interaction and its impact on forecasting accuracy is needed, as accurate forecasting is a cornerstone for hotel revenue management performance. In the next section, the aim and the objectives of the research are presented.

1.2 Aim and Objectives

To provide direction for the research, the following overall purpose statement is formulated:

The aim of this study is to improve our understanding of judgmental adjustments in hotel revenue management forecasting by exploring the impact of user overrides on the accuracy of system-generated hotel occupancy forecasts at multiple forecasting horizons.

The term *judgmental adjustments* refers to “revisions of statistically generated forecasts using relevant domain knowledge” (Lawrence et al., 2006, p. 500). These revisions include *user overrides* (i.e., manual adjustments of system-generated hotel occupancy forecasts).⁴ *Forecast accuracy* is defined as the degree to which the forecasted hotel occupancy conforms to the actual occupancy (Hyndman & Koehler, 2006). *Forecasting horizon* is defined as the length of time into the future for which the occupancy forecast is generated (Lawrence, Edmundson, & O’Connor, 1985). To achieve and clarify the research aim, three objectives are formulated:

- 1. To investigate the applicability of forecast accuracy measures to evaluate judgmental adjustments of system-generated hotel occupancy forecasts.
- 2. To examine the accuracy of non-adjusted and adjusted hotel occupancy forecasts in the context of multiple forecast horizons.
- 3. To examine the impact of different characteristics of judgmental adjustments on the accuracy of the system-generated forecasts.

4 The terms ‘judgmental adjustments’ and ‘user overrides’ will be used interchangeably throughout the dissertation, since this thesis does not investigate other judgmental adjustments except for user overrides.

In addition, two overarching research questions are formulated to guide the enquiry:

- 1. To what extent do judgmental adjustments of the system-generated occupancy forecasts impact forecasting accuracy?
- 2. How do different characteristics of judgmental adjustments impact on the accuracy of system-generated hotel occupancy forecasts?

Figure 1 illustrates the relationship between aim, objectives and research questions.⁵

1.3 Outline of the Methodology

This research employs a quantitative design utilizing a longitudinal perspective. A longitudinal study is considered appropriate as it makes it possible to observe how forecasting accuracy is affected by user interventions (i.e., judgmental adjustments) at multiple occasions over the forecasting horizon. In this way, the dynamic and iterative nature of the forecasting and re-forecasting processes can be captured, as well as the time-dependent and seasonal nature of demand, which typically differs per day, week, and month (Cetin, Demirciftci, & Bilgihan, 2016).

Three studies are executed to address the research objectives. Real-world hotel occupancy forecast data are collected which include key data such as system-generated demand forecasts, detailed data on judgmental adjustments, and actual hotel occupancies.

Study 1 comprises hotel daily occupancy forecasting data of 2,043 triples of (a) daily occupancy predictions by RMS, (b) daily manual forecast of revenue manager and (c) daily actual hotel occupancy data across multiple forecasting horizons for a mid-sized five-star hotel situated in the Netherlands, belonging to a large international hotel chain. The data include daily room night forecasts at six segment levels for a two-year period, from January 2011 to December 2012. Study 2 includes forecast data of, in total, 20,081,973 cases from 1,752 hotels for the occupancy period 24 March 2014 to 7 April 2017. Data are obtained from city, resort, airport, and casino hotels from seven geographical regions including Europe, North America, Asia, Middle East & Africa, Australasia, Central America, and South America. Study 3 data represent an extended subset of the Study 2 dataset and contain detailed forecasts and judgmental adjustments data only for those occupancy days for which one or more user overrides took place. Study 3 data consist of a total of 228,649 cases of judgmental forecast adjustments from 1,752 hotels worldwide, for the occupancy period 24 March 2014 to 7 April 2017.⁶

5 To narrow the study’s focus, sub-questions are developed; they are listed on the right side of Figure 1.

6 Full details on the data can be found in the sections 2.3.1., 3.3.1., and 4.3.1 of this thesis.

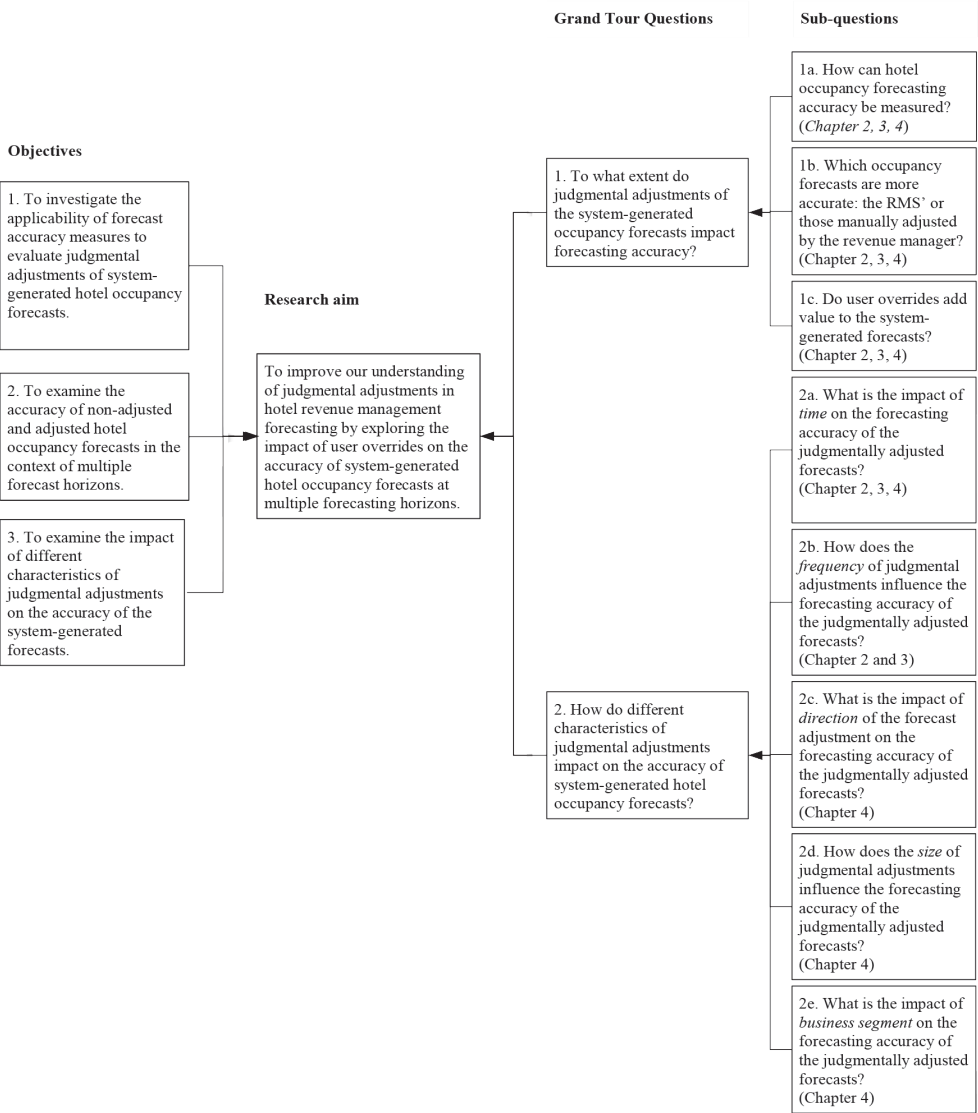


Figure 1: The relationship between aim, objectives, and research questions

The data are analyzed using the software package SPSS (version 23, release 23.0.0.3). The data analysis includes both a descriptive data analysis and a multilevel regression analysis. The multilevel analysis is chosen as the analytical approach since the data have a hierarchical or nested structure. This method is referred to in the literature under a number of names including hierarchical linear modeling (HLM), mixed linear modeling, or growth-curve modeling (Dedrick et al., 2009). The multilevel regression model can

be viewed as an extension of the classical multiple regression model as it adds the possibility to examine the relationships between variables measured at different level of the multilevel data structure. Multilevel analysis is considered appropriate as it overcomes an important limitation of conventional multiple regression analysis which relies heavily on the assumption of independence of observations; this is often violated with nested data (Hox, Moerbeek, & Schoot, 2018).

1.4 Delimitations

The main interest of this thesis is to gain an understanding of the impact of judgmental adjustments on the accuracy of system-generated hotel occupancy forecasts. While in the operations research literature the term judgmental adjustments tends to refer to the revisions of statistically generated forecasts using relevant expert domain knowledge (Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009; Flores, Olson, & Wolfe, 1992; Lawrence et al., 2006), the term is sometimes equated to all types of judgmental inputs in forecasting. These inputs, however, are usually defined more broadly than judgmental adjustments, and according to Mathews and Diamantopoulos (1986), they include three distinct forms: *a priori incorporation*, *concurrent incorporation*, and *a posteriori incorporation*. However, it is beyond the scope of this study to examine all aspects where forecasters execute their judgment related to the selection, specification and modification of the forecasting model parameters or to the blending of judgmental and statistical forecasts through a combining algorithm. Although in the context of automated RMS there are various ways in which users may influence the system (e.g., adjusting prices, changing overbooking limits), the scope of this study is the judgmental adjustment of the occupancy forecast only. Hence, strategic, operational and revenue forecasts are out of scope for this study. Also, the thesis does not examine RMS interface design features and their impact on system users' adjustment behavior.

1.5 Practical Significance

Hotel performance is heavily affected by seasonality of demand (Vives, Jacob, & Payeras, 2018). Accurate forecasting is therefore a cornerstone for financial success and impacts decisions such as pricing and operational planning (Weatherford & Kimes, 2003). Whereas a computerized revenue management system (RMS) supports the forecasting process and is indispensable for hotel performance, an autonomous (i.e., human-independent) RMS is not yet feasible (Weatherford, 2016). Accordingly, revenue managers continuously interact with their RMS, and the resulting system-supported RM decisions are an integral part of hotels' daily practices. However, the impacts of this crucial human/RMS interaction are not sufficiently investigated in the practitioners' or academic literature.

This thesis provides an important opportunity to advance our understanding of a complex and dynamic process of hotel forecast generation and adjustment. Moreover, it offers initial insights into interactions between the Revenue Manager and revenue management system within the context of hotel forecasting. The findings can be applied by hotel Revenue Managers to facilitate their systematic and periodic evaluation of their judgmental adjustments, and in this way to help them identify adjustments that add value. This study has also been designed to provide insights on how different characteristics of judgmental adjustments impact forecasting accuracy, and ultimately, to assist hotels and RMS solution providers in their quest to improve the way judgmental adjustment to system forecasts are constructed and thereby to improve their profitability.

1.6 Structure of the Thesis

The overall structure of the research takes the form of five chapters, including this introductory chapter. Figure 2 illustrates the organization and structure of the thesis.

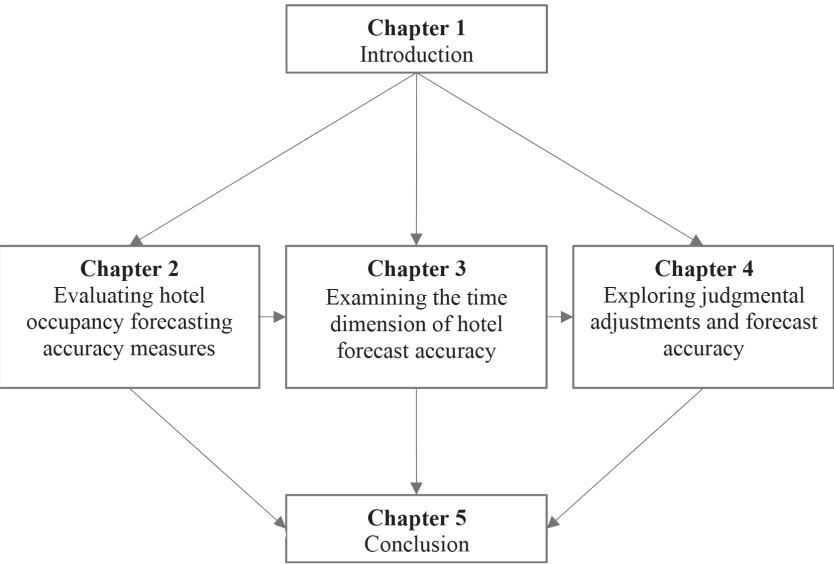


Figure 2: Structure of the thesis

Chapters 2, 3 and 4 provide the results of three empirical studies. Chapter 2 verifies the applicability of existing forecast accuracy measures for the evaluation of hotel forecasting performance. This chapter provides a critical review and discussion of forecasting accuracy measures and lays the foundation for how forecasting accuracy is measured in the subsequent two studies, also addressing

the aspects of level of data aggregation and multiple forecasting horizons. Since both hotels and customers make time-dependent decisions, Chapter 3 examines the critical relationship between forecasting horizon and forecasting accuracy for non-adjusted and adjusted hotel occupancy forecasts. It demonstrates empirically that forecast accuracy improves as the date of arrival nears and that the relationship between time and forecast accuracy is complex. Chapter 4 explores the impact of various characteristics of judgmental adjustments on improvement in forecasting accuracy, with regard to the forecasting horizon, the frequency, direction and size of adjustments, and the type of business segment. It finds that early forecast improvements are associated with lower improvements in accuracy, the judgmental adjustments to the group business are associated with larger improvements than adjustments to transient business and that the size of judgmental adjustments is positively associated with improvement in accuracy. Chapter 5 draws on the findings of the three chapters, tying up the theoretical and empirical strands, and discusses this thesis’s contribution to knowledge, its implications for practice, and directions for future research.

Chapter 2

Hotel Occupancy Forecasting Accuracy Measures⁷

It is not what you look at that matters, it is what you see.
- Henry David Thoreau -

This chapter draws attention to inherent problems in forecasting accuracy measurement, and thereby seeks to initiate a critical discussion about forecast quality assessment in hotels. An exhaustive, literature-based set of seventeen forecasting accuracy measures was applied to hotel daily occupancy forecasting data of 2,043 pairs of computer and human forecast/actuals, across multiple forecasting horizons. The empirical analysis demonstrates endemic inconsistencies across the accuracy measures, and identifies a plethora of theoretical and practical challenges with regard to total hotel-level, as well as to customer-level forecast accuracy assessment. The analysis illustrates the issues related to the level of data aggregation at multiple forecasting horizons and highlights the difficulty of interpreting conflicting forecasting accuracy results. The chapter concludes with a brief discussion of a more comprehensive approach to hotel forecasting quality assessment, and serves to warn hotel revenue managers, academics, practitioners, and solution providers against the unconsidered use of accuracy measures. As such, it provides a basis for the selection of forecasting accuracy measures (i.e., DVs) in chapters 3 and 4.

⁷ This chapter is based on: Koupriouchina, L., Van der Rest, J.I., & Schwartz, Z. (2014). On revenue management and the use of occupancy forecasting error measures, *International Journal of Hospitality Management*, 41, 104-114. An extended version appeared as: Koupriouchina, L., Van der Rest, J. I., Sierag, D. D., & Schwartz, Z. (2017). Forecasting in Hotel Revenue Management: a critical review of scale-dependent and scale-independent accuracy measures. In M. Uysal, Z. Schwartz, & E. Sirakaya-Turk (Eds.), *Management Science in Hospitality and Tourism* (pp. 269-304). Oakville: Apple Academic Press. An earlier version of this paper was presented at the 31st Annual EuroCHRIE Conference (2013). We thank all participants and reviewers for useful suggestions and comments.

2.1 Introduction

The practice of revenue management in the lodging industry has grown significantly in the past three decades, and is now considered an indispensable part of hotels' marketing and operating strategies (Cross et al., 2011). Hotels have developed revenue management policies with the aim of profitably matching or managing a fluctuating demand with the hotel's constrained and perishable capacity. This is achieved by employing a range of room pricing and allocation tools, and by addressing core revenue management concepts such as: the reservation of a portion of the capacity for higher value customers at a later date; efficient price discrimination practices to extract as much of the consumer surplus as possible; and overbooking policies to offset no shows, late cancellations, and early departures. A key ingredient for successfully implementing these core revenue management concepts is making high-quality forecasts. As Talluri and Van Ryzin (2004, p. 407, *italics added*) observe, "A revenue management system requires forecasts of quantities such as demand, price sensitivity, and cancellation probabilities, and its performance depends *critically* on the quality of these forecasts." One estimate suggests that a 20% reduction of forecast error could translate into a 1% increase in revenue generated by the revenue management system (RMS) (Pölt, 1998 cited in Talluri & Van Ryzin, 2004, p. 407). Hence, forecasting is a cornerstone element of revenue management, and the quality of forecasts is a crucial characteristic that determines the success of the RMS' capacity to generate higher revenues and profits.

According to Armstrong (1986), forecasting is an area in operations research which has grown into a discipline of its own with studies published as early as the 1960s. It continues to draw specialist research attention from a wide range of disciplines (Fildes, Nikolopoulos, Crone, & Syntetos, 2008). A relevant area with a long tradition of high quality research in forecasting is tourism, with early work conducted by Choy (1984) and Fritz, Brandon, and Xander (1984), and with seminal contributions by Martin and Witt (1988), L. W. Turner and Witt (2001), Li, Song, and Witt (2006), Song and Li (2008), Song, Li, Witt, and Fei (2010), and Song, Gao, and Lin (2013). In contrast, the literature on forecasting in hotels is underdeveloped, with notable exceptions from Kimes (1999), Weatherford, Kimes, and Scott (2001), Law (2004), El Gayar et al. (2011) and Zakhary, Atiya, El-Shishiny, and Gayar (2011).⁸

Enhancing and maintaining forecast quality is a crucial task for revenue managers, and the automated systems they employ. The quality element that revenue managers are mostly concerned with and address is, unsurprisingly, the forecast's accuracy. Within the task domain of achieving the greatest accuracy, hotels select the most accurate forecasting models available when they start using the RMS, and they continually monitor forecast accuracy (Schwartz, 1999). If the accuracy is deemed

unsatisfactory, data, forecasting models and/or their parameters, as well as the use of subjective forecasts and the practice of subjective adjustments to the forecasts, are all scrutinized and often modified, as part of the revenue management team's efforts to improve performance. One example of an area of concern where hotels often examine accuracy as part of the monitoring phase of the revenue management cycle is the black-box nature of machine-generated forecasts.

Hotels sometimes assess and verify the accuracy of their RMS' computer-generated forecasts by contrasting them with "manual" forecasts, that is, predictions made by the hotel's revenue manager or revenue management team. A common dilemma associated with this is the extent to which the RMS can be trusted to generate an accurate forecast. Interestingly, when hotels face elevated levels of risk and distress, such as a liquidity challenge or intensified competition in highly volatile markets, more pressure is placed on the revenue manager to ensure that the forecasts are accurate. In practice it is, therefore, not unusual that revenue managers override the RMS in an attempt to enhance revenue management performance. Literature provides little evidence about whether these subjective adjustments actually add value. This study, therefore, explores whether the subjective adjustments by revenue managers improve or reduce the accuracy of the sophisticated computer-generated forecasts.

Obviously, this is where the forecast quality monitoring procedures become relevant, as the question is best answered using a well-designed, systematic, and appropriately implemented forecasting quality assessment procedure. In this chapter, existing forecast accuracy measures and issues associated with their selection and use are investigated and critically discussed. Both machine and manual forecasts are used to introduce some fundamental practical issues associated with the challenge that hotels face regarding forecasting accuracy assessments in their revenue management practice. Specifically, this chapter illustrates how, when applied to real-life hotel data, this supposedly straightforward question of machine vs. manual forecasting accuracy can result in unexpected and disturbing contradictory answers. Given the importance of monitoring and assessing hotel forecasting accuracy, this emphasizes the complexity of the task ahead (in chapters 3 and 4), and the urgent need for research in this domain of hotel revenue management forecasting accuracy measures and procedures.

2.2 Literature Review

A multitude of forecasting accuracy measures have been proposed and extensively discussed in the generic (i.e., non-hospitality management specific) forecasting literature (Ahlburg, 1992; Armstrong & Collopy, 1992; Chatfield, 1992; Collopy & Armstrong, 1992; Fildes, 1992; Makridakis, 1993). Aiming to provide forecasting accuracy guidelines based on reliability, validity, outliers, sensitivity,

8 See Appendix C for an overview.

and interpretability, Armstrong and Collopy (1992) preempted this discussion in their empirical study. This early deliberation, however, did not produce any conclusive evidence, and no agreement was reached. Instead, it merely revealed the nuances of the various proposed and tested accuracy assessment methods. New measurements developed to address some of the issues described in their paper also failed to generate consensus as to which forecasting accuracy measure is “the best”.

Whilst failing to identify the ideal solution, these research efforts generated a plethora of forecasting accuracy measures. For example, in the latest ‘Makridakis competition’, i.e., the seminal large-scale M3-competition (Makridakis & Hibon, 2000), forecasting methods were evaluated using the following five different accuracy measures (1) symmetric Mean Average Percentage Error (sMAPE), (2) average ranking, (3) Percentage Better, (4) Median Absolute Percentage Error (MdAPE), and (5) Median Relative Absolute Error (MdRAE). Interestingly, a key outcome of this competition was that the answer to which forecasting *method*⁹ is best *varies* with the accuracy *measure* chosen for evaluation.

In their review of published studies on tourism demand modeling and forecasting, Song and Li (2008, p. 213) challenge the field highlighting that “forecast performance has mostly been evaluated by non-statistical measures such as Mean Absolute Percentage Error (MAPE) and means square percentage error (RSPE),” and they conclude that, “without the support of statistical testing it would be difficult to assert whether one model is significantly better than others.” The studies they reviewed generally do not justify the use of accuracy measures (e.g., Li, Song, et al., 2006; Li, Wong, Song, & Witt, 2006; Song et al., 2010).

A number of accuracy measures are also mentioned in the hospitality forecasting literature. For example, Schwartz and Hiemstra (1997) used MAPE and Mean Absolute Deviation (MAD), Schwartz (1998) used Absolute Percentage Error (APE), Yuksel (2007) used MAPE, MAD and Mean Squared Deviation (MSD), and C. Lim and Chan (2011) used Root Mean Square Error (RMSE). However, the fundamental question of which accuracy measure should be used in different hospitality forecasting tasks and circumstances, was not fully addressed in these publications. A preliminary effort to explore this question (Schwartz, 1999) has not been followed up by hospitality forecasting researchers in recent years. This topic is of utmost importance, because any evaluation of a forecasting method requires proper selection of an accuracy measures and some commonly used forecast accuracy measures have serious limitations. For example, as Armstrong and Collopy (1992) and Fildes (1992) demonstrated, MAPE puts a heavier penalty on forecasts that

exceed the actual than those that underestimate it. In addition, these studies showed that RSME is a scale-dependent measure only applicable to data sets that have the same data scales. Therefore, RMSE should not be used for comparison of forecasting accuracy of hotels of different size.

While the hospitality forecasting literature does not necessarily differ from the general literature in its failure to develop a better understanding of the topic, there is a need to justify the choice of error measures used (Ahlburg, 1992). This is an area of concern as the lodging industry has quickly moved to heavy reliance on accurate forecasts in their revenue management systems. Accordingly, in recent years, considerable effort has been directed to collecting more data, creating and testing new forecasting methods, and designing and implementing forecasting support systems that form an integral part of computerized revenue management systems. Insufficient attention has been paid to how the outcomes of all these forecasting-related efforts should be evaluated, and the cost of this “ignorance” may be high. Indeed, as demonstrated in this chapter, the error measure used may determine which, and by how much, a forecasting method is deemed more accurate. Sub-optimal levels of performance of these tests are bound to impair the performance of the hotels’ efforts to optimize revenue.

The following sections explore and illustrate some of the challenges associated with the adequate use of forecasting accuracy measures, emphasizing the need for hospitality researchers to engage in a systematic long-term effort to assess the different accuracy measures, and to justify their appropriateness for hospitality forecasting conditions and circumstances.

2.3 Method

2.3.1 Data

The data were provided by a large international chain hotel in the Netherlands. It included daily room night forecasts at six market segment levels for a two-year period, from January 2011 to December 2012. Two independent forecasts were generated during the two-year test period: daily occupancy predictions by a commercial (3rd party) RMS, and subjective (manual) assessments of the daily occupancy by the revenue management team. The occupancy for a specific date was forecasted six times before the date. Specifically, the re-forecasting occurred over two months (i.e., two months plus several days), a month, four, three, two and one week ahead. Initially, this study explored a total of 13,283 pairs of RMS-generated forecasts and their corresponding actual figures, as well as 14,245 pairs of human-generated predictions and their corresponding actual figures. For a number of reasons, including the specifics of the research questions addressed in this chapter, missing data, and horizon incompatibility between the RMS and human predictions, the final analysis included only 2,043 pairs of computer forecast/actuals and an identical

⁹ M3-Competition tested 24 forecasting methods classified into six categories: Naïve/Simple, Explicit trend models, Decomposition, ARIMA/ARARMA model, Expert system, and Neural networks.

number of human forecast/actual pairs. In particular, missing observations occurred in July 2011, as Segment one and Segment three's forecasts were redefined by the hotel (redefined in the RMS). This technical change resulted in the loss of forecast data for all previous periods for these two segments. The missing observations are indicated with symbol “-” in Table 1. In addition, while the RMS updated its forecasts daily, the revenue management team updated their forecasts on a weekly basis. Acknowledging this discrepancy in the data, only forecasts produced for the same forecasting horizon by both the RMS *and* the revenue managers were analyzed.

Table 1 presents the results of forecasting error calculations for system and manual forecasts generated one month before arrival. The forecasting accuracy was measured with the MAPE for illustrative purposes.

Table 1: MAPE of system and manual forecasts at segment level (1-month out)

Period	MAPE System							MAPE manual						
	Seg 1	Seg 2	Seg 3	Seg 4	Seg 5	Seg 6	Total	Seg 1	Seg 2	Seg 3	Seg 4	Seg 5	Seg 6	Total
Jan-11	-	#	-	#	#	#	59.3	58.9	#	70.6	#	#	#	44.8
Feb-11	-	68.4	-	#	100.0	#	49.2	40.8	61.7	81.5	#	373.9	#	36.1
Mar-11	-	#	-	#	#	267.5	42.2	24.4	#	73.2	#	#	221.1	25.2
Apr-11	-	66.5	-	#	99.7	#	64.9	31.5	240.6	45.1	#	65.3	#	21.8
May-11	-	158.1	-	#	#	#	65.9	29.8	192.7	58.1	#	#	#	22.6
Jun-11	-	65.4	-	#	#	57.4	47.1	35.3	154.4	39.4	#	#	235.5	22.9
Jul-11	-	24.6	-	#	#	#	64.1	30.2	30.7	27.1	#	#	#	19.9
Aug-11	28.8	48.5	25.9	#	#	#	20.0	24.7	29.5	22.4	#	#	#	23.8
Sep-11	32.6	189.2	36.7	#	#	#	14.1	30.8	162.7	31.9	#	#	#	10.7
Oct-11	36.6	61.2	37.0	#	#	#	175.1	28.2	70.1	41.0	#	#	284.6	22.8
Nov-11	98.2	55.1	48.0	#	#	#	83.5	53.5	47.7	50.5	#	#	57.5	20.5
Dec-11	52.7	34.9	#	#	#	#	28.4	46.1	33.0	#	#	#	#	31.5
Jan-12	43.8	41.4	51.3	#	#	#	30.7	38.5	69.2	126.5	#	#	#	36.4
Feb-12	38.6	81.3	45.3	#	#	#	29.7	38.1	69.0	59.6	#	#	#	27.6
Mar-12	37.6	#	44.0	#	#	#	19.9	35.0	#	31.6	#	#	#	15.5
Apr-12	56.4	177.1	58.1	#	#	#	40.8	51.2	86.0	61.7	#	#	#	29.0
May-12	56.6	63.8	102.7	#	#	115.8	17.6	46.8	88.5	82.4	#	#	76.8	13.5
Jun-12	21.4	#	49.3	#	#	182.0	27.7	21.3	#	39.9	#	#	189.5	23.0
Jul-12	44.4	31.9	13.9	#	#	#	21.4	57.2	29.1	19.5	#	#	#	25.1
Aug-12	25.2	16.7	27.0	#	#	#	15.4	30.0	19.8	19.2	#	#	#	16.3
Sep-12	21.1	59.9	31.7	#	#	26.0	9.1	15.7	174.0	33.7	#	#	35.2	12.8
Oct-12	23.4	133.7	34.5	#	#	40.4	14.4	28.8	155.7	43.9	#	#	54.2	13.1
Nov-12	36.9	62.9	42.2	#	#	70.5	11.8	31.9	57.6	46.9	#	#	67.1	12.4
Dec-12	32.9	73.9	53.9	#	#	#	19.5	24.4	69.4	36.8	#	#	#	16.6

Note. Symbol “-” was used in case of missing data and symbol “#” was used in case the MAPE was incomputable (division by zero).

Despite the fact that MAPE is one of the most used accuracy measures, it is not applicable to all situations (MAPE is incomputable when the actual value equals zero).

2.3.2 Accuracy Measures

Seventeen scale-dependent and independent accuracy measures were selected for this chapter based on a comprehensive review of the generic literature on forecasting accuracy measures. *Scale-dependent* measures are those where the scale of the accuracy measure depends on the scale of data and *scale-independent measures* are those where the scale of the accuracy measures does not depend on the scale of the data. The measures include: Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Error (ME), Mean Absolute Error (MAE), Median Absolute Error (MdAE), Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), Symmetric Mean Absolute Percentage Error (SMAPE), symmetric Median Absolute Percentage Error (sMdAPE), Mean Relative Absolute Error (MRAE), Median Relative Absolute Error (MdRAE), Geometric Mean Relative Absolute Error (GMRAE), Mean Absolute Scaled Error (MASE), Root Mean Square Percentage Error (RMSPE), Root Median Square Percentage Error (RMdSPE), and Relative Mean Absolute Error (RelMAE) also referred to in the literature as Cumulative Relative Absolute Error (CumRAE). See Appendix A and Appendix B where the nature of each measure is briefly described, and their relative advantages and disadvantages are outlined.

2.3.3 Computational Problems

The design of the study was modified in accordance with the nature of the data. These modifications are discussed here because, interestingly, they reflect on the reality of revenue management practices: they underscore the importance of understanding the nature of error measures, and how these measures behave differently. As such, this preliminary analysis serves to illustrate a need for revenue managers to better understand the nature of error measures, and to be cautious when choosing them. The following choices were made.

Consider, for example, the dilemma regarding the appropriate level of data aggregation. Ideally, disaggregated data should be used, in this case segment-level data, because it best reflects the reality of daily occupancy forecasting for hotel revenue management, where the decisions are segment-based and therefore call for segment-level forecast and accuracy assessments. Accordingly, the appropriate design of a cross-error measures study such as this would be to test and illustrate performance by segment. However, as will become evident in this chapter, the reality of functioning property level data suggests that this might not always be feasible due to the nature of error measures. The MAPE measure, one of the most commonly used measures for forecast accuracy in general and in the hospitality management literature in particular, has been computed to contrast human-generated predictions with system forecasts at segment level.

The results that were reported in Table 1 indicate computational issues (i.e., cells are empty). The reason is simple: the formula of the MAPE measure is given by

$$MAPE = \text{mean } |p_t| = \text{mean } \left(\left| \frac{(Y_t - F_t)}{Y_t} \times 100 \right| \right) \quad (1)$$

where p_t is percentage error, Y_t the actual value at period t , F_t the forecast at period t , $t = 1, 2, \dots, n$.

If the actual occupancy of a segment, Y_t , is zero, the MAPE is undefined, that is, it cannot be calculated because of the zero value of its denominator. Further examination of the nature of these segments is thus called for. Segments four and five are small, which explains the large number of (incomputable) days with zero actuals. However, segment six is different as it is the second largest for this hotel, and therefore, in relation to revenue, important. Despite its group-driven large size, segment six has several days with zero occupancy, i.e., when no groups stayed at the hotel. Obviously, this challenge of missing observations (of computed error measures) due to zero daily actuals not only characterizes the MAPE measure, but all forecasting accuracy measures that have a percentage error element. These additional measures include MPE, MdAPE, RMSPE, and RMdSPE. The data analysis shows that all five percentage error-based measures have shortcomings in instances with zero actuals.

There are several ways to work around this issue. The first is to ignore the incomputable periods. The major disadvantage of this strategy is that it disregards a potentially large number of forecast errors because many of these incomputable periods might have had a non-zero forecast, that is, there was a non-zero forecast error for that incomputable period that should have been part of the accuracy assessment. The second solution is to assess the forecasting error on aggregated data, that is, hotel-wide rather than by market segment. However, as noted before, and emphasized by Weatherford and Kimes (2003), the most sophisticated revenue management systems rely on accurate forecasts of demand by length of stay and rate category, that is, on forecast by segment. This is due to price discrimination, the key concept at the core of revenue management policies, is managed by applying controls at segment level, e.g., accept/reject decisions at the rate category/length-of-stay level. Ignoring this level of detail and opting for the aggregate might be detrimental for the hotel's revenue management performance because the accuracy of crucial information is not properly assessed. Finally, the hotel could simply opt to not use any of the percentage-based error measures if the number of incomputable days is large. The risk of this last approach is that, as demonstrated later, it is likely to impact which forecast is deemed most accurate, as different types of forecast error measure approaches generate different results.

2.3.4 Dealing With Multiple Forecasting Horizons

In addition to the previously explained challenge of choosing the appropriate data aggregation level and dealing with incomputable forecast accuracy measures, another important design issue relates to the nature of the forecasting horizon. In a typical revenue management environment, daily occupancy is predicted and re-predicted with a number of forecasting horizons, i.e., the forecast for the same date of stay is reproduced multiple times as time nears the actual date of stay. "Hotels using revenue management typically update forecasts on a daily basis for occupancy dates in the near future (1–2 weeks), and update on a weekly basis for dates farther away (2–8 weeks)" (Weatherford & Kimes, 2003, p. 407). Consider the challenging notion of multiple forecast-generating sources with a varying forecast frequency and time horizons. For example, with the dataset used in this example, the RMS and revenue manager forecast at different frequencies (daily vs. weekly), each for multiple horizons of one, two, three and four weeks-ahead, as well as "month plus", and "two months plus" horizons. Hence, a decision needs to be made about the lead time(s) (i.e., forecasting horizon) at which forecast accuracy should be assessed. Ideally, the hotel would assess all of them. However, for various reasons, this test-all resource-intensive approach is not always feasible. One way of tackling this conundrum is to consider how forecasts are used and to acknowledge that forecasting accuracy (independently from how it is measured) should not be the only forecast performance evaluation criterion. Weatherford and Kimes (2003) suggest that the impact on the hotel's decision-making and financial results must be considered, and Granger and Pesaran (2000, p. 537) noted:

In the real, non-academic world, forecasts are made for a purpose and the relevant purpose in economics is to help decision makers improve their decisions. It follows that the correct way to evaluate forecasts is to consider and compare the realized values of different decisions made from using alternative sets of forecasts. In the academic literature there are frequent mentions of this viewpoint, but few attempts to carry it forward into a practical example. Forecast evaluation, when considered at all, is in terms of statistical accuracy measures of point forecasts, and standard forecasting textbooks do not discuss the decision making aspects.

Accordingly, from a decision-making perspective, the question of time horizon can be treated as follows. Generally, the longer the forecast horizon, the more effective the revenue management reaction to the forecast. For example, if several months in advance it has been predicted that a hotel will face low occupancy in a certain week, the revenue management and the sales teams could use this time to create and

implement strategies to increase hotel occupancy. However, if the forecast is known only a few days in advance, only a few revenue management tactical options, such as adjusting inventory controls, remain as viable options.

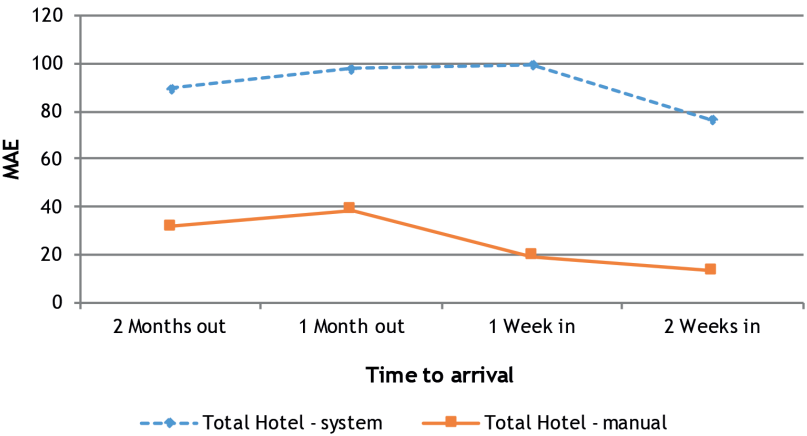


Figure 3: Mean Average Error (MAE) June 2011, all segments

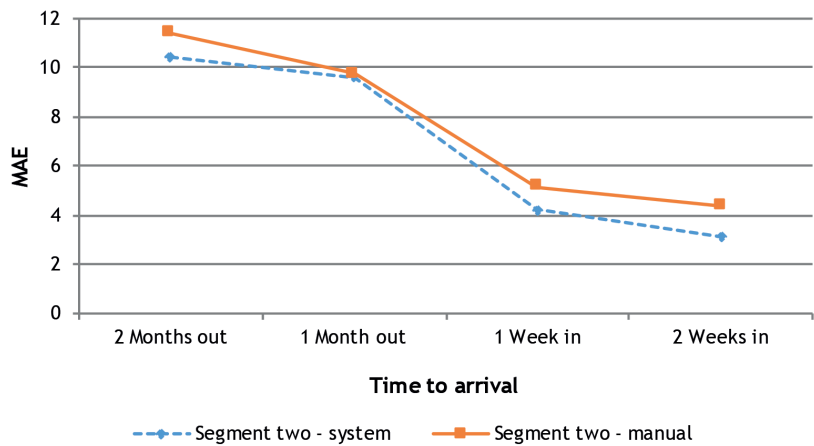


Figure 4: Mean Average Error (MAE) June 2011, at a segment level

How the two issues of aggregation and time horizon have an impact is illustrated by the following analysis of the forecasting data. According to Figure 3, when using Mean Absolute Error (MAE) as the error measure, the computer-generated forecast of June 2011’s occupancy at total hotel level was considerably less accurate than the subjective estimates by the revenue management team. It is also apparent that the

accuracy of revenue managers and the system evolves differently as the forecasting horizon shortens. However, if the accuracy is examined at segment level (see Figure 4), the computer prediction is better with the “2 Months out”, and one and two week horizons, while the revenue management subjective assessment is only comparable in terms of accuracy at the “1 Month out” forecast. As can be observed from Figure 3 and Figure 4, the same accuracy measure can thus give almost opposite results if different levels of analysis are considered (i.e., segment versus total level).

Revenue managers thus need to carefully assess how the forecasting accuracy measures are applied in relation to time and aggregation levels. With no *a priori* rules, the guiding principle should be to consider the practical implications of the forecast, the nature of the derived decisions, and data characteristics.

2.4 Illustration

To ensure a good reflection of a real and practical revenue management environment, the study explored performance of the various daily accuracy measures in relation to three analytical questions. These questions are typical of those asked by hotel revenue managers as part of the monitoring phase of the revenue management cycle. Specifically:

- (1) When are the forecasts most accurate?
- (2) Does forecasting accuracy improve over time?
- (3) Which forecasts are more accurate: the RMS’ or those of the revenue manager?

This results section provides the answers to these three questions. In answering the first question, Table 2 presents the absolute value of the error measure and the resulting accuracy ranking of that month’s forecast compared to the other 23 months, for each month and for each error measure. The numbers paint a picture of inconsistencies. Should the revenue manager wish to know which month was best in terms of forecasting accuracy, that is, in which month the hotel-wide daily occupancy forecast was most accurate, the answer depends on the accuracy measure the manager selects. For example, using the GMRAE and MRAE measures, April 2012 is ranked as the most accurate month, but when using RMSPE or RMdSPE, then April 2012 would rank in position 22 of the 24 months.

When calculating the mean ranking and standard deviations for each month, the inconsistency between measures for April 2012 is confirmed, as well as for a number of other months (e.g., Table 2: November 2011, December 2011, February 2012). As July 2011 shows, a lower standard deviation does not necessarily mean a more accurate forecast. It simply implies that there is less disagreement between accuracy measures. Taking April 2012 as an extreme case and analyzing the high

disagreement measure by measure, different explanations can be found: some are related to the data and their variability; others to the sensitivity of each formula to the data. Thus, it appears that different measures must be used for different situations. In practice, however, this may present complications such as, "how should an international hotel chain compare forecasting performance across different markets if different measurements are selected based on data for each individual property?" Moreover, even for the same property, relatively stable periods will be followed by periods of higher volatility, and forecasting data can reflect this.

Therefore, making certain measurements applicable only for a specific period of time would compromise the consistency of overall forecasting performance evaluation.

Regarding the second question, when comparing the monthly accuracies with the same month of the previous year (e.g., January 2012 vs. January 2011), there are clear indications that forecasting accuracy improves over time. However, on a month-by-month basis, for example, when January 2012 is compared to February 2012, it is evident that according to some measures (MAPE, MdAPE, sMAPE, sMdAPE, MASE, RMdSPE, RelMAE, MSE, RMSE, MAE and MdAE) accuracy improves over time while other error measures (MPE, MRAE, GMRAE, RMSPE and ME) indicate that the forecasting accuracy of occupancy in February deteriorates compared to January. In fact, in only two (July 2011 and October 2012) of the 24 months analyzed (February 2011 – December 2012) do all 17 measures agree with each other.

Answering the third question, Table 3 shows whether the RMS' or the revenue manager's forecast is more accurate. As with the previous accuracy questions, different forecasting accuracy measures give different answers. For example, in August 2011, some error measures indicate that the system outperformed the revenue manager while other measures suggest that the revenue manager was more accurate. Analysis at segment level shows similar or even higher levels of inconsistency across the error measures.

Table 2: Forecasting accuracy measures for "one month plus" segment-generated forecast, total hotel

	Jan-11	Feb-11	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12	Mar-12	Apr-12	May-12	Jun-12	Jul-12	Aug-12	Sep-12	Oct-12	Nov-12	Dec-12
MPE	58.2	47.6	41.0	64.9	65.9	46.8	64.1	9.9	7.0	1.2	4.7	19.0	2.4	28.8	12.5	33.2	7.8	17.0	10.4	1.5	4.8	13.5	1.0	10.3
MAPE	21	20	18	23	24	19	22	9	7	2	5	15	4	16	12	17	8	14	11	3	6	13	1	10
MdAPE	59.3	49.2	42.2	64.9	65.9	47.1	64.1	20.0	14.1	28.5	25.5	28.4	30.7	29.7	19.9	40.8	17.6	27.7	21.4	15.4	9.1	14.4	11.8	19.5
sMAPE	21	20	18	23	24	19	22	9	7	3	14	11	13	16	15	8	17	6	12	10	5	4	2	7
sMdAPE	71.4	54.3	45.2	79.6	68.2	47.5	72.3	15.7	10.7	26.3	24.3	18.7	27.0	25.0	15.6	16.9	14.1	16.2	17.9	11.6	6.3	8.7	6.9	14.9
MASE	22	20	18	24	21	19	23	9	4	16	14	13	17	15	8	11	6	10	12	5	1	3	2	7
RMdSPE	92.7	71.6	58.4	107.0	106.9	69.4	100.2	18.5	13.3	28.9	24.2	22.5	33.2	23.8	18.0	28.3	16.0	21.2	19.7	14.9	9.7	12.7	11.9	22.3
RMdSPE	21	20	18	24	23	19	22	8	4	16	14	12	17	13	7	15	6	10	9	5	1	3	2	11
RelMAE	11.1	74.6	58.4	132.3	103.6	62.2	113.3	14.5	10.2	26.0	25.7	17.7	29.9	22.2	16.4	18.4	14.9	16.2	19.6	12.3	6.3	8.3	6.9	15.0
MRAE	22	20	18	24	21	19	23	6	4	16	15	11	17	14	10	12	7	9	13	5	1	3	2	8
GMRAE	1.2	-	1.9	1.6	7.3	1.4	2.2	1.7	-	1.4	2.1	0.7	0.5	0.6	0.7	0.3	0.5	0.7	0.8	0.7	0.6	1.1	0.6	-
MdRAE	13	-	18	16	21	14	20	17	-	15	19	10	2	6	8	1	3	9	11	7	5	12	4	-
GMRAE	0.8	-	0.6	1.0	1.8	0.8	1.2	0.3	-	0.3	0.4	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.2	0.3	0.3	0.1	-
MdRAE	17	-	16	19	21	18	20	14	-	11	15	5	6	2	4	8	12	13	9	3	7	9	1	-
GMRAE	0.7	-	0.7	1.1	2.0	0.7	1.4	0.0	-	0.4	0.5	0.0	0.0	0.2	0.2	0.0	0.3	0.3	0.3	0.0	0.2	0.2	0.1	-
MdRAE	16	-	17	19	21	18	20	1	-	14	15	1	1	7	9	1	12	11	13	1	8	10	6	-
MdRAE	1.1	1.0	1.2	3.1	2.1	1.7	3.6	0.9	0.8	0.9	1.0	0.7	1.0	0.7	0.7	1.3	0.6	1.2	1.2	0.7	0.6	0.7	0.4	0.8
MdRAE	16	13	19	23	22	21	24	12	10	11	14	5	15	7	4	20	3	18	17	8	2	6	1	9
MdRAE	64.6	54.4	47.7	70.5	69.9	53.9	67.3	24.4	19.4	35.0	32.4	48.6	38.2	39.4	25.5	68.9	23.3	49.6	25.7	20.7	12.2	20.3	16.7	25.7
MdRAE	20	19	15	24	23	18	21	7	3	12	11	16	13	14	8	22	6	17	10	5	1	4	2	9
MdRAE	71.4	54.3	45.2	79.6	68.2	47.5	72.3	15.7	10.7	26.3	24.3	18.7	27.0	25.0	15.6	16.9	14.1	16.2	17.9	11.6	6.3	8.7	6.9	14.9
MdRAE	22	20	18	24	21	19	23	9	4	16	14	13	17	15	8	11	6	10	12	5	1	3	2	7
MdRAE	0.9	0.9	0.9	1.5	2.3	1.6	2.2	0.5	0.6	0.7	0.7	0.5	0.6	0.3	0.4	0.4	0.5	0.8	0.7	0.4	0.4	0.3	0.3	0.5
MdRAE	19	20	18	21	24	22	23	8	13	14	15	11	12	2	4	7	10	17	16	6	3	5	1	9
MdRAE	3686.7	5024.8	5572.8	23979.2	21175.8	13143.3	14605.9	1172.5	1423.4	3573.1	3095.0	801.9	1857.0	771.0	840.0	3153.3	1438.0	4263.0	1566.0	838.0	815.9	1310.5	1214.2	1854.5
MdRAE	17	19	20	24	23	21	22	6	9	16	14	2	13	1	5	15	10	18	11	4	3	8	7	12
MdRAE	60.7	70.9	74.7	154.9	145.5	114.6	120.9	34.2	37.7	59.8	55.7	28.3	43.1	27.8	29.0	56.2	37.9	65.3	39.6	28.9	28.6	36.2	34.8	43.1
MdRAE	17	19	20	24	23	21	22	6	9	16	14	2	13	1	5	15	10	18	11	4	3	8	7	12
MdRAE	49.7	53.1	58.5	133.0	134.2	97.3	114.4	9.9	11.0	11.6	4.1	8.9	6.9	21.3	11.9	26.2	7.4	18.4	7.7	2.1	11.2	25.1	3.7	18.8
MdRAE	18	19	20	23	24	21	22	8	9	11	3	7	4	15	12	17	5	13	6	1	10	16	2	14
MdRAE	52.0	56.3	60.3	133.0	134.2	98.0	114.4	28.1	28.1	48.8	48.4	21.8	29.9	22.3	23.1	43.0	32.9	45.9	34.1	23.2	20.1	27.4	22.4	29.0
MdRAE	18	19	20	23	24	21	22	8	9	17	16	2	11	3	5	14	12	15	13	6	1	7	4	10
MdRAE	45.0	42.0	56.0	119.5	129.0	86.5	109.0	22.0	17.0	43.0	46.5	16.0	20.0	17.0	20.0	41.0	36.0	33.5	33.0	18.0	12.0	20.0	11.5	21.0
MdRAE	18	16	20	23	24	21	22	11	4	17	19	3	7	4	7	15	14	13	12	6	2	7	1	10
Mean ranking	18.71	18.86	18.29	22.41	22.59	19.41	21.94	8.71	6.57	13.76	13.41	8.29	10.88	8.82	7.29	12.82	8.00	13.35	11.53	4.65	3.29	7.12	2.76	9.64
Ranking of means	19	20	18	23	24	21	22	9	4	17	16	8	12	10	6	14	7	15	13	3	2	5	1	11
Standard deviation	2.59	1.99	1.45	2.32	1.33	1.91	1.14	3.53	3.23	3.68	4.12	5.16	5.69	5.93	2.57	5.88	3.32	3.30	2.58	1.87	2.87	3.90	2.08	2.10
Ranking of StdDev's	12	6	3	9	2	5	1	17	14	18	20	21	22	24	10	23	16	15	11	4	13	19	7	8

Table 3: Forecasting accuracy measures for 1-month out system-generated vs. manual forecast, total hotel

	Jan-11	Feb-11	Mar-11	Apr-11	May-11	Jun-11	Jul-11	Aug-11	Sep-11	Oct-11	Nov-11	Dec-11	Jan-12	Feb-12	Mar-12	Apr-12	May-12	Jun-12	Jul-12	Aug-12	Sep-12	Oct-12	Nov-12	Dec-12	
MPE	system	58.2	47.6	41.0	64.9	65.9	46.8	64.1	9.9	7.0	1.2	4.7	19.0	2.4	28.8	12.5	33.2	7.8	17.0	10.4	1.5	4.8	13.5	1.0	10.3
	expert	33.5	26.6	21.7	4.2	5.8	11.5	12.2	15.5	2.7	0.5	1.4	22.6	27.6	24.4	10.2	20.0	0.3	9.7	16.2	7.7	7.1	4.9	4.0	2.4
MAPE	system	59.3	49.2	42.2	64.9	65.9	47.1	64.1	20.0	14.1	28.5	25.5	28.4	30.7	29.7	19.9	40.8	17.6	27.7	21.4	15.4	9.1	14.4	11.8	19.5
	expert	44.8	36.1	25.2	21.8	22.6	22.9	19.9	23.8	10.7	22.8	20.5	31.5	36.4	27.6	15.5	29.0	13.5	23.0	25.1	16.3	12.8	13.1	12.4	16.6
MDAPE	system	71.4	54.3	45.2	79.6	68.2	47.5	72.3	15.7	10.7	26.3	24.3	18.7	27.0	25.0	15.6	16.9	14.1	16.2	17.9	11.6	6.3	8.7	6.9	14.9
	expert	30.7	24.6	17.1	23.3	14.7	13.5	20.3	15.3	3.5	23.1	19.4	21.3	28.8	27.3	13.0	17.9	10.2	12.8	22.9	11.9	6.9	11.9	5.9	14.0
sMAPE	system	92.7	71.6	58.4	107.0	106.9	69.4	100.2	18.5	13.3	28.9	24.2	22.5	33.2	23.8	18.0	28.3	16.0	21.2	19.7	14.9	9.7	12.7	11.9	22.3
	expert	34.3	29.0	20.8	23.0	22.2	19.8	22.7	20.6	11.3	23.9	21.2	25.8	29.1	23.5	14.2	23.6	14.0	19.1	22.3	14.9	11.6	12.8	13.4	17.3
sMDAPE	system	111.1	74.6	58.4	132.3	103.6	62.2	113.3	14.5	10.2	26.0	25.7	17.7	29.9	22.2	16.4	18.4	14.9	16.2	19.6	12.3	6.3	8.3	6.9	15.0
	expert	26.6	25.6	18.8	23.4	15.8	12.6	20.4	14.2	3.4	20.7	19.6	19.3	28.9	24.0	12.2	17.8	10.7	12.8	20.5	11.8	6.6	11.3	6.1	15.0
MRAE	system	1.2	-	1.9	1.6	7.3	1.4	2.2	1.7	-	1.4	2.1	0.7	0.5	0.6	0.7	0.3	0.5	0.7	0.8	0.7	0.6	1.1	0.6	-
	expert	0.6	-	2.2	0.7	0.9	0.5	0.8	0.7	-	1.2	2.3	0.8	0.5	0.7	0.5	0.3	0.7	0.6	0.8	0.7	0.8	1.6	0.6	-
MDRAE	system	0.8	-	0.6	1.0	1.8	0.8	1.2	0.3	-	0.3	0.4	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.2	0.3	0.3	0.1	-
	expert	0.3	-	0.3	0.4	0.4	0.3	0.4	0.3	-	0.2	0.3	0.3	0.5	0.3	0.2	0.2	0.2	0.2	0.4	0.2	0.3	0.3	0.1	-
GMRAE	system	0.7	-	0.7	1.1	2.0	0.7	1.4	0.0	-	0.4	0.5	0.0	0.0	0.2	0.2	0.0	0.3	0.3	0.3	0.0	0.2	0.2	0.1	-
	expert	0.3	-	0.0	0.0	0.4	0.0	0.0	0.2	-	0.3	0.0	0.3	0.2	0.0	0.2	0.2	0.0	0.0	0.4	0.0	0.3	0.3	0.0	-
MASE	system	1.1	1.0	1.2	3.1	2.1	1.7	3.6	0.9	0.8	0.9	1.0	0.7	1.0	0.7	0.7	1.3	0.6	1.2	1.2	0.7	0.6	0.7	0.4	0.8
	expert	0.7	0.6	0.6	1.0	0.6	0.7	1.2	1.1	0.6	0.7	0.8	0.8	1.1	0.8	0.5	1.1	0.5	1.0	1.3	0.8	0.7	0.7	0.4	0.6
RMSPE	system	64.6	54.4	47.7	70.5	69.9	53.9	67.3	24.4	19.4	35.0	32.4	48.6	38.2	39.4	25.5	68.9	23.3	49.6	25.7	20.7	12.2	20.3	16.7	25.7
	expert	61.3	47.5	35.0	26.7	36.8	35.1	24.1	33.0	17.5	27.8	24.9	46.5	49.0	33.3	19.1	43.3	18.5	39.5	29.7	22.3	19.3	16.0	18.8	21.5
RMdSPE	system	71.4	54.3	45.2	79.6	68.2	47.5	72.3	15.7	10.7	26.3	24.3	18.7	27.0	25.0	15.6	16.9	14.1	16.2	17.9	11.6	6.3	8.7	6.9	14.9
	expert	30.7	24.6	17.1	23.3	14.7	13.6	20.3	15.3	3.5	23.1	19.4	21.3	28.8	27.3	13.0	17.9	10.2	12.8	22.9	11.9	6.9	11.9	5.9	14.0
RelMAE	system	0.9	0.9	0.9	1.5	2.3	1.6	2.2	0.5	0.6	0.7	0.7	0.5	0.6	0.3	0.4	0.4	0.5	0.8	0.7	0.4	0.4	0.4	0.3	0.5
	expert	0.6	0.5	0.5	0.7	0.6	0.7	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.4	0.3	0.3	0.4	0.6	0.8	0.4	0.4	0.4	0.3	0.4
MSE	system	3686.7	5024.8	5572.8	23979.2	21175.8	13143.3	14605.9	1172.5	1423.4	3573.1	3099.0	801.9	1857.0	771.0	840.0	3153.3	1438.0	4263.0	1566.0	838.0	815.9	1310.5	1214.2	1854.5
	expert	2061.8	1902.8	1516.5	2892.7	3059.7	2391.4	2138.9	2009.8	1238.6	2781.1	2498.3	969.6	1625.6	851.3	521.3	1856.6	1496.8	2863.0	1821.0	891.4	1245.2	1074.8	1856.2	1038.6
RMSE	system	60.7	70.9	74.7	154.9	145.5	114.6	120.9	34.2	37.7	59.8	55.7	28.3	43.1	27.8	29.0	56.2	37.9	65.3	39.6	28.9	28.6	36.2	34.8	43.1
	expert	45.4	43.6	38.9	53.8	55.3	48.9	46.2	44.8	35.2	52.7	50.0	31.1	40.3	29.2	22.8	43.1	38.7	53.5	42.7	29.9	35.3	32.8	43.1	32.2
ME	system	49.7	53.1	58.5	133.0	134.2	97.3	114.4	9.9	11.0	11.6	4.1	8.9	6.9	21.3	11.9	26.2	7.4	18.4	7.7	2.1	11.2	25.1	3.7	18.8
	expert	14.3	18.4	25.2	15.7	19.9	11.8	21.6	20.9	8.1	8.4	10.8	11.3	18.7	20.8	10.0	15.5	5.8	7.7	16.5	7.9	12.6	9.5	13.2	10.2
MAE	system	52.0	56.3	60.3	133.0	134.2	98.0	114.4	28.1	28.1	48.8	48.4	21.8	29.9	22.3	23.1	43.0	32.9	45.9	34.1	23.2	20.1	27.4	22.4	29.0
	expert	33.4	35.5	30.7	43.1	41.0	38.5	37.2	33.5	21.1	40.9	40.9	25.0	30.8	24.9	18.9	34.7	27.4	39.3	38.2	23.7	24.8	27.1	25.1	21.8
MDAE	system	45.0	42.0	56.0	119.5	129.0	86.5	109.0	22.0	17.0	43.0	46.5	16.0	20.0	17.0	20.0	41.0	36.0	33.5	33.0	18.0	12.0	20.0	11.5	21.0
	expert	25.0	32.0	24.0	43.0	25.0	35.5	35.0	23.0	8.0	33.0	39.5	19.0	25.0	24.0	15.0	28.5	17.0	30.0	38.0	18.0	14.0	25.0	10.0	13.0

2.5 Discussion and Future Research

Correctly evaluating the accuracy of the hotel forecasts is crucial for the effective operation of revenue management systems. These evaluations of forecast performance over time impact which forecasting models are used and the calibration of their components. Another role of accuracy measures is related to user-modifications (i.e. system overrides) of the RMS' predictions, allowing to evaluate whether these modifications improved forecasting accuracy. Given that revenue management software vendors seldom disclose the algorithms they use (Kimes & Wagner, 2001), and given the level of complexity of these forecasting models, it is unsurprising that few revenue managers are fully aware of exactly how their forecasts are generated by the RMS. The resulting so-called 'black box' approach to forecasting (McCarthy, Davis, Golobic, & Mentzer, 2006, p. 322) emphasizes the importance of regularly monitoring forecast accuracy, so as to compensate for the lack of knowledge and understanding of how the forecasting models work. Unfortunately, too little is known and understood about how to assess accuracy, and more specifically the nature of the various forecasting error measures. Therefore, this chapter proposes a critical discussion on the inherent problems of forecasting in hotel revenue management. Unconsidered use of accuracy measures is dangerous practice; the analysis of forecast errors is perhaps as important as the forecast itself.

By using a single hotel data set of machine-generated forecasts, together with human predictions and the actual hotel figures, this chapter demonstrates how detrimental this lack of knowledge and understanding might be. The data and analysis demonstrate that when applied to three typical questions that often confront revenue managers, different error measures generate contradictory answers. This emphasizes the need to better understand the mechanisms behind the evaluations of forecasting outcomes. As long as the limitations of each measure, some of which are illustrated in this chapter, are not fully understood, accuracy can easily be misjudged, leading to a potential cascade of misguided decisions related to pricing, inventory control, operational planning and even strategy formulation. The selection of an appropriate forecasting accuracy measure is a challenging task and involves a strategic capability of hotels (Van der Rest, 2006; Van der Rest & Roper, 2013). The selection of an error measure should consider its "reliability, construct validity, sensitivity to small changes, protection against outliers, and their relationship to decision making" (Armstrong & Collopy, 1992, p. 69). As argued by Ahlburg (1992, p. 99), "where the forecaster knows the loss function applicable to the problem at hand, the choice of forecasting measure is somewhat less difficult". This is seldom the case with hotel forecasting practices, where not enough is understood about the nature of the forecasting error measures, and where there is no clear understanding of the loss function.

Understanding forecasting accuracy measures, however, may only be a first step in the road to improving forecasting performance. A more comprehensive approach should be explored in which the quality of forecasting performance is evaluated. As Murphy (1993) stated in his influential essay, there are three different definitions of forecast quality, and an unambiguous definition of what constitutes a good forecast is crucial for any project designed to evaluate and enhance forecasting performance. Recently, Gilliland (2010) argued for a broader forecasting accuracy evaluation perspective, proposing a Forecast Value Added measure to identify waste and inefficiency in the forecasting process. Hence, the future of forecasting performance in hotel revenue management also includes issues that go beyond accuracy measurement, aiming for a more comprehensive forecasting performance evaluation framework.

In the context of hotel revenue management forecasting, an important component to consider for a forecasting quality evaluation framework is the level of aggregation. This is due to the notion that, from a revenue management perspective, total hotel figures might disguise important segment-level forecasting failures, and as such will limit the ability of the system to optimize based on segment-level controls. A comprehensive framework should also provide an opportunity to measure and identify biases, such as mean bias, defined as a systematic tendency for the forecast to be either less or greater than the actual (Fildes, Goodwin, Lawrence, & Nikolopoulos, 2009). Another important component of this evaluation framework should be a structured approach to evaluate effects of judgmental adjustments. The area of Human-Computer Interaction (HCI) could provide a basis for research on human decision-making to investigate whether there is a tendency to ignore or modify good advice across a number of decision areas, and to demonstrate excessive trust in the decision maker's own judgment. How much the focus should move from the narrow accuracy measures approach to a more structured, comprehensive and consistent framework of forecasting quality assessment is itself a topic for further deliberation.

In sum, the level of aggregation as well as the choice of the accuracy measure are important considerations when evaluating hotel forecasting performance. In doing so, the impact of the time dimension must be understood. The relationship between the forecasting horizon and forecasting accuracy will be examined in the next chapter of this thesis.

Chapter 3

Time Dimension of Hotel Forecast Accuracy Evaluation

Time is what prevents everything from happening at once.
- Albert Einstein -

3.1 Introduction

The forecasting horizon is an important component of revenue management, as both hotels and their customers make time-dependent decisions (C.-C. Chen & Schwartz, 2008). A thorough understanding of how demand is distributed over time in order to accurately forecast hotel occupancy along the forecasting horizon is necessary for optimal hotel revenue management, as daily occupancy predictions are used to formulate room allocation policies and price optimization decisions (Schwartz & Hiemstra, 1997). As Schwartz (2008) emphasizes, timing is important for the optimal functioning of RM algorithms. Yet, very few studies have explored the impact of the time dimension on hotel revenue management forecasting performance. Accordingly, this chapter aims to address this gap by examining the accuracy of non-adjusted and adjusted occupancy forecasts in the context of multiple forecast horizons.

3.2 Literature Review

Hotel occupancy forecasting is a complex and dynamic process where reforecasting over the time horizon continuously takes place. That is, the occupancy forecasts for each single day of the year are constantly updated and regenerated over the forecasting horizon as the date of stay approaches. Many studies have assumed that the shorter the forecasting horizon, the more accurate the forecast (Makridakis, Wheelwright, & Hyndman, 1998; Schwartz & Hiemstra, 1997; Schwartz et al., 2016). This assumption seems reasonable, since the closer the date of arrival, the more

bookings are recorded and more information is available about the arrival date itself. Yet, very little empirical work has been carried out to examine this relationship between forecasting accuracy and forecast horizon.

Much research attention has been devoted to the examination of newly proposed forecasting models rather than on real business forecasts at various forecast horizons (e.g. C. Chen & Kachani, 2007; El Gayar et al., 2011; Haensel & Koole, 2011; Law, 2004). The exception is a study by Kimes (1999) who investigated group forecasting accuracy at various days prior to arrival for approximately¹⁰ 90 hotels of a major North American hotel chain. She reported that both MAPE and MAD (and their standard deviations) decreased as the day of arrival approached. "In general, the MAPE averaged 40% at two months before arrival, dropped to about 30% at one month before arrival, and decreased to 10-15% on the day of arrival" (Kimes, 1999, p. 1106). Initially it would appear that her findings support the assumption that the longer the forecasting horizon, the less accurate the forecasts. However, the generalizability of these results is subject to certain limitations. First, her study only examined the business segment and excluded the examination of transient business, although it is an integral part of hotel operations. Second, the scope of the study was limited to a single hotel chain, and data were collected only for four months. Third, the study was carried out 20 years ago. Since then, RM technology and processes have changed significantly. However, the major limitation of the study lies in the fact that the figures report MAPE aggregated at a monthly level and not for an individual arrival date, while RM optimization decisions are made at the arrival date level. Even though it is interesting to observe that, on average, MAPE decreases as the time of arrival approaches, light still needs to be shed on what happens at the non-aggregated, daily level.

Some studies in hotel revenue management forecasting have examined forecasting models at different forecast horizons. Although they report forecast accuracy at various forecasting horizons, one major drawback of the studies is their use of simulated data for modeling and not the actual forecasts generated by the revenue management system under real hotel business conditions. For example, Schwartz (1998) applied the Curve Similarity model (Schwartz & Hiemstra, 1997) to forecast the occupancies for one hotel for 45 sample days for each of the following forecasting horizons: 1, 2, 3, 7, 14, 21, 30, 45, 60, and 99 days in advance. He evaluated generated forecasts with Absolute Percentage Error (APE) and plotted the data on the chart with the accuracy level (APE) versus the forecasting horizon for each of the 45 sample days. He examined the figures for each day and identified 8 sample days with the worst forecasting performance at a forecasting horizon of 30 days. The figures of these 8 sample days were then smoothed and plotted on the same graph.

¹⁰ The original paper does not report the exact number of hotels.

The author concluded that most curves had a similar shape, with a so-called "saddle" pattern which seemed to be caused by a sharp increase in the number of booked rooms starting around 30 days before the date of stay. He also pointed out that the model accuracy significantly improved when moving from a forecasting horizon of 14 days to 21 days, but then it deteriorated at the 30-day horizon. This evidence suggests that the relationship between forecast accuracy and forecasting horizon is not linear. However, again caution needs to be exercised when generalizing his findings, as the study was performed only for one hotel for 45 days, more than 20 years ago. Another limitation is that the curves are shown only for the worst performing 8 sample days, and no other information is provided about the relationship between forecasting horizon and forecast accuracy for the other 37 sample days.

Rajopadhye, Ben Ghalia, Wang, Baker, and Eister (2001) applied and evaluated the Holt-Winters procedure to forecast hotel room demand. They reported MAD and MAPE for the final two weeks of the simulation period. Their figures showed that MAD and MAPE values fluctuate as the arrival day nears. At 2-4 days before arrival, the forecast accuracy is lower than at 14 days before arrival. Again, this illustrates that the relationship between forecasting horizon and forecast accuracy is non-linear and complex.

In another study, Weatherford and Kimes (2003) provided a detailed overview of the main forecasting methods in Revenue Management and tested seven forecasting methods on data from Choice Hotels and Marriott Hotels. Although the forecasting horizon was not a primary focus of their research, the figures presented in their study indicated that performance of the forecasting methods impacted the forecast error over the forecasting horizon. For example, pick-up and regression forecasting methods had fairly similar forecast errors (MAE) for all days to arrival, while "the performance of the multiplicative booking method deteriorated when used more than 7 days before arrival" (Weatherford & Kimes, 2003, p. 409), confirming the nonlinear time-accuracy relationship.

More recently, Pereira (2016) introduced a new time-series forecasting method (TBAT) to forecast daily hotel room demand. The study compared the proposed model with the other five time series forecasting methods, and evaluated the results through six forecast accuracy measures (i.e., RMSE, MAPE, MdAPE, sMAPE, RelMAE, and RelRMSE) for multi-step forecasts from 1 to 56 days ahead. The relationship between forecasting horizon and forecast accuracy measures revealed a number of complex patterns, where a shorter forecasting horizon was not always associated with a lower forecast error. The study does not provide an explanation for this pattern. A possible explanation for this might be related to the complex seasonal patterns of hotel demand (containing weekly seasonal pattern and annual seasonal pattern).

While there is evidence for the relationship between the forecasting horizon and forecast accuracy, previous work indicates that the time-accuracy association is more complex than often assumed. Therefore, it is hypothesized that:

- H₁: Hotel occupancy forecast accuracy improves as the day of arrival nears.
 H₂: The relationship between forecasting horizon and forecast accuracy has linear and non-linear components.

Even less work has focused on the effect of manual forecast adjustments, although the practice of integrating human judgment to the revenue management system recommendations is common in the hotel industry (Schwartz, 1998; Schwartz & Cohen, 2004a). This practice is important as there is a pervasive belief that additional knowledge beyond 'what the system knows' is a key part of a solid approach to forecasting daily occupancies. Examples of this additional knowledge, exogenous to the system, include special local events, such as:

a large new convention in the destination, a large marketing campaign, the opening of a new tourism attraction, a reduction in the attractiveness of competing tourism attraction due to new circumstances, changes in the carrying capacity of airlines flying to the destination, ongoing or anticipated health/political/climate crisis, ... an abrupt change in hotel reputation or hotel capacity (Schwartz et al., 2016, p. 270).

One approach to incorporating this additional information is to 'inform the computer system', by adding a special event to it and indicating the anticipated impact of this event. Another option (though not necessarily optimal) is to judgmentally adjust the system forecast. Several studies have found support for these approaches; statistical forecasts can be made more accurately when experts judgmentally adjust them (Donihue, 1993; Fildes et al., 2009; McNees, 1990; D. S. Turner, 1990). The underlying cognitive mechanisms behind judgmental adjustments are, however, complex, and research has only touched the tip of the iceberg (Gönül, 2013).

The findings in the generic literature on whether judgmental adjustments improve the forecast are controversial. Although experts can add value, they are also prone to biases. A number of studies have demonstrated that forecasters often make unnecessary judgmental adjustments (Lawrence et al., 2006). Interestingly, training in forecasting methods does not seem to improve forecasters' ability to extrapolate, and forecasters continued to make adjustments even when informed they were reducing accuracy (Goodwin & Wright, 1994; J. S. Lim & Oconnor, 1995). Research on 'big losses' (a judgmental adjustment that significantly decreases the forecasting accuracy) also

identified an interesting phenomenon - after the first big loss, forecasters were still more likely to make another large adjustment, even though it increased the probability of a second big loss (Petropoulos, Fildes, & Goodwin, 2016).

Many of these results were found in the context of SKU (Stock-Keeping Unit) forecast adjustments, and their applicability to hotel occupancy forecasting should be interpreted with caution. There is a substantial difference between hotel revenue management forecasting and SKU-research forecasting. For example, in SKU research the forecast creator and the forecast user are different. Many scholars hold the view that users of the forecast sometimes feel the need to adjust the forecast in order to 'control' the forecast (Gönül, 2013). In hotel revenue management, the manager usually is a co-creator of the forecast, since he/she is involved in setting up forecast parameters by mapping special events in the system and indicating their impact on the forecast. Should substantial and consistent discrepancies be found in forecasting, revenue managers are encouraged to contact the system provider, open a 'case', and work directly with the system provider on investigating inconsistencies and re-calibrating the model (IDeaS, personal communication, July 2016). Therefore, different from managers in an SKU forecasting context, hotel revenue managers might be less inclined to make forecast adjustments solely in a pursuit of gaining the sense of control or ownership.

In the hospitality management literature, little research attention has been paid to judgmental forecast adjustments. Several attempts were made to investigate the role of user interventions in airline forecasting, a business operational context which has many similarities with hotel forecasting. An experiment at Northwest Airlines where a subset of markets was put on autopilot (not allowing human interventions to the forecasts) found that it led to the decrease in revenue. Revenue analysts on average increased revenue by 7.9 per cent (Bach, 1999). At America West airline, a study demonstrated the possibility to identify and correct flights with strong negative bias¹¹ (Loew, 2000). At US Airways, Zeni (2003) estimated that revenue analysts' interventions added up to 3 per cent incremental revenue. A study by Mukhopadhyay et al. (2007) demonstrated that user interventions were valuable, especially when competitors added flights to certain routes. After conducting a review of the history of forecasting models in revenue management, Weatherford (2016) concluded that human judgment in airline revenue management forecasting is still indispensable.

Apart from Kimes (1999), Koupriouchina, Van der Rest, and Schwartz (2014), and Schwartz and Cohen (2004a) far too little attention has been paid to judgmental adjustments in the context of hotel revenue management forecasting. The existing

11 The tendency to forecast too low or under-forecast

studies demonstrated that managerial interventions may be beneficial, although prone to biases, akin to claims in the generic forecasting literature (e.g., Bearden, Murphy, & Rapoport, 2008; Fildes et al., 2009).

Drawing on the theory of ecological rationality, some studies suggest that people use simple heuristics when making decisions (e.g., Gigerenzer, Todd, & ABC Research Group, 1999).

Human minds have evolved to act quickly, making just good enough decisions. This theoretical approach proposes that the mind has adapted its limitations to match the structures of information available in the environment. Instead of trying to process all the available information in the environment and consider all possible options, people often make surprisingly good decisions using simple 'fast and frugal' heuristics (Rogers, 2012, p. 74).

The idea that human decision makers are characterized by limited computational capacity, but are still able to make good decisions using simple heuristics dates back to the concept of "bounded rationality" (Simon, 1955). The power of this intuitive, 'fast and frugal' heuristic is demonstrated in many domains, including sports, crime and business (Goldstein & Gigerenzer, 2009; Kahneman, 2011).

Hotel revenue managers deal with complex information under significant time pressure. They need to review multiple forecasts at total hotel and at customer segment levels, and at various different forecast horizons for numerous days of arrival into the future. They use different formats and sources of information: revenue management system reports, property management system reports, STR and other competitive benchmarking information, rate shopping tools, conversations with Sales Managers, any information about events affecting their market, etc. Moreover, forecasting is only one of the numerous revenue management tasks they need to perform. Therefore, the conditions under which judgmental forecast adjustments are made favor the use of the 'fast and frugal' heuristic. According to Goldstein and Gigerenzer (2009, p. 760), "some of the fast and frugal heuristics that people use intuitively make forecasts that are as good as or better than those of knowledge-intensive procedures".

In addition, research indicates that judgmental adjustments can be beneficial if experts have good knowledge of important influencers (Fildes & Goodwin, 2007). There is some evidence to suggest that revenue managers possess contextual information, such as information about a large group booking that has not been recorded or previous experience with a specific group or one-time event that is likely to affect occupancy (Schwartz, 1998; Schwartz & Cohen, 2004a).

Kimes (1999) investigated the correlation between the average number of forecast updates for each hotel and the group forecast accuracy using four months of group forecast data of a hotel chain in North America. The study found a negative correlation between the number of updates and forecast errors (i.e., more updates were associated with more accurate forecasts) in the week before arrival. Using an ANOVA, it was concluded that the "frequency of updates was associated with increased forecast accuracy from 0-11 days before arrival" (Kimes, 1999, p. 1109).

In sum, the evidence presented in this section suggests that although humans are prone to biases, they still play an important role in the forecasting process and that their judgmental interventions can be beneficial. Therefore, it is hypothesized that:

H₃: Forecasts with a higher frequency of judgmental forecast adjustments have higher forecast accuracy.

3.3 Method

3.3.1 Data

To obtain the data, a major revenue management system provider, serving over 1.6 million hotel rooms in 124 countries, was approached. A research proposal was submitted and was subsequently discussed and approved by the management team. A non-disclosure agreement was signed and a data management plan was created (see Appendix D).

Anonymized data were extracted from the revenue management systems by the company's senior data experts. The extracted data were sent to the researcher and were accompanied by a brief explanation of the extracted fields. Periodic discussions were held with data and system experts to reconfirm the data interpretation and the understanding of the underlying forecasting processes. The data were imported in SPSS (version 23, release 23.0.0.3), and detailed code books were created and shared with the revenue management system provider to reconfirm correct interpretation of extracted data. Logical checks were performed, and all irregularities were discussed during periodic conference calls with the company.

Two distinct datasets were received. The first dataset (hereafter referred to as Dataset one) contained forecast data for the occupancy period from 24 March 2014 to 7 April 2017. The second dataset, covering the same period, contained detailed forecast data only for those occupancy days for which one or more user overrides had taken place. This second dataset (hereafter: Dataset two) was used for the analyses reported in Chapter 4.

Forecast data were extracted at the business type level, separate for group and transient business, at seven forecast horizons: 0, 1, 3, 7, 21, 84, and 252 days before arrival. The choice for these forecast horizons was complex due to the fact that the existing literature does not provide clear indications or justification and previous

studies have chosen the forecasts horizons arbitrarily. Therefore, in conjunction with the RMS system provider’s senior data experts, seven forecast horizons have been chosen to include short-term and long-term perspective. From the 196 variables extracted from the RMS, 15 were found to be applicable for the study.

Hotel_ID	Occupancy_Date	Business_Type	Days_To_Arrival	Forecast_Per_Segment	Rooms_Sold_Per_Segment	Forecast_Day_Has_Overrides	Number_Of_Overrides
2801	1/Feb/16	Transient	0	84	86	No	0
2801	1/Feb/16	Transient	1	81	86	No	0
2801	1/Feb/16	Transient	3	81	86	No	0
2801	1/Feb/16	Transient	7	84	86	No	0
2801	1/Feb/16	Transient	21	88	86	No	0
2801	1/Feb/16	Transient	84	82	86	No	0
2801	1/Feb/16	Transient	252	80	86	No	0
2801	1/Feb/16	Group	0	14	14	No	0
2801	1/Feb/16	Group	1	13	14	No	0
2801	1/Feb/16	Group	3	13	14	No	0
2801	1/Feb/16	Group	7	14	14	No	0
2801	1/Feb/16	Group	21	27	14	No	0
2801	1/Feb/16	Group	84	10	14	No	0
2801	1/Feb/16	Group	252	15	14	No	0

Figure 5: Data structure of Dataset one (partial)

Figure 5 shows an illustration of anonymized forecast data of hotel 2801 for February 1, 2016. Each row represents a separate case. For each hotel at each occupancy date there are fourteen cases (7 measurement occasions × 2 segments). For example, at 252 days before arrival the forecast occupancy for February 1, 2016 was 80 rooms for the transient segment and 15 rooms for the group segment. At 84 days before arrival, the forecasts were 82 (transient) and 10 (group). The actual occupancy on February 1, 2016, however, was 86 rooms for the transient segment and 14 rooms for the group segment. As the *Forecast_Day_Has_Overrides* column shows, there were no user overrides for hotel 2801 for occupancy on February 1, 2016. For a full overview of the variables, see Appendix E.

Table 4 summarizes Dataset one, showing that a total of 20,081,973 cases were analyzed. A case is one entry in the Dataset one as shown in Figure 5. User overrides had an impact on 1.5% of analyzed cases.

Table 4: Source data summary for dataset one

Total number of cases	25,003,446 (100%)
- Total number of incomplete pairs (% of total)	- 1,718,856 (6.87%)
- Total number of additionally excluded cases during data checks (% of total)	- 3,202,617 (12.81%)
= Total number of analyzed cases (% of total)	= 20,081,973 (80.32%)

Dataset one contained forecasting data of 1752 hotels belonging to 232 hotel chains and a variety of hotel types, including city hotels, resorts, airport hotels, and casinos. The hotels were located in seven geographical regions: Europe, North America, Asia, Middle East & Africa, Australasia, Central America and South America, as illustrated in Table 5.

Table 5: Number of hotels by region and type

	Asia		Australasia		Central America		Europe		Middle East & Africa		North America		South America	
	n	Column %	n	Column %	n	Column %	n	Column %	n	Column %	n	Column %	n	Column %
Airport	2	0.7%	1	2.4%	0	0.0%	21	2.1%	0	0.0%	2	0.8%	0	0.0%
Casino	1	0.4%	7	16.7%	0	0.0%	0	0.0%	4	2.5%	15	5.7%	0	0.0%
Casino – Native American	0	0.0%	0	0.0%	0	0.0%	0	0.0%	0	0.0%	2	0.8%	0	0.0%
Casino - Resort	2	0.7%	0	0.0%	0	0.0%	0	0.0%	4	2.5%	5	1.9%	0	0.0%
City - Capital	91	33.3%	4	9.5%	0	0.0%	259	26.0%	43	26.7%	15	5.7%	1	6.3%
City - Major	87	31.9%	14	33.3%	0	0.0%	350	35.2%	64	39.8%	155	59.2%	15	39.8%
City - Tertiary	49	17.9%	6	14.3%	0	0.0%	293	29.4%	5	3.1%	11	4.2%	0	0.0%
Resort - Destination	20	7.3%	9	21.4%	1	33.3%	18	1.8%	15	9.3%	15	5.7%	0	0.0%
Resort - Golf	0	0.0%	0	0.0%	0	0.0%	8	0.8%	3	1.9%	4	1.5%	0	0.0%
Resort - Other	5	1.8%	0	0.0%	0	0.0%	15	1.5%	9	5.6%	9	3.4%	0	0.0%
Resort - Seaside	16	5.9%	1	2.4%	2	66.7%	24	2.4%	14	8.7%	23	8.8%	0	0.0%
Resort - Ski	0	0.0%	0	0.0%	0	0.0%	7	0.7%	0	0.0%	6	2.3%	0	0.0%
Row Total	273	100%	42	100%	3	100%	995	100%	161	100%	262	100%	16	100%

The hotels ranged in size from 6 to 4,028 rooms, with an average size of 242, as illustrated in Table 6. The majority of hotels belonged to the city hotel category.

Table 6: Hotel size by hotel type

Hotel type	n	%	Size			
			Mean	Minimum	Maximum	Standard Deviation
Airport	26	1.5%	319	132	821	147
Casino	27	1.5%	383	77	1960	353
Casino - Native American	2	0.1%	358	301	416	81
Casino - Resort	11	0.6%	1366	140	4028	1433
City - Capital	413	23.6%	273	28	1261	194
City - Major	685	39.1%	242	33	1590	173
City - Tertiary	364	20.8%	162	9	1052	139
Resort – Destination	78	4.5%	230	30	1483	233
Resort - Golf	15	0.9%	218	6	401	117
Resort - Other	38	2.2%	186	25	499	106
Resort – Seaside	80	4.6%	249	30	789	161
Resort - Ski	13	0.7%	271	59	766	220
Row Total	1752	100.0%	242	6	4028	230

3.3.2 Dependent Variables

As illustrated and discussed in detail in Chapter 2, the selection and interpretation of the forecast accuracy measures is a serious challenge in itself. Limitations of forecast accuracy measures have been extensively discussed in the literature. Several studies have demonstrated that the selection of different forecast accuracy measures can lead to contradicting results (Makridakis, 1993; Trapero et al, 2011; Davydenko & Fildes, 2013; Koupriouchina et al, 2015; Pereira, 2016). Still, the selection of accuracy measures is frequently not underpinned by extensive reasoning. For example, Kimes (1999, p. 1106) chose three forecasting accuracy measures (MAPE, MAD and MFE) based on the argument that, "hotel managers are usually interested in knowing the percentage of error, the size of the error (the number of rooms sold), and the direction of error (too high or too low)." However, MAD is not the best measure for this type of study as it is scale dependent, and in her study, Kimes (1999) analyzed hotels ranging in size from 200 rooms to over 1500 rooms.

The justification for the selection of the forecast accuracy measures is based on the analysis conducted in Chapter 2. First, to transparently illustrate the consequence of using different accuracy measures on the interpretation of forecasting results, 17 forecasting accuracy measures were computed (Chapter 2). Of these measures, five were scale-dependent. These measures are described in detail in Appendix A and include (in various forms): Mean Error (ME), Mean Absolute Error (MAE), Median Absolute Error (MdAE), Mean Square Error (MSE), and Root Mean Square Error (RMSE). As can be seen from the formulae reported in Appendix A, all these measures are aggregated measures based on the forecast error e_t

$$e_t = Y_t - F_t \quad (2)$$

where Y_t is an actual value and F_t is a forecast at time t .

However, as illustrated by Figure 5, the analysis was performed on a non-aggregated, hotel/occupancy date/business segment/DTA level, where each case represented a daily forecast for a given property at a specific day to arrival for the transient or group segment. Therefore, this made a decision about the aggregation method for the computation of the accuracy measure irrelevant.

A natural question arose on whether a non-aggregated error measure e_t needed to be considered as a dependent variable at all. On the one hand, this measure was not considered advisable, as it is mostly applicable to datasets with the same data scales, which was not the case (e.g., the size of the hotels varied from 6 to 4028 rooms). Strictly following theoretical grounds and recommendations from the forecasting literature, the scale-independent accuracy measures were the only measures considered as dependent variables. On the other hand, in the field of hotel occupancy forecasting, the forecast error measure (e_t) is still widely used and reported

in studies, even where hotel properties of different sizes are compared. To draw attention to this practice, and to provide an illustration of potential consequences, the non-aggregated error measure e_t was selected (a scale-dependent measure), and analyzed and discussed in the context of other, more suitable accuracy measures.

Since e_t can assume both positive and negative values, Absolute Error (AE) was chosen as the scale-dependent dependent variable:

$$AE_t = |e_t| = |Y_t - F_t| \quad (3)$$

Second, the twelve measures belonging to the scale-independent measures category (Appendix B) were first subdivided into (1) those based on percentage errors p_t ¹²

$$p_t = \frac{(Y_t - F_t)}{Y_t} \times 100 = \frac{100 \times e_t}{Y_t} \quad (4)$$

and (2) those based on the relative errors r_t

$$r_t = \frac{e_t}{E_t} \quad (5)$$

where E_t is the forecast error obtained from the benchmark method; typically a forecast from a naïve (i.e. previous-year) method. As in the case of the scale-dependent measures, there was no need to select the best aggregated measure as the analysis was performed on the non-aggregated data.

Although the disadvantages of percentage-based accuracy measures are commonly known, many researchers continue to use this measure. Therefore, its inclusions allow to place this study in the context of existing studies. The non-aggregated equivalent typically used in other studies is APE (Absolute Percentage Error), which was selected as the first scale-independent dependent variable:

$$APE_t = |p_t| = \left| \frac{(Y_t - F_t)}{Y_t} \right| \times 100 = \left| \frac{100 \times e_t}{Y_t} \right| \quad (6)$$

12 The accuracy measures based on the percentage errors are: Mean Percentage Error (MPE), Mean Absolute Percentage Error (MAPE), Median Absolute Percentage Error (MdAPE), Symmetric Mean Absolute Percentage Error (sMAPE), symmetric Median Absolute Percentage Error (sMdAPE), Root Mean Square Percentage Error (RMSPE), Root Median Square Percentage Error (RMdSPE). The accuracy measures based on the relative errors are: Mean Relative Absolute Error (MRAE), Median Relative Absolute Error (MdRAE), Geometric Mean Relative Absolute Error (GMRAE), Mean Absolute Scaled Error (MASE), and Relative Mean Absolute Error (RelMAE) or Cumulative Relative Absolute Error (CumRAE).

As highlighted in previous studies (e.g., Armstrong & Collopy, 1992; Makridakis, 1993) this measure rewards under-forecasts and is not symmetric. To overcome its limitations, Makridakis (1993) suggested a symmetric equivalent which was selected as the second scale-independent dependent variable:

$$sAPE_t = \left| \frac{(Y_t - F_t)}{\frac{1}{2}(Y_t + F_t)} \right| \times 100 = \left| \frac{200 \times e_t}{(Y_t + F_t)} \right|$$

(7)

However, this “symmetric” measure does not always live up to its name, as illustrated in detail by Koehler (2001). Table 7 presents a simple illustration using fictional data, which extends and corrects the example presented by Koehler (2001).¹³ It shows that for the same size of error, a lower than actual forecast results in a larger sAPE compared to a higher than actual forecast, and the magnitude of this asymmetry depends on the scale of the actual.

Finally, in line with the previous chapter, the absolute relative error based on the last year’s actuals (same day of the week) as a naïve forecast was considered as a scale-independent outcome variable. However, in 34.85% of cases the last year’s actuals were missing, and in 8.41% cases the values were equal to zero, leading to an incomputable relative error measure in 43.26% of all cases. Therefore, this measure was excluded from this project. Instead, the measure proposed by Tofallis (2015), which he referred to as LnQ, was selected as the third scale-independent dependent variable:

$$LnQ_t = \ln \left(\frac{F_t}{Y_t} \right)$$

(8)

The advantage of Tofallis’ measure is that it appears to be symmetric and the results are less skewed. However, it presents the same challenges as sAPE, and for the same size of error, the results are asymmetric. Moreover, it is incomputable whenever either actual or a forecast value is equal to zero, as illustrated in Table 7.

In Table 7, both sAPE and LnQ address the issue related to APE when the forecast and actual values are reverted (as in the first example when APE equals 90% or 900%). However, for the same-size errors reported in the rest of the table, the results are asymmetric; both sAPE and LnQ favor over-forecasting, which could have disastrous consequences for real-world businesses. Moreover, as mentioned before, LnQ is not computable when either forecast or actual are equal to zero.

In sum, due to the lack of one “ideal” forecast accuracy measure, from both a theoretical and practical perspective, four measures were selected as the dependent variables for the study: one scale-dependent (AE) dependent variable and three scale-independent (APE, sAPE and LnQ) dependent variables.

13 Incorrect computations were produced by Koehler (2001) for the lines with the actual equal to 20 and equal to 100. The original calculations are corrected in Table 7.

Table 7: Illustrative comparison of AE, APE, sAPE and LnQ

Actual	Forecast	Forecast accuracy measures (= dependent variables)			
		AE	APE	sAPE	LnQ
100	10	90	90%	164%	-2.30
10	100	90	900%	164%	2.30
100	50	50	50%	66.67%	-0.69
100	150	50	50%	40.00%	0.41
50	0	50	100%	200.00%	Not computable
50	100	50	100%	66.67%	0.69
10	2	8	80%	133.33%	-1.61
10	18	8	80%	57.14%	0.59
20	5	15	75%	120.00%	-1.39
20	35	15	75%	54.55%	0.56
100	130	30	30%	26.09%	0.26
100	70	30	30%	35.29%	-0.36

3.3.3 Independent Variables

Days to Arrival was selected as the first independent variable, this refers to the number of days prior to arrival the forecast was made. Data were collected at seven available forecasting horizons: 0, 1, 3, 7, 21, 84 and 252 days before arrival. A polynomial approach for describing the change in the dependent variables was followed. The linear component was coded as 0, 1, 3, 7, 21, 84 and 252; this approach was selected because it identified the intercept as a hotel’s Absolute Error at DTA 0. For linear change, the slope represents the change occurring between each interval (i.e., between 0 and 1, between 1 and 3, etc.). Moreover, to capture changes (i.e., acceleration or deceleration) in the rate of change over the seven measurement occasions, a quadratic component of DTA (DTA quadratic) was defined and coded as 0, 1, 9, 49, 441, 7056, 63504 correspondingly.¹⁴ To ensure possibility for repeated measures testing in SPSS, an Index variable coded as 1, 2, 3, 4, 5, 6 and 7, for DTA 0, 1, 3, 7, 21, 84 and 252 correspondingly was created.

The frequency of judgmental adjustments (made for one specific day of arrival) was selected as the second independent variable. This continuous variable, which ranged from 0 – 28, was labelled as the number of overrides (*numOVER*).

14 Other higher order polynomials were also coded and tested in the models (results not reported).

In addition, five control variables of which two were at the within-hotel level and three at the hotel level were selected. The within-hotel level variables included business type and special events, coded as *SpecialEvent* (0 = no special event, 1 = special event), *BusinessType* (0 = transient, 1 = group). Special events were included due to their effect on the reliability of forecasts based on historical data, and, thus, accuracy. Inclusion of a business type control variable was considered because of differences in the characteristics of transient and group demand. The hotel-level variables were *HotelSize*, *HotelType*, and *HotelRegion*. *HotelSize* is a hotel's capacity in rooms ranging from 6 – 4028. As *HotelSize* does not have a meaningful zero, it was centered on its grand mean. *HotelType* comprised a variable with aggregated hotel type categories: airport hotels, casino hotels, city hotels, and resort hotels. This variable was coded as a dummy variable with city hotels as the reference category. Hotel size and hotel type were added as hotel-level control variables as previous studies demonstrated that these are determinants of hotel performance (Claver-Cortés, Molina-Azorín José, & Pereira-Moliner, 2007; Pine & Phillips, 2005). *HotelRegion* had several categories: Asia, Australasia, Central America, Europe, Middle East & Africa, North America and South America. It was also coded as a dummy variable, with Europe as the reference category. This variable was included since the development of revenue management might be different in geographical regions, as explained by (Kimes, 2008). Her study found that respondents from Asia were more likely to believe that the various technical aspects of RM were in need of improvement compared to North American respondents. The author suggested that this difference can be explained by slower development of revenue management in Asia compared to North America.

3.3.4 Analytical Approach

A variation of multilevel analysis commonly described as repeated measures design or longitudinal data multilevel analysis (Heck, Thomas, & Tabata, 2014; Hox et al., 2018) was applied. Forecasting accuracy at DTA was considered at level 1, nested within hotels (level 2). Two subscripts to describe hotels (*i*) and occasions of measurement (*t*) were used. It was assumed that the observed status Y_{ti} at time *t* for hotel *i* is a function of a systematic change trajectory plus random error.

The analysis strategy followed recommendations made by Hox et al. (2018), who outlined five major steps in building multilevel models. In the following section, these steps and relevant model formulations are explained.

Step 1

In this step, a model without explanatory variables, the *intercept-only model*, was built. This Model 1 permits testing whether the grand-mean intercept for forecast error (AE, APE, sAPE, LnQ) varies across hotels. It was defined at level-1 without the time-related independent variable:

$$Y_{ti} = \pi_{0i} + \varepsilon_{ti} \quad (9)$$

where π_{0i} is the average forecast error across the seven occasions, and ε_{ti} represents errors in predicting the average forecast error for hotels. Between hotels, the pooled intercept was described as

$$\pi_{0i} = \beta_{00} + u_{0i} \quad (10)$$

where β_{00} is the intercept describing the average initial mean at DTA 0 across hotels, and u_{0i} is the level-2 random component associated with describing differences in average forecast error between hotels. Combining (9) and (10) results in:

$$Y_{ti} = \beta_{00} + u_{0i} + \varepsilon_{ti} \quad (11)$$

Three parameters were thus estimated for Model 1: the fixed effect describing average forecast error (β_{00}), the between-hotel random variance (u_{0i}), and the level-1 residual variance (ε_{ti}).

This intercept-only model gives an estimate of the intra-class correlation (ICC), calculated as:

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_e^2} \quad (12)$$

The repeated measures (level-1) variance was estimated as 665.36, and the hotel-level (level-2) variance as 127.84 (Wald Z = 29.570, $p < .001$). Using Equation 12, the intra-class correlation or the proportion of variance at the hotel level was estimated for AE as $\rho = 127.84/(665.36+127.84)=0.16$ or 16.12%¹⁵, which suggests that there was sufficient variation in intercepts across hotels. About one-sixth of the variance was between hotels across time, and about five-sixths was within hotels across time. These results suggested that multilevel modeling was an adequate approach for the project's data, and, therefore, multilevel models should be developed. The

15 See Table 10.

intra-class correlation (ICC) has to be interpreted with caution; it can be imprecise as for the repeated measures design it was acknowledged that the within hotel variance could differ per measurement occasion (Heck et al., 2014).

Step 2

In this step, the three lower-level explanatory variables (DTA, DTA quadratic, *numOVER*) were added to the model. Starting with DTA, this variable was added as fixed meaning that the DTA coefficient is not allowed to differ across hotels. The predictors were added one at a time and the contribution of each individual level explanatory variable was assessed.¹⁶

Model 2 was the first to be developed; it includes the time-related level-1 variable DTA given by the equation:

$$Y_{ti} = \pi_{0i} + \pi_{1i}DTA_{ti} + \varepsilon_{ti} \quad (13)$$

In this modeling step the time-related component was treated as fixed

$$\pi_{1i} = \beta_{10} \quad (14)$$

Substituting Equations 10 and 14 into Equation 13 resulted in the following combined equation:

$$Y_{ti} = \beta_{00} + \beta_{10}DTA_{ti} + u_{0i} + \varepsilon_{ti} \quad (15)$$

Four parameters were estimated for Model 2, including two fixed effects, one random intercept (u_{0i}), and the level-1 residual (ε_{ti}).

Subsequently, Model 3 was developed in which the time-related quadratic term (DTA quadratic) was added to Model 2. The Model 3 was developed to test whether both the linear and the quadratic polynomial were significant in explaining change in forecast error, and is described with the equation for level-1 written as:

$$Y_{ti} = \pi_{0i} + \pi_{1i}DTA_{ti} + \pi_{2i}DTA_{ti}^2 + \varepsilon_{ti} \quad (16)$$

In this modeling step both time-related components were treated as fixed between hotels:

$$\pi_{1i} = \beta_{10} \quad (17)$$

¹⁶ Not all models are reported; only the most important models are described.

$$\pi_{2i} = \beta_{20} \quad (18)$$

Substituting Equations 10, 17, and 18 into Equation 16 resulted in the following combined equation for Model 3:

$$Y_{ti} = \beta_{00} + \beta_{10}DTA_{ti} + \beta_{20}DTA_{ti}^2 + u_{0i} + \varepsilon_{ti} \quad (19)$$

Five parameters were estimated for Model 3, including three fixed effects, one random intercept (u_{0i}), and the level-1 residual (ε_{ti}).

The coefficient of the quadratic component was significant but very small. For reasons of parsimony, it was excluded.

As part of the Step 2 framework three additional models were developed to include within-hotel level explanatory variables (level-1 IVs).

In Model 4, the effect of the number of overrides on the forecasting accuracy was tested. Model 4 was formulated as:

$$Y_{ti} = \pi_{0i} + \pi_{1i}DTA_{ti} + \pi_{2i}numOVER_{ti} + \varepsilon_{ti} \quad (20)$$

In this modeling step, the time (DTA) and number of overrides (*numOVER*) components were treated as fixed.

$$\pi_{1i} = \beta_{10} \quad (21)$$

$$\pi_{2i} = \beta_{20} \quad (22)$$

Substituting Equations 10, 21 and 22 into Equation 20 resulted in the following combined equation:

$$Y_{ti} = \beta_{00} + \beta_{10}DTA_{ti} + \beta_{20}numOVER_{ti} + u_{0i} + \varepsilon_{ti} \quad (23)$$

In the next model, additional level-1 control variables were added including Special Event and Business Type. The formulation of Model 5 therefore included four level-1 predictors:

$$Y_{ti} = \pi_{0i} + \pi_{1i}DTA_{ti} + \pi_{2i}numOVER_{ti} + \pi_{3i}SPECIAL_{ti} + \pi_{4i}BusinessType_{ti} + \varepsilon_{ti} \quad (24)$$

In this modeling step, the components related to special events and business type were treated as fixed.

$$\pi_{3i} = \beta_{30} \quad (25)$$

$$\pi_{4i} = \beta_{40} \quad (26)$$

Substituting Equations 10, 21, 22, 25 and 26 into Equation 24 resulted in the following combined equation:

$$Y_{ti} = \beta_{00} + \beta_{10}DTA_{ti} + \beta_{20}numOVER_{ti} + \beta_{30}SPECIAL_{ti} + \beta_{40}BusinessType_{ti} + u_{0i} + \varepsilon_{ti} \quad (27)$$

Seven parameters were estimated for Model 5, including five fixed effects, one random intercept (u_{0i}), and the level-1 residual (ε_{ti}).

An additional model was formulated to explore the interactions of these level-1 variables. No published work explores the possibility of interaction effects between the level-1 variables. Our analysis of these interaction effects is of descriptive nature. Since during special events the demand patterns might be different, the relationship between the number of overrides and the improvement in forecasting accuracy may be different depending on the presence or absence of a special event. Also, different business segments may have different demand characteristics, which could impact the relationship between the number of overrides and the improvement in forecasting accuracy. Therefore, Model 6 included the level-1 predictors and the level-1 interaction terms as follows:

$$Y_{ti} = \pi_{0i} + \pi_{1i}DTA_{ti} + \pi_{2i}numOVER_{ti} + \pi_{3i}SPECIAL_{ti} + \pi_{4i}BusinessType_{ti} + \pi_{5i}numOVER_{ti} \times SPECIAL_{ti} + \pi_{6i}numOVER_{ti} \times BusinessType_{ti} + \varepsilon_{ti} \quad (28)$$

The components for the interaction terms were treated as fixed.

$$\pi_{5i} = \beta_{50} \quad (29)$$

$$\pi_{6i} = \beta_{60} \quad (30)$$

Substituting Equations 10, 21, 22, 25, 26, 29 and 30 into Equation 28 resulted in the following level-1 combined equation:

$$Y_{ti} = \beta_{00} + \beta_{10}DTA_{ti} + \beta_{20}numOVER_{ti} + \beta_{30}SPECIAL_{ti} + \beta_{40}BusinessType_{ti} + \beta_{50}numOVER_{ti} \times SPECIAL_{ti} + \beta_{60}numOVER_{ti} \times BusinessType_{ti} + u_{0i} + \varepsilon_{ti} \quad (31)$$

Nine parameters were thus estimated for Model 6, including seven fixed effects, one random intercept (u_{0i}), and the level-1 residual (ε_{ti}).

Step 3

In this step the higher-level control variables were added to the level-1 model. In Model 7 the three level-2 variables (hotel size, hotel type and hotel geographical region) were included:

$$Y_{ti} = \pi_{0i} + \pi_{1i}DTA_{ti} + \pi_{2i}numOVER_{ti} + \pi_{3i}SPECIAL_{ti} + \pi_{4i}BusinessType_{ti} + \pi_{5i}numOVER_{ti} \times SPECIAL_{ti} + \pi_{6i}numOVER_{ti} \times BusinessType_{ti} + \varepsilon_{ti} \quad (32)^{17}$$

$$\pi_{0i} = \beta_{00} + \beta_{01}HotelSize_i + \beta_{02}HotelType_i + \beta_{03}HotelRegion_i + u_{0i} \quad (33)$$

Substituting Equations 21, 22, 25, 26, 29, 30 and 33 into Equation 32 resulted in the following equation:

$$Y_{ti} = \beta_{00} + \beta_{10}DTA_{ti} + \beta_{20}numOVER_{ti} + \beta_{30}SPECIAL_{ti} + \beta_{40}BusinessType_{ti} + \beta_{50}numOVER_{ti} \times SPECIAL_{ti} + \beta_{60}numOVER_{ti} \times BusinessType_{ti} + \beta_{01}HotelSize_i + \beta_{02}HotelType_i + \beta_{03}HotelRegion_i + u_{0i} + \varepsilon_{ti} \quad (34)$$

The hotel-level variable *HotelSize* was a continuous variable, the other two higher-level variables were categorical and coded as dummy variables. The *HotelType* variable had four categories in total, with city hotels as the reference category, and three dummy variables were created. The *HotelRegion* variable was a categorical variable with seven categories, with Europe as the reference category. In total six dummy variables were created and six parameters were estimated for β_{03} . As was explained in Model 6, seven parameters were estimated for level-1 variables. This suggest that in total nineteen parameters were estimated for Model 7, including seventeen fixed effects, one random intercept (u_{0i}), and the level-1 residual (ε_{ti}).

Step 4

In this step it was assessed whether DTA had a significant variance component between the groups. In particular, Model 8 helped to assess whether forecast errors evolved differently over time for the different hotels. This *random coefficient model* was formulated as:

$$Y_{ti} = \pi_{0i} + \pi_{1i}DTA_{ti} + \pi_{2i}numOVER_{ti} + \pi_{3i}SPECIAL_{ti} + \pi_{4i}BusinessType_{ti} + \pi_{5i}numOVER_{ti} \times SPECIAL_{ti} + \pi_{6i}numOVER_{ti} \times BusinessType_{ti} + \varepsilon_{ti} \quad (35)$$

The above formulation was exactly the same as in Model 7, but π_{1i} changed and included the random part u_{1i} . Equation 21 was re-written as:

¹⁷ The equations for π_{1i} , π_{2i} , π_{3i} , π_{4i} , π_{5i} , π_{6i} were the same as in Model 6.

$$\pi_{1i} = \beta_{10} + u_{1i} \quad (36)$$

The equations for π_{0i} , π_{2i} , π_{3i} , π_{4i} , π_{5i} , π_{6i} remained as in Model 7. Substituting Equations 22, 25, 26, 29, 30, 33 and 36 into Equation 35 resulted in the following equation:

$$Y_{ti} = \beta_{00} + \beta_{10}DTA_{ti} + \beta_{20}numOVER_{ti} + \beta_{30}SPECIAL_{ti} + \beta_{40}BusinessType_{ti} + \beta_{50}numOVER_{ti} \times SPECIAL_{ti} + \beta_{60}numOVER_{ti} \times BusinessType_{ti} + \beta_{01}HotelSize_i + \beta_{02}HotelType_i + \beta_{03}HotelRegion_i + u_{1i}DTA_{ti} + u_{0i} + \varepsilon_{ti} \quad (37)$$

This equation suggested that the intercept and linear component (DTA) vary randomly across hotels (i.e. hotels vary in forecasting accuracy and the coefficient of DTA varies across hotels).

In total, twenty-one parameters were estimated for Model 8: seventeen fixed effects, one random intercept (u_{0i}), the hotel-level residual of the slopes of the DTA (u_{1i}), the level-1 residual (ε_{ti}), and the covariance between the intercept and the slope (u_{01}).

Step 5

In this final step, the cross-level interaction between *HotelSize* (level-2 control variable) and DTA (the level-1 explanatory variable with significant slope variation in Step 4) was added to the model. This model was formulated as:

$$Y_{ti} = \pi_{0i} + \pi_{1i}DTA_{ti} + \pi_{2i}numOVER_{ti} + \pi_{3i}SPECIAL_{ti} + \pi_{4i}BusinessType_{ti} + \pi_{5i}numOVER_{ti} \times SPECIAL_{ti} + \pi_{6i}numOVER_{ti} \times BusinessType_{ti} + \varepsilon_{ti} \quad (38)$$

Equation 36 was re-written as:

$$\pi_{1i} = \beta_{10} + \beta_{11}HotelSize_i + u_{1i} \quad (39)$$

Through the substitution, the following equation for Model 9 was obtained:

$$Y_{ti} = \beta_{00} + \beta_{10}DTA_{ti} + \beta_{20}numOVER_{ti} + \beta_{30}SPECIAL_{ti} + \beta_{40}BusinessType_{ti} + \beta_{50}numOVER_{ti} \times SPECIAL_{ti} + \beta_{60}numOVER_{ti} \times BusinessType_{ti} + \beta_{01}HotelSize_i + \beta_{02}HotelType_i + \beta_{03}HotelRegion_i + \beta_{11}DTA_{ti}HotelSize_i + u_{1i}DTA_{ti} + u_{0i} + \varepsilon_{ti} \quad (40)$$

In total, twenty-two parameters were estimated for Model 9: eighteen fixed effects, one random intercept variance, the hotel-level residual of the slope of the DTA (u_{1i}), the level-1 residual variance, and the covariance between the intercept and the slope.

Scaled identity was selected as the level-1 covariance structure, and unstructured as level-2 covariance structure. To evaluate all nine models, improvements in the model fit were assessed with reference to a deviance statistic (-2 loglikelihood ratio), that is distributed as χ^2 with associated degrees of freedom, and by examining the Akaike and Bayesian Information Criteria (AIC and BIC).

3.4 Results

The results are presented in three subsections. The first includes the descriptive statistics. The second provides the results of the multilevel models for the scale-dependent dependent variable: Absolute Error (AE). In the third, the results of the scale-independent forecast accuracy measures APE, sAPE, and LnQ are reported. This last section focuses on the final model (Model 9) for all forecast accuracy measures (dependent variables), with special emphasis on analyzing the differences between the results of the AE, APE, sAPE, and LnQ models. The reporting approach for multilevel analysis follows recommendations made by Hox et al. (2018) and Dedrick et al. (2009).

3.4.1 Descriptive Statistics

Figure 6 illustrates the relationship between the mean forecast error and forecasting horizon for all forecasts. It can be observed that on average forecasting accuracy improves over time.

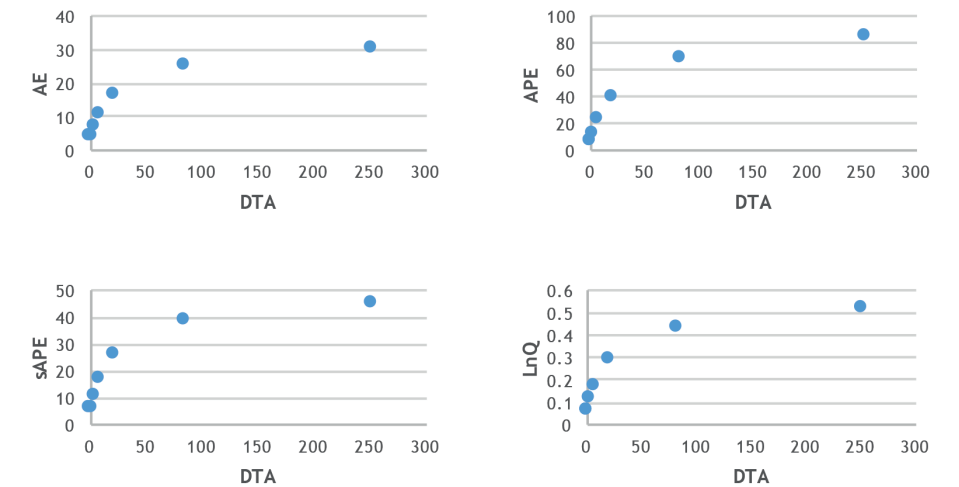


Figure 6: Relationship between mean forecast accuracy and DTA

Overall, from Figure 6 it can be observed that the forecasting error in general decreases over time and the pattern is non-linear, which provides some support for Hypothesis 1 (hotel occupancy forecast accuracy improves as the day of arrival nears) and Hypothesis 2 (the relationship between forecasting horizon and forecast accuracy has linear and non-linear components).

Table 8 shows the mean accuracy errors for different values of DTA. Thus, Table 8 reports the breakdown of forecast errors per days to arrival (DTA) for all accuracy measures, with and without overrides. Standard deviation values indicate that the data for AE, APE and sAPE show large dispersion.

Table 8: Means and standard deviations of AE, APE, sAPE and LnQ by DTA

	DTA						
	0	1	3	7	21	84	252
AE							
No Overrides	4.27 (16.65)	4.11 (9.09)	6.87 (13.3)	10.39 (17.39)	16.28 (24.57)	25.24 (36.47)	29.99 (45.58)
Overrides	11.4 (41.87)	9.46 (26.5)	12.81 (31.28)	16.05 (35.05)	23.02 (43.43)	34.49 (53.64)	42.47 (68.31)
All	4.37 (17.31)	4.19 (9.6)	6.96 (13.75)	10.47 (17.79)	16.37 (24.96)	25.38 (36.8)	30.18 (46.02)
APE							
No Overrides	6.26 (40.79)	6.99 (39.59)	12.96 (62.44)	22.19 (102.79)	39.79 (175.25)	68.82 (279.38)	85.52 (358.39)
Overrides	9.12 (63.61)	9.09 (64.9)	11.68 (68.06)	14.76 (72.71)	22.84 (96.15)	41.08 (176.02)	52.54 (235.95)
All	6.30 (41.22)	7.02 (40.08)	12.94 (62.53)	22.08 (102.42)	39.54 (174.36)	68.41 (278.15)	85.03 (356.91)
sAPE							
No Overrides	5.86 (16.62)	6.06 (12.14)	10.38 (17.76)	16.26 (24.18)	25.96 (32.36)	38.73 (40.43)	44.48 (43.35)
Overrides	7.23 (17)	6.62 (13.21)	9.07 (15.2)	11.78 (18.27)	17.81 (25.01)	27.48 (34.07)	32.16 (37.13)
All	5.88 (16.62)	6.07 (12.16)	10.36 (17.72)	16.19 (24.11)	25.84 (32.28)	38.57 (40.36)	44.29 (43.29)
LnQ							
No Overrides	0.06 (0.23)	0.06 (0.14)	0.11 (0.21)	0.17 (0.3)	0.28 (0.41)	0.44 (0.55)	0.51 (0.6)
Overrides	0.08 (0.21)	0.07 (0.16)	0.09 (0.18)	0.12 (0.22)	0.19 (0.32)	0.3 (0.45)	0.36 (0.51)
All	0.06 (0.23)	0.06 (0.14)	0.11 (0.21)	0.17 (0.3)	0.28 (0.41)	0.43 (0.55)	0.51 (0.6)

Note. Standard deviations are in parentheses.

For all cases (with and without overrides) the average AE for the first occasion (DTA = 252 days before arrival) was 30.18, while for the last occasion (DTA = 0 days) it was 4.37. This indicates a considerable improvement in forecast accuracy over time. Examining the means more closely, it can be noticed that the mean AE decreases over time from DTA 252 to DTA 1, and then slightly increases from DTA 1 to DTA 0. Also, standard deviations for all accuracy measures decrease from DTA 252 to DTA 1, while they increase from DTA 1 to DTA 0.

If the MAE are compared for forecasts with and without overrides, those without overrides are, on average, more accurate at all forecast horizons. It is important to bear in mind that AE is a scale-dependent measure and should not be used for comparing hotels of different size. Therefore, this descriptive finding has to be interpreted with caution.

Table 9: Forecast errors by hotel type

Hotel type	AE			APE			sAPE			LnQ		
	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean	Min	Max
Airport	17.78	0	502	30.33	0	15900	147.26	0	198.58	0.23	0	5.63
Casino	19.68	0	984	33.56	0	15400	183.54	0	197.71	0.22	0	5.16
Casino - Native American	17.05	0	224	33.83	0	4600	143.78	0	195.12	0.24	0	4.39
Casino - Resort	74.59	0	3661	47.47	0	110400	525.92	0	199.68	0.24	0	7.13
City - Capital	15.21	0	844	31.45	0	36300	172.69	0	198.90	0.22	0	5.90
City - Major	13.85	0	1270	35.54	0	48100	195.80	0	199.43	0.24	0	6.56
City - Tertiary	10.95	0	782	35.69	0	20800	182.55	0	198.90	0.24	0	5.89
Resort - Destination	12.14	0	1203	30.76	0	17500	169.04	0	198.92	0.22	0	5.91
Resort - Golf	12.33	0	333	45.78	0	20600	306.76	0	198.08	0.24	0	5.33
Resort - Other	11.70	0	377	37.70	0	35600	221.88	0	198.88	0.25	0	5.88
Resort - Seaside	12.56	0	411	33.80	0	14500	203.67	0	198.94	0.21	0	5.93
Resort - Ski	12.86	0	652	31.76	0	20000	180.94	0	198.02	0.21	0	5.30

If the scale-independent measures APE, sAPE and LnQ measures are considered, the results are considerably different. Forecasts with overrides are, on average, more accurate than forecasts without overrides at forecasting horizons > 1 (at DTA 3, 7, 21, 84, 252). For the same measures at DTA 0 and DTA 1, forecasts without overrides are, on average, more accurate.

All in all, it can be observed from the scale-independent accuracy measures reported in Table 8 that for the majority of the DTA points, forecasts with overrides are, on average, more accurate than forecasts without overrides, which provides some support for Hypothesis 3 (forecasts with a higher frequency of judgmental forecast adjustments have higher forecast accuracy).

A break-down of the forecast accuracies for each hotel type is reported in Table 9. The hotel type will be included in the analysis to initially explore the role of hotel type in forecast accuracy.

3.4.2 Multilevel Analysis of the Scale-Dependent Measure

In this section, the results of the nine multilevel models for Absolute Error (AE) are presented in Table 10.

Model 1 estimated the repeated measures (level-1) variance as 665.36, and the hotel-level (level-2) variance as 127.84 (Wald $Z = 29.570$, $p < .001$), which suggests that there was sufficient variation in intercepts across hotels. The total Absolute Error variance was 793.2. The interpretation of this model and its implications for analysis are explained in Step 1 of the Analytical Approach of the present chapter.

In Model 2, the time variable (DTA) was added as a linear predictor with the same coefficient for all hotels. The model predicted a value of 7.93 for DTA=0, and the error increases by 0.10 on each succeeding occasion. It means that the Absolute Error increases when DTA increases, and decreases when DTA decreases (closer to the arrival date). As shown in Table 10, when the time-related parameters are included in the initial model, the within-hotel variance is reduced from 665.36 to 595.75. The estimate of between-hotel variance is also reduced from 127.84 to 127.01.

Table 10: Results of the multilevel regression models for Absolute Error

Model:	Step 1: Null model		Step 2: Level-1 predictors		Step 3: Level-2 predictors		Step 4: Random coefficient	Step 5: Cross-level interactions	
	M_0 : Intercept only		M_1 : DTA	M_2 : DTA and DTA ²	M_3 : number of overrides	M_4 : Level-1 model with three predictors	M_5 : final Level-1 and 2 model	M_6 : Random DTA	M_7 : Cross-level interaction of DTA and Hotel Size
Fixed part									
Intercept	12.98***	(0.27)	7.93***	5.32***	(0.27)	7.88***	(0.27)	7.88***	(0.11)
DTA			0.10***	0.32***	(6.4E-5)	0.10***	(6.3E-5)	0.09***	(2.2E-3)
DTA quadratic				-0.001***	(1.1E-6)				
numOVER					1.13***	0.83***	(0.01)	0.88***	(0.01)
SPECIAL						0.57***	(0.02)	0.60***	(0.02)
numOVER×SPECIAL						-4.85***	(0.01)	-4.87***	(0.01)
numOVER×BusinessType						-0.44***	(0.03)	-0.44***	(0.03)
Hotel Size						0.09*	(0.04)	0.09*	(0.04)
Hotel Type						0.05***	(4.4E-4)	0.02***	(3.7E-4)
Airport vs City						1.70*	(0.75)	2.77***	(0.63)
Casino vs City						1.08	(0.65)	1.18*	(0.55)
Resort vs City						-1.27***	(0.29)	-2.01***	(0.24)
Hotel Region									
Asia vs Europe						2.74***	(0.27)	4.03***	(0.23)
Australasia vs Europe						-1.56*	(0.61)	-1.15*	(0.51)
Central America vs Europe						-5.55*	(2.21)	-2.39	(1.86)
Middle East & Africa vs Europe						3.61***	(0.33)	3.83***	(0.28)
North America vs Europe						-0.29	(0.28)	0.01	(0.24)
South America vs Europe						1.05	(0.96)	1.30	(0.81)
DTA×Hotel Size									3.6E-4*** (3.5E-6)
Random part									
σ_e^2	665.36***		595.75***	577.51***	595.47***	589.67***	589.66***	525.43***	525.43***
σ_{u0}^2	127.84***		127.01***	126.71***	126.40***	132.46***	14.31***	11.21***	10.25***
σ_{u1}^2								0.01***	1.1E-3***
σ_{u01}^2								0.10***	0.01***
-2*log likelihood									
AIC	187,542,836.90		185,323,783.8	184,699,517.7	185,314,300.8	183,493,533.1	183,493,485.1	181,205,701.5	181,202,299.9
BIC	187,542,840.90		185,323,787.8	184,699,521.7	185,314,304.8	183,493,537.1	183,493,489.1	181,205,709.5	181,202,307.9
	187,542,870.50		185,323,817.4	184,699,551.4	185,314,334.4	183,493,566.8	183,493,518.7	181,205,768.7	181,202,367.2

Note. Standard errors are in parentheses.
* $p < .05$, ** $p < .01$, *** $p < .001$

The results of Model 2 for AE support Hypothesis 1, which states that forecast accuracy improves as the date of arrival nears.

Model 3 added a higher order polynomial trend for time to the model. For the linear change rate (π_{1i}), the estimate (β_{10}) was 0.32 points per day. For the quadratic change rate (π_{2i}), the estimate (β_{20}) was -0.001, which indicates that, on average, hotels' change rates slow over time. Both the linear and quadratic DTA terms were significant. Table 10 shows that including the linear and quadratic polynomial in the model reduced the within-hotel variance further, from 595.75 to 577.51. The estimate of between-hotel variance was also reduced from 127.01 to 126.71. Both the linear and quadratic polynomial were significant in explaining change in Absolute Error. The turning point for DTA at which the relationship between DTA and Absolute Error switches from positive to negative was 160 days $\left(\frac{-0.32}{2 \times (-0.001)}\right)$.

Other higher-order polynomials (cubic, quartic, etc.) were tested and they were all significant. These findings support Hypothesis 2, which stated that the relationship between time and forecast accuracy is complex and has both linear and non-linear components. However, as the quadratic, cubic and other higher terms were significant, but with very small coefficients, for the reasons of model parsimony and ease of interpretation, further modeling was continued only with the linear term.¹⁸

Model 4 included the number of overrides as a predictor to test the effect of user overrides on forecasting accuracy. Table 10 shows that the number of overrides has a significant relationship with AE ($p < .001$), and for number of user overrides (π_{2i}), the estimate (β_{20}) is 1.13. As this estimate is positive, it indicates that every additional override led, on average, to an increase in Absolute Error by 1.13. Therefore, the results of Model 4 did not support Hypothesis 3, which stated that hotels with higher frequency of forecast overrides have higher forecast accuracy (i.e., lower forecast error).

Model 5 added Special event and Business type as covariates. All predictors were significantly related to hotels' Absolute Error ($p < .001$). The coefficient for user overrides ($\beta_{20} = 0.83$) suggested that for each user override, the Absolute Error was expected to increase by 0.83. Special events were positively related to the Absolute Error ($\beta_{30} = 0.57, p < .001$), meaning that if there was a special event, the Absolute Error was, on average, higher by 0.57. Since group business type was coded as 1 and transient as 0, and business type coefficient was negative ($\beta_{40} = -4.85, p < .001$), it meant that the Absolute Error for groups was, on average, 4.85 lower i.e., the forecast was more accurate than for transient business.

Additional exploratory analyses were carried out. In Model 6 two level-1 interaction terms were added. The coefficient of these interaction terms was significant. As it is important to interpret the effect of the interaction variable in combination with the direct effects of the explanatory variables that make up that interaction, following the recommendations of Hox et al. (2018) it was considered helpful to write out the regression equation. The regression equation for the effect of level-1 predictors on Absolute Error based on Model 6 was:

$$AE_{it} = 9.89 + 0.10 \times DTA_{it} + 0.88 \times numOVER_{it} + 0.60 \times SPECIAL_{it} - 4.85 \times BusinessType_{it} - 0.44 \times numOVER_{it} \times SPECIAL_{it} + 0.09 \times numOVER_{it} \times BusinessType_{it} + u_{0i} + \varepsilon_{it} \quad (41)$$

The average effect of a change of one scale point in number of user overrides was an increase in the Absolute Error with 0.88. However, this was a predicted value only for the transient segment in absence of special events. The Absolute Error of the transient segment was expected to increase by 1.04 in the presence of one override on the occupancy date with a special event. For each additional user override under the same scenario (transient business with special event), the expected increase in Absolute Error was 0.44. The Absolute Error of the group segment was expected to decrease by 3.72 in the presence of one override on the occupancy date with a special event. In case of two user overrides, the decrease in Absolute Error was expected to be 3.19. The Absolute Error of the group segment was expected to decrease by 3.88 in the presence of one override on the occupancy date without a special event. In case of two user overrides, the Absolute Error is expected to decrease by 2.91.

Model 6 for AE led to slightly different conclusions compared to Model 5. The results of Model 5 indicated that on average user overrides were damaging for forecast accuracy. However, Model 6 showed that there is a difference between the business segments: for the transient segment overrides were damaging the forecast accuracy (for the days with and without special events), while for the group segment a limited number of overrides were on average beneficial for the forecast accuracy, especially in the presence of the special events. In other words, the positive relationship between number of overrides and forecast error is weaker in the presence of special events than without special events. Also, the positive relationship between the number of overrides and forecast error is stronger for the transient than for group segment. Therefore, Model 6 provided partial support for Hypothesis 3, which stated that hotels with higher frequency of forecast overrides have higher forecast accuracy.

Model 7 added the hotel-level (time invariant) control variables: hotel size, hotel type and hotel geographical region. Hotel size had a significant relationship with AE ($p < .001$), thus larger hotels had higher Absolute Errors (this predictor was centered

¹⁸ The results of the following models (Model 4 – Model 9) reported in Table 10 remain qualitatively similar to when quadratic DTA is added to the model.

on the grand mean). As explained in section 3.3.3, Hotel Type was coded in terms of dummy variables, where city hotels is the reference category. There were significant differences between city hotels and airport hotels ($p = .024$) and between city hotels and resort hotels ($p < .001$). On average airport hotels have higher forecast errors than city hotels and resort hotels have lower forecast errors than city hotels. There was no significant difference in Absolute Errors between city hotels and casino hotels ($p = .1$).

Europe was the reference category for the geographical region. There were significant differences between Europe and the following four regions: Asia, Australasia, Central America and Middle East & Africa. Asia and Middle East & Africa had a significantly higher AE compared to Europe. Australasia and Central America had significantly lower AE compared to Europe. There were no significant differences between hotels in Europe and in North America ($p = .303$), and between Europe and South America ($p = .272$).

Models 1-7 all assumed that the DTA coefficient was similar across hotels. In Model 8, the slope of DTA was allowed to vary across hotels. The level-1 variance estimate is 525.43 (Wald $Z = 3,154.54$, $p < .001$). At level 2, an unstructured covariance matrix was specified, enabling estimate variance of the random intercept ($\sigma_{U_0}^2$), the variance of the random linear slope ($\sigma_{U_1}^2$), and the covariance between the intercept and the slope (σ_{U_01}). The results suggest that there was significant variability in random intercept to be explained between hotels (Wald $Z = 19.80$, $p < .001$), the linear slope varied significantly across hotels (Wald $Z = 29.44$, $p < .001$), and there was a significant covariance between the level-2 initial status and linear change estimates (Wald $Z = 4.98$, $p < .001$). The significant slope variance of 0.01 shows that hotels also had different DTA coefficients. Furthermore, there was a positive covariance σ_{U_01} between the status at DTA 0 and the change rate, showing that hotels with a higher Absolute Error at DTA 0 had a higher rate of change over time.

To explore whether the change in Absolute Error over time (DTA) was impacted by hotel size, the cross-level interaction between hotel size and DTA to Model 9 was added. Table 10 shows that the interaction term was positive and significant ($\beta_{11} = 0.000364$, $p < .001$), meaning that effect of DTA was bigger for larger hotels.

In all cases when comparing complex models to simple models, the more complex models were preferred on the basis of the χ^2 – difference test and on the basis of AIC and BIC statistics. All the model-fit statistics indicated Model 9 to be the best model.

It can thus be concluded that, based on Absolute Error as a dependent variable, forecast accuracy improves as the date of arrival nears, providing support for Hypothesis 1. As a result of quadratic, cubic and quartic polynomials testing, it can be concluded that the relationship between time and forecast accuracy is complex

and has linear and non-linear components, supporting Hypothesis 2. Hypothesis 3 cannot be supported; hotels with higher frequency of judgmental adjustments do not have higher forecast accuracy.

In the next section, the analysis proceeds with the multilevel models for APE, sAPE, and LnQ. The same set of model formulations is used as presented in Table 10.

3.4.3 Multilevel Analysis of Scale-Independent Accuracy Measures

This section contains a short description of the results where the final model (Model 9) is discussed for AE, APE, sAPE, and LnQ, and explanations are given on the points where conclusions for all measures are similar and for which they differ. The results for the three scale-independent accuracy measures are reported in Tables 11- 13.

Tables 10-13 show that, from the four Models 9, DTA is a significant predictor with a positive coefficient for all forecast accuracy measures. Therefore, all four dependent variables indicate that forecast accuracy improves as the day of arrival nears, supporting Hypothesis 1.

Table 11: Results of the multilevel regression models for Absolute Percentage Error

Model:	Step 1: Null model		Step 2: Level-1 predictors		Step 3: Level-2 predictors		Step 4: Random coefficient		Step 5: Cross-level interactions	
	M_0 : Intercept only	M_1 : DTA	M_2 : DTA and DTA $\wedge$ 2	M_3 : number of overrides	M_4 : Level-1 model with three predictors	M_5 : Level-1 model with interactions	M_6 : final Level-1 and 2 mode	M_7 : random DTA	M_8 : Cross-level interaction of DTA and Hotel Size	
Fixed part										
Intercept	33.07*** (0.46)	17.48*** (0.45)	9.98*** (0.45)	17.59*** (0.45)	-0.77 (0.44)	-0.80 (0.4)	-3.31*** (0.59)	-2.22*** (0.41)	-2.13*** (0.41)	
DTA	0.30*** (4.9E-4)	0.94*** (4.9E-4)	0.30*** (2.3E-3)	0.30*** (4.9E-4)	0.30*** (4.9E-4)	0.30*** (4.9E-4)	0.30*** (4.9E-4)	0.29*** (4.1E-3)	0.30*** (4.0E-3)	
DTA quadratic			-0.003*** (8.9E-6)							
numOVER				-2.41*** (0.09)	0.22* (0.09)	0.53*** (0.10)	0.53*** (0.10)	0.53*** (0.10)	0.53** (0.10)	
SPECIAL					2.87*** (0.15)	2.90*** (0.15)	2.90*** (0.15)	2.89*** (0.15)	2.89*** (0.15)	
BusinessType					44.45*** (0.09)	44.53*** (0.09)	44.53*** (0.09)	44.37*** (0.09)	44.37*** (0.09)	
numOVERxSPECIAL						-0.17 (0.24)	-0.17 (0.24)	-0.17 (0.24)	-0.17 (0.24)	
numOVERxBusinessType						-4.57*** (0.32)	-4.58*** (0.32)	-4.56*** (0.32)	-4.56*** (0.32)	
Hotel Size							-2.7E-03 (2.0E-3)	-0.02*** (1.3E-3)	-0.01*** (1.4E-3)	
Hotel Type										
Airport vs City							-3.95 (3.42)	1.48 (2.28)	1.48 (2.28)	
Casino vs City							5.66 (3.02)	7.40*** (2.09)	7.36*** (2.09)	
Resort vs City							-0.69 (1.30)	-4.08*** (0.87)	-4.09*** (0.87)	
Hotel Region										
Asia vs Europe							12.71*** (1.25)	12.32*** (0.83)	12.32*** (0.84)	
Australasia vs Europe							-8.83*** (2.78)	-7.60*** (1.85)	-7.59*** (1.85)	
Central America vs Europe							-3.66 (10.06)	0.37 (6.74)	0.37 (6.74)	
Middle East & Africa vs Europe							9.82*** (1.51)	8.15*** (1.01)	8.15*** (1.01)	
North America vs Europe							-2.05 (1.28)	-4.59 (0.87)	-4.59*** (0.87)	
South America vs Europe							9.49* (4.37)	10.14** (2.97)	10.16** (2.97)	
DTA>Hotel Size								1.7E-4*** (1.7E-5)		
Random part										
σ_e^2	36,399,14***	35,736.94***	35,586.11***	35,735.66***	35,401.73***	35,401.38***	35,401.38***	35,190.14***	35,190.14***	
σ_{U0}^2	361.82***	355.92***	354.10***	355.49***	323.16***	323.17***	293.93***	148.81***	147.71***	
σ_{U1}^2								0.03***	0.03***	
σ_{U01}^2								0.79***	0.75***	
-2*log likelihood	267,905,206.2	267,536,513.5	267,451,591.5	267,535,793.1	265,001,167.3	265,000,967.8	265,000,773.2	264,887,469.0	264,887,389.1	
AIC	267,905,210.2	267,536,517.5	267,451,601.5	267,535,797.1	265,001,171.3	265,000,971.8	265,000,777.2	264,887,477.0	264,887,397.1	
BIC	267,905,239.8	267,536,547.1	267,451,631.1	267,535,826.8	265,001,201.0	265,001,001.4	265,000,806.8	264,887,536.2	264,887,456.4	

Note. Standard errors are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 12: Results of the multilevel regression models for Symmetric Absolute Percentage Error

Model:	Step 1: Null model		Step 2: Level-1 predictors		Step 3: Level-2 predictors			Step 4: Random coefficient		Step 5: Cross-level interactions	
	M ₁ : Intercept only	M ₂ : DTA	M ₃ : DTA and DTA^2	M ₄ : number of overrides	M ₅ : Level-1 model with three predictors	M ₆ : Level-1 model with interactions	M ₇ : final Level-1 and 2 mode	M ₈ : random DTA	M ₉ : Cross-level interaction of DTA and Hotel Size		
Fixed part	20.50*** (0.15)	13.11*** (0.15)	8.95*** (0.15)	13.16*** (0.15)	6.46*** (0.14)	6.46*** (0.14)	5.36*** (0.18)	5.47*** (0.17)	5.47*** (0.17)		
	DTA	0.14*** (7.6E-5)	0.50*** (3.4E-4)	0.14*** (7.6E-5)	0.14*** (7.4E-5)	0.14*** (7.4E-5)	0.14*** (7.4E-5)	0.14*** (9.0E-4)	0.14*** (9.0E-4)		
	DTA quadratic			-0.96*** (0.01)	0.04** (0.01)	0.13*** (0.01)	0.13*** (0.01)	0.13** (0.01)	0.13*** (0.01)		
	SPECIAL				0.09*** (0.02)	0.10*** (0.02)	0.10*** (0.02)	0.10*** (0.02)	0.10*** (0.02)		
	BusinessType				16.53*** (0.01)	16.56*** (0.01)	16.56*** (0.01)	16.53*** (0.01)	16.53*** (0.01)		
	numOVERxSPECIAL					-0.09** (0.04)	-0.09** (0.04)	-0.09** (0.04)	-0.09** (0.04)		
	numOVERxBusinessType					-1.31*** (0.05)	-1.31*** (0.05)	-1.31*** (0.05)	-1.31*** (0.05)		
	Hotel Size						-0.01*** (5.9E-4)	-0.01*** (5.6E-4)	-0.01*** (5.6E-4)		
	Hotel Type						0.21 (1.02)	1.02 (0.96)	1.02 (0.96)		
	Airport vs City						2.03* (0.89)	2.51** (0.84)	2.51** (0.84)		
Casino vs City						-1.44*** (0.39)	-1.97*** (0.37)	-1.97*** (0.37)			
Resort vs City											
Hotel Region											
Asia vs Europe							5.41*** (0.37)	5.62*** (0.35)	5.62*** (0.35)		
Australasia vs Europe							-4.18*** (0.83)	-4.18*** (0.78)	-4.18*** (0.78)		
Central America vs Europe							-0.88 (3.00)	-0.17 (2.83)	-0.17 (2.83)		
Middle East & Africa vs Europe							5.16*** (0.45)	4.96*** (0.43)	4.96*** (0.43)		
North America vs Europe							-1.07** (0.38)	-1.11** (0.36)	-1.11** (0.36)		
South America vs Europe							4.32** (1.30)	4.61*** (1.22)	4.61*** (1.22)		
DTA×Hotel Size									5.7E-6 (3.9E-6)		
Random part	1,006.87***	857.98***	811.48***	857.78***	794.83***	794.80***	794.80***	786.13***	786.13***		
	σ _e ²										
	39.73***	37.98***	37.33***	37.88***	33.54***	33.54***	26.32***	23.41***	23.41***		
	σ _{U0} ²										
	σ _{U1} ²							1.4E-3***	1.4E-03***		
σ _{U01} ²							-0.01	-0.01	-0.01		
-2*log likelihood	195,859,621.4	192,646,248.6	191,527,205.47	192,641,467.1	189,433,850.5	189,433,117.5	189,432,690.6	189,222,676.6	189,222,697.5		
AIC	195,859,625.4	192,646,252.6	191,527,209.47	192,641,471.1	189,433,854.5	189,433,121.5	189,432,694.6	189,222,684.6	189,222,705.5		
BIC	195,859,655.0	192,646,282.2	191,527,239.10	192,641,500.7	189,433,884.1	189,433,151.1	189,432,724.2	189,222,743.8	189,222,764.8		

Note. Standard errors are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 13: Results of the multilevel regression models for LnQ

Model:	Step 1: Null model		Step 2: Level-1 predictors			Step 3: Level-2 predictors		Step 4: Random coefficient		Step 5: Cross-level interactions	
	M ₁ : Intercept only	M ₂ : DTA	M ₃ : DTA and DTA^2	M ₄ : number of overrides	M ₅ : Level-1 model with three predictors	M ₆ : Level -1 model with interactions	M ₇ : final Level-1 and 2 model	M ₈ : random DTA	M ₉ : Cross-level interaction of DTA and Hotel Size		
Fixed part											
Intercept	2.3E-1*** (1.8E-3)	1.4E-1*** (1.8E-3)	9.3E-2*** (1.8E-3)	1.4E-1*** (1.8E-3)	5.7E-2*** (1.7E-3)	5.7E-2*** (1.7E-3)	4.3E-2*** (2.1E-3)	4.5E-2*** (1.9E-3)	4.5E-2*** (1.9E-3)		
DTA		1.7E-3*** (1.0E-6)	5.7E-3*** (4.5E-6)	1.7E-3*** (1.0E-6)	1.7E-3*** (9.7E-7)	1.7E-3*** (9.7E-7)	1.7E-3*** (9.7E-7)	1.6E-3*** (1.2E-5)	1.6E-3*** (1.2E-5)		
DTA quadratic			-1.6E-5*** (1.8E-8)	-1.1E-2*** (1.8E-4)	1.0E-3*** (1.8E-4)	2.1E-3*** (2.0E-4)	2.1E-3*** (2.0E-4)	2.2E-3*** (1.9E-4)	2.2E-3*** (1.9E-4)		
numOVER					1.3E-3*** (3.0E-4)	1.5E-3*** (3.0E-4)	1.5E-3*** (3.0E-4)	1.4E-3*** (3.0E-4)	1.4E-3*** (3.0E-4)		
SPECIAL					2.1E-1*** (1.7E-4)	2.1E-1*** (1.7E-4)	2.1E-1*** (1.7E-4)	2.1E-1*** (1.7E-4)	2.1E-1*** (1.7E-4)		
BusinessType					-1.0E-3* (4.8E-4)	-1.0E-3* (4.8E-4)	-1.0E-3* (4.8E-4)	-1.0E-3* (4.8E-4)	-1.0E-3* (4.8E-4)		
numOVER×SPECIAL					-1.6E-2*** (6.4E-4)	-1.6E-2*** (6.4E-4)	-1.6E-2*** (6.4E-4)	-1.6E-2*** (6.4E-4)	-1.6E-2*** (6.4E-4)		
numOVER×BusinessType											
Hotel Size											
Hotel Type											
Airport vs City							-4.6E-4 (1.2E-2)	1.1E-2 (1.1E-2)	1.1E-2 (1.1E-2)		
Casino vs City							2.3E-2* (1.1E-2)	2.9E-2** (9.9E-3)	2.9E-2** (9.9E-3)		
Resort vs City							-1.5E-2** (4.7E-3)	-2.3E-2*** (4.3E-3)	-2.3E-2*** (4.3E-3)		
Hotel Region											
Asia vs Europe							6.6E-2*** (4.5E-3)	6.6E-2*** (4.1E-3)	6.6E-2*** (4.1E-3)		
Australasia vs Europe							-4.8E-2*** (1.0E-2)	-4.8E-2*** (9.2E-3)	-4.8E-2*** (9.2E-3)		
Central America vs Europe							-1.1E-2 (3.7E-2)	1.6E-3 (3.3E-2)	1.6E-3 (3.3E-2)		
Middle East & Africa vs Europe							6.2E-2*** (5.5E-3)	5.6E-2*** (5.0E-3)	5.6E-2*** (5.0E-3)		
North America vs Europe							-1.0E-2* (4.6E-3)	-1.1E-2* (4.2E-3)	-1.1E-2* (4.2E-3)		
South America vs Europe							5.2E-2** (1.6E-2)	5.4E-2*** (1.4E-2)	5.4E-2*** (1.4E-2)		
DTA×Hotel Size									8.0E-8 (5.1E-8)		
Random part											
σ _e ²	1.7E-1***	1.5E-1***	1.4E-1***	1.5E-1***	1.4E-1***	1.4E-1***	1.4E-1***	1.4E-1***	1.4E-1***		
σ _{u0} ²	5.7E-3***	5.5E-3***	5.4E-3***	5.5E-3***	4.9E-3***	4.9E-3***	3.9E-3***	3.2E-3***	3.2E-3***		
σ _{u1} ²							2.4E-7***	2.4E-7***	2.4E-7***		
σ _{U01} ²							6.5E-7	6.5E-7	6.5E-7		
-2*log likelihood	21,299,003.3	18,736,705.4	17,920,563.0	18,732,806.9	17,209,936.4	17,209,338.3	17,209,022.6	16,997,942.6	16,997,971.9		
AIC	21,299,007.3	18,736,709.4	17,920,567.0	18,732,810.9	17,209,940.4	17,209,342.3	17,209,026.6	16,997,950.6	16,997,979.9		
BIC	21,299,036.9	18,736,739.0	17,920,596.6	18,732,840.5	17,209,970.0	17,209,371.9	17,209,056.2	16,998,009.9	16,998,039.2		
Note. Standard errors are in parentheses. * p < .05, ** p < .01, ***p < .001											

Note. Standard errors are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$

When all the relevant predictors and interaction terms are considered for the interpretation of the impact of the number of user overrides on forecast accuracy, it can be concluded that according to all forecast accuracy measures, the number of overrides has a significant negative relationship with accuracy, but the impact differs based on the business type. On average, overrides are beneficial for the group business and damaging for the transient business. Therefore, Hypothesis 3, which does not take into account the business type, is not supported. Hotels with a higher frequency of forecast overrides have higher forecast errors, but this effect depends on the business type.

Observing the results for Model 9 for all four dependent variables, slight differences can be seen between scale-dependent and scale-independent accuracy measures with regards to the coefficients for special events. There is a negative coefficient for AE and positive coefficients are noted for APE, sAPE, LnQ. As demonstrated in Chapter 2, inconsistencies can be found between the measures. However, AE is a scale-dependent measure and therefore not the most suitable measure. It is reasonable to argue that the results of APE, sAPE and LnQ are more valid and that the coefficient for special event is positive. If this is true, it would mean that on average, forecast errors are higher in the presence of special events.

Another difference apparent from comparing the results of the four different accuracy measures of Model 9, is that different accuracy measures had different significance of the interaction terms even though a rather consistent pattern can be observed. For AE, sAPE, and LnQ, both level-1 interaction terms ($numOVER \times SPECIAL$ and $numOVER \times BusinessType$) were significant, while for APE only $numOVER \times BusinessType$ interaction term was significant. Also, hotel size is significant and positive for AE and significant and negative for all other measures.

In terms of hotel type there are also some disagreements across forecast accuracy measures. For example, according to AE there are significant differences between airport and city hotels, while according to APE, sAPE and LnQ the differences between them are not significant. All four error measures agree that there is a significant difference between casino hotels and city hotels, and that casino hotels on average have higher forecast errors than city hotels. Moreover, all measures agree that there is a significant difference between resort hotels and city hotels and that, on average, resort hotels have more accurate forecasts.

In terms of hotel geographical regions, all accuracy measures agree that there is a significant difference between Europe and Asia, Europe and Australasia, and between Europe and Middle East & Africa. In addition, all measures indicate that there is no significant difference between Europe and Central America. There is an inconsistency across forecast accuracy measures for the difference between North America and Europe and between South America and Europe.

The measures show disagreement on the significance of the cross-level interaction: according to sAPE and LnQ the cross-level interaction term is not significant, while AE and APE show a significant, positive interaction term, suggesting that a conclusion on whether the change in forecast accuracy over time (DTA) is impacted by the hotel size depends on the accuracy measure. A likely explanation for this contradiction in the results is that AE is a scale-dependent measure, and APE is highly sensitive to the outliers.

All forecast accuracy measures agree that there is significant variance within hotels, and that the intercepts and slopes are different for hotels. The measures disagree on the significance of the covariance between intercept and the slope. According to sAPE and LnQ, covariance is not significant while AE and APE show significant, positive covariance.

This section has analyzed the results of the final model (Model 9) for the scale-independent accuracy measures and discussed them in light of the results reported in the previous section. All four dependent variables indicate that forecast accuracy improves as the day of arrival nears and that the number of overrides has a significant impact on accuracy and that the impact differs based on the business type. It can be argued based on the scale-independent accuracy measures that forecast errors are higher in the presence of special events and that hotel size, type and region are associated with forecast accuracy. Table 14 gives an overview of the results for the three hypotheses based on multilevel analyses using all four dependent variables.

Table 14: Summary results hypotheses 1-3

Hypothesis	APE	sAPE	LnQ	Conclusion	AE
H ₁ : Forecast accuracy improves as the date of arrival nears.	Yes	Yes	Yes	Supported	Yes
H ₂ : The relationship between time and forecast accuracy is complex and has linear and non-linear components.	Yes	Yes	Yes	Supported	Yes
H ₃ : Forecasts with a higher frequency of judgmental forecast adjustments have higher forecast accuracy.	No	No	No	Not supported	No

Note. Conclusions on the hypotheses testing were drawn on the basis of the three scale-independent measures.

3.5 Discussion and Future Research

The findings indicate that the forecasting horizon (i.e., Days To Arrival) has a significant relationship with forecasting accuracy. That is, forecasting accuracy improves as the day of arrival nears. Moreover, it was demonstrated that the relationship between time and accuracy is complex and has linear and non-linear components. The findings regarding the impact of time are in agreement with those of Pereira (2016), Rajopadhye et al. (2001), and Schwartz (1998), who suggest that the patterns of change over time are complex. In addition, the frequency of judgmental adjustments (i.e., number of overrides) has a significant negative impact

on forecasting accuracy ($p < .001$ for AE, sAPE, and LnQ; $p < 0.01$ for APE). A novel finding however, is that the relationship between the number of user overrides and forecast accuracy depends on the business type. On average, increasing the number of user overrides is beneficial for group business, but damaging for transient business forecasts.

While the findings for the hypotheses are consistent across the accuracy measures, slightly different conclusions can be drawn when other variables are considered. For example, according to AE, there is no significant difference in forecast accuracy between Europe and North America, while according to APE, sAPE and LnQ the forecasts in North America are significantly more accurate than in Europe. Therefore, this chapter confirms the importance of using the multiple forecast measures when examining forecast accuracy as demonstrated by Chapter 2 and published in Koupriouchina et al. (2014).

This study shows that the hotel size influences forecast accuracy, although a uniform result was not found across all dependent variables. According to the scale-independent measures, hotels that are larger than average have lower forecast errors compared to smaller hotels. This finding is similar to those of Kimes (1999) who stated that larger hotels had lower group forecasting errors. She hypothesized that larger hotels may have more experienced staff dealing with groups and that these experienced employees may have learned from frequent forecast updates.

The results indicate that there are significant differences between city hotels and resorts, where on average, resorts have more accurate forecasts than city hotels. This result can be explained by the typical characteristics of the resort business: these hotels normally receive guest reservations in advance, work regularly with tour operators, and have controlled lengths of stay. Changes to the bookings are usually also communicated well in advance. Each year, resorts typically sign contracts on the daily number of rooms allocated to them for the coming season, and the tour-operators advertise and attract customers (Hadjinicola & Panayi, 1997). As tour-operators normally package the hotel stay with airplane tickets, the packages tend to have a fixed duration which makes it easier to forecast. Moreover, as Vinod (2004) mentioned, resorts typically receive early bookings, which also makes it easier to forecast compared to other type of hotels. Therefore, overall it is reasonable to believe that market demand is easier to forecast at resort hotels.

Significant differences were found between several geographical regions to which the hotels belonged, which can be potentially explained by differences in training and the development of RM practice. In terms of the development of RM across the regions, various levels of maturity are reached. RM was ‘born’ in the USA (e.g., Marriott adopted it in the 90s) and was also adopted at an early stage in Europe. Other regions are currently implementing RM practices, but at a different pace, and there are different capabilities and levels of professionalization in the regions.

Overall, this chapter shows that multilevel modeling can provide additional insights into the hotel forecast accuracy at multiple forecast horizons. Research on (hotel) revenue management forecasting at the business type level and for different types of hotels in different geographical regions is scarce. This study provides initial insights in this important area. It also highlights the need to be extremely cautious about how to measure and interpret forecasting accuracy in the context of hotel revenue management.

Further research should investigate the relationship between forecasting horizon and accuracy in more depth, for example by including additional data points. For the current study, due to the extensive data, it was unfeasible to extract complete forecast data for every DTA. The forecasts were extracted at seven occasions, as determined by the company's experts (i.e. DTA 0, 1, 3, 7, 21, 84 and 252). With more powerful and advanced computing resources and storage facilities, it would be feasible to extract and analyze more data points, ideally for each DTA. Inclusion of this additional data can help to improve the models, especially with regard to estimating the linear and non-linear components.

Another important issue for future research is related to the polynomial functions. Heck et al. (2014) warned about the disadvantage of using polynomial functions in defining change trajectories, as there are strong correlations between the components of the function. This type of strong correlation can potentially cause problems when optimally estimating the model's parameters; they propose to change the polynomial functions into orthogonal ones. However, they mention that this approach only works if the occasions are equally spaced, which is not the case for this dataset. They mention that Hox (2010) suggests using orthogonal polynomials even in data situations where the repeated measurements may not be exactly spaced. In future studies, analyses including orthogonal polynomials should be explored.

This chapter focused on the difference between user overrides performed for the group and for the transient business segments. Another possible area of future research would be to consider sub-segments and to investigate judgmental adjustments at a more disaggregated level. Due to the limitations in the dataset, overrides were analyzed only in terms of the number of overrides per arrival date. This chapter did not explore the characteristics of these overrides in detail. The next chapter examines judgmental adjustments more closely in order to understand whether, for example, the size of overrides, their direction, or their interaction impact forecast accuracy.

Chapter 4

Judgmental Adjustments and Forecast Accuracy

Good judgment comes from experience, and a lot of that comes from bad judgment.

- Will Rogers -

4.1 Introduction

Since the introduction of revenue management in the early 1990s, the roles and responsibilities of revenue management professionals have evolved, and their reliance on automated solutions has progressively increased (Ferguson & Smith, 2014). Today, hotel revenue managers frequently interact with revenue management systems (RMS), and the resulting system-supported RM decisions have become an integral part of a hotels' daily practices.

Even though RMSs are rapidly becoming indispensable for hotel operations, few users have a solid understanding of their algorithms. This phenomenon has been identified as the 'black box' problem of forecasting practice (McCarthy et al., 2006). While the forecasting literature focuses on developing sophisticated models which are then transposed by software providers into algorithms for their RMS, research on how these systems are used to make real-world business decisions is often ignored. Belobaba (2002) warns about the growing gap between RM system complexity and actual decision making. Due to the high complexity and proprietary nature of system algorithms, RMS users often do not have a clear understanding of how the system-generated forecasts are created. Yet, revenue mangers hold final responsibility for the hotel property's revenue management decisions, in which accurate forecasts play a crucial role. They are expected to anticipate any change in market conditions, and to identify and correct any RMS-forecast anomalies to facilitate optimal pricing and inventory control decisions. Consequently, they manually adjust the

system-generated forecasts using their professional judgment. However, whether these adjustments are based on additional information exogenous to the revenue management system (Schwartz et al., 2016), and whether they add value to system-generated hotel occupancy forecasts, and how, remains unknown. Accordingly, this chapter aims to examine the impact of the different characteristics of judgmental forecast adjustments on the forecasting accuracy of the system-generated forecasts.

4.2 Literature Review

Very little research has been conducted in the hotel revenue management domain on judgmental forecast adjustments at single or multiple forecast horizons. Kimes (1999) investigated the correlation between the timing of forecast updates and hotel forecast accuracy for group forecasts of one hotel chain in North America. She found that “early forecast updates had little impact on forecast accuracy, but frequent updates during the month before arrival were associated with reduced forecast error” (Kimes, 1999, p. 1109).

The airline industry shares important business characteristics with hotel management. The key studies on judgmental adjustments at multiple forecast horizons in the airline industry were reviewed. Mukhopadhyay et al. (2007) analyzed 103 flight departures of one major U.S. airline in the seven-day period April 20 - 26, 2001 to explore the complexity and non-linearity of the relationship between days left to departure (DL) and forecast accuracy. They concluded that for the top 25% critical booking threshold (CBT) group:

from DL = 56 to DL = 21, RMAs improved machine forecasts more than 50% of the time they adjusted. The performance of analysts close to departure (DL = 7) is inferior to system performance. Interestingly, adjustment accuracy declined when RMAs had more information about the flights at DL < 21 (Mukhopadhyay et al., 2007, p. 318).

Similar patterns were observed for 50% and 75% CBT groups. Therefore, it can be hypothesized:

H₁: Early judgmental adjustments on the forecast horizon are associated with lower improvements in forecast accuracy than late adjustments.

Group business plays a key role in terms of both occupancy and revenue, and represents an important segment for many hotels. According to Anderson and Xie (2010, p. 57) “many city hotels have upwards of 50 percent of their reservations blocked far in advance by large groups at discounted rates for conferences or

special events”. Cross et al. (2009, p. 69) estimate that for many full-service and luxury hotels, group business is the source of 60 percent of revenue. For full-service Marriott brands, group business “can represent more than half the hotel’s revenue” (Hormby, Morrison, Dave, Meyers, & Tenca, 2010, p. 49). Moreover, group business decisions have a direct impact on the room availability for the transient segment. Once rooms are blocked for a group, they become unavailable for other segments. Therefore, decisions about group acceptance directly impact the revenue management optimization cycle for the entire hotel.

Despite its widely acknowledged importance, group business is an emerging body of research of (Choi, 2006; Cross et al., 2009; El Gayar et al., 2011; Farley, 2003; Hormby et al., 2010; Kimes, 1999; Noone & Hultberg, 2011; Schwartz & Cohen, 2003). However, the literature is fragmented, as each study addresses different aspects of group revenue management. Group forecasting, which lies at the core of group revenue management, has received even less research attention (El Gayar et al., 2011). So far only one systematic study on group forecast accuracy in hotels has been conducted (Kimes, 1999). She concluded that more accurate group forecasts were achieved in larger hotels with a higher dependence on group business and with more frequent forecasts updates in the month before arrival.

This paucity of research on how to optimize room revenue while simultaneously taking into account group and non-group decisions, presents a serious limitation to the concept of total revenue management, where all hotel revenue streams are optimized holistically (Cross et al., 2009; Josephi et al., 2015; Kimes, 2011, 2017). This limitation poses a serious issue for revenue management practice.

There are a number of reasons why modeling group forecasts is difficult. The primary reason is linked to the nature of the group business: “groups demand blocks of rooms, introducing a combinatorial aspect to the optimization problem” (Hormby et al., 2010, p. 49). A related reason is that the nature of group cancellations is more variable and statistically uncertain (Sierag, Van der Rest, Koole, Van der Mei, & Zwart, 2017). Another important reason is that the information necessary to calculate the unconstrained demand is not always available and/or reliable. For transient demand, statistical methods can be used to infer unconstrained demand. However, for group demand, lost and turndown data need to be recorded; in practice these are not always logged or completed by the sales team (Hormby et al., 2010).

These modeling challenges may have significant consequences. El Gayar et al. (2011) point out that the majority of existing forecasting models in hotel revenue management ignore group reservations and assume a pre-defined probability distribution to represent group arrivals. They argue that ignoring the group is a critical limitation as groups can sometimes constitute the main guest segment. Also, assuming a probability distribution for group demand can lead to inaccurate revenue management decisions, as hotel arrival characteristics tend to have a dynamically

changing and evolving nature. For example, in a single small hotel case study, Sierag et al. (2017) found evidence for an inhomogeneous Poisson nature of the probability distribution function.

As group business represents an important segment in terms of occupancy and revenue, it is known that revenue managers spend time on monitoring and adjusting group forecasts. Noone and Hultberg (2011) surveyed revenue managers and sales staff and found evidence that revenue managers use multiple sources of information for group forecasting, including information about *specific group characteristics* (e.g., meeting purpose and schedule, mandatory or non-mandatory event, citywide block versus in-house rooming list, contract attrition terms, etc.) and *external information* (e.g., general demand trends, rate and inventory availability across hotels in the market, rates and availability in other sectors, such as airlines). Their study also revealed that revenue managers explicitly refer to communication with the sales team as part of their forecasting activities. Insights obtained from the sales team, who are in direct contact with the group, contain important and the most updated information about the final group size or any changes affecting the final group demand.

Therefore, since group demand modelling has important limitations, and revenue managers analyze multiple sources of information exogenous to the RMS, the following can be hypothesized:¹⁹

H₂: Judgmental forecast adjustments to group business are associated with larger improvements in accuracy than adjustments to transient business.

The size of forecast adjustment is one of the key characteristics of the forecast adjustment process, it has an effect on the resulting forecasting accuracy. Several studies have been undertaken on the topic. When examining judgmental revision of short-term multi-item sales forecasts, Diamantopoulos and Mathews (1989) found that larger judgmental adjustments were more effective in improving forecast accuracy than smaller ones. In a study limited to a single company and a specific statistical method (i.e., Holt's linear trend [exponential smoothing] method) with only six, quarterly observations, Fildes et al. (2009) further examined the relationship between override magnitude and accuracy. They argued that the judgmental

¹⁹ In the absence of judgmental adjustments, there are indications that modeling challenges result in higher forecasting errors for group forecasts. In one hotel chain study, Kimes (1999) reports substantial group forecasting errors depending on the time horizon with MAPE up to 40% two months before arrival. However, no direct comparison was made with the transient segment, although the errors she found for group business are much higher than the forecast errors reported by other studies at the total hotel level (Schwartz et al., 2016; Yuksel, 2007), or by room type (Pereira, 2016).

adjustment size was positively related to accuracy improvement. Furthermore, Fildes et al. (2009, p. 5) based the size-accuracy hypothesis on the assumption that, "large adjustments were likely to be associated with reliable information about events which will have large anticipated effects not reflected in the system forecast", while the opposite would be true for small adjustments. In addition, they drew attention to research on human decision making, where people demonstrated excessive trust in their own judgment (e.g., Bonaccio & Dalal, 2006; Yaniv, 2004), "even when an impossibly good forecast was provided" (Fildes et al., 2009, p. 15), so that small adjustments reduced accuracy (J. S. Lim & Oconnor, 1995). Through their study, Fildes et al. (2009) confirmed a highly significant and positive relationship between the size of adjustment and the size of accuracy improvement. The authors stated that, "small adjustments were often ineffective in reducing forecast error" (Fildes et al., 2009, p. 15). They provided some tentative explanations for this phenomenon related to the unreliability of the information, smaller effects of anticipated special events, and the possibility that forecasters saw false patterns in noise. Moreover, they suggested that small adjustments may be made simply to signal that a user was 'at work', implying that small modifications may not always have a sound basis. They even suggested avoiding making smaller adjustments as these only have marginal effects on the accuracy, thereby making more time available for the forecasters. Trapero, Fildes, and Davydenko (2011) also found that the manually adjusted final forecast is more accurate than a system forecast for larger adjustments, and that there are no large differences between system forecasts and final forecasts for small adjustments.

In contrast, Syntetos, Nikolopoulos, Boylan, Fildes, and Goodwin (2009) found that for intermittent demand and considering zero forecasts:

small absolute adjustments are likely to be beneficial. This contrasts with results from research involving non-intermittent demand goods where small adjustments tended to have limited value. A possible explanation lies in the nature of the information being used by the forecaster to make the adjustment (Syntetos et al., 2009, p. 80).

Trapero, Pedregal, Fildes, and Kourentzes (2013) concluded that larger adjustments tended to be made when they were needed, but were generally overoptimistic, particularly for promotional events which were perceived as high impact.

Moreover, Davydenko and Fildes (2013) acknowledged that the measure used for the size of adjustment in Fildes et al. (2009) was asymmetrical, and they proposed a new approach to measuring the size of adjustments. Therefore, conclusions from the 2009 study should be taken with caution, as additional investigation is needed.

Although some conflicting evidence exists on the influence of the size of adjustment on the forecast accuracy, most studies appear to support the notion that larger manual adjustments are more likely to improve algorithmic predictions' accuracy than smaller adjustments. Accordingly, it can be hypothesized that in the revenue management-specific domain of daily hotel occupancy forecasts, human adjustments are more effective when the adjustment is larger.

H₃: The size of judgmental adjustments is positively associated with improvement in accuracy.

Much of the current literature on forecast adjustments pays special attention to the direction of adjustment. Traditionally, it has been argued that forecasters are biased towards optimism (Fildes et al., 2009; Franses & Legerstee, 2011; Kimes, 1999; Syntetos et al., 2009; Trapero et al., 2011; Trapero et al., 2013).

The study by Trapero et al. (2011) within supply chain management found that 62% of positive adjustments were too large or in the wrong direction, while only 51.7% of negative adjustments were overoptimistic. Fildes et al. (2009) suggest that negative adjustments are based on more reliable information, and therefore lead to more accurate forecasts than positive adjustments, and suggested a model for the optimal weights of statistical and judgmental forecasts based on the sign of adjustments. Syntetos et al. (2009, p. 78) conclude that "the results show quite conclusively that negative adjustments perform better than the positive ones. Positive adjustments seem to be beneficial only when they are superimposed on zero forecasts, in which case negative adjustments cannot be considered by definition." Therefore, it is hypothesized that:

H₄: Negative judgmental adjustments are associated with higher improvements in accuracy than positive.

The commonly held belief that positive adjustments often reduce forecasting accuracy has been challenged by Davydenko and Fildes (2013), who suggested that it may depend on the error measure used when comparing accuracy levels. Another aspect that might lead to the cautious interpretation of findings in the literature is the use of normalized versus non-normalized data as illustrated by Trapero et al. (2011). According to the non-normalized data, final forecasts were less accurate than system forecasts for positive adjustments. For the same dataset after the normalization procedure, the final forecasts were more accurate than system forecasts, even for positive adjustments. The authors argued that this difference can be explained by the different weight given to larger errors and by the fact that APE punishes over-forecasting more than it does under-forecasting. As illustrated

in Table 7 and explained in Chapter 3, APE favors under-forecasting and sAPE and LnQ favor over-forecasting. It therefore is expected that the directional effect can be different depending on the accuracy measure used. Measures that penalize over-forecasting more than under-forecasting will most likely lead to the conclusion that positive overrides are not beneficial. This is an important consideration for the interpretation of the findings.

4.3 Method

4.3.1 Data

Chapter 3 explained that two distinct datasets were received. The first dataset, which formed the basis of analysis in Chapter 3, contained forecast data for the occupancy period from March 24, 2014 to April 7, 2017. The second dataset, covering the same time period, contained detailed data on the judgmental forecast adjustments of 1,752 hotels belonging to 195 hotel chains and forms the basis for the present chapter. Table 15 summarizes dataset 2; in total 228,649 cases were analyzed.

Table 15: Source data summary for dataset two

Total number of cases	273,203 (100%)
- Total number of incomplete cases (% of total)	- 44,554 (16.31%)
= Total number of analyzed cases (% of total)	= 228,649 (83.69%)

The data were sourced from a variety of hotel types and regions: 1,462 city hotels, 224 resorts, 26 airport hotels, and 40 casino hotels from seven geographical regions: Europe, North America, Asia, Middle East & Africa, Australasia, Central America and South America.

Forecast data were extracted at the business type level, separately for group and transient business. The data only included those records for hotels that had overrides in the analyzed period and the occupancy dates for the overrides that had been performed. From the 88 extracted variables, 19 were found applicable for this study.

Property ID	Occupancy Date	Override Date	DTA	Business Type	Mod Num	Occupancy Forecast Group	Occupancy Forecast Group Prior Day	Occupancy Forecast Transient	Occupancy Forecast Transient Prior Day	Value Type	Latest or History	Override Status	Override Type	Special Event
16	3/ Jun/14	6/ May/14	28	1	0	43	36	147	159	TotDmd	history	Active	Demand	0
16	3/ Jun/14	28/ May/14	6	1	1	65	65	125	126	TotDmd	latest	Reverted	Demand	0
16	4/ Jun/14	6/ May/14	29	1	0	37	33	156	162	TotDmd	history	Active	Demand	0
16	4/ Jun/14	28/ May/14	7	1	1	59	53	132	139	TotDmd	latest	Reverted	Demand	0
16	5/ Jun/14	6/ May/14	30	1	0	42	37	153	158	TotDmd	history	Active	Demand	0
16	5/ Jun/14	28/ May/14	8	1	1	43	45	147	146	TotDmd	latest	Reverted	Demand	0
16	9/ Jun/14	6/ May/14	34	1	0	42	39	152	155	TotDmd	history	Active	Demand	0
16	9/ Jun/14	28/ May/14	12	1	1	42	44	152	152	TotDmd	latest	Reverted	Demand	0
16	10/ Jun/14	6/ May/14	35	1	0	40	39	147	156	TotDmd	history	Active	Demand	0
16	10/ Jun/14	28/ May/14	13	1	1	43	44	152	151	TotDmd	latest	Reverted	Demand	0
16	11/ Jun/14	6/ May/14	36	1	0	36	35	154	160	TotDmd	history	Active	Demand	0
16	11/ Jun/14	28/ May/14	14	1	1	35	39	160	156	TotDmd	latest	Reverted	Demand	0
16	16/ Jun/14	15/ May/14	32	1	0	38	37	121	156	TotDmd	history	Active	Demand	0
16	16/ Jun/14	28/ May/14	19	1	1	46	49	125	118	TotDmd	latest	Reverted	Demand	0
16	17/ Jun/14	6/ May/14	42	1	0	37	35	152	160	TotDmd	history	Active	Demand	0
16	17/ Jun/14	15/ May/14	33	1	1	38	36	135	159	TotDmd	history	Active	Demand	0
16	17/ Jun/14	28/ May/14	20	1	2	54	58	140	128	TotDmd	latest	Reverted	Demand	0
16	18/ Jun/14	15/ May/14	34	1	0	25	23	140	171	TotDmd	history	Active	Demand	0
16	18/ Jun/14	28/ May/14	21	1	1	24	25	166	138	TotDmd	latest	Reverted	Demand	0

Figure 7: Data structure of Dataset two (partial)

The most important variables are shown in Figure 7, which illustrates part of the anonymized forecast adjustments data of hotel 16 for the period June 3, 2014 to June 18, 2014. Each row represents a separate case. For each hotel at each occupancy date the judgmental adjustments at DTA are reported. For example, at 28 days before arrival the forecast occupancy for June 3, 2014 for group was 43 and for transient it was 147. There was one override for this occupancy date at DTA 28 (Mod Num is a modification number, it starts at zero and is incremented each time the override values are modified). As can be seen, several occupancy dates are missing, which means that for June 6, 7, 8, 12, 13, 14, 15 no adjustments were made to the RMS forecasts. Each row represents a separate forecast adjustment at different DTA. For a full overview of the variables, see Appendix F.

4.3.2 Dependent Variables

This chapter builds on the measures selected in Chapter 3 (i.e., AE, APE, sAPE, and LnQ). However, the focus of the chapter is on user overrides and on their impact on forecast accuracy. Therefore, a comparison between the system and user forecasts is of principal interest. This comparison is possible as a result of the computation of so-called forecast improvement metrics. In total, four forecast improvement measures were chosen as the dependent variables: APE_{imp} , $sAPE_{imp}$, LnQ_{imp} , and AE_{imp} .

First, APE_{imp} is the improvement measure, which Fildes et al. (2009) referred to as *Forecast Improvement* (FCIMP). In this chapter the choice was made to refer to it as APE_{imp} and not FCIMP, since more than one forecast improvement measures will be formulated following the logic of FCIMP. The definitions of FCIMP and APE_{imp} are identical as the following formula shows:

$$FCIMP = 100 \times \frac{(|Actual - System forecast| - |Actual - Final forecast|)}{Actual} \quad (42)$$

To make the formulation of all forecast improvement metrics transparent, equation 42 can be re-written as:

$$APE_{imp_t} = \left| \frac{Y_t - SF_t}{Y_t} \right| \times 100 - \left| \frac{Y_t - UF_t}{Y_t} \right| \times 100 = System APE_t - User APE_t \quad (43)$$

where, Y_t is the actual value, SF_t is the system forecast at time t , and UF_t is the user forecast at time t . The APE_{imp} measure can thus be interpreted as: a positive value of *APE Forecast Improvement* means that the user improved the forecast, a negative value means that the user deteriorated the forecast.

Second, a forecast improvement measure based on sAPE was derived through the following series of equations:

$$sAPE_{imp_t} = System sAPE_t - User sAPE_t \quad (44)$$

$$System sAPE_t = \left| \frac{(Y_t - SF_t)}{\frac{1}{2}(Y_t + SF_t)} \right| \times 100 = \left| \frac{200 \times System e_t}{(Y_t + SF_t)} \right| \quad (45)$$

$$User sAPE_t = \left| \frac{(Y_t - UF_t)}{\frac{1}{2}(Y_t + UF_t)} \right| \times 100 = \left| \frac{200 \times User e_t}{(Y_t + UF_t)} \right| \quad (46)$$

The third forecast improvement measure selected for analysis was based on the scale-independent LnQ as defined below:

$$LnQ_{imp_t} = |System Ln Q_t| - |User Ln Q_t| \quad (47)$$

$$System LnQ_t = \ln \left(\frac{SF_t}{Y_t} \right) \quad (48)$$

$$User LnQ_t = \ln \left(\frac{UF_t}{Y_t} \right) \quad (49)$$

The fourth forecast improvement measure was based on scale-dependent AE and is defined as follows:

$$AE_{imp_t} = |System\ e_t| - |User\ e_t| = |Y_t - SF_t| - |Y_t - UF_t| \quad (50)$$

For all forecast improvement measures a positive value means that the user improved the forecast and a negative value means the opposite. AE measure is scale-dependent, and therefore, is not ideal for the datasets with hotels of different sizes. In general, scale-dependent measures are more applicable for datasets with similar scales. AE is however widely used in hotel revenue management research and, therefore, was included as a dependent variable in the analysis. However, hypothesis testing was only calculated for $sAPE_{imp}$, LnQ_{imp} , and APE_{imp} .

4.3.3 Independent Variables

A number of independent variables were selected for the study. Because of the hierarchical structure of the data, multilevel analysis was chosen as the analytical approach. For this reason, the independent variables are explained according to the nested structure of the data, with the forecast overrides at level 1 and hotels at level 2.

The first level-1 predictor was Days to Arrival (DTA). This variable is similar to the one determined in Chapter 3, however the Chapter 3 forecasts were extracted at seven fixed forecast horizons (0, 1, 3, 7, 21, 84, 252 days to arrival). The overrides in this chapter were extracted whenever the overrides took place, and DTA is thus a continuous variable ranging from 0 to 596. The second level-1 predictor was Business Type (0 = transient, 1 = group). Adjustment Size was included as a third predictor to indicate the size of the judgmental adjustment. Fildes et al. (2009) analyzed the size of judgmental adjustments using a measure defined as $100 \times (\text{Final forecast} - \text{System forecast}) / \text{System forecast}$. However, later this approach was changed as Davydenko and Fildes (2013, p. 512) acknowledged that the measure was asymmetrical:

if an expert doubles a statistical forecast (say from 10 units to 20 units), he/she increases it by 100%, but if he/she halves a statistical forecast (say from 20 units to 10 units), he/she decreases it by 50% (not 100%).

To correct for this, Davydenko and Fildes (2013) proposed a measure using the natural logarithm of the Final forecast/System forecast ratio. The authors argued that it would help to restore symmetry with ratio data since $\ln(A/B) = -\ln(B/A)$. The result of the proposed formula can take positive and negative values, depending on whether the user forecast was higher or lower than the system forecast. As the

hypotheses about the size of the adjustments and the direction are tested separately, the **absolute value** of the *Adjustment Size*_t was taken for the purpose of this chapter and defined as:

$$Absolute\ Adjustment\ Size_t = \left| \ln \left(\frac{UF_t}{SF_t} \right) \right| \quad (51)$$

As a fourth level-1 predictor, Override Direction (0 = negative override, 1 = positive override) was chosen. A positive adjustment (value 1) means that the user increased the forecast occupancy; a negative adjustment (value 0) means that the user decreased the forecast occupancy.

In addition to the four level-1 predictors which are directly related to the hypotheses, five level-1 control variables were selected, including Special Event, Number of Overrides, Override Status, Override Recency Status, and Override Type. Special event is "a date or set of contiguous dates for which the hotel data cannot be forecasted in the normal fashion, or period(s) where the transient business pattern is different than normal" (IdeaS, 2019). *SpecialEvent* is a dichotomous variable (0 = no special event, 1 = special event), representing a special event on the occupancy day that is configured to impact forecasts, such as national holidays, conventions, music festivals, car races, etc. Number of Overrides is a continuous variable (ranging from 0 – 28), representing the number of user overrides for an arrival date at a segment level.

System overrides can be reverted (deactivated) by revenue managers. Override Status (0 = reverted, 1 = active) and Override Recency Status (0 = history, 1 = latest) are dichotomous variables indicating the override status. The combination of these variables helps to distinguish between active and reverted overrides. Active demand overrides have Override Recency Status = 'latest' and Override Status = 'active'.

Override Type is a dichotomous variable (0 = wash, 1 = demand) indicating the type of override. Wash overrides are typically applicable to the group business segment with unusual wash patterns. When a group decides to book a hotel, they usually contact hotel via a Request For Proposal and the process of negotiation starts. The hotel checks whether requested rooms are available for the group in the requested period and if the outcome is positive, the hotel sales department responds with a formal proposal of rooms and facilities and related rates and charges. If the proposal is accepted, then both parties sign a group sales agreement (DeKay, Yates, & Toh, 2004). The key issue with the agreed numbers of group rooms is that it is subject to uncertainty. Group typically book at least several months or even years before the event, and a number of factors can influence final occupancy (e.g., client's company budget cuts, layoffs, sickness of the meeting attendees). For example, if 100 rooms were blocked for 3 nights for the group event, but only 80 attendees came to the event, there is a risk of 60 (20×3) unsold room nights and

loss of associated revenues. Since optimization of perishable assets is one of the key goals of Revenue Management, hotel sales department and revenue managers pro-actively manage rooms block for the groups. The full discussion of this complex and interesting issue and a complete list of strategies utilized by the hotels is out of the scope of this dissertation. For the purpose of this dissertation, it is worthwhile to mention that one of the key strategies is to forecast room block utilization, based on the group characteristics, the purpose of the event, any historical information available on that group or a group with similar characteristics. For example, if the last year 100 rooms were booked for a group and only 80 rooms were occupied by that group, the associated wash percentage was 20%. An accurate estimate of the wash percentage helps to accurately forecast group demand. Different group types may have different wash percentages and it is important to anticipate any changes in wash percentages for the optimization of revenue. For example, if a group has a room block of 100 rooms, but the sales manager has heard from trusted sources that many attendees are booking another hotel closer to the event, the wash percentage must be changed to avoid keeping the rooms blocked unnecessarily.

Similar to Chapter 3, three control variables at the hotel level were analyzed: Hotel Size, Hotel Type and Hotel Region. The first level-2 control variable, Hotel Size, represented the hotel's capacity in room number, ranging from 6 – 4028. As Hotel Size does not have a meaningful zero, it was centered on its grand mean. The second level-2 control variable, Hotel Type, has aggregated hotel type categories: airport hotels, casino hotels, city hotels, and resort hotels. Hotel Type was coded as dummy variables with city hotels as the reference category. The third level-2 control variable, Hotel Region has seven categories: Asia, Australasia, Central America, Europe, Middle East & Africa, North America and South America. It was also coded in terms of dummy variables, with Europe as the reference category.

4.3.4 Analytical Approach

The analysis strategy followed recommendations by Hox et al. (2018), who outlined five major steps in building multilevel models. These steps and relevant model formulations are described below. This chapter applies a classical multilevel approach in contrast to the variation of the multilevel modeling (Repeated measures design) used in Chapter 3. To distinguish level-1 and level-2 variables, two subscripts to describe override cases (*i*) and hotels (*j*) are used.

Step 1

In this step a model without explanatory variables (intercept-only model) is built. Model 1 makes it possible to test whether the grand-mean intercept for the improvement of forecast accuracy ($sAPE_{imp}$, LnQ_{imp} , APE_{imp} , AE_{imp}) varies across hotels. It was defined at level-1 as follows:

$$Y_{ij} = \beta_{0j} + e_{ij} \quad (52)$$

where β_{0j} is the intercept, and e_{ij} represents errors in predicting the average improvement in forecast error for hotels. This intercept can be defined as:

$$\beta_{0j} = \gamma_{00} + u_{0j} \quad (53)$$

These equations were combined by substituting Equation 53 into Equation 52:

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad (54)$$

Three parameters were estimated for Model 1: the fixed effect describing average improvement in forecast error (γ_{00}), the between-hotel random variance (u_{0j}), and the level-1 residual variance (e_{ij}). The pooled intercepts for $sAPE_{imp}$, LnQ_{imp} , APE_{imp} , AE_{imp} were 0.96, 0.015, 3.50, and 1.52 and respectively with $p < .001$ for all measures.

This intercept-only model was useful as it gave an estimate of the intra-class correlation (ICC): $sAPE_{imp} = 7.4\%$, $LnQ_{imp} = 8.6\%$, $APE_{imp} = 1.0\%$, and $AE_{imp} = 3.1\%$.²⁰ The ICC results for $sAPE_{imp}$ and LnQ_{imp} were above the recommended 5% benchmark (Heck et al., 2014), suggesting that multilevel modeling was an adequate approach for the project's data. The ICC for AE_{imp} and APE_{imp} were below the recommended 5%. However, for consistency, multilevel regressions for these forecast improvement accuracy measures were also performed.

Step 2

In models 2 and 3, the lower-level explanatory variables were added to Model 1. They were added as fixed, so that the corresponding variance components of the slopes were fixed at zero. This means that the effects of the explanatory variables were assumed to be similar across level-2 units (i.e., hotels). The main predictors were added one at a time as well as all four simultaneously.²¹ Model 2 included all fixed effects in one model, given by the equation:

$$Y_{ij} = \gamma_{00} + \gamma_{10} DTA_{ij} + \gamma_{20} BusinessType_{ij} + \gamma_{30} SizeAdjustmentAbs_{ij} + \gamma_{40} Direction_{ij} + u_{0j} + e_{ij} \quad (55)$$

Seven parameters were estimated for Model 2, including five fixed effects, one random intercept (u_{0j}), and the level-1 residual (e_{ij}). In Model 3, the level-1 control variables were added as shown in the following equation:

20 The ICCs were calculated based on variances at level one and two which are reported in the results section in the Tables 19 - 22.

21 For parsimony reasons, only the simultaneous model was reported in the result section.

$$Y_{ij} = \gamma_{00} + \gamma_{10}DTA_{ij} + \gamma_{20}BusinessType_{ij} + \gamma_{30}SizeAdjustmentAbs_{ij} + \gamma_{40}Direction_{ij} + \gamma_{50}Special_{ij} + \gamma_{60}OverrideNumber_{ij} + \gamma_{70}OverrideType_{ij} + \gamma_{80}OverrideStatus_{ij} + \gamma_{90}LatestOrHistory_{ij} + u_{0j} + e_{ij} \quad (56)$$

Twelve parameters were estimated for Model 3, including ten fixed effects, one random intercept (u_{0j}), and the level-1 residual (e_{ij}).

Step 3

In this step, the higher-level control variables were added to the model. Model 4 includes three level-2 covariates (i.e., hotel size, hotel type, hotel geographical region) and is formulated as:

$$Y_{ij} = \gamma_{00} + \gamma_{10}DTA_{ij} + \gamma_{20}BusinessType_{ij} + \gamma_{30}SizeAdjustmentAbs_{ij} + \gamma_{40}Direction_{ij} + \gamma_{50}Special_{ij} + \gamma_{60}OverrideNumber_{ij} + \gamma_{70}OverrideType_{ij} + \gamma_{80}OverrideStatus_{ij} + \gamma_{90}LatestOrHistory_{ij} + \gamma_{01}HotelSize_j + \gamma_{02}HotelType_j + \gamma_{03}HotelRegion_j + u_{0j} + e_{ij} \quad (57)$$

The hotel-level variable *HotelSize* is a continuous variable and only one parameter needed to be estimated for γ_{01} . However, the other two higher-level variables were categorical and coded as dummy variables.²² In total, ten parameters were estimated for level-2 variables. As explained in Model 3, twelve parameters were estimated for level-1 variables. This suggests that in total, twenty-two parameters were estimated for Model 4, including twenty fixed effects, one random intercept (u_{0j}), and the level-1 residual (e_{ij}).

Step 4

In this step, whether the slope of DTA had a significant variance component between the hotels was assessed. This model, the *random coefficient model*, is given by:

$$Y_{ij} = \gamma_{00} + \gamma_{10}DTA_{ij} + \gamma_{20}BusinessType_{ij} + \gamma_{30}SizeAdjustmentAbs_{ij} + \gamma_{40}Direction_{ij} + \gamma_{50}Special_{ij} + \gamma_{60}OverrideNumber_{ij} + \gamma_{70}OverrideType_{ij} + \gamma_{80}OverrideStatus_{ij} + \gamma_{90}LatestOrHistory_{ij} + \gamma_{01}HotelSize_j + \gamma_{02}HotelType_j + \gamma_{03}HotelRegion_j + u_{1j}DTA_{ij} + u_{0j} + e_{ij} \quad (58)$$

In total twenty-four parameters were estimated for Model 5: twenty fixed effects, one random intercept (u_{0j}), the hotel-level residual of the slopes of the DTA (u_{1j}), the level-1 residual (e_{ij}), and the covariance between the intercept and the slope (u_{01}).

22 The *HotelType* variable had four categories in total and three dummy variables were created. Therefore, three parameters were estimated for γ_{02} . The *HotelRegion* variable was also a categorical variable with seven categories; in total six dummy variables were created and six parameters were estimated for γ_{03} .

Step 5

In this step, the cross-level interaction between *HotelSize* (a level-2 control variable) and *DTA* (the level-1 predictor that had significant slope variation in Step 4) was added to the model in order to test whether hotel size explains the variance in the slope of DTA across hotels. This model was formulated as:

$$Y_{ij} = \gamma_{00} + \gamma_{10}DTA_{ij} + \gamma_{20}BusinessType_{ij} + \gamma_{30}SizeAdjustmentAbs_{ij} + \gamma_{40}Direction_{ij} + \gamma_{50}Special_{ij} + \gamma_{60}OverrideNumber_{ij} + \gamma_{70}OverrideType_{ij} + \gamma_{80}OverrideStatus_{ij} + \gamma_{90}LatestOrHistory_{ij} + \gamma_{01}HotelSize_j + \gamma_{02}HotelType_j + \gamma_{03}HotelRegion_j + \gamma_{11}DTA_{ij}HotelSize_j + u_{1j}DTA_{ij} + u_{0j} + e_{ij} \quad (59)$$

In total, twenty-five parameters were estimated for Model 6: twenty fixed effects, one random intercept (u_{0j}), the hotel-level residual of the slopes of the DTA (u_{1j}), the level-1 residual (e_{ij}), and the covariance between the intercept and the slope (u_{01}).

In-line with Chapter 3, SPSS, version 23, release 23.0.0.3 was used for analysis. For Steps 1 – 2 scaled identity was selected as level-1 covariance structure. For Steps 3 – 5, scaled identity²³ was again selected as level-1 covariance structure, and unstructured covariance structure was selected for level-2. The reporting approach followed recommendations made by Hox et al. (2018) and Dedrick et al. (2009). Improvements in the model fit were assessed with reference to a deviance statistic (–2 loglikelihood ratio) distributed as χ^2 with associated degrees of freedom. Additionally, AIC and BIC statistics were examined to compare the model fit.

4.4 Results

The results are organized in three subsections. In the first, the descriptive statistics are reported. The second contains a detailed interpretation of the results of the five multilevel modeling steps for $sAPE_{imp}$, after which the results of LnQ_{imp} , APE_{imp} , AE_{imp} are reported and compared. In the final subsection, the additional analyses exploring interaction effects are reported.

4.4.1 Descriptive Statistics

Figure 8 shows the percentage of judgmental adjustments at different DTA categories²⁴. It reveals that the majority of judgmental adjustments (80%) happened in an 84-day forecast window (i.e., the period of time between the day on which the forecast is made and the day of arrival). Figure 9 reports the mean improvement

23 This structure has constant variance. It is assumed that there is no correlation between any element.

24 DTA was transformed into a categorical variable to facilitate comparison with Chapter 3 (DTA 0, 1, 3, 7, 21, 84, and 252). Multilevel models were formulated and estimated based on DTA as a continuous variable.

in forecast accuracy per DTA category and with a polynomial trend line of the third order. As illustrated in Figure 9, for six of the seven forecast windows applied in Chapter 3, judgmental adjustments improved the forecasting accuracy according to three forecast improvement measures. For the DTA category ‘between 85 – 252’, $sAPE_{imp}$ indicated a deterioration. All measures indicated that manual adjustments reduced the accuracy for forecast windows above 252 days. Therefore, from Figure 9 it appears that early judgmental adjustments are not only associated with lower improvement in accuracy, but that those done more than 252 days before arrival decrease the accuracy of the forecasts.

Table 16 presents the breakdown of judgmental adjustments according to DTA, override type and business type. It is interesting to observe that for transient segment, 95.5% of all overrides were demand overrides and for group segment 63.7% were wash overrides. As Table 17 shows, after analysis of all forecasts (total), user-adjusted forecasts are, on average, more accurate than system forecasts for all accuracy measures (see the last row, where the User Forecasts errors are lower than System forecast errors). If the direction of the user override is considered, a slightly different picture emerges. While negative judgmental adjustments improve the system forecast according to all measures, for positive overrides the results differ per measure. Table 18 shows that on average, city hotels had a higher improvement in forecast accuracy compared to other hotel types. APE_{imp} also had higher Standard Deviations; it appears that $sAPE_{imp}$ and LnQ_{imp} are more balanced measures.

In summary, these results show that the majority of all judgmental adjustments (80%) are performed 84 day before arrival or earlier. It seems that early judgmental adjustments on the forecast horizon are associated with lower improvements in forecast accuracy than late adjustments. The majority of overrides for transient segment are demand overrides and the majority of the overrides for the group segment are wash overrides. Negative judgmental adjustments improve the system forecast according to all forecast improvement measures, which is not the case for positive adjustments.

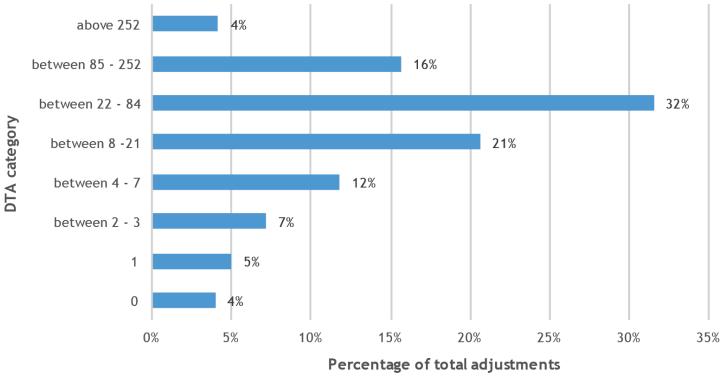


Figure 8: Judgmental adjustments per DTA category

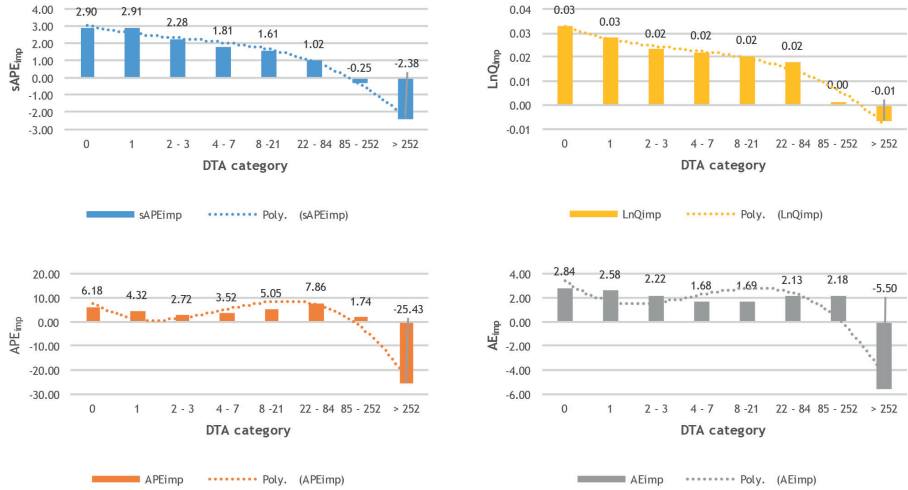


Figure 9: Mean improvement in forecast accuracy per DTA category

4.4.2 Multilevel Analysis

The results are reported in two parts. First, a detailed interpretation is given of $sAPE_{imp}$ for all six multilevel models ($M_1 - M_6$). Second, the results of the multilevel regressions are reported for LnQ_{imp} , APE_{imp} , and AE_{imp} , and a comparison is made between the results of the final models (M_6) for each of the four measures.

Table 16: Number of judgmental adjustments per DTA per override type for Transient and Group

Overrides	0		1		between 2 - 3		between 4 - 7		DTA between 8-21		between 22 - 84		between 85 - 252		above 252		Total	
	Count	N %	Count	N %	Count	N %	Count	N %	Count	N %	Count	N %	Count	N %	Count	N %	Count	N %
Transient																		
Wash	437	7.3%	528	6.4%	653	5.3%	938	4.6%	1,416	4.0%	1,908	3.6%	1,298	5.2%	310	4.8%	7,488	4.5%
Demand	5,544	92.7%	7,746	93.6%	11,736	94.7%	19,369	95.4%	34,123	96.0%	50,528	96.4%	23,477	94.8%	6,130	95.2%	158,653	95.5%
Total	5,981	100.0%	8,274	100.0%	12,389	100.0%	20,307	100.0%	35,539	100.0%	52,436	100.0%	24,775	100.0%	6,440	100.0%	166,141	100.0%
Group																		
Wash	4,144	85.3%	4,366	81.2%	5,652	78.7%	9,058	75.4%	13,959	66.9%	19,126	56.7%	9,528	52.8%	2,316	46.5%	68,149	63.7%
Demand	714	14.7%	1,010	18.8%	1,530	21.3%	2,958	24.6%	6,917	33.1%	14,592	43.3%	8,525	47.2%	2,667	53.5%	38,913	36.3%
Total	4,858	100.0%	5,376	100.0%	7,182	100.0%	12,016	100.0%	20,876	100.0%	33,718	100.0%	18,053	100.0%	4,983	100.0%	107,062	100.0%
Total (Transient and Group)																		
Wash	4,581	42.3%	4,894	35.9%	6,305	32.2%	9,996	30.9%	15,375	27.3%	21,034	24.4%	10,826	25.3%	2,626	23.0%	75,637	27.7%
Demand	6,258	57.7%	8,756	64.1%	13,266	67.8%	22,327	69.1%	41,040	72.7%	65,120	75.6%	32,002	74.7%	8,797	77.0%	197,566	72.3%
Total	10,839	100.0%	13,650	100.0%	19,571	100.0%	32,323	100.0%	56,415	100.0%	86,154	100.0%	42,828	100.0%	11,423	100.0%	273,203	100.0%

Note. DTA was transformed into a categorical variable here as well to facilitate comparison with Chapter 3.

Table 17: Forecast error per adjustment direction

Adjustments	N	Mean sAPE		Mean LnQ		Mean APE		Mean AE		MdAPE		MdAE		
		SF	UF	SF	UF	SF	UF	SF	UF	SF	UF	SF	UF	
Negative	122,771	54%	21.01	20.36	0.15	-0.04	45.81	27.66	25.62	22.81	10.42	9.38	12.00	11.00
Positive	105,878	46%	21.91	19.29	-0.06	0.13	30.51	44.14	24.26	24.16	9.90	8.82	11.00	10.00
Total	228,649	100%	21.43	19.86	0.05	0.04	38.60	35.43	24.98	23.45	10.17	9.09	12.00	10.00

Note. SF is System Forecast and UF is User Forecast.

Table 18: Improvement in forecast errors by hotel type

Hotel type	sAPE _{imp}				LnQ _{imp}				APE _{imp}				AE _{imp}			
	Mean	Min	Max	SD	Mean	Min	Max	SD	Mean	Min	Max	SD	Mean	Min	Max	SD
	Count	N %	Count	N %	Count	N %	Count	N %	Count	N %	Count	N %	Count	N %	Count	N %
Airport	0.4	-162.0	170.4	17.5	4.1E-03	-2.4	3.3	0.2	0.1	-3,200.0	1,100.0	79.3	-0.4	-166.0	133.0	20.9
Casino	0.0	-148.2	158.2	14.7	8.5E-04	-2.5	2.2	0.2	1.0	-500.0	2,094.7	49.1	-3.6	-1,524.0	1,446.0	137.8
City	1.5	-194.4	196.1	22.6	1.6E-02	-5.2	5.5	0.3	3.0	-15,000.0	27,300.0	194.1	1.4	-548.0	753.0	24.9
Resort	0.5	-190.9	192.7	23.0	2.1E-04	-4.0	4.0	0.3	0.9	-19,200.0	12,000.0	247.7	1.7	-406.0	452.0	25.7
Total	1.3	-194.4	196.1	22.4	1.4E-02	-5.2	5.5	0.3	2.7	-19,200.0	27,300.0	196.9	1.3	-1,524.0	1,446.0	33.7

Table 19: Fixed effects estimates and variance-covariance estimates for improvement in sAPE

Model:	Step 1: Null model		Step 2: Level-1 predictors		Step 3: Level-2 predictors		Step 4: Random coefficient	Step 5: Cross-level interactions				
DV = sAPE _{imp}	M ₀ : Intercept only		M ₁ : Level -1 model with main predictors	M ₂ : Level -1 model with all predictors	M ₃ : final Level-1 and 2 model		M ₄ : random DTA	M ₅ : Cross-level interaction of DTA and Hotel Size				
Fixed part	1.11****	(0.24)	-0.61**	(0.23)	-1.21****	(0.33)	-1.20**	(0.39)	-0.42	(0.36)	-0.43	(0.36)
Intercept												
Level-1												
Main predictors												
DTA			-0.02****	(7.5E-04)	-0.02****	(7.8E-04)	-0.02****	(7.8E-04)	-0.02****	(3.8E-03)	-0.02****	(3.9E-03)
Business Type			2.28****	(0.13)	4.13****	(0.16)	4.12****	(0.16)	3.65****	(0.16)	3.65****	(0.16)
Size Adjustment Abs			6.19****	(0.15)	5.70****	(0.15)	5.70****	(0.15)	6.47****	(0.15)	6.47****	(0.15)
Direction			1.85**	(0.10)	1.68****	(0.10)	1.68****	(0.10)	1.58****	(0.10)	1.58****	(0.10)
Control variables												
Special					0.19	(0.14)	0.19	(0.14)	0.27	(0.14)	0.27	(0.14)
Override Number					0.04**	(0.01)	0.04**	(0.01)	0.03*	(0.01)	0.03*	(0.01)
Override Type					2.94****	(0.18)	2.95****	(0.18)	2.43****	(0.18)	2.43****	(0.18)
Override Status					-3.08****	(0.13)	-3.07****	(0.13)	-3.21****	(0.13)	-3.21****	(0.13)
Override Recency Status					0.10	(0.13)	0.10	(0.13)	0.09	(0.13)	0.10	(0.13)
Level-2												
Hotel Size							6.8E-04	(9.1E-04)	5.0E-04	(7.7E-04)	3.9E-04	(7.9E-04)
Hotel Type												
Airport vs City							-1.03	(1.73)	-0.90	(1.53)	-0.90	(1.53)
Casino vs City							-1.13	(1.62)	-2.12	(1.43)	-2.11	(1.43)
Resort vs City							0.21	(0.69)	0.37	(0.61)	0.37	(0.61)
Hotel Region												
Asia vs Europe							-0.59	(0.67)	0.22	(0.58)	0.22	(0.58)
Australasia vs Europe							0.84	(1.30)	-0.03	(1.13)	-0.04	(1.13)
Central America vs Europe							3.12	(3.69)	9.29**	(3.38)	9.27**	(3.38)
Middle East & Africa vs Europe							-0.05	(0.90)	-0.70	(0.79)	-0.71	(0.79)
North America vs Europe							0.20	(0.56)	-0.38	(0.48)	-0.39	(0.48)
South America vs Europe							3.25	(2.36)	4.43*	(2.10)	4.43*	(2.10)

Table 19: Continued

Cross-level interactions		9.9E-06		(1.4E-05)	
DTA×Hotel Size					
Random part					
σ_e^2	721.61***	559.45***	556.03***	539.00***	539.00***
σ_{u0}^2	57.49***	46.51***	43.32***	28.98***	28.97***
σ_{u1}^2				0.01***	0.01***
σ_{u01}^2				-0.09***	-0.09**
-2*log likelihood	2,123,693.04	2,048,657.84	2,046,562.43	2,041,507.06	2,041,506.55
AIC	2,123,699.04	2,048,671.84	2,046,586.43	2,041,555.06	2,041,556.55
BIC	2,123,730.02	2,048,744.06	2,046,710.22	2,041,802.64	2,041,814.44

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 19 presents the results of six multilevel models related to the improvement in Symmetric Absolute Percentage Error. The intercept-only model 1 for $sAPE_{imp}$ estimated the within-hotel variance as 721.61 and the hotel level variance as 57.49 (Wald $Z = 17.71, p < .001$), suggesting variation in intercepts across hotels, i.e. hotels differ in forecasting accuracy.

In Model 2, the four level-1 predictors (i.e., DTA, Business Type, Size of Adjustment, Direction) were added while keeping the slope fixed for all hotels. As Table 19 shows, the regression coefficients for all four variables were significant. The regression coefficient for DTA is -0.02, which means that for each extra day added to the forecast window, the improvement of forecast accuracy is expected to decrease by 0.02. Better performance of the user becomes apparent as the date of arrival approaches. The DTA in the data for this study range from 0 to 596 days, so the predicted difference in improvement between the override for the closest to the day of arrival and the furthest was $(596 - 0) \times 0.02 = 11.92$ (percentage points). Hence, the further away from the occupancy day that an override was made, the smaller the expected improvement in accuracy. This outcome provides initial support for Hypothesis 1, which stated that early forecast adjustments are associated with lower improvements in accuracy than the late adjustments.

The regression coefficient for Business type was 2.28 ($p < .001$). Since group was coded as 1, this means that on average, group businesses are associated with larger improvements in accuracy (an increase of $sAPE$ by 2.28) than transient businesses, when all other variables are kept constant. This provides initial support for Hypothesis 2, which states that overrides to group business are associated with larger improvements in accuracy than overrides to transient business.

The regression coefficient for size of adjustment was 6.19 ($p < .001$). The coefficient of 6.19 indicates that on average, every unit increase in size of adjustment is associated with a 6.19 percentage points increase in the improvement in forecast accuracy ($sAPE_{imp}$). This provides initial support for Hypothesis 3, which states that the size of judgmental adjustments is positively associated with improvement in accuracy.

The regression coefficient for direction of override was 1.85 ($p < .001$). Since the direction of override was coded as negative = 0, positive = 1, this means that on average, the positive judgmental adjustments have a 1.85 percentage points higher increase in forecast improvement than negative, when all other variables are kept constant. This does not provide initial support for Hypothesis 4 which states that negative judgmental adjustments are associated with larger improvements in accuracy than positive.

If all main level-1 predictors are included in the initial intercept-only model, the within-hotel variance decreased from 721.61 to 559.45. The estimate of between-hotel variance decreased from 57.49 to 46.51. The chi-square difference test comparing Model 2 to Model 1, indicated that adding the main level-1 predictors led to a significant improvement in the model ($p < .001$).

In Model 3, control variables at level-1 (i.e., Special, Override Number, Override Type, Override Status, Override Recency Status) are added. As Table 19 shows, the main effects remain similar when compared to Model 2. The regression coefficient for special events was not significant ($p = .194$), which means that there was no significant difference in improvement in accuracy for the overrides related to days of arrival with and without special events. The regression coefficient for the override number was 0.04 ($p = .008$), i.e. for each additional (re)adjustment for the same occupancy date, the average improvement in forecast accuracy increased by 0.04 percentage points. The regression coefficient for override type was 2.94 ($p < .001$). As override type was coded as wash = 0, demand = 1, this means that on average, demand overrides had a 2.94-point higher improvement in accuracy than wash overrides, when all other variables were kept constant. The regression coefficient for override status was -3.08 ($p < .001$). As the override status was coded as reverted = 0, active = 1, on average, reverted overrides had a 3.08-point higher improvement in accuracy than active overrides. The regression coefficient for the second indicator of the override status (i.e., override *recency* status), however, was not significant. This indicates that there was no difference between the past or the latest adjustments. When additional level-1 predictors were included in Model 3, the within-hotel variance decreased from 559.45 to 556.03. The estimate of between-hotel variance decreased from 46.51 to 43.32. The model fit test comparing Model 3 to Model 2 indicated that inclusion of control variables at level-1 significantly improved the model ($p < .001$).

Model 4 added level-2 or hotel-level control variables: hotel size, hotel type and hotel geographical region. None of these predictors had significant relationships, showing that the improvement in forecast accuracy did not depend on the hotels' size, type, or geographical regions. The model fit test which compared Model 4 to Model 3 resulted in $p = 0.87$, showing that the addition of the level-2 covariates did not improve the model.

Models 1 – 4 assumed that the slope was the same for all hotels. In Model 5, the DTA slope was allowed to vary across hotels. The level-1 variance estimate was 539.00 (Wald $Z = 331.78$, $p < .001$). At level-2, an unstructured covariance matrix was specified enabling an estimation of the variance of the random intercept ($\sigma_{u_0}^2$), the variance of the random linear slope ($\sigma_{u_1}^2$), and the covariance between the intercept and the slope ($\sigma_{u_{01}}$). The results suggest significant variability in

random intercept between hotels (Wald $Z = 14.70$, $p < .001$), the linear slope varied significantly across hotels (Wald $Z = 14.51$, $p < .001$), and a significant covariance between the level-2 initial status and linear change estimates (Wald $Z = -3.33$, $p < .01$). The chi-square difference test, which compared Model 5 to Model 4, indicated that adding the random slope for DTA led to a significant improvement in the model ($p < .001$). Thus, hotels significantly differed in how the forecast improvement measure ($sAPE_{imp}$) was affected by the date to arrival.

In Model 6, the cross-level interaction between hotel size and DTA was added to explore whether the change in $sAPE_{imp}$ over time (DTA) was impacted by hotel size. As Table 19 shows, the interaction term was not significant ($p = 0.48$), which means that DTA does not affect forecast accuracy differently for larger than average hotels compared to smaller than average hotels (hotel size was centered on its grand mean). Thus, hotel size cannot explain why there is a difference in the effect of DTA on $sAPE_{imp}$ across hotels. The model fit test which compared Model 6 to Model 5 resulted in $p = 0.48$; the addition of the cross-level interaction did not improve the model.

The results for Model 6 for LnQ_{imp} , APE_{imp} , and AE_{imp} are reported in Tables 20 - 22. It should be noted that the interaction between Hotel size and DTA was not significant for all three scale-independent accuracy measures, but it was significant for AE_{imp} .

Table 20: Fixed effects estimates and variance-covariance estimates for models of the predictors of improvement in LnQ

Model:	Step 1: Null model		Step 2: Level-1 predictors		Step 3: Level-2 predictors		Step 4: Random coefficient	Step 5: Cross-level interactions		
	M ₁ : Intercept only	M ₂ : Level -1 model with main predictors	M ₃ : Level -1 model with all predictors	M ₄ : final Level-1 and 2 model	M ₅ : random DTA	M ₆ : Cross-level interaction of DTA and Hotel Size				
DV = LnQ _{emp}										
Fixed part										
Intercept	1.8E-2***	(3.0E-3)	-8.5E-3**	(3.1E-3)	-2.0E-2***	(4.5E-3)	-1.1E-2*	(4.8E-3)	-1.1E-2* (4.8E-3)	
Level-1										
Main predictors										
DTA			-2.6E-4***	(1.0E-5)	-2.3E-4***	(1.0E-5)	-2.6E-4***	(5.3E-5)	-2.6E-4*** (5.4E-5)	
Business Type			3.0E-2***	(1.7E-3)	5.6E-2***	(2.2E-3)	4.9E-2***	(2.2E-3)	4.9E-2*** (2.2E-3)	
Size Adjustment Abs			8.1E-2***	(2.1E-3)	7.3E-2***	(2.2E-3)	8.6E-2***	(2.2E-3)	8.6E-2*** (2.2E-3)	
Direction			2.1E-2***	(1.4E-3)	1.9E-2***	(1.4E-3)	1.9E-2***	(1.4E-3)	1.9E-2*** (1.4E-3)	
Control variables										
Special					3.3E-3	(1.9E-3)	4.5E-3*	(1.9E-3)	4.5E-3* (1.9E-3)	
Override Number					4.0E-4*	(1.8E-4)	3.6E-4*	(1.8E-4)	3.6E-4* (1.8E-4)	
Override Type					4.1E-2***	(2.5E-3)	3.4E-2***	(2.4E-3)	3.4E-2*** (2.4E-3)	
Override Recency Status					-3.7E-2***	(1.8E-3)	-4.0E-2***	(1.7E-3)	-4.0E-2*** (1.7E-3)	
Override Recency Status					3.7E-4	(1.7E-3)	7.1E-4	(1.7E-3)	7.2E-4 (1.7E-3)	
Level-2										
Hotel Size						6.2E-6	(1.3E-5)	4.4E-6	(1.0E-5)	3.5E-6 (1.0E-5)
Hotel Type										
Airport vs City						-1.1E-2	(2.4E-2)	-7.1E-3	(2.0E-2)	-7.1E-3 (2.0E-2)
Casino vs City						-6.5E-3	(2.2E-2)	-2.4E-2	(1.9E-2)	-2.4E-2 (1.9E-2)
Resort vs City						3.1E-3	(9.7E-3)	7.0E-3	(8.1E-3)	7.1E-3 (8.1E-3)
Hotel Region										
Asia vs Europe						-1.1E-2	(9.3E-3)	3.2E-3	(7.6E-3)	3.2E-3 (7.6E-3)
Australasia vs Europe						1.3E-2	(1.8E-2)	-2.2E-3	(1.5E-2)	-2.2E-3 (1.5E-2)
Central America vs Europe						2.8E-2	(5.1E-2)	1.2E-1*	(4.4E-2)	1.2E-1* (4.4E-2)
Middle East & Africa vs Europe						-5.2E-4	(1.2E-2)	-8.6E-3	(1.0E-2)	-8.7E-3 (1.0E-2)
North America vs Europe						2.6E-3	(7.8E-3)	-5.1E-3	(6.4E-3)	-5.2E-3 (6.4E-3)
South America vs Europe						3.5E-2	(3.3E-2)	5.4E-2	(2.8E-2)	5.4E-2 (2.8E-2)

Table 20: Continued

Cross-level interactions					
DTA×Hotel Size					
Random part					
σ_e^2	9.9E-2***	9.8E-2***	9.8E-2***	9.4E-2***	9.4E-2***
σ_{u0}^2	9.4E-3***	9.1E-3***	8.6E-3***	4.8E-3***	4.8E-3***
σ_{u1}^2				2.3E-6***	2.3E-6***
σ_{U01}^2				-1.2E-5***	-1.2E-5*
-2*log likelihood					
AIC	119,597.26	117,047.31	115,713.53	109,421.85	109,421.56
	119,603.26	117,061.31	115,737.53	109,469.85	109,471.56
BIC	119,634.18	117,133.45	115,861.20	109,717.19	109,729.20

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group), Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand), Override Status (0 = reverted, 1 = active), Override Recency Status (0 = history, 1 = latest), Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39). * $p < .05$, ** $p < .01$, *** $p < .001$

Table 21: Fixed effects estimates and variance-covariance estimates for models of the predictors of improvement in APE

Model:	Step 1: Null model		Step 2: Level-1 predictors		Step 3: Level-2 predictors		Step 4: Random coefficient	Step 5: Cross-level interactions	
	M ₁ : Intercept only		M ₂ : Level-1 model with main predictors		M ₃ : Level-1 model with all predictors		M ₄ : final Level-1 and 2 model	M ₅ : random DTA	M ₆ : Cross-level interaction of DTA and Hotel Size
DV = APE _{Emp}									
Fixed part									
Intercept	4.09***	(0.90)	13.93***	(0.96)	25.96***	(2.30)	27.98***	(2.37)	22.17*** (2.28)
Level-1									
Main predictors									
DTA			-0.12***	(6.5E-03)	-0.09***	(6.7E-03)	-0.09***	(6.8E-03)	-0.11*** (2.7E-02)
Business Type			5.57***	(1.05)	3.80**	(1.42)	3.92**	(1.42)	5.51*** (1.39)
Size Adjustment Abs			39.08***	(1.41)	39.39***	(1.44)	39.48***	(1.44)	46.33*** (1.46)
Direction			-31.30***	(0.93)	-32.56***	(0.94)	-32.65***	(0.94)	-30.04*** (0.94)
Control variables									
Special					-1.52	(1.26)	-1.44	(1.26)	0.48 (1.25)
Override Number					0.11	(0.12)	0.11	(0.12)	0.14 (0.10)
Override Type					-4.44**	(1.53)	-3.94*	(1.54)	-2.27 (1.50)
Override Status					-14.83***	(1.14)	-14.80***	(1.14)	-14.27*** (1.11)
Override Recency Status					2.39*	(1.11)	2.38*	(1.11)	2.93** (1.09)
Level-2									
Hotel Size							3.6E-03	(2.5E-03)	1.4E-03 (1.9E-03)
Hotel Type									
Airport vs City							-5.88	(6.13)	-4.37 (5.21)
Casino vs City							-0.07	(5.69)	0.35 (4.23)
Resort vs City							-0.04	(2.23)	-1.50 (1.82)
Hotel Region									
Asia vs Europe							-5.32*	(2.22)	-1.85 (1.75)
Australasia vs Europe							-2.03	(4.42)	-0.72 (3.70)
Central America vs Europe							-26.29*	(11.85)	69.54*** (11.88)
Middle East & Africa vs Europe							-10.47**	(3.03)	-3.29 (2.43)
North America vs Europe							-3.22	(1.71)	-0.59 (1.25)
South America vs Europe							-18.15*	(8.85)	-4.40 (8.69)

Table 21: Continued

Cross-level interactions					
DTA× Hotel Size					
Random part					
σ_e^2	46,713.04***	45,346.47***	45,316.92***	44,276.69***	44,276.64***
σ_{u0}^2	448.64***	170.18***	143.51***	250.84***	251.00***
σ_{u1}^2				0.53***	0.53***
σ_{U01}^2				-11.50***	-11.50***
-2*log likelihood					
AIC	3,031,177.83	2,997,865.17	2,996,688.14	2,993,081.59	2,993,080.88
	3,031,183.83	2,997,879.17	2,996,712.14	2,993,129.59	2,993,130.88
BIC	3,031,214.77	2,997,951.32	2,996,835.81	2,993,376.93	2,993,388.53

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 22: Fixed effects estimates and variance-covariance estimates for models of the predictors of improvement in AE

Model:	Step 1: Null model		Step 2: Level-1 predictors		Step 3: Level-2 predictors		Step 4: Random coefficient		Step 5: Cross-level interactions		
	M ₁ : Intercept only		M ₂ : Level -1 model with main predictors		M ₃ : Level -1 model with all predictors		M ₄ : final Level-1 and 2 model		M ₅ : Cross-level interaction of DTA and Hotel Size		
DV = AE _{imp}											
Fixed part	Intercept	1.80***	(0.22)	3.05***	(0.25)	3.74***	(0.44)	4.09***	(0.49)	4.46***	(0.47)
	Level-1										
	Main predictors										
	DTA	-0.03***	(1.1E-03)	-0.03***	(1.2E-03)	-0.03***	(1.2E-03)	-0.03***	(1.2E-03)	-0.02***	(4.5E-03)
	Business Type	-0.07	(0.19)	0.84**	(0.25)	0.84**	(0.25)	0.82**	(0.25)	0.29	(0.24)
Size Adjustment Abs	7.97***	(0.22)	7.68***	(0.23)	7.68***	(0.23)	7.69***	(0.23)	8.21***	(0.23)	
	Direction	-2.81***	(0.16)	-2.99***	(0.16)	-2.99***	(0.16)	-2.98***	(0.16)	-2.91***	(0.16)
Control variables											
Special			-0.15	(0.22)	-0.16	(0.22)	-0.16	(0.22)	0.22	(0.22)	
Override Number			0.08***	(0.02)	0.08***	(0.02)	0.08***	(0.02)	0.07**	(0.02)	
Override Type			1.30***	(0.27)	1.30***	(0.27)	1.33***	(0.27)	0.62*	(0.27)	
Override Status			-3.27***	(0.20)	-3.26***	(0.20)	-3.26***	(0.20)	-3.23***	(0.20)	
Override Recency Status			0.15	(0.19)	0.16	(0.19)	0.16	(0.19)	0.58**	(0.19)	
Level-2											
Hotel Size					3.7E-03***	(9.0E-04)	1.1E-02***	(8.1E-04)	1.2E-02***	(8.4E-04)	
Hotel Type											
Airport vs City					-1.53	(1.83)	-1.53	(1.83)	-1.40	(1.71)	
Casino vs City					-9.08***	(1.72)	-9.08***	(1.72)	-6.11***	(1.60)	
Resort vs City					0.69	(0.71)	0.69	(0.71)	0.04	(0.67)	
Hotel Region											
Asia vs Europe					-0.84	(0.69)	-0.84	(0.69)	-1.71**	(0.63)	
Australasia vs Europe					1.43	(1.35)	1.43	(1.35)	-0.46	(1.24)	
Central America vs Europe					4.65	(3.77)	4.65	(3.77)	15.20***	(3.71)	

Table 22: Continued

Middle East & Africa vs Europe	0.17	(0.95)	-0.19	(0.88)	-0.15	(0.88)
North America vs Europe	-0.50	(0.57)	-0.59	(0.52)	-0.57	(0.52)
South America vs Europe	2.97	(2.48)	4.13	(2.40)	4.14	(2.40)
Cross-level interactions						
DTA×Hotel Size					-8.5E-05***	(1.6E-05)
Random part						
σ_e^2	1,311.57***	1,304.74***	1,283.00***	1,283.05***	1,230.54***	1,230.63***
σ_{u0}^2	41.68***	40.68***	38.32***	36.18***	27.43***	27.03***
σ_{u1}^2					0.01***	0.01***
σ_{U01}^2					-0.16***	-0.14***
-2*log likelihood						
AIC	2,292,121.09	2,236,789.86	2,232,282.70	2,232,244.62	2,224,559.18	2,224,531.40
BIC	2,292,127.09	2,236,803.86	2,232,306.70	2,232,288.62	2,224,607.18	2,224,581.40
	2,292,158.11	2,236,876.07	2,232,430.49	2,232,515.56	2,224,854.75	2,224,839.29

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).
* $p < .05$, ** $p < .01$, *** $p < .001$

Focusing on Model 6 of the three scale-independent measures, the following was observed. First, DTA was a significant predictor with a negative coefficient for all forecast accuracy measures. Hence, support was found for Hypothesis 1: early judgmental forecast adjustments are associated with lower improvements in accuracy than late forecast adjustments. Second, business type was a significant predictor with a positive coefficient for $sAPE_{imp}$, LnQ_{imp} , and APE_{imp} .²⁵ Hypothesis 2 was supported which states that overrides to group business are associated with larger improvements in accuracy than overrides to transient business. Third, the size of adjustment was a significant predictor with a positive coefficient for all measures. Thus on average, larger overrides are associated with higher improvements in forecast accuracy, which supports Hypothesis 3: the size of judgmental adjustments is positively associated with improvements in accuracy. Fourth, the direction of the judgmental adjustment was a significant predictor for all forecast improvement measures. However, for APE_{imp} and AE_{imp} , the coefficient was negative, while for $sAPE_{imp}$ and LnQ_{imp} it was positive. The latter results can be interpreted as follows: positive judgmental adjustments lead to larger improvements in forecast accuracy compared to negative adjustments. The results for APE_{imp} and AE_{imp} lead to the opposite conclusion. Therefore, these results do not provide conclusive support for Hypothesis 4, which states that negative user overrides are associated with a higher improvement in accuracy than positive overrides. This finding supports the outcomes of Chapter 2: important conclusions might differ depending on the forecast accuracy measure used and results obtained with different measures can lead to contradictory conclusions.

Moreover, the four control variables at level-1 (i.e., Special, Override Number, Override Type, Override Status, Override Recency Status) were explored. Special events, coded as no special event = 0, special event = 1, were only significant for LnQ_{imp} ($p = 0.02$), not for APE_{imp} and $sAPE_{imp}$. For LnQ_{imp} the coefficient was positive, which means that in the presence of special events, the improvement in forecasting accuracy is higher when compared to adjustments made for days without special events. Number of overrides had a significant effect with positive coefficients for $sAPE_{imp}$, LnQ_{imp} , and AE_{imp} . This means that repetitive overrides for the same occupancy date are, on average, associated with a higher improvement in accuracy. The number of overrides was, however, not a significant predictor for APE_{imp} . Override type, coded as wash overrides = 0, and demand overrides = 1, was significant with a positive coefficient for $sAPE_{imp}$, LnQ_{imp} , and AE_{imp} , indicating that demand overrides were associated with higher improvement in accuracy compared to wash overrides. It was not significant for APE_{imp} . The negative coefficient for override status, coded

25 Business type was not significant for AE_{imp} . When Model 3 and Model 4 are reviewed for AE_{imp} (Table 22), it can be noticed that Business Type was a significant predictor with a positive coefficient.

as reverted = 0, active = 1, indicated that active overrides were associated with lower improvements in accuracy compared to the reverted overrides. This predictor was significant for APE_{imp} , $sAPE_{imp}$, LnQ_{imp} , and AE_{imp} . Override Recency Status was not significant for $sAPE_{imp}$ and LnQ_{imp} , but was significant for APE_{imp} and AE_{imp} with a positive coefficient, meaning that no conclusion can be drawn for the Recency Status predictor.

With reference to the three level-2 control variables, hotel size and hotel type were not significant for $sAPE_{imp}$, LnQ_{imp} , and APE_{imp} . Table 22 shows that hotel size was significant for AE_{imp} , for the casino hotel type when compared to city hotels. In terms of the impact of hotel region, all forecast accuracy measures agree that there was significant difference between Central America and Europe, where the former had a higher improvement in forecast accuracy. According to $sAPE_{imp}$, significant differences were also found between South America and Europe, where overrides in South America led to greater improvements in accuracy compared to Europe. According to AE_{imp} , significant differences were also found between Asia and Europe.

Table 23 provides a summary of the results for each hypothesis for all four forecast improvement measures. As can be observed in the table, there is sufficient evidence to support hypotheses 1, 2, and 3. As was anticipated in Table 7, various accuracy measures perform differently for the same magnitude of error in case of overforecasting versus underforecasting. This measure behavior may have impacted the analysis of positive and negative judgmental adjustments and is in line with the findings of Chapter 2. Further insights should be obtained for Hypothesis 4.

Table 23: Summary results hypotheses 1-4

Hypothesis	$sAPE_{imp}$	LnQ_{imp}	APE_{imp}	Conclusion	AE_{imp}
H ₁ : Early forecast adjustments are associated with lower improvements in accuracy than the late adjustments.	Yes	Yes	Yes	Supported	Yes
H ₂ : Judgmental forecast adjustments to group business are associated with larger improvements in accuracy than adjustments to transient business.	Yes	Yes	Yes	Supported	No
H ₃ : Size of judgmental adjustments is positively associated with improvement in accuracy.	Yes	Yes	Yes	Supported	Yes
H ₄ : Negative judgmental adjustments are associated with higher improvement in accuracy than positive.	No	No	Yes	Not supported	Yes

Note. Conclusions on the hypotheses testing were drawn on the basis of the three scale-independent measures.

4.4.3 Additional Analysis

Thus far, the chapter has focused on the main effects. However, the underlying processes may be even more complex. For example, the impact of the main predictor may depend on the value of another predictor. The presence of such interaction effects can have important implications for the interpretation of the statistical

Table 24: Models with level-1 interaction terms to further explore the relationship between DTA and sAPE_{imp}

Model: DV = sAPE _{imp}	M _{6,1,1} [*]	M _{6,1,2} [*]	M _{6,1,3} [*]	M _{6,1,4} [*]	M _{6,1,5} [*]	M _{6,1,6} [*]
Fixed part						
Intercept	-0.32 (0.37)	-1.42*** (3.9E-3)	0.52 (0.37)	-0.36 (0.36)	-0.35 (0.37)	0.96* (0.39)
DTA	-0.02*** (0.19)	-0.01 (0.16)	-0.04*** (3.9E-3)	-0.02*** (0.16)	-0.02*** (3.9E-3)	-0.05*** (4.5E-3)
Business Type	3.47*** (0.15)	3.29*** (0.21)	3.59*** (0.15)	3.65*** (0.15)	3.64*** (0.15)	1.93*** (0.20)
Size Adjustment Abs	6.47*** (0.10)	12.01*** (0.10)	6.57*** (0.13)	6.48*** (0.10)	6.47*** (0.10)	12.52*** (0.21)
Direction	1.58*** (0.14)	1.67*** (0.14)	0.16 (0.23)	1.58*** (0.13)	1.58*** (0.10)	0.26* (0.13)
Special	0.27 (0.01)	0.26 (0.01)	0.23 (0.03)	-0.05 (0.01)	0.27 (0.01)	-0.09 (0.01)
Override Number	0.03*	0.04**	0.03*	0.03*	0.03*	0.04**
Override Type	2.39*** (0.18)	2.30*** (0.18)	2.34*** (0.18)	2.43*** (0.18)	2.32*** (0.18)	0.82*** (0.22)
Override Status	-3.21*** (0.13)	-3.15*** (0.13)	-3.33*** (0.13)	-3.21*** (0.13)	-3.21*** (0.13)	-3.29*** (0.13)
Override Recency Status	0.09 (0.13)	0.24 (0.13)	0.08 (0.13)	0.09 (0.13)	0.09 (0.13)	0.21 (0.13)
DTA×Business Type	3.1E-3 (1.7E-3)					0.02*** (2.1E-3)
DTA×Size Adjustment Abs		-0.06*** (1.6E-3)				-0.06*** (1.6E-3)
DTA×Direction			0.03*** (1.5E-3)			0.03*** (1.5E-3)
DTA×Special				0.01** (1.7E-3)		0.01** (1.7E-3)
DTA×Override Type					2.3E-3 (7.9E-4)	0.02*** (2.5E-3)
Hotel Type						
Hotel Size	4.2E-4 (7.9E-4)	8.5E-4 (7.9E-4)	4.1E-4 (7.9E-4)	4.2E-4 (7.9E-4)	3.7E-4 (7.9E-4)	9.1E-4 (7.9E-4)
Airport vs City	-0.90 (1.53)	-0.65 (1.53)	-0.96 (1.53)	-0.90 (1.53)	-0.92 (1.53)	-0.84 (1.54)
Casino vs City	-2.15 (1.43)	-2.34 (1.43)	-2.16 (1.43)	-2.13 (1.43)	-2.09 (1.43)	-2.45 (1.44)
Resort vs City	0.38 (0.61)	0.27 (0.62)	0.41 (0.62)	0.37 (0.62)	0.37 (0.61)	0.29 (0.62)
Hotel Region						
Asia vs Europe	0.21 (0.58)	0.20 (0.58)	0.10 (0.58)	0.21 (0.58)	0.23 (0.58)	0.19 (0.59)
Australasia vs Europe	-0.04 (1.13)	0.45 (1.14)	-0.10 (1.14)	-0.06 (1.13)	-0.04 (1.13)	0.39 (1.14)
Central America vs Europe	9.25** (3.38)	8.88** (3.39)	9.05** (3.39)	9.28** (3.38)	9.27** (3.37)	8.61* (3.41)
Middle East & Africa vs Europe	-0.72 (0.79)	-0.70 (0.80)	-0.77 (0.80)	-0.70 (0.79)	-0.70 (0.79)	-0.73 (0.80)
North America vs Europe	-0.40 (0.48)	-0.29 (0.49)	-0.48 (0.49)	-0.38 (0.48)	-0.37 (0.48)	-0.33 (0.49)
South America vs Europe	4.44* (2.10)	3.82 (2.11)	4.45* (2.11)	4.46* (2.11)	4.44* (2.10)	4.05 (2.12)

Table 24: Continued

DTA×Hotel Size	9.7E-6 (1.4E-5)	6.2E-6 (1.4E-5)	1.3E-5 (1.4E-5)	9.8E-6 (1.4E-5)	1.0E-5 (1.4E-5)	7.6E-6 (1.4E-5)
Random part						
σ_e^2	538.99***	535.44***	538.03***	538.97***	539.00***	534.12***
σ_{u0}^2	29.05***	28.82***	29.23***	29.01***	28.94***	29.28***
σ_{u1}^2	0.01***	0.01***	0.01***	0.01***	0.01***	0.01***
σ_{U01}^2	-0.10**	-0.07*	-0.09**	-0.09**	-0.09**	-0.07*
-2*log likelihood						
AIC	2,041,503.18	2,040,078.83	2,041,105.33	2,041,496.75	2,041,505.30	2,039,514.87
BIC	2,041,555.18	2,040,130.83	2,041,157.33	2,041,548.75	2,041,557.30	2,039,574.87
	2,041,823.39	2,040,399.03	2,041,425.53	2,041,816.96	2,041,825.51	2,039,884.34

Note. N = 228,649. Standard errors are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).

models in Tables 19-22 (Dawson, 2014). Therefore, to gain a deeper understanding of possible interaction effects, additional analyses were performed to further explore each of the four hypotheses. The significant results are reported below.

- H₁: Early forecast adjustments are associated with lower improvements in accuracy than the late forecast adjustments.
- 1.1. Does the relationship between DTA and forecast improvement depend on business type?
 - 1.2. Does the relationship between DTA and forecast improvement depend on adjustment size?
 - 1.3. Does the relationship between DTA and forecast improvement depend on override direction?
 - 1.4. Does the relationship between DTA and forecast improvement depend on special events?
 - 1.5. Does the relationship between DTA and forecast improvement depend on override type?

Table 24 reports each interaction term separately (Models 6.1.1 – 6.1.5) and then simultaneously (Model 6.1.6). Taking one interaction term at a time, it can be observed that the impact of DTA on forecast improvement depends on the size of the adjustment (Model 6.1.2), the direction (Model 6.1.3), the presence of a special event (Model 6.1.4), but not on the business type (Model 6.1.1) and the override type (Model 6.1.5). However, when the moderators are added simultaneously (Model 6.1.6), all five interaction terms are significant.²⁶

As the dotted baseline in Figure 10 illustrates, forecast adjustments made further away from the day of arrival are associated with lower improvements in accuracy than those made closer to the day of arrival. However, this relationship depends on the business type, the direction, the presence of a special event, and the override type. The negative relationship between the DTA and forecast improvement is weaker for group segment compared with the transient segment. Similar interaction effects can be observed for positive (versus negative) overrides, special events (versus no special events), and demand (versus wash) overrides. The negative impact of DTA on forecast improvement is thus weaker for group business forecasts, positive overrides, special events, and demand overrides compared with the baseline. As Figure 11 shows, the opposite effect can be observed for the size of the judgmental adjustments. The negative relationship between DTA and forecast improvement is stronger for larger than for smaller adjustments.

26 According to the model fit tests, Model 6.1.6 has a superior model fit compared to Models 6.1.1 – 6.1.5. Therefore, in the summary of the additional analysis at the end of this section, the outcome of Model 6.1.6 is reported.

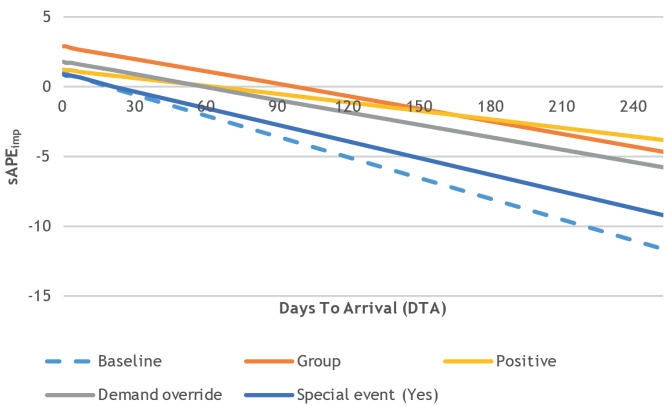


Figure 10: Regression lines for improvement in sAPE by DTA, for business type, direction, override type and special event based on Model 6.1.6²⁷

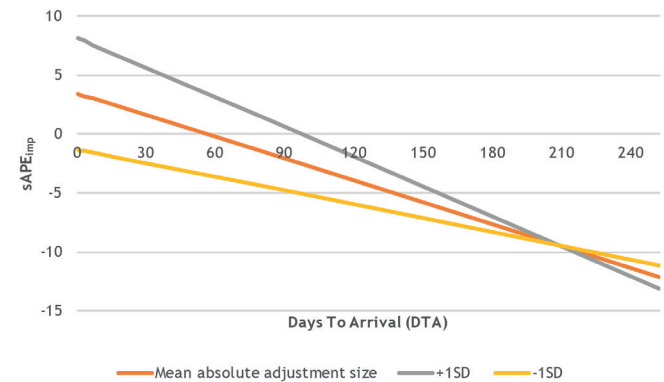


Figure 11: Regression lines for improvement in sAPE by DTA, for override size based on Model 6.1.6

27 In Figure 10 the baseline represents the effect of DTA on accuracy improvement when the values of all other predictors are zero.

Table 25: Models with level-1 interaction terms for additional analysis of relationship between DTA and improvement in accuracy

Model:	$M_{6.1.6}^{\cdot}$ DV = sAPE _{imp}		$M_{6.2.6}^{\cdot}$ DV = LnQ _{imp}		$M_{6.3.6}^{\cdot}$ DV = APE _{imp}		$M_{6.4.6}^{\cdot}$ DV = AE _{imp}	
Fixed part								
Intercept	0.96*	(0.39)	8.5E-3	(5.1E-3)	-1.22	(2.55)	2.96***	(0.51)
DTA	-0.05***	(4.5E-3)	-6.6E-4***	(6.3E-5)	0.31***	(0.03)	-1.6E-3	(0.01)
Business Type	1.93***	(0.20)	1.7E-2***	(2.7E-3)	6.81***	(1.72)	0.17	(0.30)
Size Adjustment Abs	12.52***	(0.21)	2.0E-1***	(3.0E-3)	95.30***	(1.99)	12.54***	(0.32)
Direction	0.26*	(0.13)	2.7E-3	(1.7E-3)	-10.51***	(1.13)	-0.95***	(0.19)
Special	-0.09	(0.18)	-8.5E-4	(2.3E-3)	0.21	(1.52)	-0.14	(0.26)
Override Number	0.04**	(0.01)	5.4E-4**	(1.8E-4)	0.23*	(0.10)	0.08***	(0.02)
Override Type	0.82***	(0.22)	5.4E-3	(3.0E-3)	0.36	(1.83)	-0.13	(0.33)
Override Status	-3.29***	(0.13)	-4.0E-2***	(1.7E-3)	-12.10***	(1.11)	-3.06***	(0.20)
Override Recency Status	0.21	(0.13)	3.0E-3	(1.7E-3)	4.19***	(1.09)	0.69***	(0.19)
DTA×Business Type	0.02***	(2.1E-3)	4.1E-4***	(2.8E-5)	-0.09***	(0.02)	-2.8E-3	(3.1E-3)
DTA×Size Adjustment Abs	-0.06***	(1.6E-3)	-1.3E-3***	(2.3E-5)	-0.55***	(0.02)	-0.05***	(2.4E-3)
DTA×Direction	0.03***	(1.5E-3)	3.6E-4***	(2.0E-5)	-0.40***	(0.01)	-0.04***	(2.3E-3)
DTA×Special	0.01**	(1.7E-3)	7.7E-5**	(2.3E-5)	0.02	(0.02)	0.01*	(2.6E-3)
DTA×Override Type	0.02***	(2.5E-3)	4.4E-4***	(3.3E-5)	-0.04*	(0.02)	0.02***	(3.7E-3)
Hotel Size	9.1E-4	(7.9E-4)	1.3E-5	(1.0E-5)	0.01*	(2.8E-3)	0.01***	(8.3E-4)
Hotel Type								
Airport vs City	-0.84	(1.54)	-5.8E-3	(2.0E-2)	-3.48	(5.28)	-1.25	(1.69)
Casino vs City	-2.45	(1.44)	-3.0E-2	(1.9E-2)	-0.16	(4.35)	-5.93***	(1.59)
Resort vs City	0.29	(0.62)	6.4E-3	(8.2E-3)	-1.88	(1.86)	-0.19	(0.66)
Hotel Region								
Asia vs Europe	0.19	(0.59)	2.7E-3	(7.8E-3)	-1.67	(1.80)	-1.48*	(0.63)
Australasia vs Europe	0.39	(1.14)	5.9E-3	(1.5E-2)	3.25	(3.74)	-0.02	(1.23)
Central America vs Europe	8.61*	(3.41)	9.9E-2*	(4.5E-2)	70.04***	(12.04)	15.56***	(3.67)
Middle East & Africa vs Europe	-0.73	(0.80)	-9.5E-3	(1.1E-2)	-2.43	(2.48)	0.06	(0.87)
North America vs Europe	-0.33	(0.49)	-4.2E-3	(6.5E-3)	0.09	(1.31)	-0.24	(0.51)
South America vs Europe	4.05	(2.12)	4.6E-2	(2.8E-2)	-11.52	(8.77)	3.75	(2.38)
DTA×Hotel Size	7.6E-6	(1.4E-5)	4.7E-8	(2.0E-7)	-1.5E-4	(9.5E-5)	-9.1E-5***	(1.6E-5)
Random part								
σ_e^2	534.12***		9.3E-2***		43,787.82***		1,226.35	
σ_{u0}^2	29.28***		5.0E-3***		212.78***		25.62	
σ_{u1}^2	0.01***		2.3E-6***		0.51***		0.01	
$\sigma_{U_{01}}^2$	-0.07*		-6.2E-6		-10.24***		-0.11	
-2*log likelihood	2,039,514.87		105,982.37		2,990,640.51		2,223,735.73	
AIC	2,039,574.87		106,042.37		2,990,700.51		2,223,795.73	
BIC	2,039,884.34		106,351.54		2,991,009.68		2,224,105.20	

Note. N = 228,649. Standard errors are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 25 reports the final models with interactions for sAPE_{imp}, LnQ_{imp}, APE_{imp}, and AE_{imp} for additional analysis of H₁ (the results of sAPE_{imp} are reported again to facilitate the comparison). It shows inconsistencies among the measures on the impact of the interaction terms. For the majority of the interaction terms, forecast accuracy measures disagree on the significance or on the sign of the coefficients, or on both. The exception is DTA × Size Adjustment, showing that the impact of DTA on forecast improvement depends on the adjustment size. For larger adjustments, DTA has a stronger negative relationship with forecast improvement than for smaller adjustments (Figure 11).

A second subset of additional research questions were formulated to explore the interaction effects of override direction, special events, and override type on the relationship between business type and accuracy improvement:

H₂: Judgmental forecast adjustments to group business are associated with larger improvement in accuracy than adjustments to transient business.

- 2.1. Does the relationship between business type and forecast improvement depend on override direction?
- 2.2. Does the relationship between business type and forecast improvement depend on special events?
- 2.3. Does the relationship between business type and forecast improvement depend on override type?

As can be seen from Model 6.1.10 in Table 26, the impact of Business Type on improvement in forecast accuracy does not depend on Special events or on Override type, but it does depend on the direction of the override. The difference between the business types in terms of forecast improvement is bigger when it concerns negative overrides, as illustrated in Figure 12.

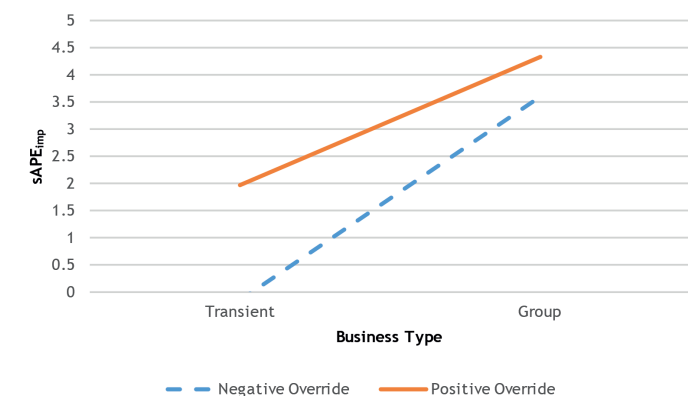
**Figure 12:** Regression lines for improvement in sAPE by Business Type, for override direction

Table 26: Models without and with level-1 interaction terms for additional analysis of relationship between business type and improvement in accuracy

Model:	$M_{6.1.7}^2$		$M_{6.1.8}^2$		$M_{6.1.9}^2$		$M_{6.1.10}^2$	
DV = sAPE _{imp}								
Fixed part								
Intercept	-0.66	(0.37)	-0.43	(0.36)	0.17	(0.47)	-0.15	(0.47)
DTA	-0.02***	(3.9E-3)	-0.02***	(3.9E-3)	-0.02***	(3.9E-3)	-0.02***	(3.9E-3)
Business Type	4.30***	(0.19)	3.66***	(0.17)	3.00***	(0.36)	3.75***	(0.38)
Size Adjustment Abs	6.47***	(0.15)	6.47***	(0.15)	6.43***	(0.15)	6.44***	(0.15)
Direction	2.12***	(0.13)	1.58***	(0.10)	1.59***	(0.10)	2.12***	(0.13)
Special	0.27	(0.14)	0.31	(0.18)	0.27	(0.14)	0.31	(0.18)
Override Number	0.03*	(0.01)	0.03*	(0.01)	0.03*	(0.01)	0.03*	(0.01)
Override Type	2.38***	(0.18)	2.43***	(0.18)	1.80**	(0.36)	1.84***	(0.36)
Override Status	-3.10***	(0.13)	-3.21***	(0.13)	-3.20***	(0.13)	-3.10***	(0.13)
Override Recency Status	0.14	(0.13)	0.09	(0.13)	0.10	(0.13)	0.15	(0.13)
Business Type×Direction	-1.40***	(0.22)					-1.39***	(0.22)
Business Type×Special			-0.09	(0.29)			-0.08	(0.29)
Business Type×Override Type					0.81*	(0.41)	0.69	(0.41)
Hotel Size	3.6E-4	(7.9E-4)	3.9E-4	(7.9E-4)	3.8E-4	(7.9E-4)	3.5E-4	(7.9E-4)
Hotel Type								
Airport vs City	-0.89	(1.53)	-0.90	(1.53)	-0.95	(1.53)	-0.93	(1.53)
Casino vs City	-2.08	(1.43)	-2.11	(1.43)	-2.12	(1.43)	-2.09	(1.43)
Resort vs City	0.41	(0.61)	0.37	(0.61)	0.36	(0.61)	0.40	(0.61)
Hotel Region								
Asia vs Europe	0.14	(0.58)	0.22	(0.58)	0.20	(0.58)	0.13	(0.58)
Australasia vs Europe	-0.09	(1.13)	-0.04	(1.13)	-0.06	(1.13)	-0.10	(1.13)
Central America vs Europe	9.18**	(3.37)	9.26**	(3.38)	9.23**	(3.37)	9.15**	(3.37)
Middle East & Africa vs Europe	-0.75	(0.79)	-0.71	(0.79)	-0.71	(0.79)	-0.75	(0.79)
North America vs Europe	-0.42	(0.48)	-0.39	(0.48)	-0.37	(0.48)	-0.41	(0.48)
South America vs Europe	4.45*	(2.10)	4.43*	(2.10)	4.42*	(2.10)	4.44*	(2.10)
DTA×Hotel Size	1.0E-5	(1.4E-5)	9.9E-6	(1.4E-5)	9.8E-6	(1.4E-5)	1.0E-5	(1.4E-5)
Random part								
σ_e^2	538.90***		539.00***		539.00***		538.90***	
σ_{u0}^2	28.95***		28.96***		28.92***		28.91***	
σ_{u1}^2	0.01***		0.01***		0.01***		0.01***	
σ_{u01}^2	-0.10**		-0.09**		-0.09**		-0.09**	
-2*log likelihood	2,041,464.66		2,041,506.44		2,041,502.55		2,041,461.68	
AIC	2,041,516.66		2,041,558.44		2,041,554.55		2,041,517.68	
BIC	2,041,784.87		2,041,826.65		2,041,822.76		2,041,806.52	

Note. N = 228,649. Standard errors are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$

In Table 27, the interaction effects on the relationship between Business Type and forecast accuracy are presented for each dependent variable for additional analysis of H_2 . The results for all interaction terms are consistent among all forecast accuracy measures.

Table 27: Models without and with level-1 interaction terms for additional analysis of relationship between business type and improvement in accuracy

Model:	$M_{6.1.10}^2$ DV = sAPE _{imp}		$M_{6.2.10}^2$ DV = LnQ _{imp}		$M_{6.3.10}^2$ DV = APE _{imp}		$M_{6.4.10}^2$ DV = AE _{imp}	
Fixed part								
Intercept	-0.15	(0.47)	-1.1E-2	(6.3E-3)	6.91*	(3.39)	3.86***	(0.65)
DTA	-0.02***	(3.9E-3)	-2.6E-4***	(5.4E-5)	-0.11***	(0.03)	-0.02***	(4.5E-3)
Business Type	3.75***	(0.38)	5.7E-2***	(5.1E-3)	40.11***	(3.23)	2.28***	(0.57)
Size Adjustment Abs	6.44***	(0.15)	8.6E-2***	(2.2E-3)	46.56***	(1.46)	8.19***	(0.23)
Direction	2.12***	(0.13)	2.9E-2***	(1.8E-3)	-5.33***	(1.19)	-1.21***	(0.20)
Special	0.31	(0.18)	5.3E-3*	(2.4E-3)	1.18	(1.57)	0.45	(0.27)
Override Number	0.03*	(0.01)	3.7E-4*	(1.8E-4)	0.08	(0.10)	0.07**	(0.02)
Override Type	1.84***	(0.36)	2.8E-2***	(4.9E-3)	-1.07	(2.98)	0.30	(0.54)
Override Status	-3.10***	(0.13)	-3.7E-2***	(1.8E-3)	-9.09***	(1.12)	-2.90***	(0.20)
Override Recency Status	0.15	(0.13)	1.7E-3	(1.7E-3)	5.66***	(1.09)	0.74***	(0.19)
Business Type×Direction	-1.39***	(0.22)	-2.8E-2***	(2.9E-3)	-64.50***	(1.94)	-4.43***	(0.33)
Business Type×Special	-0.08	(0.29)	-2.0E-3	(3.8E-3)	-1.42	(2.54)	-0.60	(0.43)
Business Type×Override Type	0.69	(0.41)	6.2E-3	(5.5E-3)	-5.43	(3.41)	0.22	(0.61)
Hotel Size	3.5E-4	(7.9E-4)	2.8E-6	(1.0E-5)	2.3E-3	(2.9E-3)	0.01***	(8.4E-4)
Hotel Type								
Airport vs City	-0.93	(1.53)	-7.1E-3	(2.0E-2)	-3.62	(5.22)	-1.33	(1.71)
Casino vs City	-2.09	(1.43)	-2.3E-2	(1.9E-2)	1.33	(4.26)	-5.99***	(1.60)
Resort vs City	0.40	(0.61)	7.7E-3	(8.1E-3)	-0.37	(1.83)	0.10	(0.67)
Hotel Region								
Asia vs Europe	0.13	(0.58)	1.7E-3	(7.6E-3)	-3.66*	(1.77)	-1.91**	(0.64)
Australasia vs Europe	-0.10	(1.13)	-3.3E-3	(1.5E-2)	-3.06	(3.71)	-0.61	(1.24)
Central America vs Europe	9.15**	(3.37)	1.1E-1*	(4.4E-2)	66.34***	(11.89)	15.19***	(3.71)
Middle East & Africa vs Europe	-0.75	(0.79)	-9.5E-3	(1.0E-2)	-5.20*	(2.44)	-0.28	(0.88)
North America vs Europe	-0.41	(0.48)	-5.8E-3	(6.3E-3)	-1.52	(1.27)	-0.67	(0.52)
South America vs Europe	4.44*	(2.10)	5.4E-2	(2.8E-2)	-3.58	(8.69)	4.21	(2.40)
DTA×Hotel Size	1.0E-5	(1.4E-5)	1.1E-7	(2.0E-7)	-6.8E-5	(9.5E-5)	-8.4E-5***	(1.6E-5)
Random part								
σ_e^2	538.90***		9.4E-2***		44,056.84***		1,229.58***	
σ_{u0}^2	28.91***		4.8E-3***		242.07***		27.08***	
σ_{u1}^2	0.01***		2.3E-6***		0.52***		0.01***	
σ_{u01}^2	-0.09**		-1.2E-5*		-11.14***		-0.13***	
-2*log likelihood	2,041,461.68		109,327.53		2,991,975.77		2,224,344.11	
AIC	2,041,517.68		109,383.53		2,992,031.77		2,224,400.11	
BIC	2,041,806.52		109,672.09		2,992,320.33		2,224,688.95	

Note. N = 228,649. Standard errors are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$

A third subset of additional research questions was formulated to explore the interaction effects of business type, override direction, special events, and override type on the relationship between the size of adjustment and the improvement in forecast accuracy:

H₃: Size of judgmental adjustments is positively associated with improvements in accuracy.

- 3.1. Does the relationship between the size of adjustment and forecast improvement depend on business type?
- 3.2. Does the relationship between the size of adjustment and forecast improvement depend on override direction?
- 3.3. Does the relationship between the size of adjustment and forecast improvement depend on special events?
- 3.4. Does the relationship between the size of adjustment and forecast improvement depend on override type?

Model 6.1.15 in Table 28 shows that for $sAPE_{imp}$ the impact of adjustment size on accuracy improvement depends on business type, override direction, special events, and override type. For adjustments to group business, the effect of adjustment size is stronger. Similar interaction effects can be observed for positive overrides and for special events and for wash overrides.

As Table 29 indicates, these interaction results do not show consistently across the four accuracy measures, except for the interaction term Size of Adjustment $\times$ Business Type.

As Figure 13 illustrates, for transient business, increasing the size of the adjustment has a profound negative impact while for the group segment, adjusting the size basically has no effect.

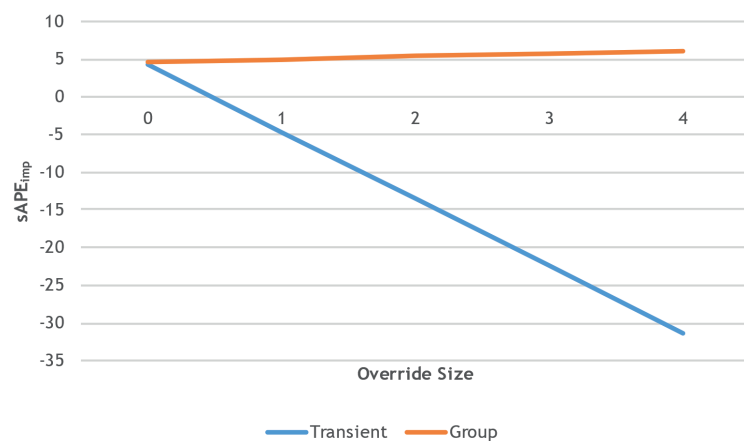


Figure 13: Regression lines for improvement in sAPE by judgmental adjustment size, for business type

Table 28: Models without and with level-1 interaction terms for additional analysis of relationship between size of adjustment and improvements in accuracy

Model:	M _{6.1.7} :		M _{6.1.8} :		M _{6.1.9} :		M _{6.1.10} :		M _{6.1.11} :	
DV = sAPE _{imp}										
Fixed part										
Intercept	2.69***	(0.36)	3.30***	(0.35)	-0.49	(0.36)	-1.33***	(0.37)	4.29***	(0.36)
DTA	-0.02***	(3.7E-3)	-0.03***	(3.6E-3)	-0.02***	(3.9E-3)	-0.02***	(3.9E-3)	-0.02***	(3.6E-3)
Business Type	-0.78***	(0.18)	2.49***	(0.16)	3.62***	(0.16)	3.96***	(0.16)	0.33	(0.18)
Size Adjustment Abs	-3.50***	(0.23)	-5.87***	(0.18)	6.90***	(0.17)	9.73***	(0.29)	-8.92***	(0.37)
Direction	1.39***	(0.10)	-4.28***	(0.11)	1.58***	(0.10)	1.60***	(0.10)	-4.03***	(0.11)
Special	0.35**	(0.14)	0.24	(0.14)	0.62***	(0.16)	0.30*	(0.14)	0.04	(0.15)
Override Number	0.01	(0.01)	0.03*	(0.01)	0.03*	(0.01)	0.03*	(0.01)	0.01	(0.01)
Override Type	0.83***	(0.18)	1.28***	(0.18)	2.42***	(0.18)	3.52***	(0.20)	1.15***	(0.20)
Override Status	-3.19***	(0.13)	-3.23***	(0.13)	-3.20***	(0.13)	-3.18***	(0.13)	-3.20***	(0.13)
Override Recency Status	0.32*	(0.13)	0.66***	(0.12)	0.11	(0.13)	0.12	(0.13)	0.75***	(0.12)
Size Adjustment Abs×Business Type	16.85***	(0.30)							9.28***	(0.31)
Size Adjustment Abs×Direction			32.29***	(0.27)					30.40***	(0.28)
Size Adjustment Abs×Special					-1.73***	(0.32)			1.28***	(0.32)
Size Adjustment Abs×Override Type							-4.39***	(0.33)	-2.74***	(0.33)
Hotel Size	6.5E-4	(7.8E-4)	7.2E-4	(7.3E-4)	4.0E-4	(7.9E-4)	5.1E-4	(7.9E-4)	9.2E-4	(7.3E-4)
Hotel Type										
Airport vs City	-0.71	(1.50)	-0.96	(1.41)	-0.87	(1.53)	-0.78	(1.52)	-0.79	(1.41)
Casino vs City	-2.39	(1.40)	-2.01	(1.32)	-2.10	(1.43)	-2.27	(1.42)	-2.28	(1.32)
Resort vs City	0.49	(0.60)	0.44	(0.56)	0.37	(0.61)	0.38	(0.61)	0.52	(0.56)
Hotel Region										
Asia vs Europe	-0.21	(0.57)	-0.40	(0.53)	0.21	(0.58)	0.12	(0.58)	-0.65	(0.53)
Australasia vs Europe	-0.34	(1.11)	-0.18	(1.04)	-0.02	(1.13)	-0.02	(1.12)	-0.35	(1.04)
Central America vs Europe	8.62*	(3.32)	10.56**	(3.11)	9.27**	(3.37)	9.37**	(3.36)	10.26**	(3.11)
Middle East & Africa vs Europe	-0.98	(0.78)	-0.92	(0.73)	-0.70	(0.79)	-0.75	(0.79)	-1.11	(0.73)
North America vs Europe	-0.72	(0.48)	-0.65	(0.44)	-0.40	(0.48)	-0.44	(0.48)	-0.84	(0.44)
South America vs Europe	3.28	(2.07)	4.45*	(1.95)	4.39*	(2.10)	4.43*	(2.09)	3.85*	(1.95)
DTA×Hotel Size	6.3E-6	(1.3E-5)	8.4E-6	(1.3E-5)	1.0E-5	(1.4E-5)	1.0E-5	(1.4E-5)	6.8E-6	(1.3E-5)
Random part										
σ _e ²	531.60***		507.88***		538.91***		538.59***		505.30***	
σ _{u0} ²	27.82***		24.67***		28.89***		28.57***		24.57***	
σ _{u1} ²	0.01***		0.01***		0.01***		0.01***		0.01***	
σ _{u01} ²	-0.08**		-0.11***		-0.09**		-0.10**		-0.11***	
-2*log likelihood	2,038,324.21		2,028,011.29		2,041,477.98		2,041,331.62		2,026,876.69	
AIC	2,038,376.21		2,028,063.29		2,041,529.98		2,041,383.62		2,026,934.69	
BIC	2,038,644.42		2,028,331.50		2,041,798.19		2,041,651.83		2,027,233.84	

Note. N = 228,649. Standard errors are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$

Table 29: Models without and with level-1 interaction terms for additional analysis of relationship between size of adjustment and improvements in accuracy

Model:	$M_{6.1.15^*}$ DV = sAPE _{imp}	$M_{6.2.15^*}$ DV = LnQ _{imp}	$M_{6.3.15^*}$ DV = APE _{imp}	$M_{6.4.15^*}$ DV = AE _{imp}
Fixed part				
Intercept	4.29*** (0.36)	8.1E-2*** (4.6E-3)	2.51 (2.36)	3.46*** (0.50)
DTA	-0.02*** (3.6E-3)	-2.6E-4*** (4.9E-5)	-0.04 (0.02)	-0.02*** (4.5E-3)
Business Type	0.33 (0.18)	-1.5E-2*** (2.4E-3)	-8.58*** (1.55)	0.56* (0.28)
Size Adjustment Abs	-8.92*** (0.37)	-2.6E-1*** (5.6E-3)	127.38*** (3.82)	11.56*** (0.58)
Direction	-4.03*** (0.11)	-6.6E-2*** (1.5E-3)	23.04*** (1.02)	-2.99*** (0.18)
Special	0.04 (0.15)	2.3E-3 (2.0E-3)	2.87* (1.35)	-0.57* (0.24)
Override Number	0.01 (0.01)	-1.0E-4 (1.7E-4)	0.04 (0.09)	0.06** (0.02)
Override Type	1.15*** (0.20)	3.5E-3 (2.6E-3)	-1.38 (1.66)	2.26*** (0.31)
Override Status	-3.20*** (0.13)	-4.2E-2*** (1.7E-3)	-13.26*** (1.08)	-3.19*** (0.20)
Override Recency Status	0.75*** (0.12)	7.6E-3*** (1.6E-3)	0.09 (1.06)	0.62** (0.19)
Size Adjustment Abs×Business Type	9.28*** (0.31)	2.3E-1*** (4.4E-3)	88.92*** (3.00)	1.03* (0.47)
Size Adjustment Abs×Direction	30.40*** (0.28)	4.7E-1*** (3.9E-3)	-305.72*** (2.66)	0.57 (0.44)
Size Adjustment Abs×Special	1.28*** (0.32)	7.1E-3 (4.6E-3)	-8.58** (3.14)	4.16*** (0.49)
Size Adjustment Abs×Override Type	-2.74*** (0.33)	8.2E-4 (4.8E-3)	-2.27 (3.24)	-7.00*** (0.51)
Hotel Size	9.2E-4 (7.3E-4)	8.7E-6 (9.1E-6)	2.8E-3 (2.6E-3)	0.01*** (8.5E-4)
Hotel Type				
Airport vs City	-0.79 (1.41)	-1.1E-2 (1.8E-2)	-1.10 (5.03)	-1.30 (1.71)
Casino vs City	-2.28 (1.32)	-2.6E-2 (1.7E-2)	-1.00 (4.07)	-6.42*** (1.61)
Resort vs City	0.52 (0.56)	3.1E-3 (7.1E-3)	0.49 (1.77)	0.05 (0.67)
Hotel Region				
Asia vs Europe	-0.65 (0.53)	-7.8E-3 (6.7E-3)	-0.74 (1.70)	-1.87** (0.64)
Australasia vs Europe	-0.35 (1.04)	-6.9E-3 (1.3E-2)	-0.78 (3.55)	-0.48 (1.25)
Central America vs Europe	10.26** (3.11)	1.4E-1*** (3.9E-2)	55.98*** (11.69)	15.55*** (3.72)
Middle East & Africa vs Europe	-1.11 (0.73)	-9.8E-3 (9.2E-3)	-2.08 (2.36)	-0.27 (0.89)
North America vs Europe	-0.84 (0.44)	-9.4E-3 (5.6E-3)	-0.95 (1.20)	-0.65 (0.52)
South America vs Europe	3.85* (1.95)	4.7E-2 (2.5E-2)	-7.31 (8.52)	4.25 (2.41)
DTA×Hotel Size	6.8E-6 (1.3E-5)	7.2E-8 (1.7E-7)	-1.0E-4 (8.6E-5)	-8.4E-5*** (1.6E-5)
Random part				
σ_e^2	505.30***	8.6E-2***	41,763.97***	1,229.10***
σ_{u0}^2	24.57***	3.7E-3***	146.76***	27.41***
σ_{u1}^2	0.01***	1.8E-6***	0.42***	0.01***
σ_{u01}^2	-0.11***	-1.7E-5***	-7.84***	-0.14***
-2*log likelihood	2,026,876.69	89,934.15	2,980,101.04	2,224,258.14
AIC	2,026,934.69	89,992.15	2,980,159.04	2,224,316.14
BIC	2,027,233.84	90,291.02	2,980,457.91	2,224,615.29

Note. N = 228,649. Standard errors are in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$ **Table 30:** Summary results for main hypotheses and additional research questions

Hypotheses / Research questions	sAPE _{imp}	LnQ _{imp}	APE _{imp}	Conclusion	AE _{imp}
H ₁ : Early forecast adjustments are associated with lower improvements in accuracy than the late adjustments.	Yes	Yes	Yes	Supported	Yes
1.1. Does the relationship between DTA and forecast improvement depend on business type ? ($M_{6.1.6^*}$, $M_{6.2.6^*}$, $M_{6.3.6^*}$, $M_{6.4.6^*}$)	Yes, stronger	Yes, stronger	Yes, weaker	Further investigation is needed	No
1.2. Does the relationship between DTA and forecast improvement depend on adjustment size ? ($M_{6.1.6^*}$, $M_{6.2.6^*}$, $M_{6.3.6^*}$, $M_{6.4.6^*}$)	Yes, weaker	Yes, weaker	Yes, weaker	For larger adjustments, the effect of DTA is bigger	Yes, weaker
1.3. Does the relationship between DTA and forecast improvement depend on override direction ? ($M_{6.1.6^*}$, $M_{6.2.6^*}$, $M_{6.3.6^*}$, $M_{6.4.6^*}$)	Yes, stronger	Yes, stronger	Yes, weaker	Further investigation is needed	Yes, weaker
1.4. Does the relationship between DTA and forecast improvement depend on special events ? ($M_{6.1.6^*}$, $M_{6.2.6^*}$, $M_{6.3.6^*}$, $M_{6.4.6^*}$)	Yes, stronger	Yes, stronger	No	Further investigation is needed	Yes, stronger
1.5. Does the relationship between DTA and forecast improvement depend on override type ? ($M_{6.1.6^*}$, $M_{6.2.6^*}$, $M_{6.3.6^*}$, $M_{6.4.6^*}$)	Yes, stronger	Yes, stronger	Yes, weaker	Further investigation is needed	Yes, stronger
H ₂ : Judgmental forecast adjustments to group business are associated with larger improvements in accuracy than adjustments to transient business.	Yes	Yes	Yes	Supported	Yes
2.1. Does the relationship between business type and forecast improvement depend on override direction ? ($M_{6.1.10^*}$, $M_{6.2.10^*}$, $M_{6.3.10^*}$, $M_{6.4.10^*}$)	Yes	Yes	Yes	For negative judgmental adjustments, business type has a stronger effect than for positive overrides.	Yes
2.2. Does the relationship between business type and forecast improvement depend on special events ? ($M_{6.1.10^*}$, $M_{6.2.10^*}$, $M_{6.3.10^*}$, $M_{6.4.10^*}$)	No	No	No		No
2.3. Does the relationship between business type and forecast improvement depend on override type ? ($M_{6.1.10^*}$, $M_{6.2.10^*}$, $M_{6.3.10^*}$, $M_{6.4.10^*}$)	No	No	No		No
H ₃ : Size of judgmental adjustments is positively associated with improvement in accuracy.	Yes	Yes	Yes	Supported	Yes
3.1. Does the relationship between the size of adjustment and forecast improvement depend on business type ? ($M_{6.1.15^*}$, $M_{6.2.15^*}$, $M_{6.3.15^*}$, $M_{6.4.15^*}$)	Yes	Yes	Yes	For adjustments to transient business forecasts, adjustment size has a stronger negative effect	Yes
3.2. Does the relationship between the size of adjustment and forecast improvement depend on override direction ? ($M_{6.1.15^*}$, $M_{6.2.15^*}$, $M_{6.3.15^*}$, $M_{6.4.15^*}$)	Yes, positive	Yes, positive	Yes, negative	Further investigation is needed	No
3.3. Does the relationship between the size of adjustment and forecast improvement depend on special events ? ($M_{6.1.15^*}$, $M_{6.2.15^*}$, $M_{6.3.15^*}$, $M_{6.4.15^*}$)	Yes, positive	No	Yes, negative	Further investigation is needed	Yes, positive
3.4. Does the relationship between the size of adjustment and forecast improvement depend on override type ? ($M_{6.1.15^*}$, $M_{6.2.15^*}$, $M_{6.3.15^*}$, $M_{6.4.15^*}$)	Yes, negative	No	No	Further investigation is needed	Yes, negative
H ₄ : Negative judgmental adjustments are associated with higher improvement in accuracy than positive. (See above 1.3: $M_{6.1.6^*} - M_{6.4.6^*}$; 2.1: $M_{6.1.10^*} - M_{6.4.10^*}$; 3.2: $M_{6.1.15^*} - M_{6.4.15^*}$)	No	No	Yes	Further investigation is needed	Yes

Table 30 summarizes the results for each additional research question, and merges it with the results of the main hypotheses testing given in Table 23. As can be observed, there are three hypotheses and five interaction terms for which all forecast accuracy measures provide consistent findings.

4.5 Discussion and Future Research

The findings indicate that judgmental forecast adjustments impact forecast accuracy, and that characteristics of these adjustments affect accuracy differently. Support was found for hypotheses 1, 2, and 3, but not 4. When interactions were explored, additional insights were obtained. The analysis shows that late forecast adjustments are associated with a higher improvement in accuracy than early adjustments (H_1). This finding differs from Mukhopadhyay et al. (2007) but is in agreement with Kimes (1999) who showed that updates closer to the date of arrival were associated with a reduced forecast error. An interaction effect was also found: the larger the adjustment, the stronger the negative relationship between DTA and accuracy improvement. Further research may shed light on a possible explanation for these results. A related finding that also requires further investigation is related to the inclusion of a random slope in Model 5, which suggests that hotels differ in how forecast improvement changes over the forecasting window.

Another important finding is that adjustments to group business are associated with larger improvements in accuracy than those to the transient segment (H_2). This effect is stronger for negative than for positive adjustments. While these results have not previously been described in the literature, they provide support for the findings of Chapter 3, which found that increasing the number of judgmental adjustments is beneficial for group segment, but damaging for transient segment forecasts. Additional research should explore the role of intra-organizational collaboration (Van der Rest, Roper & Wang, 2018), as Revenue Managers may receive updated information from Sales, based on the most recent interactions with the group clients.

The size of the judgmental adjustment was found to be positively related to accuracy improvement (H_3). This finding is consistent with research into the SKU-forecasting of supply-chain companies (Diamantopoulos & Mathews, 1989; Fildes et al., 2009; Trapero et al., 2011). Moreover, a moderating effect of business type was found: increasing the size of a transient business forecast adjustment has a profound negative impact, whereas adjusting the size of a group segment forecast adjustment has a small positive effect.

However, no consistent support was found for the expectation that there is a relationship between direction of the forecast adjustment and forecasting accuracy (H_4). While Syntetos et al. (2009, p. 78) found that “negative adjustments perform better than the positive ones”, this finding was challenged by Davydenko and Fildes (2013) who suggested that the relationship between the direction of the override

and forecasting accuracy may depend on the error measure used when comparing accuracy levels (as also demonstrated in this chapter). Further studies should focus the analysis on the accuracy measure, as sAPE and LnQ place a heavier penalty when forecasts are low compared to when forecasts are high, and therefore favor over-forecasting, while APE places a higher penalty on positive errors than on negative errors, thereby favoring under-forecasting.²⁸

The findings were controlled for by hotel size, hotel type, and hotel region. While no significant relationships were found for size and type, a significant difference was found between hotels in Central America and Europe, where the former showed higher improvements in forecast accuracy.

The findings have important practical implications for improving the human/machine link in the hotel revenue management cycle, in particular the manner in which subjective adjustments to algorithmic output are constructed and managed profitably. This study serves as a first step in the investigation of how to contextualize users’ and system providers’ decisions to subjectively adjust algorithmic forecasting output. It could be used to enhance the training of existing and new system users, and to inform them on how to better make override decisions. Furthermore, by drawing attention to the influence and importance of user overrides, the findings could assist vendors in improving their revenue management systems. For example, vendors could add machine procedures that monitor various aspects of users’ overriding activities (e.g., when overrides are done, their size, direction) and measure the overrides’ effectiveness. This enhancement could be used to generate alarms, as well as periodic reports for users, system administrators, and solution providers. As the findings indicate that active overrides are associated with lower improvements in accuracy (compared to the reverted overrides), and that demand overrides are associated with higher improvements in accuracy (compared to wash overrides), it is important that these monitors distinguish between override status and override type.

²⁸ See Table 7 for a detailed illustration.

Chapter 5

Conclusion

I cannot teach anybody anything. I can only make them think.

- Socrates -

5.1 General Contribution

The aim of this thesis was to improve our understanding of judgmental adjustments in hotel revenue management forecasting by exploring the impact of user overrides on the accuracy of system-generated hotel occupancy forecasts at multiple forecasting horizons. To achieve this, three empirical studies were conducted. Collectively, the studies demonstrate that the assessment of hotel forecasting accuracy is a non-trivial and complex task because of lack of one ideal measure. Chapter 2 shows that different accuracy measures can lead to conflicting results when analyzed on a single-hotel dataset. Reexamining these inconsistencies in two subsequent chapters using two large datasets with forecasts from multiple hotels confirms this finding. These studies also reveal that the forecasting horizon and the number of overrides have a significant impact on forecast accuracy. In chapter 3 the number of overrides increased the forecast error, while in chapter 4 it decreased the error. This contradiction could be explained by differences in methodological approach. Chapter 3 analyzed the forecasts with and without overrides, while chapter 4 examined only forecasts with overrides. Specifically, chapter 4 identifies the characteristics of the judgmental adjustments (e.g., timing, size, direction) and their impact on the change in forecasting accuracy. Moreover, it shows that the effect of these overrides depends on the business type for which a forecast is made. That is, overrides of group forecasts improve forecast accuracy more than judgmental adjustments of transient forecasts.

Overall, the thesis contributes to the current body of knowledge in a number of ways. First, the notion that many of the existing forecast accuracy measures are not universally applicable, that they can be incomputable, and that they can

produce misleading results, is well established in the generic forecasting literature (Ahlburg, 1992; Armstrong & Collopy, 1992; Chatfield, 1992; Collopy & Armstrong, 1992; Davydenko & Fildes, 2013; Fildes, 1992; Goodwin & Lawton, 1999; Hyndman & Koehler, 2006; Koehler, 2001; Makridakis, 1993). Investigating this issue in a hotel revenue management context, however, is new. A preliminary call made by Schwartz (1999) was not followed up. For this reason, key problems can be observed in the hotel forecasting literature. First, some forecasting studies have continued to use forecast accuracy measures while ignoring their widely acknowledged limitations, such as measures that favor biases (e.g., MAPE), or measures that are scale-dependent (e.g., MAE, RMSE), even when hotel samples comprised hotels with different sizes. Wu, Song, and Shen (2017) reviewed 171 articles on tourism and hotel demand modeling and forecasting, and found that the most widely used measurements of forecasting accuracy were the mean absolute percentage error (MAPE), the root mean square error (RMSE), the root mean square percentage error (RMSPE), and the mean absolute error (MAE); 66, 36, 33, and 19 studies adopted these measures, respectively. However, as mentioned earlier, all these measures have important limitations. MAE and RMSE are scale-dependent and therefore should not be used for comparison of hotels of different size. MAPE favors under-forecasting and is sensitive to outliers. Neither MAPE nor RMSPE can be computed when the time series contain zero. A second issue in the hotel forecasting literature is that the majority of studies use forecasting accuracy measures without properly justifying their selection. This thesis confirms these key problems, and provides empirical evidence on the notion that different forecasting accuracy measures can produce different and even contradictory outcomes in hotel revenue management forecasting. This is an important contribution, as hotel revenue management is characterized by a highly complex forecasting context (e.g., multiple booking horizons, continuous reforecasting, etc.). The empirical support for this contribution was established through three separate studies. The findings of the first study, which were published as Koupriouchina et al. (2014), have been adopted by Pereira (2016, p. 20) who confirms that different accuracy measures produce contradictory answers, “not only within a specific hotel’s segment, but also across segments.” Moreover, the three studies contribute to the existing body of knowledge by generating a warning that the use of forecast accuracy measures, without a deep understanding of the characteristics and limitations of each measure, can lead to misjudged hotel forecasting override decisions. Future research should explore how this may fuel sub-optimal pricing, inventory controls, operational planning, and even strategy decisions. A natural progression of this work is to further analyze the consequences of the inconsiderate use of the forecasting accuracy measures: for instance, reduced reliability of the forecasting performance assessment, undetected wrong interventions, failure to learn from critical mistakes, undesired behavior

encouraged by incorrectly designed bonuses, negative impacts on the effectiveness of revenue management systems, and other resulting impediments to the hotels’ efforts to optimize revenue.

Second, in this research, the importance of the time dimension was explored. Prior research either ignored or simplified the role of the time dimension in hotel revenue management forecasting research. Only a few studies analyzed forecasts at multiple forecasting horizons, and in the best case, they included few arbitrarily selected fixed data points, while forecasting is a dynamic process and re-forecasting happens continuously. This is the first time, to our knowledge, that the time variable has been considered as a predictor in the analysis of overrides in a hotel context, providing a more accurate representation of the dynamic reforecasting process. Despite its exploratory nature, chapter 3 identified that the relationship between time and forecast accuracy is complex, with both linear and non-linear components. Moreover, chapter 4 offered insights into the impact of the timing of judgmental adjustments on forecasting accuracy. It was found that early forecast adjustments are associated with lower improvement in accuracy than late forecast adjustments. Furthermore, an interaction effect of the adjustment size on the relationship between time and forecast improvement was observed - time was more important for larger adjustments. To date only two studies (i.e., Claveria, Monte, & Torra, 2015; Pereira, 2016) have evaluated forecast accuracy for multiple forecasting horizons.

While previous work acknowledges the importance of judgmental interventions in forecasting, there are also conflicting views on the impact of these interventions. Some research in this field has demonstrated that although experts can add value, they are also prone to biases, often make unnecessary adjustments, ignore warnings that their adjustments reduce accuracy, and continue making large adjustments even in a presence of high probability of a significant loss (Goodwin & Wright, 1994; Lawrence et al., 2006; J. S. Lim & Oconnor, 1995; Petropoulos et al., 2016). By investigating user interventions in a multiple forecasting horizon context, support for these conflicting views is provided by demonstrating that judgmental adjustments can add value to system-generated forecasts (depending on their characteristics).

Third, this research extends knowledge of the impact of judgmental adjustments on forecast accuracy at a disaggregated level. An interesting finding was that the overrides impact forecast accuracy in a complex manner, depending on the business type. Judgmental forecast adjustments to group segment were associated with larger improvements in accuracy than adjustments to transient segment. Moreover, the size of judgmental adjustments was positively associated with improvements in accuracy and, for the adjustments to the group segment, the size of adjustment had a more profound effect than for the adjustments to the transient segment. These findings are important, as little is known about group forecasting despite its importance for hotel performance. Prior research has mostly reported findings at

the total hotel level (Yuksel, 2007) or only at a group level (Kimes, 1999). However, segmentation is a key characteristic of profitable revenue management, which facilitates tailoring the strategy to match the needs of various customers and to increase bookings, revenue and profit. The empirical findings of this study provide a new understanding of judgmental adjustments at the segment level. The work by Sierag et al. (2017, p. 46) followed our suggestions and concluded that, “As a whole, the study finds support for the work of Koupriouchina et al. (2014) who argue that research in forecasting should take place at a more granular level”. The study challenged previous findings that positive adjustments, “were much less likely to improve accuracy than negative adjustments” (Fildes et al., 2009, p. 3). In fact, it was demonstrated that the selection of the accuracy measure for this evaluation was also critical: different measures lead to opposite and contradictory conclusions.

Finally, an important contribution is the empirical work itself, which demonstrates how to analyze large real-world datasets on hotel occupancy forecasts through multilevel regression. By examining extensive datasets, this study responds to calls for more work in forecasting based on non-simulated, real-world data (Bodea, Ferguson, & Garrow, 2009; Sierag et al., 2017). This study led to the realization that the analysis of high volumes of real forecasting data is not only time-consuming, but requires several pre-conditions: powerful computers for analysis, good and clear communication with the data provider about the functionality of the RMS, and a solid understanding of the structure and nature of the extracted data. Using a multilevel modeling approach, this study addressed the need for more rigorous and fruitful research in hospitality “by controlling for the contextual and potentially confounding variables that exist at multiple levels” (Wong, 2016, p. 7). This study provides important empirical evidence of the current state of real-world judgmental forecast adjustments considering a wide variety of different hotel types, sizes, and geographical regions over a pluriennial forecasting period.

5.2 Limitations of the Research

A review of previous work indicated a number of limitations. Most studies concentrated on the examination of new forecasting methods using simulated data, rather than an evaluation of real business forecasts and their judgmental adjustments. When real business data were used, many studies investigated only one hotel or one hotel chain. Moreover, most of the analyses were performed within a limited timeframe (e.g., several weeks/months), and studies often concentrated on one segment only (e.g., Kimes, 1999). Importantly, the majority of the studies did not justify the selection of the accuracy measures. Even inappropriate measures were used (e.g., MAPE: sensitive to outliers, favoring under-forecasting, MAE: not suitable for comparisons of the series with different scales) without fully disclosing the implications. Also, as demonstrated by Davydenko & Fildes (2013) some studies

applied asymmetrical measures for analyzing the size of judgmental adjustments, thereby undermining previous findings in the literature about the negative relationship between the size of judgmental adjustments and forecasting accuracy. This dissertation overcomes many of these important limitation by exploring the real-business forecast data of 1,752 hotels for two segments over approximately a 3-year period. It analyzed data using four different forecast accuracy measures, justifying its selection and emphasizing the limitations of each measure. Also, it applies Davydenko & Fildes’ (2013) suggestion, which helps to restore symmetry in the adjustment-size measure, and, therefore, makes the findings about the relationship between the size of judgmental adjustments and forecasting accuracy more robust.

However, the findings of this study are still subject to at least three limitations. The first of these relates to the generalizability of the results. Although the volume of extracted data was extensive, the data were only available for a limited time period (24 months for study 1, and 36 months for studies 2 and 3). Extending the time periods in the analysis would have reinforced the change patterns identified, or perhaps uncovered others (i.e., periodical cycles over several years, etc.). Moreover, as previous-year data were not available, the relative forecast accuracy measures could not be computed as they were based on two-year data, requiring one more previous year of data for studies 2 and 3.

Second, even though an extensive and comprehensive dataset was obtained, data were obtained from a single revenue management system (i.e., IDEaS revenue management system – G2), and, therefore, the conclusions may not apply to other revenue management systems. Inclusion of other revenue management systems in the study (e.g., Duetto, Infor EzRMS, Rainmaker) would have increased robustness. This choice was made for practical purposes – it took several weeks of specialized training to understand how the RMS forecasting module worked and several more months to learn how to combine and interpret extracted data. Moreover, different systems have different functionalities and data structures. Considering the limited empirical evidence on hotel revenue management forecasting overrides, it was decided to opt for an in-depth revelatory study of one revenue management system using data from more than a thousand hotels around the world.

Third, and perhaps the most important limitation, lies in the absence of specific methods developed for hotel revenue management forecasting research. Approaches and methods of this study were inspired by product SKU-forecast research. SKU forecasting in manufacturing has some important differences with RM forecasting in hotels. Therefore, great care must be taken when adopting the SKU research methods in a hotel Revenue Management context. For example, a lack of a well-calibrated and well-tested methodology for analyzing the multiple overrides over the booking horizon was identified. In revenue management systems, the status of the

override can change from active to reverted, and a reverted override can be (re) activated later. While override status was included as a variable in the multilevel models, it was not analyzed at a granular level because the time dimension of the changes in the override statuses could not be precisely analyzed i.e., it was uncertain how long specific overrides remained active in the system, and how critical the timing of changes in their statuses were. As the hotels in the datasets varied in the way the majority of the hotel demand was received (e.g., some received it within the last 90 days before arrival, whereas for others it was spread out across 365 days), it was difficult to determine how uncertainty in the time dimension of the override status changes would affect the results at the individual hotel property level.

5.3 Directions for Future Research

This thesis opens up at least six interesting and novel areas for future research. The following research suggestions provide possible directions to extend the results of this study, and thus to increase the body of knowledge on the hotel revenue management forecasting adjustment process.

1. This research incorporated a limited number of hotel-level variables (e.g., geographical region, hotel type, size of hotel). Future research could include additional hotel-related variables. For example: Are there any common characteristics that top performers (in terms of forecasting) share? How long have they used (the same version of) the RMS? Do they operate in highly competitive and volatile markets, or in fairly stable markets? Are these hotels predominantly chain-affiliated or independent? Moreover, according to a review conducted by Lawrence et al. (2006), judgmental forecast accuracy also depends on the characteristics of time series (e.g., trend, seasonality, randomness, and discontinuities). For instance, previous research showed that subjects reacted differently to downward trends than to upward trends. Andreassen and Kraus (1990), found that noise also impacts the forecaster's ability to detect a trend, and this was confirmed in later work. Including these and similar variables would improve our understanding of how organizational characteristics and market contexts impact judgmental adjustments, which would in turn affect hotel revenue management forecasting performance.
2. A richer theoretical framework including variables at different levels should be developed and tested. According to I.-P. C. Chen (2010), the interaction between a revenue manager's judgment and RMS is not the only element that affects the accuracy of forecasting. Other influential factors should be taken into consideration. The author distinguishes between the external environment (e.g., the economy, social trend, or competitors' action) and internal factors

(e.g., RMS, Human judgment, Model selection, Forecasting method, Data collection, Goal of forecasting). The author also states that "professional's judgment may be affected by work experience, available information, and related knowledge within the area, as well as by motive, belief, and forecasting system" (I.-P. C. Chen, 2010, p. 16). It is suggested to advance research on judgmental adjustment in hotel revenue management forecasting, with multilevel modeling including variables at several levels and their interactions. This is similar to Wong (2016) who proposed a research agenda in which multilevel methods could help to investigate the linkage between service providers, employees, and customers.

3. The mechanisms of the decision-making processes and users' behaviors associated with each override are virtually unknown. A greater focus on psychological and behavioral aspects of forecast overrides could lead to a better understanding of how and why certain decisions are made. Further research might explore, for example, research questions such as: How do Revenue Managers review forecasts? How do they make a final decision on whether to override? How do they determine the size of override? Do they periodically monitor the override, and adjust or revert it if necessary? How do they evaluate their forecasting performance? Other more fundamental research questions that might be explored are: Why do some revenue managers override more than others? How do perceptions and trust in the RMS forecast impact forecast adjustment behavior? Are there personality traits that explain judgmental adjustment behavior? Currently, very little research has been done to clarify these issues. Schwartz and Cohen (2004b) conducted a study involving 57 experienced hotel revenue managers who were exposed to simulated forecasting software. Gender of the revenue manager was found to be significant for the level of expressed confidence in the forecast accuracy; the women expressed about 10% less confidence in their forecasts than the men. Another variable influencing the confidence according to the study was experience. "For every additional year of experience, the level of confidence in forecast increases by 0.37%" (Schwartz & Cohen, 2004b, p. 64). Since confidence in the own forecast/expertise may have a direct impact on the propensity to override, it may be worthwhile hypothesizing whether, for example, personality affects propensity to adjust system-generated forecasts through its effect on the level of confidence. Insights from the Human-Computer Interaction (HCI) literature, which specifically deals with the ways humans interact with information and technologies (Rogers, 2012), could help to identify relevant theories, and to formulate these hypotheses related to the interaction between RM professionals and the RMS, the reasons behind their forecast overrides, and how trust, behavior, and attitude affect accuracy and

consequently financial performance. In addition, the theory of *ecological rationality* (Gigerenzer et al., 1999), which suggests that the mind adapts its limitations to match the structures of information available in the environment, may be useful.

4. In the Human-Computer Interaction field a systematic research has been done to select relevant cognitive theories that could be useful for the interface design. The research on this topic in the hotel RMS context is virtually non-existent. According to Schwartz and Cohen (2004a, p. 85), “the nature of the user interface influenced the way the revenue managers adjusted the computers’ forecasts”. Since the user interface impacts the behavior related to the judgmental adjustments, the impact of the revenue management system interface on the user forecast adjustment behavior deserves additional attention.
5. Forecast adjustments are not the only way in which users can influence the system. Apart from forecast overrides, revenue managers can perform price overrides or overbooking levels overrides; they can activate or modify length of stay controls, register special events, and perform other system activities that may influence decisions suggested by the revenue management system. Future work should explore the attributes of these overrides and their impact. A natural progression of this study is to analyze pricing overrides and their impact on, for example, Average Daily Rate (ADR) or Average Daily Rate Index (ARI). In a similar vein, further research can explore whether the overrides of overbooking levels can lead to improvements in occupancy. It would be interesting to investigate whether some of the conclusions related to forecast overrides are, to some degree, applicable to the price and overbooking level overrides.
6. Building on the previous suggestion, it can be proposed, that although forecast adjustments have a direct impact on forecast accuracy, there may also be cascade effects in the complex system of interrelated key revenue management decisions (e.g., capacity allocation, length of stay controls, pricing). The RMS under analysis in this thesis consisted of four interconnected modules: (a) forecasting, (b) optimization, (c) control, and (d) monitoring. The results of the forecasting module are fed into an optimization module which determines the optimal mix of demand for each day. Once the optimal business mix is defined by the optimization module, its results are used in the control module responsible for the proposal of the controls that are necessary to ensure that an optimal mix of demand is achieved. This control module is connected to the monitoring module; the circle is completed through the connection of the monitoring module with the forecast module. The process is dynamically repeated multiple times for the same arrival date. But, apart from the forecast overrides, users can perform

price overrides or override the overbooking levels. Future work should also study all overrides as part of a complex, dynamic, and interconnected system, where one decision simultaneously impacts many others. If the debate on the impact of user influences on revenue management system is to be moved forward, a more holistic understanding and examination of major user influences and their impact on topline (e.g., RevPAR, TrevPAR), bottom line (e.g., GOPPAR) and competitive set performance (e.g., MPI, ARI, RGI) is needed. With regard to the latter, a deeper understanding could be gained of the flow-through effects of user overrides on operational decisions and expenses (e.g., payroll).

5.4 Implications

The findings of this study have a number of implications for practice, and are of interest to hotels as well as to RMS software providers.

This study has illustrated potential deficiencies and dangers of the inconsiderate use of forecasting accuracy measures. Different error measures generate contradictory answers and the forecasting accuracy can be misjudged and, as a consequence, potentially undermine decision-making in other important hotel management areas: pricing, inventory control, operational planning, distribution, and strategy. The findings also demonstrate the need to further educate the Revenue Managers about the pitfalls and biases of each accuracy measure, so that they are in a position to carefully select the forecasting accuracy measures applicable for their conditions. This cautious choice should be based on the theoretical nature of error measures, together with the practical implications of the forecast, the nature of the derived decisions, and the data characteristics.

Given the importance of forecasting in the hotel revenue management optimization cycle, a more fruitful approach may be to expand the common set of narrow accuracy measures approach with a more structured, comprehensive and consistent framework of forecasting quality assessment. This framework should incorporate multiple forecast accuracy measures, as no single measure can be preferred. One of the important elements of this framework could be the collaborative development and implementation of a set of automated or semi-automated forecast quality monitoring procedures, including feedback mechanisms that allow users to learn from their past decisions and actions. To make these feedback mechanisms meaningful, additional efforts will be required from the hotels and their Revenue Managers. For example, it would be extremely helpful if Revenue Managers kept a consistent log in RMS of the reasons for overrides, especially for large and frequent overrides, making it possible to systematically assess these reasons and incorporate the outcomes of this analysis in the feedback loop.

RMS providers, on the other hand, could add automated procedures to continuously monitor judgmental adjustments introduced by the Revenue Managers, and analyze and report on various important aspects such as size, direction, frequency, timing, segments to which they are applied, type of override (demand or wash), and so forth. This analysis could include measurements of override effectiveness and the impact on forecasting performance, categorized by type of override and by reason. This enhancement to the RMS could be used to generate alarms, as well as periodic reports to users, system administrators and solution providers. Moreover, by leveraging the immense number of exchanges with the thousands of hotels operating in different conditions, RMS software providers could further shape their software with a deeper understanding of the users' behavior. These insights could influence how new versions of the software are designed and launched, how the users are trained (e.g., informing them on how to better make override decisions, which pitfalls to avoid in interpretation of the forecasting performance), and how customers' requests are handled (e.g., in case of complaints or requests to fine-tune the forecasting module in order to improve forecasting accuracy).

Another practical implication is that both forecasting horizons and aggregation levels should be taken into account when designing a comprehensive forecasting quality assessment framework. Current industry practice is to evaluate forecasting performance at the total hotel level only 30 days before arrival. There are two major concerns with this approach. First, as highlighted in this study, the relationship between time and forecasting accuracy is complex. It is therefore suggested to monitor forecasting accuracy at multiple time horizons, taking into account hotel-specific time frame(s) at which forecasts influence important managerial decisions. Second, decisions in revenue management are segment-based, and reporting accuracy at the hotel level only could potentially distort the real picture (e.g., 'compensating' under-forecasting in one segment over-forecasting in another segment). Therefore, a reasonable approach would be to use more granular, segment-level data for segment-level forecast accuracy monitoring.

In sum, a well-designed, systematic and appropriately implemented framework could help Revenue Managers make better judgments about forecasting performance, avoiding misleading interpretations, encouraging helpful override behavior, and minimizing override behaviors that are damaging for the forecasting performance. Regular forecast monitoring could also help to compensate for the 'black box' effect related to an incomplete understanding by the Revenue Managers of how the algorithms behind RMS really function. Due to the high numbers of forecasts, and the dynamic and complex nature of forecasting and re-forecasting processes, this framework, however, cannot be developed in isolation. Close cooperation and continuous communication between hotels and RMS software providers is a key condition for the continuing success of hotel revenue managers.

REFERENCES

- Aghazadeh, S. (2007). Revenue forecasting models for hotel management. *Journal of Business Forecasting*, 26(3), 33-37.
- Ahlburg, D. A. (1992). Error measures and the choice of a forecast method. *International Journal of Forecasting*, 8(1), 99-100.
- Anderson, C. K., & Xie, X. (2010). Improving hospitality industry sales. *Cornell Hospitality Quarterly*, 51(1), 53-67.
- Andreassen, P. B., & Kraus, S. J. (1990). Judgmental extrapolation and the salience of change. *Journal of Forecasting*, 9(4), 347-372.
- Armstrong, J. S. (1985). *Long range forecasting: from crystal ball to computer* (2nd ed.). New York: John Wiley and Sons.
- Armstrong, J. S. (1986). The ombudsman: Research on forecasting: A quarter-century review, 1960-1984. *Interfaces*, 16(1), 89-103.
- Armstrong, J. S. (Ed.) (2001). *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers.
- Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting*, 8(1), 69-80.
- Bach, T. (1999). *Maximizing the value of analyst intervention*. Paper presented at the AGIFORS Reservations and Yield Management Study Group Proceedings, London.
- Baker, T., Murthy, N. N., & Jayaraman, V. (2002). Service package switching in hotel revenue management systems. *Decision Sciences*, 33(1), 109-132.
- Bearden, J. N., Murphy, R. O., & Rapoport, A. (2008). Decision biases in revenue management: Some behavioral evidence. *Manufacturing & Service Operations Management*, 10(4), 625-636.
- Belobaba, P. D. (2002). Future of revenue management. *Journal of Revenue & Pricing Management*, 1(1), 87.
- Bendoly, E. (2013). Real-time feedback and booking behavior in the hospitality industry: Moderating the balance between imperfect judgment and imperfect prescription. *Journal of Operations Management*, 31(1), 62-71.
- Bermúdez, J. D., Corberán-Vallet, A., & Vercher, E. (2009). Multivariate exponential smoothing: A Bayesian forecast approach based on simulation. *Mathematics and Computers in Simulation*, 79(5), 1761-1769.
- Bobb, L. M., & Veral, E. (2008). Open issues and future directions in revenue management. *Journal of Revenue & Pricing Management*, 7(3), 291-301.
- Bodea, T., Ferguson, M., & Garrow, L. (2009). Data set choice-based revenue management: Data from a major hotel chain. *Manufacturing & Service Operations Management*, 11(2), 356-361.
- Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, 101(2), 127-151.
- Boyd, A. (1998). Airline alliance revenue management: global alliances within the airline industry add complexity to the yield management problem. *OR/MS today*, 25, 28.
- Brännäs, K., Hellström, J., & Nordström, J. (2002). A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting*, 18(1), 19-30.
- Bunn, D. W., & Taylor, J. W. (2001). Setting accuracy targets for short-term judgemental sales forecasting. *International Journal of Forecasting*, 17(2), 159-169.
- Carbone, R., & Armstrong, J. S. (1982). Note. Evaluation of extrapolative forecasting methods: Results of a survey of academicians and practitioners. *Journal of Forecasting*, 1(2), 215-217.
- Cetin, G., Demirciftci, T., & Bilgihan, A. (2016). Meeting revenue management challenges: Knowledge, skills and abilities. *International Journal of Hospitality Management*, 57, 132-142.
- Chatfield, C. (1992). A commentary on error measures. *International Journal of Forecasting*, 8(1), 100-102.
- Chen, C.-C., & Schwartz, Z. (2008). Timing matters: Travelers' advanced-booking expectations and decisions. *Journal of Travel Research*, 47(1), 35-42.
- Chen, C., & Kachani, S. (2007). Forecasting and optimisation for hotel revenue management. *Journal of Revenue & Pricing Management*, 6(3), 163-174.
- Chen, I.-P. C. (2010). *Hotel revenue management: Investigating the interaction of information technology and judgmental forecasting*. (Master of Hospitality Administration), University of Nevada, Las Vegas.
- Chiang, W.-C., Chen, J. C. H., & Xu, X. (2007). An overview of research on revenue management: current issues and future research. *International Journal of Revenue Management*, 1(1), 97-128.
- Choi, S. (2006). Group revenue management: A model for evaluating group profitability. *Cornell Hotel & Restaurant Administration Quarterly*, 47(3), 206-206.

- Choy, D. J. L. (1984). Forecasting tourism revisited. *Tourism Management*, 5(3), 171-176.
- Choy, D. J. L. (1985). Forecasting hotel-industry performance. *Tourism Management*, 6(1), 4-7.
- Claver-Cortés, E., Molina-Azorín José, F., & Pereira-Moliner, J. (2007). The impact of strategic behaviours on hotel performance. *International Journal of Contemporary Hospitality Management*, 19(1), 6-20.
- Claveria, O., Monte, E., & Torra, S. (2015). A new forecasting approach for the hospitality industry. *International Journal of Contemporary Hospitality Management*, 27(7), 1520-1538.
- Collopy, F., & Armstrong, J. S. (1992). Generalization and communication issues in the use of error measures: A reply. *International Journal of Forecasting*, 8(1), 107-109.
- Cranage, D. (2003). Practical time series forecasting for the hospitality manager. *International Journal of Contemporary Hospitality Management*, 15(2), 86-93.
- Cross, R. (1997a). Launching the revenue rocket. *Cornell Hotel and Restaurant Administration Quarterly*, 38(2), 32-43.
- Cross, R. (1997b). *Revenue management: hard-core tactics for market domination*. New York: Broadway Books.
- Cross, R., Higbie, J. A., & Cross, D. Q. (2009). Revenue management's renaissance. *Cornell Hospitality Quarterly*, 50(1), 56-81.
- Cross, R., Higbie, J. A., & Cross, Z. (2011). Milestones in the application of analytical pricing and revenue management. *Journal of Revenue & Pricing Management*, 10(1), 8-18.
- Cullen, K. (2015). *The evolving dynamics of revenue management* (2nd ed.). USA: HSMIA Foundation.
- Davydenko, A., & Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to SKU-level demand forecasts. *International Journal of Forecasting*, 29(3), 510-522.
- Dawson, J. F. (2014). Moderation in Management Research: What, Why, When, and How. *Journal of Business and Psychology*, 29(1), 1-19. doi:10.1007/s10869-013-9308-7
- Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., Lee, R. S. (2009). Multilevel modeling: A review of methodological issues and applications. *Review of Educational Research*, 79(1), 69-102.
- DeKay, F., Yates, B., & Toh, R. S. (2004). Non-performance penalties in the hotel industry. *International Journal of Hospitality Management*, 23(3), 273-286.
- Denizci Guillet, B., & Mohammed, I. (2015). Revenue management research in hospitality and tourism: A critical review of current literature and suggestions for future research. *International Journal of Contemporary Hospitality Management*, 27(4), 526-560.
- Diamantopoulos, A., & Mathews, B. (1989). Factors affecting the nature and effectiveness of subjective revision in sales forecasting: An empirical study. *Managerial and Decision Economics*, 10(1), 51-59.
- Diebold, F. X. (2001). *Elements of forecasting* (2nd ed.). Cincinnati, OH: South-Western.
- Donihue, M. R. (1993). Evaluating the role judgment plays in forecast accuracy. *Journal of Forecasting*, 12, 81-92.
- El Gayar, N. F., Saleh, M., Atiya, A., El-Shishiny, H., Fayed Zakhary, A. A. Y., & Mohammed Habib, H. A. A. (2011). An integrated framework for advanced hotel revenue management. *International Journal of Contemporary Hospitality Management*, 23(1), 84-98.
- Erdem, M., & Jiang, L. (2016). An overview of hotel revenue management research and emerging key patterns in the third millennium. *Journal of Hospitality and Tourism Technology*, 7(3), 300-312.
- Farley, T. (2003). Groups need revenue management too. *Journal of Revenue & Pricing Management*, 2(2), 153.
- Ferguson, M., & Smith, S. (2014). The changing landscape of hotel revenue management and the role of the hotel revenue manager. *Journal of Revenue & Pricing Management*, 13(3), 224-232.
- Fildes, R. (1992). On error measures: A response to the commentators — the best error measure? *International Journal of Forecasting*, 8(1), 109-111.
- Fildes, R., & Goodwin, P. (2007). Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces*, 37(6), 570-576.
- 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.
- Fildes, R., Nikolopoulos, K., Crone, S. F., & Syntetos, A. A. (2008). Forecasting and operational research: a review. *Journal of the Operational Research Society*, 59(9), 1150-1172.
- Franses, P. H., & Legerstee, R. (2011). Combining SKU-level sales forecasts from models and experts. *Expert Systems with Applications*, 38(3), 2365-2370.
- Fritz, R. G., Brandon, C., & Xander, J. (1984). Combining time-series and econometric forecast of tourism activity. *Annals of Tourism Research*, 11(2), 219-229.
- Geraghty, M., & Johnson, E. (1997). Revenue management saves national car rental. *Interfaces*, 27(1), 107-127.
- Ghalia, M. B., & Wang, P. P. (2000). Intelligent system to support judgmental business forecasting: The case of estimating hotel room demand. *Ieee Transactions on Fuzzy Systems*, 8(4), 380-397.
- Gigerenzer, G., Todd, P. M., & ABC Research Group. (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Gilliland, M. (2010). *The business forecasting deal*. Hoboken, NJ: Wiley.
- Goldstein, D. G., & Gigerenzer, G. (2009). Fast and frugal forecasting. *International Journal of Forecasting*, 25(4), 760-772.
- Gönül, M. S. (2013). A popular and important tool for forecast users: judgmental adjustments. *Tahmin Kullanıcılarının Hizmetindeki Popüler ve Önemli Bir Araç: 'Yargısal Düzeltmeler'*. 24(2), 73-93.
- 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., & Lawton, R. (1999). On the asymmetry of the symmetric MAPE. *International Journal of Forecasting*, 15(4), 405-408.
- Goodwin, P., & Wright, G. (1994). Heuristics, biases and improvement strategies in judgmental time-series forecasting. *Omega-International Journal of Management Science*, 22(6), 553-568.
- Granger, C. W. J., & Pesaran, M. H. (2000). Economic and statistical measures of forecast accuracy. *Journal of Forecasting*, 19(7), 537-560.
- Guadix, J., Cortés, P., Onieva, L., & Muñuzuri, J. (2010). Technology revenue management system for customer groups in hotels. *Journal of Business Research*, 63(5), 519-527.
- Haensel, A., & Koole, G. (2011). Booking horizon forecasting with dynamic updating: A case study of hotel reservation data. *International Journal of Forecasting*, 27(3), 942-960.
- Hanks, R. D., Cross, R. G., & Noland, R. P. (2002). Discounting in the hotel industry: a new approach. *The Cornell Hotel and Restaurant Administration Quarterly*, 43(4), 94-103.
- Heck, R. H., Thomas, S. L., & Tabata, L. N. (2014). *Multilevel and longitudinal modelling with IBM SPSS*. New York, NY: Routledge.
- Hormby, S., Morrison, J., Dave, P., Meyers, M., & Tenca, T. (2010). Marriott International increases revenue by implementing a group pricing optimizer. *Interfaces*, 40(1), 47-57.
- Hox, J. J. (2010). *Multilevel analysis: techniques and applications* (2nd ed. ed.). New York, NY: Routledge Academic.
- Hox, J. J., Moerbeek, M., & Schoot, R. v. d. (2018). *Multilevel analysis: techniques and applications* (3rd ed. ed.). New York, NY: Routledge.
- Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, 22(4), 679-688.
- IDEAS. (2019). Glossary. Retrieved from <https://ideas.com/tools-resources/glossary/>
- Ivanov, S., & Zhechev, V. (2012). Hotel revenue management a critical literature review. *Tourism*, 60(2), 175-197.
- Jones, P., & Hamilton, D. (1992). Yield management: Putting people in the big picture. *Cornell Hotel and Restaurant Administration Quarterly*, 33(1), 89-95.
- Joseph, S., Stierand, M., & Van Mourik, A. (2015). Hotel revenue management: Then, now and tomorrow. *Journal of Revenue and Pricing Management*.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Kimes, S. E. (1989). The basics of yield management. *Cornell Hotel and Restaurant Administration Quarterly*, 30(3), 14-19.
- Kimes, S. E. (1999). Group forecasting accuracy in hotels. *Journal of the Operational Research Society*, 50(11), 1104-1110.
- Kimes, S. E. (2003). Revenue management: A retrospective. *Cornell Hotel and Restaurant Administration Quarterly*, 44(5-6), 131-138.
- Kimes, S. E. (2008). Hotel revenue management: Today and tomorrow [Electronic article]. *Cornell Hospitality Report*, 8(14), 6-15.
- Kimes, S. E. (2011). The future of hotel revenue management. *Journal of Revenue & Pricing Management*, 10(1), 62-72.
- Kimes, S. E. (2016). The evolution of hotel revenue management. *Journal of Revenue and Pricing Management*, 15(3), 247-251.
- Kimes, S. E. (2017). *The future of hotel revenue management*. Retrieved from Cornell Hospitality Report: <http://scholarship.sha.cornell.edu/chrpubs/235/#.WLarnZqItWA.email>
- Kimes, S. E., & Wagner, P. E. (2001). Preserving your revenue-management system as a trade secret. *Cornell Hotel and Restaurant Administration Quarterly*, 42(5), 8-15.
- Koehler, A. B. (2001). The asymmetry of the SAPE measure and other comments on the M3-Competition. *International Journal of Forecasting*, 17(4), 570-574.
- Koupiouchina, L., Van der Rest, J. I., & Schwartz, Z. (2014). On revenue management and the use of occupancy forecasting error measures. *International Journal of Hospitality Management*, 41, 104-114.
- Lai, K.-K., & Ng, W.-L. (2005). A stochastic approach to hotel revenue optimization. *Computers and Operations Research*, 32(5), 1059-1072.

- Law, R. (1998). Room occupancy rate forecasting: a neural network approach. *International Journal of Contemporary Hospitality Management*, 10(6), 234-239.
- Law, R. (2004). Initially testing an improved extrapolative hotel room occupancy rate forecasting technique. *Journal of Travel & Tourism Marketing*, 16(2-3), 71-77.
- Lawrence, M., Edmundson, R. H., & O'Connor, M. J. (1985). An examination of the accuracy of judgmental extrapolation of time series. *International Journal of Forecasting*, 1(1), 25-35.
- Lawrence, M., Goodwin, P., O'Connor, M., & Onkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, 22(3), 493-518.
- Lawrence, M., & O'Connor, M. (2005). Judgmental forecasting in the presence of loss functions. *International Journal of Forecasting*, 21(1), 3-14.
- Lee, M. (2012). *Improving the forecasting accuracy of hotel arrivals: A new non-homogeneous poisson approach*. Paper presented at the Annual Meeting of the 43rd Decision Sciences Institute, San Francisco, CA.
- Lee, M. (2018). Modeling and forecasting hotel room demand based on advance booking information. *Tourism Management*, 66, 62-71.
- Li, G., Song, H., & Witt, S. F. (2006). Time varying parameter and fixed parameter linear AIDS: An application to tourism demand forecasting. *International Journal of Forecasting*, 22(1), 57-71.
- Li, G., Wong, K. K. F., Song, H., & Witt, S. F. (2006). Tourism demand forecasting: A time varying parameter error correction model. *Journal of Travel Research*, 45(2), 175-185.
- Lim, C., & Chan, F. (2011). An econometric analysis of hotel-motel room nights in New Zealand with stochastic seasonality. *International Journal of Revenue Management*, 5(1), 63-83.
- Lim, C., Chang, C., & McAleer, M. (2009). Forecasting h(m)otel guest nights in New Zealand. *International Journal of Hospitality Management*, 28(2), 228-235.
- Lim, J. S., & O'Connor, M. (1995). Judgmental adjustment of final forecasts - its effectiveness and biases. *Journal of Behavioral Decision Making*, 8(3), 149-168.
- Loew, J. (2000). *Forecast versus actual analysis*. Paper presented at the AGIFORS Reservations and Yield Management Study Group Proceedings, New York City.
- Makridakis, S. (1993). Accuracy measures: theoretical and practical concerns. *International Journal of Forecasting*, 9(4), 527-529.
- Makridakis, S., & Hibon, M. (2000). The M3-Competition: Results, conclusions and implications. *International Journal of Forecasting*, 16, 451-476.
- Makridakis, S., Wheelwright, S., & Hyndman, R. (1998). *Forecasting methods and applications*. USA: John Wiley & Sons, Inc.
- Marriott, J. J., & Cross, R. (2000). Room at the revenue inn. In P. Krass (Ed.), *Book of Management Wisdom: Classic Writings by Legendary Managers* (pp. 199 - 208). New York: Wiley.
- Martin, C. A., & Witt, S. F. (1988). Forecasting performance. *Tourism Management*, 9(4), 326-329.
- McCarthy, T. M., Davis, D. F., Golobic, S. L., & Mentzer, J. T. (2006). The evolution of sales forecasting management: A 20-year longitudinal study of forecasting practices. *Journal of Forecasting*, 25(5), 303-324.
- McKenzie, J. (2011). Mean absolute percentage error and bias in economic forecasting. *Economics Letters*, 113(3), 259-262.
- McNees, S. K. (1990). The role of judgment in macroeconomic forecasting accuracy. *International Journal of Forecasting*, 6(3), 287-299.
- Mei, H., & Zhan, Z. (2013). An analysis of customer room choice model and revenue management practices in the hotel industry. *International Journal of Hospitality Management*, 33, 178-183.
- Mukhopadhyay, S., Samaddar, S., & Colville, G. (2007). Improving revenue management decision making for airlines by evaluating analyst-adjusted passenger demand forecasts. *Decision Sciences*, 38(2), 309-327.
- Murphy, A. H. (1993). What is a good forecast - an essay on the nature of goodness in weather forecasting. *Weather and forecasting*, 8(2), 281-293.
- Noone, B. M., & Hultberg, T. (2011). Profiting through teamwork: the role of the revenue management and sales functions in group revenue management. *Cornell Hospitality Quarterly*, 52(4), 407-420.
- Pan, B., & Yang, Y. (2016). Forecasting destination weekly hotel occupancy with big data. *Journal of Travel Research*, 56(7), 957-970.
- Pereira, L. N. (2016). An introduction to helpful forecasting methods for hotel revenue management. *International Journal of Hospitality Management*, 58, 13-23.
- Petropoulos, F., Fildes, R., & Goodwin, P. (2016). Do 'big losses' in judgmental adjustments to statistical forecasts affect experts' behaviour? *European Journal of Operational Research*, 249(3), 842-852.
- Pine, R., & Phillips, P. (2005). Performance comparisons of hotels in China. *International Journal of Hospitality Management*, 24(1), 57-73.
- Rajopadhye, M., Ben Ghalia, M., Wang, P. P., Baker, T., & Eister, C. V. (2001). Forecasting uncertain hotel room demand. *Information Sciences*, 132(1-4), 1-11.
- Rogers, Y. (2012). HCI Theory: Classical, Modern, and Contemporary. *Synthesis Lectures on Human-Centered Informatics*, 5(2), 1-129.
- Schwartz, Z. (1998). Convention and conference facilities: A framework of statistical predictions and judgmental adjustments for daily occupancy forecasts. *Journal of Convention & Exhibition Management*, 1(1), 71-88.
- Schwartz, Z. (1999). Monitoring the accuracy of multiple occupancy forecasts. *Hospitality Review*, 17(1), 29-42.
- Schwartz, Z. (2008). Time, price, and advanced booking of hotel rooms. *International Journal of Hospitality & Tourism Administration*, 9(2), 128-146.
- Schwartz, Z., & Cohen, E. (2003). Hotel revenue management with group discount room rates. *Journal of Hospitality and Tourism Research*, 27(1), 24-47.
- Schwartz, Z., & Cohen, E. (2004a). Hotel revenue-management forecasting: Evidence of expert-judgment bias. *Cornell Hotel and Restaurant Administration Quarterly*, 45(1), 85-98.
- Schwartz, Z., & Cohen, E. (2004b). Subjective estimates of occupancy forecast uncertainty by hotel revenue managers. *Journal of Travel & Tourism Marketing*, 16(4), 59-66.
- Schwartz, Z., & Hiemstra, S. (1997). Improving the accuracy of hotel reservations forecasting: Curves similarity approach. *Journal of Travel Research*, 36(1), 3-14.
- Schwartz, Z., Uysal, M., Webb, T., & Altin, M. (2016). Hotel daily occupancy forecasting with competitive sets: a recursive algorithm. *International Journal of Contemporary Hospitality Management*, 28(2), 267-285.
- Sierag, D., Van der Rest, J. I., Koole, Van der Mei, R. D., & Zwart. (2017). A call for exploratory data analysis in revenue management forecasting: a case study of a small and independent hotel in The Netherlands. *International Journal of Revenue Management*, 10(1), 28-51.
- Simon, H. A. (1955). A behavioral model of rational choice. *The Quarterly Journal of Economics*, 69(1), 99-118.
- 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., & Li, G. (2008). Tourism demand modelling and forecasting - A review of recent research. *Tourism Management*, 29(2), 203-220.
- Song, H., Li, G., Witt, S. F., & Fei, B. (2010). Tourism demand modelling and forecasting: How should demand be measured? *Tourism Economics*, 16(1), 63-81.
- Steed, E., & Gu, Z. (2009). Hotel management company forecasting and budgeting practices: A survey-based analysis. *International Journal of Contemporary Hospitality Management*, 21(6), 676-697.
- STR. (2019). Glossary. Retrieved from <https://www.strglobal.com/resources/glossary/en-gb#G>
- Syntetos, A. A., Nikolopoulos, K., Boylan, J. E., Fildes, R., & Goodwin, P. (2009). The effects of integrating management judgement into intermittent demand forecasts. *International Journal of Production Economics*, 118(1), 72-81.
- Talluri, K. T., & Van Ryzin, G. (2004). *The theory and practice of revenue management*. Hamburg: Springer.
- Tang, C. M. F., King, B. E. M., & Kulendran, N. (2015). Estimating future room occupancy fluctuations to optimize hotel revenues. *Journal of Travel & Tourism Marketing*, 32(7), 870-885.
- Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66, 1352-1362.
- Trapero, J. R., Fildes, R., & Davydenko, A. (2011). Nonlinear identification of judgmental forecasts effects at SKU level. *Journal of Forecasting*, 30(5), 490-508.
- Trapero, J. R., Pedregal, D. J., Fildes, R., & Kourentzes, N. (2013). Analysis of judgmental adjustments in the presence of promotions. *International Journal of Forecasting*, 29(2), 234-243.
- Tse, T. S. M., & Poon, Y. T. (2015). Analyzing the use of an advance booking curve in forecasting hotel reservations. *Journal of Travel & Tourism Marketing*, 32(7), 852-869.
- Turner, D. S. (1990). The role of judgment in macroeconomic forecasting. *Journal of Forecasting*, 9, 315-345.
- Turner, L. W., & Witt, S. F. (2001). Forecasting tourism using univariate and multivariate structural time series models. *Tourism Economics*, 7(2), 135-147.
- Van der Rest, J. I. (2006). Room rate pricing: a resource-advantage perspective. In P. J. Harris & M. Mongiello (Eds.), *Accounting and financial management: Developments in the international hospitality industry* (pp. 211-239). Oxford: Elsevier-Butterworth-Heinemann.
- Van der Rest, J. I., & Roper, A. (2013). A resource-advantage perspective on pricing: shifting the focus from ends to means-end in pricing research? *Journal of Strategic Marketing*, 21(6), 484-498.
- Van der Rest, J. I., Roper, A., & Wang, X. L. (2018). Why is a change of company pricing policy so hard to implement? *International Journal of Hospitality Management*, 69, 30-40.
- Vives, A., Jacob, M., & Payeras, M. (2018). Revenue management and price optimization techniques in the hotel sector: A critical literature review. *Tourism Economics*, 24(6), 720-752.

- Weatherford, L. R. (2016). The history of forecasting models in revenue management. *Journal of Revenue and Pricing Management*, 15(3-4), 212-221.
- Weatherford, L. R., & Kimes, S. E. (2003). A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting*, 19(3), 401-415.
- Weatherford, L. R., Kimes, S. E., & Scott, D. A. (2001). Forecasting for hotel Revenue management: Testing aggregation against disaggregation. *Cornell Hotel and Restaurant Administration Quarterly*, 42(4), 53-64.
- Wong, I. A. (2016). Linking firms, employees, and customers: A multilevel research agenda for hospitality studies. *Cornell Hospitality Quarterly*, 57(1), 7-20.
- Wu, D. C. G., Song, H. Y., & Shen, S. J. (2017). New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management*, 29(1), 507-529.
- Yang, Y., Pan, B., & Song, H. (2013). Predicting hotel demand using destination marketing organization's web traffic data. *Journal of Travel Research*, 53(4), 433-447.
- Yaniv, I. (2004). Receiving other people's advice: Influence and benefit. *Organizational Behavior and Human Decision Processes*, 93(1), 1-13.
- Yeoman, I. (2011). Ten years of revenue management. *Journal of Revenue & Pricing Management*, 10(1), 1-7.
- Yeoman, I., & McMahon-Beattie, U. (2017). The turning points of revenue management: a brief history of future evolution. *Journal of Tourism Futures*, 3(1), 66-72.
- Yuksel, S. (2007). An integrated forecasting approach to hotel demand. *Mathematical and Computer Modelling*, 46(7), 1063-1070.
- Zakhary, A., Atiya, A. F., El-Shishiny, H., & Gayar, N. E. (2011). Forecasting hotel arrivals and occupancy using Monte Carlo simulation. *Journal of Revenue & Pricing Management*, 10(4), 344-366.
- Zeni, R. H. (2003). The value of analyst interaction with revenue management systems. *Journal of Revenue & Pricing Management*, 2(1), 37-46.

Summary

JUDGMENTAL ADJUSTMENTS IN REVENUE MANAGEMENT

Exploring the Impact of User Overrides on the Accuracy of System-Generated Occupancy Forecasts at Multiple Forecasting Horizons

The field of hotel revenue management has developed rapidly in the past decades. Automated revenue management systems with complex and protected algorithms have greatly changed the Revenue Management profession, demanding a new set of skills including but not limited to technical know-how, critical thinking and problem-solving skills, and business acumen. Today's Revenue Managers increasingly need to understand the broader organizational perspective in order to stimulate and enable interdepartmental co-operation. The business environment in which hotels operate has also become much more complex and is characterized by the entrance of new players, channels, and disruptive technologies. To create accurate occupancy forecasts under these conditions is difficult, and hotels have become increasingly reliant on advanced computer systems that continually reforecast occupancies on a forecasting horizon. However, the performance of these systems is not flawless, and revenue managers sometimes choose to override their revenue management systems. Whether these user overrides actually add value has not yet been subjected to rigorous academic study. The aim of this thesis, therefore, is to explore the relationship between judgmental adjustments and the accuracy of system-generated hotel occupancy forecasts at multiple forecasting horizons. To this purpose, three studies have been conducted.

The first study investigates the applicability of existing forecast accuracy measures and provides a critical review and discussion of forecasting accuracy measures for the evaluation of hotel forecasting performance. By computing 17 forecast accuracy measures on 2,043 pairs of machine-generated forecasts, revenue manager's predictions, and on the actual hotel occupancy figures, the study demonstrates that different accuracy measures generate conflicting results. Since forecast accuracy evaluations impact the calibration of forecasting accuracy models, this empirical study draws attention to the importance of choosing the appropriate accuracy measure to analyze forecast errors and related issues.

The second study examines the relationship between the forecasting horizon and the forecasting accuracy for non-adjusted and adjusted hotel occupancy forecasts. A longitudinal multilevel analysis of 20,081,973 cases from 1,752 hotels shows that forecast accuracy significantly improves as the day of arrival nears, and that this relationship between forecasting horizon and forecasting accuracy has linear and non-linear components. While the frequency of judgmental adjustments significantly impacts forecasting accuracy, this relationship depends on the business type: frequent adjustments are beneficial to the forecasts of the group business segment, but damaging for the transient business segment.

The third study explores the association between specific characteristics of judgmental adjustments and forecasting accuracy. Multilevel analysis of 228,649 cases of judgmental forecast adjustments from 1,752 hotels shows that adjustments made late on the forecasting horizon are associated with a bigger improvement in accuracy than early adjustments. A random slope specification reveals that there is substantial variation in the relationship between the forecast horizon and forecast improvement across hotels. Moreover, the size of judgmental adjustments is positively related to accuracy improvement, meaning that larger judgmental adjustments are associated with bigger improvements. Contrary to previous research, no conclusive evidence was found for the relationship between the direction of the forecast adjustment (below or above the system-generated forecast) and forecasting accuracy. Also, adjustments to the group business segment are associated with larger accuracy improvements than those to the transient business segment.

Overall, the research extends the current knowledge of the relationship between judgmental adjustments and forecast accuracy. First, all three studies demonstrate that the use of different forecast accuracy measures can lead to conflicting results. This finding is important as the inappropriate use of forecasting accuracy measures can bring about a cascade of suboptimal decisions, for example in terms of pricing, inventory controls, operational planning, and strategic decisions. Second, using a disaggregated and longitudinal reforecasting perspective, the studies improve our understanding of how timing, as well as other adjustment characteristics, impact the accuracy of adjusted forecasts, showing how and when judgmental adjustments add value to system-generated forecasts.

The research has practical implications for evaluating and improving the human-machine link in the hotel revenue management cycle. The findings can be used to enhance the training of existing and new revenue management system users. By drawing attention to the influence and importance of user overrides, the findings can assist software vendors in improving their revenue management systems through the implementation of automated procedures to monitor users' overriding activities and to measure the overrides' effectiveness.

Samenvatting (Dutch Summary)

OORDEELSKUNDIGE AANPASSINGEN IN REVENUE MANAGEMENT

Een Verkennend Onderzoek naar het Effect van Handmatige Aanpassingen op de Nauwkeurigheid van Systeem-gegenereerde Hotelbezettingsgraadprognoses voor Meerdere Prognosehorizonten

Het vakgebied hotel revenue management heeft zich snel ontwikkeld in het afgelopen decennium. Geautomatiseerde revenue-managementsystemen met complexe algoritmen hebben het werk van de revenue manager ingrijpend veranderd. De revenue manager van vandaag dient niet alleen over vaktechnische kennis te beschikken, maar ook over een behoorlijke dosis zakelijk inzicht en het vermogen om kritisch te kunnen denken, complexe problemen op te lossen en de samenwerking tussen de verschillende organisatieonderdelen te bevorderen. De zakelijke omgeving waarin hotels opereren is ook veel complexer geworden, onder meer door de komst van nieuwe spelers, nieuwe distributiekanalen en technologische innovaties. Het is daardoor moeilijker geworden nauwkeurige prognoses te maken van de bezettingsgraden. Hotel revenue managers worden steeds afhankelijker van geavanceerde computersystemen die de bezettingsgraden voortdurend opnieuw bepalen voor steeds veranderende prognosehorizonten. De prognoses van deze computersystemen komen lang niet altijd overeen met de feitelijke bezettingscijfers en omzetmanagers kiezen er daarom soms voor de prognoses van hun revenue-managementsystemen handmatig aan te passen. Of deze aanpassingen daadwerkelijk het gewenste effect ressorteren, is nog niet eerder onderwerp van systematisch wetenschappelijk onderzoek geweest. Dit proefschrift onderzoekt de invloed van handmatige aanpassingen op de nauwkeurigheid van systeem-gegenereerde hotelbezettingsgraadprognoses voor meerdere prognosehorizonten. Daartoe zijn drie deelstudies uitgevoerd.

In een eerste deelstudie zijn de maatstaven voor het beoordelen van de nauwkeurigheid van de bezettingsgraadvoorspellingen aan een kritisch onderzoek onderworpen. Daartoe is voor 17 verschillende maatstaven de nauwkeurigheid van 2.043 geautomatiseerde bezettingsgraadprognoses berekend. Deze computerprognoses werden twee aan twee vergeleken met handmatige voorspellingen van de revenue manager. Het onderzoek laat zien dat de maatstaven tegenstrijdige antwoorden genereren wanneer de nauwkeurigheid van de computerprognoses wordt vergeleken met de handmatige voorspellingen

van de revenue manager. Daarmee vestigt deze empirische studie de aandacht op het belang van het selecteren van de juiste nauwkeurigheidsmaatstaf voor het analyseren van prognosefouten en daaraan gerelateerde problemen.

In de tweede studie is de relatie tussen de prognosehorizon en de nauwkeurigheid van niet-aangepaste en aangepaste bezettingsgraadprognoses onderzocht. Longitudinale multilevel regressieanalyse (met herhaalde metingen) op basis van 20.081.973 prognoses van 1.752 hotels wereldwijd toont aan dat de nauwkeurigheid van de prognose aanzienlijk verbetert naarmate de prognosehorizon korter wordt, en dat deze relatie tussen prognosehorizon en nauwkeurigheid zowel lineaire als niet-lineaire componenten heeft. Hoewel de hoeveelheid handmatige aanpassingen van de computerprognoses een groot effect heeft op de nauwkeurigheid van een bezettingsgraadvoorspelling, hangt dit effect mede af van het segmenttype: frequente aanpassingen zijn gunstig voor de prognoses van groepsboekingen, maar schadelijk voor het zogenaamde transient segment.

In de derde studie zijn de kenmerken van de handmatige aanpassingen onderzocht. Multilevel analyse van 228.649 aanpassingen van de bezettingsprognose van 1.752 hotels toont aan dat handmatige aanpassingen die in een laat stadium ten opzichte van de prognosehorizon plaatsvinden, een groter positief effect hebben op het verbeteren van de nauwkeurigheid van de computervoorspelling, dan aanpassingen in een vroeg stadium. De mate waarin de voorspelling verbetert ten opzichte van de prognosehorizon verschilt echter per hotel. Bovendien is de grootte van de aanpassing positief gerelateerd aan een toename van de nauwkeurigheid, terwijl in tegenstelling tot eerder onderzoek de hypothese dat de op- of neerwaardse aanpassingsrichting van invloed is op de nauwkeurigheidsverbetering niet wordt ondersteund door het onderzoek.

Het onderzoek draagt in twee opzichten bij aan de bestaande kennis over het effect van handmatige aanpassingen op het verbeteren van geautomatiseerde voorspellingen. Ten eerste blijkt uit alledrie de deelstudies dat verschillende nauwkeurigheidsmaatstaven tot tegenstrijdige en inconsistente resultaten leiden. Deze bevinding is van groot belang omdat handmatige aanpassingen een opeenstapeling van suboptimale computerbeslissingen tot gevolg kunnen hebben, waaronder prijsbepaling, voorraad-beheer, operationele planning en zelfs strategische beslissingen. Ten tweede verbetert het proefschrift met behulp van een longitudinaal herhaald voorspellingsperspectief, ons begrip van hoe timing, evenals andere aanpassingskenmerken, de nauwkeurigheid van computerprognoses beïnvloeden, en laat het zien hoe en wanneer aanpassingen de computerprognoses verbeteren. Het effect van timing heeft lineaire en niet-lineaire componenten en is afhankelijk van het segmenttype. Het onderzoek is bovendien van belang voor verbeteren van de mens-machine-interactie in de hotel-revenue-managementcyclus.

De bevindingen kunnen bijvoorbeeld worden gebruikt voor het trainen van huidige en toekomstige gebruikers die hun revenue managementsystemen beter willen leren beheersen, en door softwareleveranciers om de systemen zelf te verbeteren.

CURRICULUM VITAE

Larissa Koupriouchina currently works as a senior lecturer in revenue management, and is a research fellow at Hotelschool The Hague, an international school specialized in hospitality management. While studying for her Bachelor's in Economics at the Novosibirsk State Academy of Management in Novosibirsk (Russia), Larissa won a scholarship for a one-year exchange program to attend the Northern Arizona University (United States) in 1999-2000. After graduation, she moved to Italy and in 2001 she obtained a Master's in Business Controlling and Management of Production from the Politecnico di Torino (Italy). In Milan she started a career in consulting at Deloitte & Touche, where she gained professional working experience at several international companies specializing in technology risk domain and data mining. Larissa worked on creating new approaches and procedures for automating manual audit activities and developing new IT-based methodologies for profit recovery analysis. After almost ten years of working with clients from various industries, she decided to focus on her favorite industry – hospitality. In 2009, Larissa was awarded her second master's degree, this time in Hospitality Management from the Erasmus University Rotterdam (The Netherlands), and on graduation, she was invited to become a lecturer at the Hotelschool The Hague.

Larissa's ambition has always been to grow academically, and to work on projects together with industry partners. In 2012, Larissa started to explore a research topic for her doctoral study at the Leiden University Dual PhD Center. In 2013, under the guidance of her supervisor at Leiden Law School of Leiden University, she obtained a five-year competitive research grant from the Netherlands Organisation for Scientific Research (NWO). As a part-time external PhD candidate, Larissa has given several industry presentations related to her research, and has published a book chapter and several journal articles on her research. She uses her research insights in her revenue management classes and, for two consecutive years (2015/2016 and 2016/2017), she was nominated by students in the category of Most Dedicated Teacher.

Appendices

APPENDIX A: SCALE-DEPENDENT FORECASTING ACCURACY MEASURES

Let us denote Y_t as actual value and F_t as forecast. Forecast error can be defined as $e_t = Y_t - F_t$, where $t = 1, 2, \dots, n$.

Scale-dependent measures are those where the scale of the accuracy measure depends on the scale of data. Scale-dependent measures are applicable for data sets that have the same data scales.

Accuracy measure	Advantages	Disadvantages
Mean Square Error (MSE) $MSE = mean(e_t^2)$	• The most preferred measure of forecast accuracy according to results of a survey of practitioners and academicians (Carbone & Armstrong, 1982). • Avoids some issues related to ME where positive and negative errors may offset each other. • Compared to MAE it is easier to handle mathematically (and is often used in statistical optimization) (Makridakis et al., 1998).	• Has been broadly criticized for use in comparing forecasting methods across series, as it can be disastrous to average MSE from different series as it is scale-dependent and has poor protection to outliers (Armstrong, 2001; Armstrong & Collopy, 1992). • Compared to MAE is not easily interpretable and not easily explained to non-specialists (Makridakis, 1993). • It squares the errors and therefore it emphasizes the highest errors (Makridakis, 1993).
Root Mean Square Error (RMSE) $RMSE = \sqrt{MSE} = \sqrt{mean(e_t^2)}$	• Historically popular due to theoretical relevance in statistical modeling. • Often is preferred to MSE as it is on the same scale as the data (Hyndman & Koehler, 2006). • Rated “good” on Sensitivity and Relationship to decisions (Armstrong & Collopy, 1992).	• “more sensitive to outliers than MAE or MdAE, which has led some authors (e.g., Armstrong, 2001) to recommend against its use in forecast accuracy evaluation.” (Hyndman & Koehler, 2006, p. 682). • Rated “poor” on Reliability and Outlier protection (Armstrong & Collopy, 1992).
Mean Error (ME) $ME = \frac{1}{n} \sum_{t=1}^n e_t$	None were found in the literature review.	• Positive and negative errors tend to offset one another. • Provides information only about systematic under- or over-forecasting (forecasting bias). • Does not give much information about the size of the typical errors (Makridakis et al., 1998).
Mean Absolute Error (MAE) $MAE = mean(e_t) = \frac{1}{n} \sum_{t=1}^n e_t $	• By taking the absolute value of error, it avoids some issues related to ME where positive and negative errors may offset each other. • Easy to explain to non-specialists (Makridakis et al., 1998).	None were found in the literature review.
Median Absolute Error (MdAE) $MdAE = median(e_t)$	None were found in the literature review.	None were found in the literature review.

APPENDIX B: SCALE-INDEPENDENT FORECASTING ACCURACY MEASURES

For the measures based on percentage errors, the percentage error is given by $\frac{(Y_t - F_t)}{Y_t} \times 100 = \frac{100 \times e_t}{Y_t}$. For measures based on relative errors, divide the error by the error obtained using another standard forecasting method. The relative error can be defined as $r_t = e_t/E_t$, where E_t is the forecast error obtained from the benchmark error.

Scale-independent measures are those where the scale of the accuracy measures does not depend on the scale of the data. Scale-independent measures are applicable for different sized data sets.

Accuracy measure	Advantages	Disadvantages
Mean Percentage Error (MPE) $MPE = \frac{1}{n} \sum_{t=1}^n p_t$	None were found in the literature review.	• Positive and negative percentage errors tend to offset one another (Makridakis et al., 1998). • Difficulties when the time series contain zeros, since the percentage error cannot be computed and when the time series values are very close to zero, the computation involving percentage error can be meaningless (Makridakis et al., 1998).
Mean Absolute Percentage Error (MAPE) $MAPE = mean(p_t)$	• Is similar to MAD, but it is dimensionless, which makes it easier for communication purposes and comparing different forecasts. • “Incorporates the best characteristics among the various accuracy criteria. Moreover it is the only measure (in addition to Percent Better) that means something to decision makers” (Makridakis, 1993, p. 528). • Most frequently used in practice. • Easy to communicate, useful in comparing forecasts from different situations (Armstrong, 1985). • Rated “good” on Construct validity and Sensitivity (Armstrong & Collopy, 1992).	• A bias favoring estimates of forecast that are below the actual values which places a “heavier penalty on positive errors than on negative errors” (Hyndman & Koehler, 2006, p. 683). See McKenzie (2011) for an intuitive explanation on how MAPE puts a higher penalty on positive errors than on negative errors. • Difficulties when the time series contain zeros, since the percentage error cannot be computed and when the time series values are very close to zero, the computation involving percentage error can result in very large numbers and can be meaningless (Hyndman & Koehler, 2006; Makridakis et al., 1998). • MAPEs cannot be compared directly with Naïve 1 (random walk) or Naïve 2 (de-seasonalized random walk) models as is the case with RAE and therefore Geometric Means and Relative Medians which summarize them. • Rated “poor” on Outlier protection (Armstrong & Collopy, 1992).
Median Absolute Percentage Error (MdAPE) $MdAPE = median(p_t)$	• A closer relationship to decision making (Bunn & Taylor, 2001). • Rated “good” on Construct Validity and Outlier protection (Armstrong & Collopy, 1992).	• Can lead to infinite values occurring due to division by zero or could result in very large numbers if divided by numbers close to zero (Hyndman & Koehler, 2006). • Places a “heavier penalty on positive errors than on negative errors” (Hyndman & Koehler, 2006, p. 683). • Rated “poor” on Sensitivity (Armstrong & Collopy, 1992).

Accuracy measure	Advantages	Disadvantages
Symmetric Mean Absolute Percentage Error (sMAPE) $sMAPE = mean\left(\frac{200 Y_t - F_t }{Y_t + F_t}\right)$	• Avoids the problem of large errors when the actual values are close to zero and the large difference between the absolute percentage errors when <i>actual</i> is greater than <i>forecast</i> and vice versa. • Fluctuates between - 200% and 200% while the non- symmetric measure does not have limits (Makridakis & Hibon, 2000).	• Hyndman and Koehler (2006) report that sMAPE can take negative values, although it is meant to be an “absolute percentage error”. • sMAPE, proposed by Makridakis (1993) to overcome asymmetry of MAPE places a heavier penalty when forecasts are low compared to when forecasts are high and therefore is not as “symmetric” as the name suggests (Bunn & Taylor, 2001; Goodwin & Lawton, 1999; Hyndman & Koehler, 2006; Koehler, 2001).
Symmetric Median Absolute Percentage Error (SMdAPE) $sMdAPE = median\left(\frac{200 Y_t - F_t }{Y_t + F_t}\right)$	• Not influenced by extreme values and is more robust than the average absolute percentage error (Makridakis & Hibon, 2000).	• Can lead to infinite values occurring due to division by zero or could result in very large numbers if divided by numbers close to zero (Hyndman & Koehler, 2006). • Similarly to sMAPE, it is not as “symmetric” as the name suggests.
Mean Relative Absolute Error (MRAE) $MRAE = mean(r_t)$	• Ease of interpretation relative to Theil’s <i>U</i> (Armstrong & Collopy, 1992).	• An error of benchmark method can be small and relative error can have infinite variance (Hyndman & Koehler, 2006). • “RAE-based measures mean <i>absolutely</i> nothing to decision makers who cannot understand their meaning” (Makridakis, 1993, p. 528).
Median Relative Absolute Error (MdRAE) $MdRAE = median(r_t)$	• It is designed to be easy to interpret and it lends itself easily to summarizing across horizons and across series as it controls for scale and for the difficulty of forecasting. The Median RAE (MdRAE) is recommended for comparing the accuracy of alternative models as it also controls for outliers (for information on the performance of this measure, see Armstrong & Collopy, 1992; Makridakis & Hibon, 2000). • Rated “good” on Construct validity and Outlier protection (Armstrong & Collopy, 1992).	• “RAE-based measures mean <i>absolutely</i> nothing to decision makers who cannot understand their meaning” (Makridakis, 1993, p. 528). • An error of benchmark method can be small and relative error can have infinite variance (Hyndman & Koehler, 2006). • May be appropriate only in “specific cases when the number or series involved is neither very small nor very large, when <i>no</i> winsorizing is required <i>and</i> when the results are reported exclusively for statistical audiences” (Makridakis, 1993, p. 528). • MdRAE does not perform better than MAPE (see exhibit 9 in Armstrong & Collopy, 1992) over the criteria of reliability, validity, outlier protection, sensitivity, and value to decision making.” (Makridakis, 1993). • Rated “poor” on Sensitivity and Relationship to decisions (Armstrong & Collopy, 1992).
Geometric Mean Relative Absolute Error (GMRAE) $GMRAE = gmean(r_t)$ $= \sqrt[n]{\frac{AE_1}{AEB_1} \times \dots \times \frac{AE_n}{AEB_n}}$	• Rated “good” on Construct validity and Sensitivity (Armstrong & Collopy, 1992).	• “RAE-based measures mean <i>absolutely</i> nothing to decision makers who cannot understand their meaning” (Makridakis, 1993, p. 528). • An error of benchmark method can be small and relative error can have infinite variance (Hyndman & Koehler, 2006). • May be appropriate only in “specific cases when the number or series involved is neither very small nor very large, when <i>no</i> winsorizing is required <i>and</i> when the results are reported exclusively for statistical audiences” (Makridakis, 1993, p. 528). • GMRAE does not perform better than MAPE (see exhibit 9 in Armstrong & Collopy, 1992) over the criteria of reliability, validity, outlier protection, sensitivity, and value to decision making.” (Makridakis, 1993, p. 528). • Rated “poor” on Relationship to decisions (Armstrong & Collopy, 1992).

Accuracy measure	Advantages	Disadvantages
Mean Absolute Scaled Error (MASE) $MASE = mean(q_t)$ where $q_t = \frac{e_t}{\frac{1}{n-1} \sum_{i=2}^n Y_i - Y_{i-1} }$	• Has meaningful scale and is easy to interpret. “When $MASE < 1$, the proposed method gives, on average, smaller errors than the one-step errors from the naïve method. If multi-step forecasts are being computed, it is possible to scale by the in- sample MAE computed from multi-step naïve forecasts.” Values of MASE greater than one indicate that the forecasts are worse, on average, than in-sample one-step forecasts from the naïve method. (Hyndman & Koehler, 2006, p. 685). • Not subject to degeneracy problems (Hyndman & Koehler, 2006). • “The only circumstance under which . . . would be infinite or undefined is when <i>all</i> historical observations are equal” (Hyndman & Koehler, 2006, p. 686). • “Less sensitive to the outliers and more easily interpreted than RMSSE, and less variable on small samples than MdASE” (Hyndman & Koehler, 2006, p. 686).	• “In some circumstances an asymmetric loss function may be preferred (see e.g., Lawrence & O’Connor, 2005), in which case some other (asymmetric) function of the scaled errors may be appropriate. Diebold (2001) gives some examples of suitable asymmetric functions” (Hyndman & Koehler, 2006, p. 686).
Root Mean Square Percentage Error (RMSPE) $RMSPE = \sqrt{mean(p_i^2)}$	None were found in the literature review.	• Can be infinite or undefined if actual equals to zero in the period of interest, and having an “extremely skewed distribution” (Hyndman & Koehler, 2006, p. 683).
Root Median Square Percentage Error (RMdSPE) $RMdSPE = \sqrt{median(p_i^2)}$	None were found in the literature review.	• Can be infinite or undefined if actual equals to zero in the period of interest, and having an “extremely skewed distribution” (Hyndman & Koehler, 2006, p. 683).
RelMAE or CumRAE RelMAE = MAE / MAEb, Where MAE _b is MAE from benchmark method. The most commonly used benchmarks methods are the random walk or “naïve” and “naïve 2”	• “As advantage of these methods is their interpretability. For example, relative MAE measures the improvements possible from the proposed forecast method relative to the benchmark forecast method. When RelMAE < 1, the proposed method is better than benchmark method, and when RelMAE > 1, the proposed method is worse than the benchmark method” (Hyndman & Koehler, 2006, p. 684).	• “However, they require several forecasts on the same series to enable a MAE (or MSE) to be computed. One common situation where it is not possible to use such measures is where one is measuring the out-of-sample forecast accuracy at a single forecast horizon across multiple series. It makes no sense to compute the MAE across series (due to their different scales)” (Hyndman & Koehler, 2006, p. 684).

APPENDIX C: LITERATURE OVERVIEW OF FORECASTING IN HOTEL RM

Author	Focus	Publication	Topic
Lee (2018)	Modeling	Tourism Management	Proposes a stochastic approach to the short-term forecasting of hotel booking arrivals
Pan and Yang (2016)	Modeling	Journal of Travel Research	Develops a forecasting model of weekly hotel occupancy for a destination, incorporating several tourism big data sources (i.e., search engine queries, website traffic, and weekly weather information).
Pereira (2016)	Modeling	Journal of Hospitality Management	Introduces new time series forecasting models capable of tackling seasonal complexities.
Schwartz et al. (2016)	Modeling	International Journal of Contemporary Hospitality Management	Develops a new approach to hotel daily occupancy forecasting by integrating adding the hotel competitive-set’s predicted occupancy.
Tang, King, and Kulendran (2015)	Modeling	Journal of Travel and Tourism Marketing	Proposes an estimation mechanism for the assessment of the likelihood of forthcoming occupancy peaks and troughs of hotel demand.
Tse and Poon (2015)	Modeling	Journal of Travel and Tourism Marketing	Demonstrates empirically the use of an advance booking curve for forecasting hotel reservations at property level.
Bendoly (2013)	Decision-making	Journal of Operations Management	Uses a laboratory experiment to investigate how different forms of continuous performance feedback influence agent decisions
Mei and Zhan (2013)	Modeling	International Journal of Hospitality Management	Proposes a model to measure the probability of room choice by analyzing the characteristics and preferences of customers, allowing hotels to predict customer demand for each rate class.
Yang, Pan, and Song (2013)	Modeling	Journal of Travel Research	Proposes a model for predicting hotel demand in which the web traffic volume data are incorporated.
Lee (2012)	Modeling	Inform	Develops a new non-homogenous Poisson process approach to improve short-term demand forecast
El Gayar et al. (2011)	Modeling	International Journal of Contemporary Hospitality Management	Presents an advanced model on hotel arrivals as an interrelated process of stochastic parameters and simulates forward in time the actual process of reservations to obtain the forecast
Haensel and Koole (2011)	Modeling	International Journal of Forecasting	Forecasts the accumulated booking curve, as well as the number of reservations expected for each day in the booking horizon
Zakhary et al. (2011)	Modeling	Journal of Revenue and Pricing Management	Proposes a new Monte Carlo simulation approach for the arrivals and occupancy forecasting problem
Guadix, Cortés, Onieva, and Muñuzuri (2010)	Modeling	Journal of Business Research	Introduces a new decision support system that sets the revenue maximization criteria for a hotel
Bermúdez, Corberán-Vallet, and Vercher (2009)	Modeling	Mathematics and Computers in Simulation	Proposes prediction of time series with correlated errors at each time point using a Bayesian forecast approach based on the multivariate Holt–Winters model
C. Lim, Chang, and McAleer (2009)	Modeling	International Journal of Hospitality Management	Highlights time series models that hotel and motel industry practitioners can use to forecast guest nights

Author	Focus	Publication	Topic
Steed and Gu (2009)	Practices	International Journal of Contemporary Hospitality Management	Finds that forecasting methods used in the industry present opportunities for improving accuracy
Aghazadeh (2007)	Modeling	Journal of Business Forecasting	Discusses forecasting methods used in the hotel industry
C. Chen and Kachani (2007)	Modeling	Journal of Revenue and Pricing Management	Argues that there is scant work on hotel forecasting. Present a framework for forecasting and optimization
Yuksel (2007)	Modeling	Mathematical and Computer Modeling	Forecasts demand fluctuations in the hotel business that lead to crises and creates a systematic and dynamic process that could be re-used
Law (2004)	Modeling	Journal of Travel and Tourism Marketing	Proposes an improved extrapolative time series forecasting technique to compute future hotel occupancy rates
Schwartz and Cohen (2004a)	Decision-making	Cornell Hotel and Restaurant Administration Quarterly	Finds that the nature of the user interface influences the way revenue managers adjust the computers' forecasts
Schwartz and Cohen (2004b)	Decision-making	Journal of Travel and Tourism Marketing	Underscores the subjective nature of forecast uncertainty
Cranage (2003)	Modeling	International Journal of Contemporary Hospitality Management	Tests time series models to forecast sales
Weatherford and Kimes (2003)	Modeling	International Journal of Forecasting	Researches the best forecasting method
Baker, Murthy, and Jayaraman (2002)	Modeling	Decision Sciences	Develops a new forecasting-allocation approach to explicitly account for differences in service packages
Brännäs, Hellström, and Nordström (2002)	Modeling	International Journal of Forecasting	Develops an integer-valued moving average model by cross-sectional and temporal aggregation
Rajopadhye et al. (2001)	Modeling	Information Sciences	Discusses the Holt-Winters forecasting procedure and its application to forecasting unconstrained hotel room demand
Weatherford et al. (2001)	Modeling	Cornell Hotel and Restaurant Administration Quarterly	Tests four different approaches to obtaining a forecast
Kimes (1999)	Accuracy	Journal of the Operational Research Society	Determines the accuracy of group forecasts and to identify factors associated with accurate forecasts
Schwartz (1999)	Accuracy	Hospitality Review	Lists the criteria for selecting an appropriate error measure
Law (1998)	Modeling	International Journal of Contemporary Hospitality Management	Applies neural networks to the forecasting of room occupancy rates
Schwartz and Hiemstra (1997)	Modeling	Journal of Travel Research	Develops and test a new type of an extrapolative forecasting model
Choy (1985)	Modeling	Tourism Management	Provides a method to forecast the annual hotel-industry occupancy rate for specific localities

APPENDIX D: DATA MANAGEMENT PLAN

Leiden University Data Management Plan²⁹

The Research Data Management Regulations Leiden University requires researchers to write a data management plan at the start of a long-term research project³⁰. Please contact the Centre for Digital Scholarship at the University Libraries Leiden if you need help: datamanagement@library.leidenuniv.nl

Name and contact details	Larissa Koupriouchina (l.koupriouchina@hotelschool.nl ; +31 88 028 18 00)
Name of project and group	Prof. Dr. Jean-Pierre van der Rest
Description of your research	This is a study about Human-computer interaction between revenue management professionals and forecasting support systems
Project duration	Start: 01-01-2013 End: 12-11-2019
Names of people and their responsibilities for data management	Larissa Koupriouchina is responsible for data collecting, data storing, documenting, sharing and archiving.
Funding body(ies)	NWO
Grant number	023.002.090
Partner organizations	Not applicable

About this Data Management Plan

Date written	26-02-2018
Date last update	12-11-2019
Version	No. 1 (retrospectively)

Changes in this version of the Data Management Plan

Component	Progress / Execution
Please describe briefly what progress you have made, any questions or issues you have encountered and want to discuss, etc.	
1. Data collection	
2. Data storage and back-up	
3. Data documentation	
4. Data access, sharing and reuse	
5. Data preservation and archiving	

29 This template is based on the 3TU data management plan, the University of Bath data management plan, and the Data Management Checklist of the University of Western Sydney.

30 <http://regulations.leiden.edu/research/research-data-management-regulations-leiden-university.html>

1. Data collection
Describing the data you will be creating/collecting

1.1
Will the project use existing or third party data?

☐ No
☐ Own/group previous research
☐ Academic collaborators
☒ Commercial collaborators
☐ Publicly available database / archive
☐ Specialist commercial data provider
☐ Other (please specify)

Describe briefly provenance, type and format of this data. Are there any restrictions or requirements for use of third party data such as licensing conditions?

Anonymized hotel data were received from a major revenue management system provider. The data are protected by a Non-disclosure agreement.

1.2
What type(s) of data will you collect or create, in what file format(s)?³¹

Note that not all formats are long-lived. For sustainable access you best use the formats recommended by data archives, see for examples: http://researchdata.4tu.nl/fileadmin/editor_upload/File_formats/Digital_Preservation_Support_levels.pdf or <http://www.dans.knaw.nl/en/deposit/information-about-depositing-data/DANSpreferredformatsUK.pdf>

Data are received in .csv format.

1.3
How will you collect and/or create your data?

Please describe briefly. Name any relevant protocols and/or standard in your area of expertise.

Hotel data are extracted by the company and send to the researcher in the .csv format.

1.4
What tools, instruments, equipment, hardware or software will you use to capture, produce, collect or create the data?

Please give the names of the tools and state if they are already available. If not, state how you intend to acquire them. If applicable, describe whether you use a paper or electronic lab journal.

The data are imported in SPSS and stored in .sav format.

1.5
What is the estimated size of the data?

Please describe briefly. Stages to be adopted if relevant.

Data stage	Specification of type of research data	Software choice and file format	Data size now	Data size when project is finished
Raw data	Data directly obtained from the company (dataset one and two)	Software: Excel File format: .csv	19.3 GB	
Processed data	Forecast data processed in SPSS	Software: SPSS File format: .sav	6.55 GB (dataset one) 140 MB (dataset two)	
Results	Detailed results in the form of SPSS output and excel sheets	Software: SPSS, Excel File format: .spv, .xlsx	2.3 GB	
Other...				

Raw data are imported and analyzed in SPSS.

31 Data types can be: documents (text, MS Word), spreadsheets, field notebooks, diaries, questionnaires, transcripts, surveys, codebooks, audiotapes, videotapes, photographs, (transcribed) test responses, models, algorithms, measurements, simulations, observations, software source code, computational model output, etc.

2. Data storage and security
Ensuring that all research data are stored securely and backed up or copied regularly during your research

2.1
Where will you store your data?

Please describe how safe storage is guaranteed. Specify your method if your data is collected and / or transported in different locations / countries.

☒ On university departmental network storage (J:)
☐ On university personal network storage (P-)
☐ In a Virtual Research Environment (Sharepoint)
☒ Physical storage (e.g. USB, external hard drive)
☐ Cloud service (e.g. SURFdrive)
☐ Other, namely: ...

Data are stored on the restricted J: drive and an external hard drive for the backup.

2.2
Will your data be backed up?

Please specify briefly for each storage device frequency, location of backups and who is responsible. Describe how you can restore your data in the event of data loss and who is responsible.

Larissa Koupriouchina is responsible for data storage. University network (J:) is backed up every night. Data stored on the J-drive are copied to the external hard drive at the end of each business day.

2.3
Are there any commercialization, ethical or confidentiality restrictions about handling your data?

Please specify briefly.

☐ Contractual obligations
☐ Requirements by law : protection of personal data (e.g. privacy law) : specify in 4.1
☐ Requirements by law : copyright, intellectual property : specify in 4.1
☐ Ethical restrictions (e.g. ethical review): specify in 4.1
☐ Commercial considerations (e.g. patentability)
☐ Formal security standards
☐ No requirements
☒ Other, namely: Nondisclosure agreement signed with the revenue management system provider.

"Recipient acknowledges and agrees that Recipient's access to and use of the Confidential Information is solely for the PhD project of Mrs. Koupriouchina, which includes publishing a research paper in a scientific journal and a PhD thesis as required by her university. Furthermore, parties agree that Confidential Information will only be used anonymously to respect the privacy of the owner of the Confidential Information."

2.4
How will access to the data be managed during the project?

Please specify for each storage device, from different locations / countries.

Only collaborators on the project are allowed access to the data.

2.5
What are the main risks to data security?

Please list risks, e.g. accidental deletion, falling into the wrong hands. Please describe what would happen if the data get lost or become unusable.

Accidental deletion, loss of the external hard disk, theft. Loss of data could prove highly detrimental for the project.

2.6
What measures do you take to comply with the security requirements and to mitigate the risks?

Describe how you can restore your data in the event of data loss and who is responsible. If applicable, please describe procedures to ensure personal data are handled confidentially and who is responsible.

☒ Access restrictions
☒ Encryptions
☐ Data processing
☒ De-identification / Anonymization
☒ Regular back-ups
☒ Master copy stored on university network storage
☒ Master copy stored elsewhere
☐ Other, namely: ...

Data are stored in locations that are encrypted and accessible only to authorized personnel. All data are anonymized.

2.7	How do you differentiate between raw and processed data? <i>Please explain briefly why you (do not) differentiate.</i>
	☐ I will not differentiate ☐ I will create a new file for processed data ☒ I will create a new file for processed data and I will lock raw data ☐ Other, namely: ...
	Data needs to be processed and restructured for analysis, resulting in multiple files.
2.8	Is there any non-digital data or outputs that the project will generate? Where will these outputs be stored? <i>Please specify briefly and describe who is responsible for storage of these outputs.</i>
	No
2.9	Do you expect to have any supplementary costs for storage not covered by the project budget? <i>Please specify</i>
	No
3. Data documentation Documenting your data to help future users to understand and reuse it	
3.1	How will files be named? <i>Please describe briefly.</i>
	Files will be names as follows: "DatasetName_version".
3.2	How will folders be named and structured? <i>You are invited to draw a folder structure and describe it briefly.</i>
	Raw data file and processed data file for analysis. 1. Data (Folder) 1.1. Raw data (Subfolder) 1.2. Processed data (Subfolder) 1.3. Syntax (Subfolder) 1.4. Output (Subfolder) 1.4.1. SPSS output (Subfolder) 1.4.2. Excel output (Subfolder)
3.3	How do you handle version control to maintain all changes that are made to the data? <i>Please explain your choice briefly. Remember to also document any deletion of data, if applicable.</i>
	☐ No version control (e.g. original files are overwritten) ☐ Version control software, namely: ... ☒ Data/version number in filename/folder ☐ 'Track changes' feature in software ☒ By saving the script with which I process my data ☐ Other, namely: ...
	For important changes I also make a manual note in red in my research notebook (paper).
3.4	What metadata standard will be used, if any? ²² <i>Please explain why you use this standard (most used in my discipline, required by the data archive where I will deposit my data). Please outline how the metadata will be created (read me file, spreadsheet, in the data). If no standard exist, please specify which metadata is needed to understand the data.</i>
	☒ No metadata standard is used ☐ Generic metadata standard (e.g. Dublin Core) ☐ Standard automatic Windows metadata (e.g. from Word, Excel) ☐ Specialized metadata standard, namely: ... ☐ Other metadata standard, namely:

32 See <http://www.dcc.ac.uk/resources/metadata-standards> or http://en.wikipedia.org/wiki/Metadata_standards or the relevant repository.

3.5	What supporting information / documentation will you create to enhance understanding of the data? <i>Please describe briefly how peers should be able to understand the data. Examples are a readme.txt, lab journals, a codebook, survey questions etc. Is there a standard for documentation in your field? Describe at what moment in your research process you will add the documentation necessary to make sure the data is understandable for peers.</i>						
	Codebooks were created for each dataset and labels are used in the dataset whenever applicable to enhance understanding of the data. Important data processing steps are described in a paper research journal.						
4. Data access, sharing and reuse Managing access and security, sharing your data							
4.1	Are there any restrictions placed on sharing / reuse of some / all of your data? <i>Please account for not sharing your data. Reasons may be ethical, commercial, security-related, protection of personal data rules, intellectual property, copyright,</i>						
	Access to data is restricted by the nondisclosure agreement signed with the revenue management system.						
4.2	With whom will you share your data at which stage in your research? You can use the table below. <i>Please state any sharing requirements, e.g. funder data sharing policy. Please describe briefly how you will share your data: on request, pro-actively, etc. Please specify how your data can be accessed.</i>						
		Would not share with anyone	Would share with my immediate collaborators	Would share with others in my research Centre or at my institution	Would share with scientists in my field	Would share with scientists outside of my field	Would share with anyone
	Immediately after the data has been generated	N/A	Yes – for advice and guidance	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement
	After the data has been normalized and/or corrected for errors	N/A	Yes – for advice and guidance	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement
	After the data has been processed for analysis	N/A	Yes – for advice and guidance	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement
	After the data has been analyzed	N/A	Yes – for advice and guidance	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement
	Immediately before publication	N/A	Yes – for advice and guidance	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement
	Immediately after the findings derived from this data have been published	N/A	Yes – for advice and guidance	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement	No – access to data by the nondisclosure agreement
Based on: Interview worksheet, Jake Carlson, Purdue University Libraries / Distributed Data Curation Center							

	
4.3	If intending to share any part of the data, do your participant consent forms include information about intentions for sharing, retention of data and steps taken to protect participant's privacy and confidentiality? ☒ Not applicable. ☐ Yes. <i>Please specify the relevant formula in the consent form.</i>
4.4	Who has authority to grant (additional) access to your data? <i>Please describe briefly.</i> ☐ Only you ☐ A colleague from the project, namely: ... ☐ Supervisor ☐ Funder ☐ Collaborator / research partner organization ☒ Other, namely: ... The revenue management system provider has authority to grant additional access.
4.5	How will you manage copyright and Intellectual Property Rights issues? <i>Who owns the data? How will the data be licensed for reuse? Please describe briefly your choices and their consequences.</i> The principal investigator owns the data and remains copyright.
4.6	What is the audience for reuse? <i>Please list possible audiences and purposes. Consider who might use it now and who might use it later.</i> Not applicable, the data can be reused only after a written consent from the revenue management system provider.
5. Data preservation and archiving	
Preserving your data	
5.1	Which criteria will you use to decide which data has to be archived? <i>Please briefly describe your choices.</i> ☒ Type of data (raw, processed) and how easy it is to reproduce it ☐ Relevance of content for others ☐ Usability of format for others ☒ Data underlying publications ☒ Verification of research ☐ Available time ☐ Available money ☐ Other, namely:
5.2	How long should your data be preserved? Are there any requirements regarding the disposal of data? <i>State obligations you have by law, funder, university, etc. if any.</i> <i>Describe how you will dispose of the data, e.g. how you will get approval, what people and/or tools you need, etc.</i> According to the Research Data Management Regulations of Leiden University, data should be preserved for 10 years. Data disposal is required, according to the nondisclosure agreement data have to be destroyed after two years since the agreement has been signed.

5.3	Which data repository is appropriate for archiving your data? <i>Please describe briefly. Does this archive have a 'data seal of approval' or another form of certification?</i> ☐ Discipline specific (international) repository, namely ... ☐ 4TU.Centre for Research Data ☐ SurfSara ☐ DANS Easy ☐ Other (international) repository, namely : ☒ Other, namely: currently unknown
5.4	Does the archive have specific requirements concerning file formats, metadata etc.? <i>Provide relevant urls to the documentation on these requirements. Describe how you intend to meet those requirements, e.g. converting the file formats, providing supplementary documentation.</i> <i>Will there be extra costs to prepare your data for archiving? Please specify. See http://www.data-archive.ac.uk/media/247429/costingtool.pdf</i>
5.5	What costs (if any) will your selected repository charge? Who pays? <i>Please state the costs in euro's and the institution that pays for it.</i> Currently unknown. Universiteit Leiden will be responsible for payment should this be necessary.
5.6	Who is responsible for the data after the project ends? <i>Please state a position and the current person in that position.</i> Larissa Koupriouchina, PhD candidate.

APPENDIX E: DATASET ONE DATA DEFINITION

Table 31: Data definition of the dataset one variables selected for analysis

Variable	Label	Level	Description	Values	Measurement
<i>P_uniqueChainID</i>		Chain level	Chain ID to which hotel belongs.	Integer	Nominal
<i>uniquePropertyID</i>		Hotel level	Unique property ID. Displayed as “Hotel_ID” in Figure 5.	Integer	Nominal
<i>S_roomCapacity</i>	HotelSize	Hotel level	Hotel room capacity.	6 to 4028	Scale
<i>P_geographicalRegion</i>	HotelRegion	Hotel level	Geographical region of the hotel.	1 = Asia 2 = Australasia 3 = Central America 4 = Europe 5 = Middle East & Africa 6 = North America 7 = South America	Nominal
<i>P_hotelType</i>		Hotel level	Type of hotel.	1 = Airport 2 = Casino 3 = Casino - Native American 4 = Casino – Resort 5 = City – Capital 6 = City – Major 7 = City – Tertiary 8 = Resort – Destination 9 = Resort – Golf 10 = Resort – Other 11 = Resort – Seaside 12 = Resort - Ski	Nominal
<i>Agg_P_hotelType</i>	HotelType	Hotel level	Aggregated hotel type, recoded variable.	1 = Airport 2 = Casino 3 = City 4 = Resort	Nominal
<i>OccupancyDate</i>		Within hotel	Occupancy date for which forecast is made. Displayed as “Occupancy_Date” in Figure 5.	24-MAR-2014 to 07-APR-2017	Scale
<i>businessTypeID</i>	BusinessType	Within hotel	Dichotomous variable representing business segment. Displayed as “Business_Type” in Figure 5.	0 = Transient 1 = Group	Nominal
<i>dta</i>	DTA	Within hotel	Days to arrival. Displayed as “Days_To_Arrival” in Figure 5.	0, 1, 3, 7, 21, 84, 252	Scale
<i>Fp_OccupancyForecast</i>		Within hotel	Forecast-pace data for a certain number of days-to-arrival before an occupancy date. Displayed as “Forecast_Per_Segment” in Figure 5.	1 to 4023	Scale

Table 31: Continued

Variable	Label	Level	Description	Values	Measurement
<i>SM_roomsSold</i>		Within hotel	The final rooms sold at the business segment level. Displayed as “Rooms_Sold_Per_Segment” in Figure 5.	1 to 6940	Scale
<i>S_roomsSold</i>		Within hotel	The final rooms sold for the hotel for this occupancy date for all segments.	0 to 4021	Scale
<i>F_DayHasUserOverrides</i>		Within hotel	Dichotomous variable representing whether forecast days has user overrides. Displayed as “Forecast_Day_Has_Overrides” in Figure 5.	0 = No demand override 1 = Demand override	Nominal
<i>F_NumUserOverrides</i>	numOVER	Within hotel	Number of User Forecast overrides for the occupancy date. Displayed as “Number_of_Overrides” in Figure 5.	0 to 28	Scale
<i>S_SpecialEventImpactsForecast</i>	SpecialEvent	Within hotel	Dichotomous variable representing special event on the occupancy day that is configured to impact forecasts.	0 =No special event 1 = Special event	Nominal
<i>SM_LY_roomsSold</i>		Within hotel	Last year’s rooms sold at business segment level.	0 to 4017	Scale

APPENDIX F: DATASET TWO DATA DEFINITION

Table 32: Data definition of the dataset two variables selected for analysis

Variable	Level	Description	Values ³³	Measurement
<i>P_uniqueChainID</i>	Hotel	Chain ID to which hotel belongs.	Integer	Nominal
<i>uniquePropertyID</i>	Hotel	Unique property ID.	Integer	Nominal
<i>S_roomCapacity</i>	Hotel	Hotel room capacity.	17 to 4028	Scale
<i>P_geographicalRegion</i>	Hotel	Geographical region of the hotel.	1 = Asia 2 = Australasia 3 = Central America 4 = Europe 5 = Middle East & Africa 6 = North America 7 = South America	Nominal
<i>P_hotelType</i>	Hotel	Type of hotel.	1 = Airport 2 = Casino 3 = Casino - Native American 4 = Casino – Resort 5 = City – Capital 6 = City – Major 7 = City – Tertiary 8 = Resort – Destination 9 = Resort – Golf 10 = Resort – Other 11 = Resort – Seaside 12 = Resort - Ski	Nominal
<i>OccupancyDate</i>	Business Type	Occupancy date for which forecast is made.	20-MAR-2015 to 31-AUG-2016	Scale
<i>businessTypeID</i>	Business Type	Dichotomous variable representing business segment	Original values: 1 = Transient 2 = Group Re-coded values: 0 = Group 1 = Transient	Nominal
<i>dta</i>	Business Type	Days to arrival, difference between Occupancy date and Override date.	0 to 548	Scale
<i>SM_roomsSold</i>	Business Type	The final rooms sold at the business segment level.	12 to 944	Scale
<i>S_roomsSold</i>	Hotel	The final rooms sold for the hotel for this occupancy date for all segments.	12 to 3995	Scale
<i>S_SpecialEventImpactsForecast</i>	Business Type	Dichotomous variable representing special event on the occupancy day that is configured to impact forecasts.	0 =No special event 1 = Special event	Nominal
<i>OverrideUpdatedBy</i>	Business Type	User ID who performed an override.	Integer	Nominal
<i>ForecastedDate</i>	Business Type	Date on which override took place.	12-NOV-2013 to 18-MAY-2016	Scale

33 Values were reported after the initial data check and outlier removal (N = 120,105).

Table 32: Continued

Variable	Level	Description	Values ³³	Measurement
<i>OverrideUpdatedOn</i>	Business Type	Date and time of override.	12-NOV-2013 05:24:30 to 18-MAY-2016 22:12:41	Scale
<i>Override Recency Status</i>	Business Type	One of indicators of override status. Demand Overrides with <i>Override Recency Status</i> = 'latest' and <i>Override Status</i> = 'Active' are currently active.	Original values: latest history	Nominal
<i>Override Status</i>	Business Type	One of indicators of override status. Demand Overrides with <i>Override Recency Status</i> = 'latest' and <i>Override Status</i> = 'Active' are currently active.	Original values: Reverted Active	Nominal
<i>OverrideModNum</i>	Business Type	Modification number. Starts at zero and is incremented each time the Override values are modified.	0 to 45	Scale
<i>OverrideType</i>	Business Type	Override type, Demand or Wash	Original values: Wash Demand Re-coded values: 0 = Wash 1 = Demand	Nominal
<i>RoomCapacityCurrentDay</i>	Hotel	Room Capacity Current Day.	0 to 4028	Scale
<i>RoomsSoldCurrentDay</i>	Business Type	The number of rooms sold on current day.	0 to 3746	Scale
<i>RoomsSoldPriorDay</i>	Business Type	The number of room sold for the entire hotel on prior day.	0 to 3683	Scale
<i>RoomsSoldForMktSeg</i>	Business Type	The # of rooms sold for the Market Segments with Overrides.	0 to 950	Scale
<i>OccForecastTransientPriorDay</i>	Business Type	Previous day's Forecast Transient for this occupancyDate. <i>OccForecastTransientPriorDay</i> is the previous day's <i>OccForecastTransient</i> value. Subtracting these two provides the change in the forecast as a result of any overrides added that day, along with any changes in rooms sold, days to arrival, etc.	0 to 3971	Scale

Table 32: Continued

Variable	Level	Description	Values ³³	Measurement
<i>OccForecastGroupPriorDay</i>	Business Type	Previous day's Forecast Group for this occupancy date.	0 to 2861	Scale
<i>OccForecastTransient</i>	Business Type	The total Forecast for Transient after Overrides are applied. <i>OccForecastTransient</i> is the current system forecast for all of the Transient business on the day the override was added (this is from the processing run after the override was added).	0 to 3963	Scale
<i>OccForecastGroup</i>	Business Type	The total Forecast for Group after Overrides are applied.	0 to 2954	Scale
<i>S_LY_roomsSold</i>	Hotel	The final Rooms Sold for the Hotel for this <i>occupancyDate</i> Last Year, same-day-of-week aligned.	0 to 4016	Scale
<i>SM_LY_roomsSold</i>	Business Type	Last year's rooms sold at business segment level.	0 to 1834	Scale
<i>SystemDemandOrWash</i>	Business Type	Total Demand Unconstrained System aggregated by Business Type across Market Segment Groups.	0 to 642	Scale
		Value is System Demand if <i>OverrideType</i> is 'Demand', or it is (Wash% * RoomsSoldForMktSeg) if <i>OverrideType</i> is 'Wash'.		
		Corresponds to the Total Demand Unconstrained System column on RMS printscreen.		
<i>UserDemandOrWash</i>	Business Type	Total Demand – Unconstrained User Demand for Transient MSG and Constrained User Demand for Group MSG.	0 to 7030	Scale
		Similar rule as for <i>SystemDemandOrWash</i> but for User Demand.		
		Corresponds to the Total Demand Unconstrained(Transient)/Constrained(Group) User column on RMS printscreen.		

Table 32: Continued

Variable	Level	Description	Values ³³	Measurement
<i>SystemForecastForMktSeg</i>	Business Type	The System Forecast for the Market Segments where Overrides were applied, but not influenced by user overrides.	0 to 944	Scale
		Corresponds to the Occupancy Forecast Current column on RMS printscreen.		
		Note from email: The <i>SystemForecastForMktSeg</i> value must be from the prior day's forecast process, so it will also not include any changes in Rooms Sold or Cancellations that have changed in the current day.		
		This can be observed in the data when an estimate was performed, and the user override lowers the remaining demand (and the Total Demand) but the Estimate value (<i>UserForecastForMktSeg</i>) is higher than the System value.		

APPENDIX G: DETAILED ADDITIONAL ANALYSIS FROM CHAPTER 4

Table 33: Models with level-1 interaction terms for additional analysis of relationship between DTA and improvement in accuracy

Model:	M _{6,2,1} [†]	M _{6,2,2} [†]	M _{6,2,3} [†]	M _{6,2,4} [†]	M _{6,2,5} [†]	M _{6,2,6} [†]		
DV = LnQ _{imp}								
Fixed part								
Intercept	-8.9E-3	(4.9E-3)	-2.9E-2***	(4.8E-3)	-1.0E-2*	(4.9E-3)	8.5E-3	(5.1E-3)
DTA	-2.8E-4***	(5.5E-5)	2.1E-5	(5.5E-5)	-2.8E-4***	(5.7E-5)	-6.6E-4***	(6.3E-5)
Business Type	4.6E-2***	(2.5E-3)	4.2E-2***	(2.2E-3)	4.8E-2***	(2.2E-3)	4.9E-2***	(2.2E-3)
Size Adjustment Abs	8.6E-2***	(2.2E-3)	1.9E-1***	(3.0E-3)	8.7E-2***	(2.2E-3)	8.6E-2***	(2.2E-3)
Direction	1.9E-2***	(1.4E-3)	2.0E-2***	(1.4E-3)	1.4E-3	(1.7E-3)	1.9E-2***	(1.4E-3)
Special	4.5E-3*	(1.9E-3)	4.2E-3*	(1.9E-3)	-6.5E-5	(2.3E-3)	4.4E-3*	(1.9E-3)
Override Number	3.6E-4*	(1.8E-4)	5.3E-4**	(1.8E-4)	3.7E-4*	(1.8E-4)	3.7E-4*	(1.8E-4)
Override Type	3.3E-2***	(2.5E-3)	3.2E-2***	(2.4E-3)	3.3E-2***	(2.4E-3)	3.3E-2***	(2.7E-3)
Override Status	-4.0E-2***	(1.7E-3)	-3.8E-2***	(1.7E-3)	-4.1E-2***	(1.7E-3)	-4.0E-2***	(1.7E-3)
Override Recency Status	6.9E-4	(1.7E-3)	3.6E-3*	(1.7E-3)	5.1E-4	(1.7E-3)	6.9E-4	(1.7E-3)
DTA×Business Type	6.1E-5**	(2.3E-5)						4.1E-4***
DTA×Size Adjustment Abs			-1.2E-3***	(2.2E-5)				-1.3E-3***
DTA×Direction					3.7E-4***	(2.0E-5)		3.6E-4***
DTA×Special							7.8E-5***	(2.3E-5)
DTA×Override Type							2.8E-5	7.7E-5**
Hotel Size	4.1E-6	(1.0E-5)	1.2E-5	(1.0E-5)	3.8E-6	(1.0E-5)	3.3E-6	4.4E-4***
Hotel Type								1.3E-5
Airport vs City	-7.0E-3	(2.0E-2)	-2.3E-3	(2.0E-2)	-8.1E-3	(2.0E-2)	-7.3E-3	(2.0E-2)
Casino vs City	-2.5E-2	(1.9E-2)	-2.8E-2	(1.9E-2)	-2.4E-2	(1.9E-2)	-2.4E-2	(1.9E-2)
Resort vs City	7.2E-3	(8.1E-3)	6.2E-3	(8.2E-3)	7.5E-3	(8.1E-3)	7.0E-3	(8.1E-3)
Hotel Region								6.4E-3
Asia vs Europe	3.0E-3	(7.6E-3)	2.3E-3	(7.7E-3)	1.6E-3	(7.7E-3)	3.4E-3	(7.6E-3)
Australasia vs Europe	-2.2E-3	(1.5E-2)	6.6E-3	(1.5E-2)	-2.9E-3	(1.5E-2)	-2.2E-3	(1.5E-2)
Central America vs Europe	1.1E-1*	(4.4E-2)	1.0E-1*	(4.5E-2)	1.1E-1*	(4.5E-2)	1.2E-1*	(4.4E-2)
Middle East & Africa vs Europe	-8.8E-3	(1.0E-2)	-9.1E-3	(1.1E-2)	-9.5E-3	(1.0E-2)	-8.6E-3	(1.0E-2)
Europe								-9.5E-3
North America vs Europe	-5.4E-3	(6.4E-3)	-4.0E-3	(6.4E-3)	-6.3E-3	(6.4E-3)	-5.0E-3	(6.4E-3)
South America vs Europe	5.4E-2	(2.8E-2)	4.2E-2	(2.8E-2)	5.4E-2	(2.8E-2)	5.4E-2	(2.8E-2)

Table 33: Continued

DTA×Hotel Size	1.0E-7	(2.0E-7)	3.9E-8	(2.0E-7)	1.4E-7	(1.9E-7)	1.0E-7	(2.0E-7)	1.1E-7	(2.0E-7)	4.7E-8	(2.0E-7)
Random part												
σ_e^2	9.4E-2***		9.3E-2***		9.4E-2***		9.4E-2***		9.4E-2***		9.3E-2***	
σ_{u0}^2	4.8E-3***		4.9E-3***		4.9E-3***		4.8E-3***		4.8E-3***		5.0E-3***	
$\sigma_{\delta1}^2$	2.3E-6***		2.4E-6***		2.3E-6***		2.3E-6***		2.3E-6***		2.3E-6***	
σ_{U01}^2	-1.2E-5*		-7.4E-6		-1.1E-5*		-1.2E-5*		-1.2E-5*		-6.2E-6	
-2*log likelihood												
AIC	109,414.58		106,637.20		109,081.14		109,410.12		109,420.51		105,982.37	
BIC	109,466.58		106,689.20		109,133.14		109,462.12		109,472.51		106,042.37	
	109,734.53		106,957.15		109,401.09		109,730.07		109,740.46		106,351.54	

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).

* $p < .05$, ** $p < .01$, *** $p < .001$



Table 34: Models with level-1 interaction terms for additional analysis of relationship between DTA and improvement in accuracy

Model:	M _{6,3,1} [†]	M _{6,3,2} [†]	M _{6,3,3} [†]	M _{6,3,4} [†]	M _{6,3,5} [†]	M _{6,3,6} [†]
DV = APE _{imp}						
Fixed part						
Intercept	16.48*** (2.33)	14.89*** (2.31)	9.03*** (2.29)	22.14*** (2.29)	21.06*** (2.35)	-1.22 (2.55)
DTA	-0.03 (0.03)	0.02 (0.03)	0.12*** (0.03)	-0.11*** (0.03)	-0.09** (0.03)	0.31*** (0.03)
Business Type	15.33*** (1.59)	1.51 (1.40)	6.05*** (1.39)	5.51*** (1.39)	5.66*** (1.39)	6.81*** (1.72)
Size Adjustment Abs	46.33*** (1.46)	98.19*** (1.97)	45.56*** (1.97)	46.32*** (1.46)	46.45*** (1.46)	95.30*** (1.99)
Direction	-30.03*** (0.94)	-29.16*** (0.93)	-10.60*** (0.93)	-30.04*** (0.94)	-29.99*** (0.94)	-10.51*** (1.13)
Special	0.37 (1.25)	0.55 (1.25)	1.04 (1.25)	0.95 (1.52)	0.52 (1.25)	0.21 (1.52)
Override Number	0.16 (0.10)	0.21* (0.10)	0.16 (0.10)	0.14 (0.10)	0.14 (0.10)	0.23* (0.10)
Override Type	0.15 (1.51)	-4.15*** (1.50)	-0.65 (1.50)	-2.25 (1.50)	-0.69 (1.67)	0.36 (1.83)
Override Status	-14.36*** (1.11)	-13.77*** (1.11)	-12.53*** (1.11)	-14.26*** (1.11)	-14.18*** (1.11)	-12.10*** (1.11)
Override Recency Status	2.99** (1.09)	3.94*** (1.09)	3.18** (1.09)	2.94** (1.09)	2.96** (1.09)	4.19*** (1.09)
DTA×Business Type						
DTA×Size Adjustment Abs	-0.20*** (0.02)					-0.09*** (0.02)
DTA×Direction		-0.58*** (0.01)				-0.55*** (0.02)
DTA×Special			-0.42*** (0.01)	0.01		-0.40*** (0.01)
DTA×Override Type				-0.01 (0.02)		0.02 (0.02)
					-0.04* (0.02)	-0.04* (0.02)
Hotel Size	2.2E-3 (2.9E-3)	0.01* (2.8E-3)	3.7E-3 (2.8E-3)	3.2E-3 (2.9E-3)	3.5E-3 (2.9E-3)	0.01* (2.8E-3)
Hotel Type						
Airport vs City	-5.12 (5.22)	-3.61 (5.30)	-4.15 (5.21)	-4.38 (5.21)	-4.24 (5.21)	-3.48 (5.28)
Casino vs City	1.75 (4.24)	-0.83 (4.38)	0.55 (4.23)	0.38 (4.23)	0.20 (4.23)	-0.16 (4.35)
Resort vs City	-1.70 (1.82)	-1.61 (1.87)	-1.76 (1.82)	-1.50 (1.82)	-1.45 (1.82)	-1.88 (1.86)
Hotel Region						
Asia vs Europe	-1.59 (1.76)	-2.46 (1.81)	-1.11 (1.75)	-1.88 (1.75)	-2.04 (1.76)	-1.67 (1.80)
Australasia vs Europe	-0.60 (3.70)	2.47 (3.75)	0.22 (3.70)	-0.72 (3.70)	-0.77 (3.70)	3.25 (3.74)
Central America vs Europe	69.79*** (11.89)	66.22*** (12.07)	73.66*** (11.86)	69.60*** (11.88)	69.59*** (11.88)	70.04*** (12.04)
Middle East & Africa vs Europe	-2.76 (2.44)	-3.33 (2.49)	-2.50 (2.43)	-3.29 (2.43)	-3.39 (2.43)	-2.43 (2.48)
North America vs Europe	-0.35 (1.25)	-0.64 (1.32)	0.12 (1.25)	-0.60 (1.25)	-0.69 (1.25)	0.09 (1.31)
South America vs Europe	-5.44 (8.70)	-12.21 (8.79)	-3.34 (8.68)	-4.42 (8.69)	-4.59 (8.69)	-11.52 (8.77)

Table 34: Continued

DTA×Hotel Size	-6.2E-5 (9.6E-5)	-1.2E-4 (9.5E-5)	-1.2E-4 (9.6E-5)	-8.0E-5 (9.6E-5)	-8.3E-5 (9.6E-5)	-1.5E-4 (9.5E-5)
Random part						
σ _e ²	44,243.85***	43971.90***	44,088.11***	44,276.59***	44,275.91***	43,787.82***
σ _{u0} ²	247.13***	217.34***	239.66***	250.75***	250.30***	212.78***
σ _{u1} ²	0.53***	0.51***	0.52***	0.53***	0.53***	0.51***
σ _{u01} ²	-11.39***	-10.36***	-11.18***	-11.50***	-11.48***	-10.24***
-2*log likelihood						
AIC	2,992,918.22	2,991,569.50	2,992,131.70	2,993,080.59	2,993,076.35	2,990,640.51
BIC	2,992,970.22	2,991,621.50	2,992,183.70	2,993,132.59	2,993,128.35	2,990,700.51
	2,993,238.17	2,991,889.46	2,992,451.66	2,993,400.54	2,993,396.30	2,991,009.68

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).

* p < .05, ** p < .01, *** p < .001

Table 35: Models with level-1 interaction terms for additional analysis of relationship between DTA and improvement in accuracy

Model:	M _{6,4,1} ²	M _{6,4,2} ²	M _{6,4,3} ²	M _{6,4,4} ²	M _{6,4,5} ²	M _{6,4,6} ²
DV = AE _{imp}						
Fixed part						
Intercept	3.70*** (0.48)	3.68*** (0.47)	3.16*** (0.48)	4.53*** (0.47)	4.98*** (0.48)	2.96*** (0.51)
DTA	-0.01** (4.6E-3)	-0.01* (4.5E-3)	-2.2E-4 (4.7E-3)	-0.03*** (4.5E-3)	-0.03*** (5.0E-3)	-1.6E-3 (0.01)
Business Type	1.58*** (0.28)	0.01 (0.24)	0.38 (0.24)	0.30 (0.24)	0.24 (0.24)	0.17 (0.30)
Size Adjustment Abs	8.21*** (0.23)	12.67*** (0.32)	8.07*** (0.23)	8.21*** (0.23)	8.16*** (0.23)	12.54*** (0.32)
Direction	-2.90*** (0.16)	-2.84*** (0.16)	-0.95*** (0.19)	-2.91*** (0.16)	-2.93*** (0.16)	-0.95*** (0.19)
Special	0.20 (0.22)	0.21 (0.22)	0.27 (0.22)	-0.05 (0.26)	0.21 (0.22)	-0.14 (0.26)
Override Number	0.07** (0.02)	0.07*** (0.02)	0.07** (0.02)	0.07** (0.02)	0.07** (0.02)	0.08*** (0.02)
Override Type	0.93** (0.27)	0.52 (0.27)	0.77** (0.27)	0.64* (0.27)	-0.07 (0.30)	-0.13 (0.33)
Override Status	-3.25*** (0.20)	-3.18*** (0.20)	-3.07*** (0.20)	-3.23*** (0.20)	-3.26*** (0.20)	-3.06*** (0.20)
Override Recency Status	0.59** (0.19)	0.70*** (0.19)	0.60** (0.19)	0.57** (0.19)	0.57** (0.19)	0.69*** (0.19)
DTA×Business Type	-0.02*** (2.6E-3)					-2.8E-3 (3.1E-3)
DTA×Size Adjustment Abs		-0.05*** (2.4E-3)				-0.05*** (2.4E-3)
DTA×Direction			-0.04*** (2.2E-3)			-0.04*** (2.3E-3)
DTA×Special				4.7E-3 (2.6E-3)		0.01* (2.6E-3)
DTA×Override Type					0.02*** (3.1E-3)	0.02*** (3.7E-3)
Hotel Type						
Hotel Size	0.01*** (8.4E-4)	0.01*** (8.4E-4)	0.01*** (8.4E-4)	0.01*** (8.4E-4)	0.01*** (8.4E-4)	0.01*** (8.3E-4)
Airport vs City	-1.40 (1.70)	-1.22 (1.71)	-1.28 (1.70)	-1.37 (1.71)	-1.46 (1.71)	-1.25 (1.69)
Casino vs City	-5.81*** (1.59)	-6.17*** (1.60)	-6.02*** (1.59)	-6.13*** (1.60)	-5.98*** (1.60)	-5.93*** (1.59)
Resort vs City	-0.05 (0.67)	-0.09 (0.67)	-0.06 (0.67)	1.5E-3 (0.67)	-0.02 (0.67)	-0.19 (0.66)
Hotel Region						
Asia vs Europe	-1.66** (0.63)	-1.74** (0.63)	-1.56* (0.63)	-1.71** (0.64)	-1.60 (0.63)	-1.48* (0.63)
Australasia vs Europe	-0.44 (1.24)	-0.07 (1.24)	-0.36 (1.24)	-0.46 (1.24)	-0.45 (1.24)	-0.02 (1.23)
Central America vs Europe	15.57*** (3.69)	15.12*** (3.70)	15.81*** (3.69)	15.43*** (3.71)	15.49*** (3.70)	15.56*** (3.67)
Middle East & Africa vs Europe	-0.07 (0.88)	-0.12 (0.88)	-0.05 (0.88)	-0.15 (0.88)	-0.09 (0.88)	0.06 (0.87)
North America vs Europe	-0.47 (0.52)	-0.47 (0.52)	-0.44 (0.52)	-0.56 (0.52)	-0.49 (0.52)	-0.24 (0.51)
South America vs Europe	4.03 (2.39)	3.62 (2.40)	4.15 (2.39)	4.17 (2.40)	4.23 (2.40)	3.75 (2.38)

Table 35: Continued

DTA×Hotel Size	-8.3E-5*** (1.6E-5)	-8.8E-5*** (1.6E-5)	-8.9E-5*** (1.6E-5)	-8.5E-5*** (1.6E-5)	-8.4E-5*** (1.6E-5)	-9.1E-5*** (1.6E-5)
Random part						
σ ² _e	1,230.17***	1,228.42***	1,228.81***	1,230.60***	1,230.50***	1,226.35
σ ² _{u0}	26.50***	26.37***	26.45***	27.12***	26.80***	25.62
σ ² _{u1}	0.01***	0.01***	0.01***	0.01***	0.01***	0.01
σ ² _{u01}	-0.13***	-0.12***	-0.13***	-0.14***	-0.13***	-0.11
-2*log likelihood						
AIC	2,224,447.17	2,224,114.61	2,224,194.61	2,224,528.15	2,224,505.17	2,223,735.73
BIC	2,224,499.17	2,224,166.61	2,224,246.61	2,224,580.15	2,224,557.17	2,223,795.73
	2,224,767.38	2,224,434.82	2,224,514.82	2,224,848.36	2,224,825.38	2,224,105.20

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).

* *p* < .05, ** *p* < .01, *** *p* < .001

Table 36: Models without and with level-1 interaction terms for additional analysis of relationship between size of adjustment and improvement in accuracy

Model:	M _{6.2.7} [†]	M _{6.2.8} [†]	M _{6.2.9} [†]	M _{6.2.10} [†]	M _{6.2.11} [†]
DV = LnQ_{imp}					
Fixed part					
Intercept	4.4E-2*** (4.8E-3)	4.5E-2*** (4.5E-3)	-1.2E-2* (4.8E-3)	-2.3E-2*** (4.9E-3)	8.1E-2*** (4.6E-3)
DTA	-1.7E-4** (5.1E-5)	-3.3E-4*** (4.9E-5)	-2.6E-4*** (4.9E-5)	-2.5E-4*** (5.4E-5)	-2.6E-4*** (4.9E-5)
Business Type	-2.5E-2*** (2.4E-3)	3.8E-2*** (2.1E-3)	4.9E-2*** (2.1E-3)	5.3E-2*** (2.2E-3)	-1.5E-2*** (2.4E-3)
Size Adjustment Abs	-1.1E-1*** (3.5E-3)	-1.3E-1*** (2.7E-3)	9.0E-2*** (2.7E-3)	1.3E-1*** (4.2E-3)	-2.6E-1*** (5.6E-3)
Direction	1.6E-2*** (1.4E-3)	-7.0E-2*** (1.5E-3)	1.9E-2*** (1.5E-3)	1.9E-2*** (1.4E-3)	-6.6E-2*** (1.5E-3)
Special	5.1E-3** (1.9E-3)	3.1E-3 (1.8E-3)	7.8E-3*** (1.8E-3)	4.9E-3* (1.9E-3)	2.3E-3 (2.0E-3)
Override Number	-6.5E-5 (1.8E-4)	1.9E-4 (1.7E-4)	3.6E-4 (1.8E-4)	3.3E-4 (1.8E-4)	-1.0E-4 (1.7E-4)
Override Type	9.2E-3*** (2.4E-3)	2.1E-2*** (2.4E-3)	3.4E-2*** (2.4E-3)	4.8E-2*** (2.7E-3)	3.5E-3 (2.6E-3)
Override Status	-4.1E-2*** (1.7E-3)	-4.2E-2*** (1.7E-3)	-4.0E-2*** (1.7E-3)	-3.9E-2*** (1.7E-3)	-4.2E-2*** (1.7E-3)
Override Recency Status	2.9E-3 (1.7E-3)	6.5E-3*** (1.6E-3)	7.9E-4 (1.7E-3)	1.1E-3 (1.7E-3)	7.6E-3*** (1.6E-3)
Size Adjustment Abs×Business Type	3.1E-1*** (4.3E-3)				2.3E-1*** (4.4E-3)
Size Adjustment Abs×Direction		5.0E-1*** (3.8E-3)			4.7E-1*** (3.9E-3)
Size Adjustment Abs×Special			-1.8E-2*** (4.8E-3)		7.1E-3 (4.6E-3)
Size Adjustment Abs×Override Type				-6.3E-2*** (4.8E-3)	8.2E-4 (4.8E-3)
Hotel Size	6.9E-6 (1.0E-5)	6.5E-6 (9.2E-6)	3.5E-6 (1.0E-5)	5.1E-6 (1.0E-5)	8.7E-6 (9.1E-6)
Hotel Type					
Airport vs City	-5.3E-3 (2.0E-2)	-1.3E-2 (1.8E-2)	-7.1E-3 (1.8E-2)	-5.1E-3 (2.0E-2)	-1.1E-2 (1.8E-2)
Casino vs City	-2.9E-2 (1.8E-2)	-2.2E-2 (1.7E-2)	-2.4E-2 (1.9E-2)	-2.6E-2 (1.9E-2)	-2.6E-2 (1.7E-2)
Resort vs City	5.9E-3 (8.0E-3)	3.6E-3 (7.2E-3)	7.1E-3 (8.1E-3)	7.0E-3 (8.1E-3)	3.1E-3 (7.1E-3)
Hotel Region					
Asia vs Europe	-3.4E-3 (7.5E-3)	-3.6E-3 (6.8E-3)	3.2E-3 (7.6E-3)	2.0E-3 (7.6E-3)	-7.8E-3 (6.7E-3)
Australasia vs Europe	-7.0E-3 (1.5E-2)	-3.3E-3 (1.3E-2)	-2.1E-3 (1.5E-2)	-1.8E-3 (1.5E-2)	-6.9E-3 (1.3E-2)
Central America vs Europe	1.1E-1* (4.4E-2)	1.5E-1*** (3.9E-2)	1.1E-1* (4.4E-2)	1.2E-1** (4.4E-2)	1.4E-1*** (3.9E-2)
Middle East & Africa vs Europe	-1.1E-2 (1.0E-2)	-7.8E-3 (9.3E-3)	-8.5E-3 (9.3E-3)	-9.1E-3 (1.0E-2)	-9.8E-3 (9.2E-3)
North America vs Europe	-9.6E-3 (6.2E-3)	-6.7E-3 (5.6E-3)	-5.3E-3 (5.6E-3)	-5.9E-3 (6.3E-3)	-9.4E-3 (5.6E-3)
South America vs Europe	4.2E-2 (2.7E-2)	5.5E-2* (2.5E-2)	5.3E-2 (2.8E-2)	5.5E-2 (2.8E-2)	4.7E-2 (2.5E-2)

Table 36: Continued

DTA×Hotel Size	5.9E-8 (1.9E-7)	1.1E-7 (1.8E-7)	1.1E-7 (2.0E-7)	1.1E-7* (2.0E-7)	7.2E-8 (1.7E-7)
Random part					
σ _e ²	9.2E-2***	8.7E-2***	9.4E-2***	9.4E-2***	8.6E-2***
σ _{u0} ²	4.6E-3***	3.7E-3***	4.8E-3***	4.8E-3***	3.7E-3***
σ _{u1} ²	2.1E-6***	1.9E-6***	2.3E-6***	2.3E-6***	1.8E-6***
σ _{U01} ²	-8.5E-6	-1.7E-5***	-1.2E-5*	-1.2E-5*	-1.7E-5***
-2*log likelihood	104,265.30	92,704.62	109,408.13	109,253.50	89,934.15
AIC	104,317.30	92,756.62	109,460.13	109,305.50	89,992.15
BIC	104,585.25	93,024.57	109,728.08	109,573.45	90,291.02

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive), SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).
* *p* < .05, ** *p* < .01, *** *p* < .001

Table 37: Models without and with level-1 interaction terms for additional analysis of relationship between size of adjustment and improvement in accuracy

Model:	$M_{6,3,7}$	$M_{6,3,8}$	$M_{6,3,9}$	$M_{6,3,10}$	$M_{6,3,11}$
DV = APE_{imp}					
Fixed part					
Intercept	27.24*** (2.33)	-9.05*** (2.22)	21.71*** (2.29)	17.24*** (2.37)	2.51 (2.36)
DTA	-0.11*** (0.03)	-0.07** (0.03)	-0.12*** (0.03)	-0.11*** (0.03)	-0.04 (0.02)
Business Type	-1.71 (1.54)	11.45*** (1.35)	5.41*** (1.39)	7.11*** (1.40)	-8.58*** (1.55)
Size Adjustment Abs	26.51*** (2.35)	173.42*** (1.82)	49.43*** (1.62)	66.37*** (2.80)	127.38*** (3.82)
Direction	-30.32*** (0.94)	21.23*** (1.02)	-30.01*** (0.94)	-29.90*** (0.94)	23.04*** (1.02)
Special	0.59 (1.25)	0.87 (1.22)	3.14* (1.39)	0.72 (1.25)	2.87* (1.35)
Override Number	0.09 (0.10)	0.18 (0.09)	0.13 (0.10)	0.11 (0.10)	0.04 (0.09)
Override Type	-4.78** (1.52)	4.86** (1.46)	-2.26 (1.50)	3.56* (1.65)	-1.38 (1.66)
Override Status	-14.32*** (1.11)	-13.30*** (1.08)	-14.24*** (1.11)	-14.05*** (1.11)	-13.26*** (1.08)
Override Recency Status	3.23** (1.09)	-0.71 (1.06)	2.98** (1.09)	3.09** (1.09)	0.09 (1.06)
Size Adjustment Abs×Business Type	31.43*** (2.92)				88.92*** (3.00)
Size Adjustment Abs×Direction		-291.02*** (2.62)			-305.72*** (2.66)
Size Adjustment Abs×Special			-14.07*** (3.22)		-8.58** (3.14)
Size Adjustment Abs×Override Type				-26.97*** (3.21)	-2.27 (3.24)
Hotel Size	3.6E-3 (2.9E-3)	1.9E-3 (2.7E-3)	3.2E-3 (2.9E-3)	3.7E-3 (2.9E-3)	2.8E-3 (2.6E-3)
Hotel Type					
Airport vs City	-3.65 (5.22)	-3.09 (5.04)	-4.39 (5.21)	-3.88 (5.22)	-1.10 (5.03)
Casino vs City	-0.66 (4.23)	1.89 (4.08)	0.54 (4.23)	0.02 (4.24)	-1.00 (4.07)
Resort vs City	-1.43 (1.82)	0.24 (1.76)	-1.47 (1.82)	-1.36 (1.82)	0.49 (1.77)
Hotel Region					
Asia vs Europe	-2.45 (1.76)	0.73 (1.70)	-1.97 (1.76)	-2.22 (1.76)	-0.74 (1.70)
Australasia vs Europe	-1.24 (3.70)	0.40 (3.57)	-0.67 (3.70)	-0.76 (3.71)	-0.78 (3.55)
Central America vs Europe	69.63*** (11.87)	53.98*** (11.59)	69.47*** (11.87)	69.59*** (11.86)	55.98*** (11.69)
Middle East & Africa vs Europe	-3.34 (2.43)	-2.20 (2.36)	-3.21 (2.44)	-3.35 (2.44)	-2.08 (2.36)
North America vs Europe	-0.58 (1.25)	-1.09 (1.20)	-0.64 (1.25)	-0.74 (1.25)	-0.95 (1.20)
South America vs Europe	-5.92 (8.69)	-3.17 (8.47)	-4.87 (8.69)	-3.39 (8.68)	-7.31 (8.52)

Table 37: Continued

DTA×Hotel Size	-8.6E-5 (9.7E-5)	-9.4E-5 (8.9E-5)	-8.1E-5 (9.6E-5)	-8.0E-5 (9.6E-5)	-1.0E-4 (8.6E-5)
Random part					
σ_e^2	44,251.52***	41,953.29***	44,273.19***	44,263.81***	41,763.97***
σ_{u0}^2	255.65***	190.00***	253.35***	255.67***	146.76***
σ_{u1}^2	0.54***	0.44***	0.53***	0.52***	0.42***
σ_{u01}^2	-11.67***	-9.19***	-11.53***	-11.54***	-7.84***
-2*log likelihood					
AIC	2,992,965.22	2,981,079.89	2,993,061.80	2,993,010.27	2,980,101.04
BIC	2,993,017.22	2,981,131.89	2,993,113.80	2,993,062.27	2,980,159.04
	2,993,285.17	2,981,399.84	2,993,381.75	2,993,330.22	2,980,457.91

Note. N = 228,649. Standard errors are in parentheses.

DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 38: Models without and with level-1 interaction terms for additional analysis of relationship between size of adjustment and improvement in accuracy

Model:	M _{6,4,7} ²	M _{6,4,8} ²	M _{6,4,9} ²	M _{6,4,10} ²	M _{6,4,11} ²
DV = AE _{imp}					
Fixed part					
Intercept	4.83*** (0.48)	4.54*** (0.47)	4.62*** (0.47)	3.03*** (0.48)	3.46*** (0.50)
DTA	-0.02*** (4.5E-3)	-0.02*** (4.5E-3)	-0.02*** (4.5E-3)	-0.02*** (4.5E-3)	-0.02*** (4.5E-3)
Business Type	-0.21 (0.27)	0.28 (0.24)	0.35 (0.24)	0.80** (0.25)	0.56* (0.28)
Size Adjustment Abs	7.07*** (0.35)	7.98*** (0.28)	7.32*** (0.26)	13.40*** (0.43)	11.56*** (0.58)
Direction	-2.93*** (0.16)	-3.02*** (0.18)	-2.91*** (0.16)	-2.87*** (0.16)	-2.99*** (0.18)
Special	0.23 (0.22)	0.22 (0.22)	-0.52* (0.24)	0.27 (0.22)	-0.57* (0.24)
Override Number	0.06** (0.02)	0.07** (0.02)	0.07** (0.02)	0.06** (0.02)	0.06** (0.02)
Override Type	0.45 (0.27)	0.61* (0.27)	0.65* (0.27)	2.36*** (0.30)	2.26*** (0.31)
Override Status	-3.23*** (0.20)	-3.23*** (0.20)	-3.25*** (0.20)	-3.18*** (0.20)	-3.19*** (0.20)
Override Recency Status	0.61** (0.19)	0.59** (0.19)	0.56** (0.19)	0.62** (0.19)	0.62** (0.19)
Size Adjustment Abs×Business Type	1.92*** (0.45)				1.03* (0.47)
Size Adjustment Abs×Direction		0.59 (0.42)			0.57 (0.44)
Size Adjustment Abs×Special			3.63*** (0.48)		4.16*** (0.49)
Size Adjustment Abs×Override Type				-7.00*** (0.50)	-7.00*** (0.51)
Hotel Size	0.01*** (8.4E-4)	0.01*** (8.4E-4)	0.01*** (8.4E-4)	0.01*** (8.5E-4)	0.01*** (8.5E-4)
Hotel Type					
Airport vs City	-1.34 (1.71)	-1.37 (1.71)	-1.44 (1.71)	-1.23 (1.71)	-1.30 (1.71)
Casino vs City	-6.16*** (1.60)	-6.11*** (1.60)	-6.13*** (1.60)	-6.37*** (1.61)	-6.42*** (1.61)
Resort vs City	0.02 (0.67)	0.01 (0.67)	0.01 (0.67)	0.03 (0.67)	0.05 (0.67)
Hotel Region					
Asia vs Europe	-1.75** (0.64)	-1.71** (0.63)	-1.69** (0.64)	-1.85** (0.64)	-1.87** (0.64)
Australasia vs Europe	-0.48 (1.24)	-0.45 (1.24)	-0.49 (1.24)	-0.41 (1.25)	-0.48 (1.25)
Central America vs Europe	15.35*** (3.71)	15.45*** (3.71)	15.47*** (3.71)	15.52*** (3.72)	15.55*** (3.72)
Middle East & Africa vs Europe	-0.19 (0.88)	-0.16 (0.88)	-0.18 (0.88)	-0.21 (0.89)	-0.27 (0.89)
North America vs Europe	-0.60 (0.52)	-0.57 (0.52)	-0.55 (0.52)	-0.65 (0.52)	-0.65 (0.52)
South America vs Europe	4.01 (2.40)	4.14 (2.40)	4.26 (2.40)	4.19 (2.41)	4.25 (2.41)

Table 38: Continued

DTA×Hotel Size	-8.5E-5*** (1.6E-5)	-8.5E-5*** (1.6E-5)	-8.5E-5*** (1.6E-5)	-8.4E-5*** (1.6E-5)	-8.4E-5*** (1.6E-5)
Random part					
σ ² _e	1,230.54***	1,230.62***	1,230.30***	1,229.54***	1,229.10***
σ ² _{u0}	27.08***	27.03***	27.03***	27.40***	27.41***
σ ² _{u1}	0.01***	0.01***	0.01***	0.01***	0.01***
σ ² _{u01}	-0.14***	-0.14***	-0.13***	-0.14***	-0.14***
-2*log likelihood	2,224,513.06	2,224,529.45	2,224,475.40	2,224,333.94	2,224,258.14
AIC	2,224,565.06	2,224,581.45	2,224,527.40	2,224,385.94	2,224,316.14
BIC	2,224,833.27	2,224,849.66	2,224,795.61	2,224,654.15	2,224,615.29

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).
* *p* < .05, ** *p* < .01, *** *p* < .001

Table 39: Models without and with level-1 interaction terms for additional analysis of relationship between business type and improvement in accuracy

Model:	$M_{6.2.12}^{\cdot}$		$M_{6.2.13}^{\cdot}$		$M_{6.2.14}^{\cdot}$		$M_{6.2.15}^{\cdot}$	
DV = LnQ_{imp}								
Fixed part								
Intercept	-1.6E-2**	(4.8E-3)	-1.1E-2*	(4.8E-3)	-4.6E-3	(6.2E-3)	-1.1E-2	(6.3E-3)
DTA	-2.6E-4***	(5.4E-5)	-2.6E-4***	(5.4E-5)	-2.6E-4***	(5.4E-5)	-2.6E-4***	(5.4E-5)
Business Type	6.2E-2***	(2.6E-3)	4.9E-2***	(2.3E-3)	4.2E-2***	(4.9E-3)	5.7E-2***	(5.1E-3)
Size Adjustment Abs	8.6E-2***	(2.2E-3)	8.6E-2***	(2.2E-3)	8.6E-2***	(2.2E-3)	8.6E-2***	(2.2E-3)
Direction	2.9E-2***	(1.8E-3)	1.9E-2***	(1.4E-3)	1.9E-2***	(1.4E-3)	2.9E-2***	(1.8E-3)
Special	4.5E-3*	(1.9E-3)	5.3E-3*	(2.4E-3)	4.5E-3*	(1.9E-3)	5.3E-3*	(2.4E-3)
Override Number	3.6E-4*	(1.8E-4)	3.6E-4*	(1.8E-4)	3.8E-4*	(1.8E-4)	3.7E-4*	(1.8E-4)
Override Type	3.3E-2***	(2.4E-3)	3.4E-2*	(2.4E-3)	2.7E-2***	(4.9E-3)	2.8E-2***	(4.9E-3)
Override Status	-3.7E-2***	(1.8E-3)	-4.0E-2*	(1.7E-3)	-4.0E-2***	(1.7E-3)	-3.7E-2***	(1.8E-3)
Override Recency Status	1.7E-3	(1.7E-3)	7.0E-4	(1.7E-3)	7.5E-4	(1.7E-3)	1.7E-3	(1.7E-3)
Business Type×Direction	-2.8E-2***	(2.9E-3)					-2.8E-2***	(2.9E-3)
Business Type×Special			-2.3E-3	(3.8E-3)			-2.0E-3	(3.8E-3)
Business Type×Override Type					8.5E-3	(5.4E-3)	6.2E-3	(5.5E-3)
Hotel Size	2.8E-6	(1.0E-5)	3.5E-6	(1.0E-5)	3.4E-6	(1.0E-5)	2.8E-6	(1.0E-5)
Hotel Type								
Airport vs City	-6.7E-3	(2.0E-2)	-7.1E-3	(2.0E-2)	-7.6E-3	(2.0E-2)	-7.1E-3	(2.0E-2)
Casino vs City	-2.3E-2	(1.9E-2)	-2.4E-2	(1.9E-2)	-2.4E-2	(1.9E-2)	-2.3E-2	(1.9E-2)
Resort vs City	7.7E-3	(8.1E-3)	7.1E-3	(8.1E-3)	7.0E-3	(8.1E-3)	7.7E-3	(8.1E-3)
Hotel Region								
Asia vs Europe	1.8E-3	(7.6E-3)	3.2E-3	(7.6E-3)	3.0E-3	(7.6E-3)	1.7E-3	(7.6E-3)
Australasia vs Europe	-3.2E-3	(1.5E-2)	-2.2E-3	(1.5E-2)	-2.4E-3	(1.5E-2)	-3.3E-3	(1.5E-2)
Central America vs Europe	1.1E-1*	(4.4E-2)	1.1E-1*	(4.4E-2)	1.1E-1*	(4.4E-2)	1.1E-1*	(4.4E-2)
Middle East & Africa vs Europe	-9.5E-3	(1.0E-2)	-8.6E-3	(1.0E-2)	-8.7E-3	(1.0E-2)	-9.5E-3`	(1.0E-2)
North America vs Europe	-5.9E-3	(6.4E-3)	-5.2E-3	(6.4E-3)	-5.0E-3	(6.3E-3)	-5.8E-3	(6.3E-3)
South America vs Europe	5.4E-2	(2.8E-2)	5.4E-2	(2.8E-2)	5.4E-2	(2.8E-2)	5.4E-2	(2.8E-2)
DTA×Hotel Size	1.1E-7	(2.0E-7)	1.1E-7	(2.0E-7)	1.0E-7	(2.0E-7)	1.1E-7	(2.0E-7)
Random part								
σ_e^2	9.4E-2		9.4E-2		9.4E-2***		9.4E-2***	
σ_{u0}^2	4.8E-3		4.8E-3		4.8E-3***		4.8E-3***	
σ_{u1}^2	2.3E-6		2.3E-6		2.3E-6***		2.3E-6***	
σ_{u01}^2	-1.2E-5*		-1.2E-5*		-1.2E-5*		-1.2E-5*	
-2*log likelihood	109,329.07		109,421.19		109,419.10		109,327.53	
AIC	109,381.07		109,473.19		109,471.10		109,383.53	
BIC	109,649.02		109,741.14		109,739.05		109,672.09	

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 40: Models without and with level-1 interaction terms for additional analysis of relationship between business type and improvement in accuracy

Model:	$M_{6.3.12}^{\cdot}$		$M_{6.3.13}^{\cdot}$		$M_{6.3.14}^{\cdot}$		$M_{6.3.15}^{\cdot}$	
DV = APE_{imp}								
Fixed part								
Intercept	10.88***	(2.31)	22.15***	(2.29)	22.30***	(3.36)	6.91*	(3.39)
DTA	-0.11***	(0.03)	-0.12***	(0.03)	-0.12***	(0.03)	-0.11***	(0.03)
Business Type	35.52***	(1.65)	5.81***	(1.44)	5.45	(3.04)	40.11***	(3.23)
Size Adjustment Abs	46.30***	(1.45)	46.33***	(1.46)	46.33***	(1.47)	46.56***	(1.46)
Direction	-5.31***	(1.19)	-30.04***	(0.94)	-30.04***	(0.94)	-5.33***	(1.19)
Special	0.68	(1.25)	1.17	(1.57)	0.48	(1.25)	1.18	(1.57)
Override Number	0.09	(0.10)	0.14	(0.10)	0.14	(0.10)	0.08	(0.10)
Override Type	-5.17**	(1.50)	-2.26	(1.50)	-2.32	(2.99)	-1.07	(2.98)
Override Status	-9.09***	(1.12)	-14.27***	(1.11)	-14.26***	(1.11)	-9.09***	(1.12)
Override Recency Status	5.65***	(1.09)	2.91**	(1.09)	2.92**	(1.09)	5.66***	(1.09)
Business Type×Direction	-64.35***	(1.94)					-64.50***	(1.94)
Business Type×Special			-1.88	(2.55)			-1.42	(2.54)
Business Type×Override Type					0.08	(3.41)	-5.43	(3.41)
Hotel Size	2.3E-3	(2.9E-3)	3.2E-3	(2.9E-3)	3.2E-3	(2.9E-3)	2.3E-3	(2.9E-3)
Hotel Type								
Airport vs City	-3.59	(5.21)	-4.41	(5.21)	-4.37	(5.21)	-3.62	(5.22)
Casino vs City	1.06	(4.24)	0.38	(4.23)	0.34	(4.23)	1.33	(4.26)
Resort vs City	-0.50	(1.82)	-1.51	(1.82)	-1.51	(1.82)	-0.37	(1.83)
Hotel Region								
Asia vs Europe	-3.76*	(1.76)	-1.88	(1.76)	-1.89	(1.76)	-3.66*	(1.77)
Australasia vs Europe	-3.15	(3.70)	-0.76	(3.70)	-0.75	(3.70)	-3.06	(3.71)
Central America vs Europe	66.58***	(11.88)	69.55***	(11.88)	69.58***	(11.88)	66.34***	(11.89)
Middle East & Africa vs Europe	-5.11*	(2.44)	-3.28	(2.43)	-3.28	(2.43)	-5.20*	(2.44)
North America vs Europe	-1.38	(1.26)	-0.59	(1.25)	-0.59	(1.25)	-1.52	(1.27)
South America vs Europe	-3.74	(8.69)	-4.40	(8.69)	-4.41	(8.69)	-3.58	(8.69)
DTA×Hotel Size	-7.0E-5	(9.5E-5)	-8.1E-5	(9.6E-5)	-8.1E-5	(9.6E-5)	-6.8E-5	(9.5E-5)
Random part								
σ_e^2	44,057.60***		44,276.55***		44,276.64***		44,056.84***	
σ_{u0}^2	241.26***		251.04***		250.99***		242.07***	
σ_{u1}^2	0.52***		0.53***		0.53***		0.52***	
σ_{u01}^2	-11.13***		-11.50***		-11.50***		-11.14***	
-2*log likelihood	2,991,978.63		2,993,080.34		2,993,080.88		2,991,975.77	
AIC	2,992,030.63		2,993,132.34		2,993,132.88		2,992,031.77	
BIC	2,992,298.58		2,993,400.29		2,993,400.83		2,992,320.33	

Note. N = 228,649. Standard errors are in parentheses.
DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).
* $p < .05$, ** $p < .01$, *** $p < .001$

Table 41: Models without and with level-1 interaction terms for additional analysis of relationship between business type and improvement in accuracy

Model:	$M_{6.4.12}^2$		$M_{6.4.13}^2$		$M_{6.4.14}^2$		$M_{6.4.15}^2$	
DV = AE_{imp}								
Fixed part								
Intercept	3.73***	(0.48)	4.44***	(0.47)	4.91***	(0.65)	3.86***	(0.65)
DTA	-0.02***	(4.5E-3)	-0.02***	(4.5E-3)	-0.02***	(4.5E-3)	-0.02***	(4.5E-3)
Business Type	2.36***	(0.29)	0.40	(0.25)	-0.17	(0.54)	2.28***	(0.57)
Size Adjustment Abs	8.20***	(0.23)	8.21***	(0.23)	8.18***	(0.23)	8.19***	(0.23)
Direction	-1.21***	(0.20)	-2.91***	(0.16)	-2.90***	(0.16)	-1.21***	(0.20)
Special	0.23	(0.22)	0.46	(0.27)	0.22	(0.22)	0.45	(0.27)
Override Number	0.07**	(0.02)	0.07**	(0.02)	0.07**	(0.02)	0.07**	(0.02)
Override Type	0.47	(0.27)	0.63*	(0.27)	0.18	(0.54)	0.30	(0.54)
Override Status	-2.90***	(0.20)	-3.23***	(0.20)	-3.23***	(0.20)	-2.90***	(0.20)
Override Recency Status	0.74***	(0.19)	0.58**	(0.19)	0.58**	(0.19)	0.74***	(0.19)
Business Type×Direction	-4.44***	(0.33)					-4.43***	(0.33)
Business Type×Special			-0.65	(0.43)			-0.60	(0.43)
Business Type×Override Type					0.59	(0.61)	0.22	(0.61)
Hotel Size	0.01***	(8.4E-4)	0.01***	(8.4E-4)	0.01***	(8.4E-4)	0.01***	(8.4E-4)
Hotel Type								
Airport vs City	-1.32	(1.71)	-1.37	(1.71)	-1.40	(1.71)	-1.33	(1.71)
Casino vs City	-6.00***	(1.60)	-6.10***	(1.60)	-6.12***	(1.60)	-5.99***	(1.60)
Resort vs City	0.10	(0.67)	0.00	(0.67)	0.00	(0.67)	0.10	(0.67)
Hotel Region								
Asia vs Europe	-1.91**	(0.64)	-1.70**	(0.63)	-1.71**	(0.64)	-1.91**	(0.64)
Australasia vs Europe	-0.60	(1.24)	-0.45	(1.24)	-0.46	(1.24)	-0.61	(1.24)
Central America vs Europe	15.21***	(3.71)	15.42***	(3.71)	15.42***	(3.71)	15.19***	(3.71)
Middle East & Africa vs Europe	-0.29	(0.88)	-0.14	(0.88)	-0.15	(0.88)	-0.28	(0.88)
North America vs Europe	-0.67	(0.52)	-0.57	(0.52)	-0.55	(0.52)	-0.67	(0.52)
South America vs Europe	4.21	(2.40)	4.15	(2.40)	4.13	(2.40)	4.21	(2.40)
DTA×Hotel Size	-8.4E-5***	(1.6E-5)	-8.5E-5***	(1.6E-5)	-8.5E-5***	(1.6E-5)	-8.4E-5***	(1.6E-5)
Random part								
σ_e^2	1,229.59***		1,230.62***		1,230.63***		1,229.58***	
σ_{u0}^2	27.11***		27.01***		27.02***		27.08***	
σ_{u1}^2	0.01***		0.01***		0.01***		0.01***	
σ_{u01}^2	-0.13***		-0.13***		-0.14***		-0.13***	
-2*log likelihood	2,224,346.20		2,224,529.14		2,224,530.45		2,224,344.11	
AIC	2,224,398.20		2,224,581.14		2,224,582.45		2,224,400.11	
BIC	2,224,666.41		2,224,849.34		2,224,850.66		2,224,688.95	

Note. N = 228,649. Standard errors are in parentheses.

DTA (continuous variable ranging from 0 to 596); Business Type (0 = transient, 1 = group); Size Adjustment Abs (continuous variable ranging from 0 to 5.46); Direction (0 = negative, 1 = positive); SPECIAL (0 = no special event, 1 = special event); Override Number (continuous variable ranging from 1 to 52); Override Type (0 = wash, 1 = demand); Override Status (0 = reverted, 1 = active); Override Recency Status (0 = history, 1 = latest); Hotel Size (continuous variable, centered on grand mean, ranging from -302.61 to 3,725.39).

* $p < .05$, ** $p < .01$, *** $p < .001$