A Call for Exploratory Data Analysis in Revenue Management Forecasting: a case study of a small and independent hotel in The Netherlands

Dirk Sierag
CWI Amsterdam / Vrije Universiteit Amsterdam, d.d.sierag@cwi.nl

Jean-Pierre van der Rest
Leiden University, j.i.van.der.rest@law.leidenuniv.nl

Ger Koole
Vrije Universiteit Amsterdam, ger.koole@vu.nl

Rob van der Mei
CWI Amsterdam / Vrije Universiteit Amsterdam, r.d.van.der.Mei@cwi.nl

Bert Zwart
CWI Amsterdam / Eindhoven University of Technology, bert.zwart@cwi.nl

To appear in *International Journal of Revenue Management*, 2016

Abstract

Using five years of data collected from a small and independent hotel in The Netherlands this case study explores RMS data as a means to seek new insights into occupancy forecasting. The study provides an insight into the random nature of group cancellations, an important but neglected aspect in hotel revenue management modeling. The empirical study also shows that in a local market context demand differs significantly per point of time during the day, in addition to a seasonal monthly and weekly demand pattern. Moreover, the study presents evidence on the inhomogeneous Poisson nature of the probability distribution function that demand follows, a crucial characteristic for forecasting modeling that is generally assumed but not reported in the hotel forecasting literature. This implies that demand is more uncertain for smaller than for larger hotels. By reporting the results of an in-depth case study, this paper seeks to draw attention to the critical and often overlooked role of exploratory data analysis in hotel revenue management forecasting. Implications for theory and directions for future research are provided.

Key words: hotel; revenue management; forecasting; data analysis; SME; independent; small

Acknowledgements
This research was partly funded by the Ministry of Education, Culture and Science (OCW) via the Regional Attention and Action for Knowledge circulation (RAAK), Foundation Innovation Alliance (SIA).

A Call for Exploratory Data Analysis in Revenue Management Forecasting: a case study of a small and independent hotel in The Netherlands

1. Introduction

Since the early 1990s hotel revenue management practice has evolved gradually (Ferguson & Smith, 2014) setting off large investments in sophisticated revenue management systems (RMS). Whilst varying in structure these RMS essentially calculate and update booking limits within a reservation system, extracting and processing information from various other systems (Phillips, 2005). One of these systems, lying at the heart of each RMS, is forecasting (Lemke, Riedel & Gabrys, 2013). As Talluri and van Ryzin (2004a, p. 407) 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." While there is ample research on forecasting, a major weakness of work in hotel revenue management is its focus on the *model* selection aspect of hotel forecasting, with notable exceptions such as Schwartz and Hiemstra (1997), Kimes (1999), Schwartz (2003), Schwartz and Cohen (2004), Bendoly (2013) and Koupriouchina et al. (2014). Forecasting comprises multiple facets including (a) problem definition, (b) information gathering, (c) preliminary (exploratory) data analysis, (d) choosing and fitting models, and (e) evaluating and adjusting the model (Makridakis, Wheelwright & Hyndman, 1998). In hotel revenue management all steps are critical and overlooking any of these steps can lead to underperformance of the RMS. Moreover, even after an initial round of model selection and evaluation, new data analysis will be required: hotels operate in a changing environment affecting the nature of the data, and thus adjustments to the model analysis may be required. Yet, most research focuses on defining a forecasting problem, developing or selecting a forecasting model, and testing the model. The crucial steps of information gathering and (preliminary) data analysis are often overlooked.

This article, therefore, aims to draw attention to the importance of regular data analysis by demonstrating how a real-life hotel can gain new forecasting insights by exploring and analyzing data from its RMS. To this purpose, key factors of hotel demand, price sensitivity, and cancellations are identified, by analyzing data from a small and independent hotel in The Netherlands. In particular, group cancellation behavior, the effects of uncertainty in demand, and different dimensions of seasonality are studied. The remainder of the article is organized as follows. Section two explains the background of the research problem. In section three, the data set and hotel case study are described. Sections four, five and six, respectively, provide the case study findings, in particular insights into (1) different levels of seasonality; (2) group cancellation behavior; and (3) uncertainty in demand and cancellations. Finally, the article discusses the findings, the limitations of the research and provides directions for future research.

2. Background

Forecasting is an area in operations research which over the years has grown into a whole discipline of its own with specialist research attention from a wide range of disciplines and sectors (Fildes, Nikolopoulos, Crone & Syntetos, 2008). For example, forecasting has received continuous research attention in tourism with work as early as Fritz, Brandon and Xander (1984), and with advanced contributions such as Li, Song and Witt (2006), Li, Wong, Song and Witt (2006), and Song, Gao and Lin (2013). As Li, Song and Witt (2005) and Song and Li (2008) identified in two comprehensive literature reviews, 451 studies on tourism demand modeling and forecasting were published during the period 1960-2008. The hospitality literature has traditionally paid little attention to forecasting in hotel revenue management with the exception of

Law (1998), Weatherford, Kimes and Scott (2001), Cranage (2003), Law (2004), Lim, Chang and McAleer (2009), El Gayar et al. (2011), Yang, Pan and Song (2014), and Kouprouchina et al. (2014). In the operations research literature a stream of forecasting applications in hotels can be observed with work from Rajopadhye et al. (2001), Baker, Murthy and Jayaraman (2002), Brännäs, Hellström and Nordström (2002), Weatherford and Kimes (2003), Aghazadeh (2007), Chen and Kachani (2007), Yüksel (2007), Bermúdez, Corberán-Vallet and Vercher (2009), Guadix, Cortés, Onieva and Muñuzuri (2010), Haensel and Koole (2011a), Zakhary, Atiya, El-Shishiny and Gayar (2011) and Lee (2012).

Hotel RMS traditionally assume that demand for each rate class is distinct and independent of the alternative options hotel guests have when booking a room. To challenge this common assumption and to incorporate other important buying behavior aspects, customer choice models have been proposed in the revenue management literature (e.g. Talluri & van Ryzin, 2004; Liu & van Ryzin, 2008; Erdely & Topaloglu, 2010; Aydin, Birbil, Frenk & Noyan, 2012; Meissner & Strauss, 2012; Sierag, Koole, Van der Mei, Van der Rest & Zwart, 2015). When customer choice behavior is incorporated, data analysis research will be especially important as in order to apply customer choice modeling to hotel revenue management practice successfully the appropriate choice (and estimation) of model parameters is crucial (e.g. van Ryzin, 2005; van Ryzin & Vulcano, 2013; Newman, Garrow, Ferguson & Jacobs, 2014). Moreover, Bodea, Ferguson and Garrow (2009, p. 356) criticize the literature as "the measurement of revenue benefits associated with choice-based RM has been based primarily on simulated data." They argue that there is a need to test these models on real data sets to see if the customer choice concept really works as:

> *"choice-based systems are not simply an incremental improvement or "add-on" to existing product-based systems, but are fundamentally different. Consequently, successful implementation of these systems will require a company to invest significant resources in*

developing new data collection procedures, RM algorithms, and user support systems" (Bodea et al. 2009, p. 357).

Bodea et al. (2009) describe the laborious process of data collection and validation in order to provide a data set that could be used to benchmark the choice-based models proposed in the revenue management literature. They developed a data set based on five hotel properties and discuss its potential uses including 'proofing of concepts' and 'benchmarking'. Their study shows how crucial data collection is especially as a precursor to demand and forecasting model development. Studies focusing exclusively on data *analysis*, such as graphing data (for visual inspection), computing statistics (for relationships), decomposition analysis (for trends, unusual or extreme data points), however, are virtually nonexistent. This is an important omission as exploratory analysis is key to the selection of the class of quantitative models (Makridakis et al., 1998). Moreover, the academic literature on forecasting in hotel revenue management – with an inclination for modeling – makes many assumptions about the properties and nature of data, but which often are not supported by preliminary empirical research.

3. Case Description

Five years of data (2008-2012) was collected from a small and independent hotel. The utilization of such data is of theoretical and practical importance as little is known about revenue management in this type of hotel, which makes up the majority of all hotel properties in Europe (Luciani, 1999; Holverson & Revaz, 2006). Moreover, small and independent hotels generally do not employ a revenue manager who interacts with the RMS (Lee Ross & Johns, 1997). This is an important criterion as the data was thus not limited by endogenous system effects.

The hotel is located in the countryside in The Netherlands and attracts business as well as leisure clients. As Table 1 illustrates, the hotel has 34 rooms which are divided into six room types each with a typical price.

Table 1: Overview of Room Types and Prices

Room type	Abbreviation	# Rooms	Typical price
Standard	STD	8	119
Garden view	GV	8	127
Large garden view	LGV	6	134
Old	STO	6	103
Private garden	PG	5	140
Bridal Suite	BRD	1	140

All rooms have a maximal capacity of two persons. The hotel has other facilities such as conference rooms and a restaurant. The restaurant not only serves hotel guests but also locals and tourists from the area.

Collecting the data was a lengthy process. Interaction with the hotel owner, the property system vendor, and two RM experts were needed to ensure data integrity. The data set had the following structure. Each data entry was a reservation for one hotel room. As a result, group bookings were recorded as separate reservations and further examination was required to identify which reservations were part of group bookings. Within each reservation several characteristics were recorded. First of all, the arrival date and the departure date of the booking was recorded, along with the check-in time and check-out time once the guest had stayed in the hotel. Also, the day and time of the booking were recorded. This characteristic proved to be essential for the data-analysis. If the reservation was cancelled, the cancellation date was recorded. The room type for which the reservation was made was present. The hotel regularly upgraded guests for free if a better room was available, but this was not recorded. The price that was paid for the reservation (room only) was recorded. The number of occupants of the room was also recorded, and it was even specified how many adults and children the room was booked for. Finally, the travel purpose (business or leisure), the name of the guest and if applicable the company name were present. A sample of the data set is presented in Table 2.

Table 2: Overview of Data Set Properties

Booking		**Arrival**		**Departure**		**Segment**	**Cancellation**	**Room**		**Occupancy**	
Time	Date	Date	Time	Date	Time		Date	Type	Price	adults	children
03:13	2007-04-04	2008-02-14	14:00	2008-02-15	11:00	Business	2008-02-06	PG	140	1	1

The following statistics were computed per room type (STD, GV, LGV, STO, PG, BRD) and for all data (TOTAL): (a) total number of reservations, (b) average occupancy, (c) average number of reservations, (d) average price that was paid for a room for one night, (e) percentage of nights that the hotel or room type was fully occupied, (f) percentage of rooms that was sold to groups, (g) average number of days before arrival, (g) average length of stay, (h) percentage of guests that stayed more than one night, and (i) the percentage of bookings that was cancelled.

Table 3: Example of Key Statistics

	Total March 2008
Total # reservations	443
Average occupation	1.38
Average # reservations	14.29
Average price	€121.74
% maximal occupancy	3.23%
% groups	65.99%
# days before arrival	46.03
LOS	1.48
% LOS > 1	28.91%
Total Revenu	€53,152
% cancellations	29.67%

Table 3 provides an example of these statistics (TOTAL, one month). These five-year statistics were determined for each month of the year, to capture average changes during the year, and per year, to capture changes from year to year. The hotel suffered from the recent economic crisis. Whereas the total number of reservations was 6,747in 2008, in 2012 this was reduced by 26.6% to 4,952. Total revenue reduced from €840,858.57 to €644,919.20.

4. Exploring the Data: insight into seasonality

An important aspect of demand is seasonality; the recurring pattern of demand across the year, week, or even during the day. In this section seasonality is analyzed on these three different levels, with promising results.

4.1 Annual Seasonality

Changes in demand were first explored at the annual level. To compensate for seasonality within a week, the demand of different weekdays were aggregated in a week. Figures 1, 2 and 3 show the average annual demand pattern for all guests, for the leisure guests, and for business guests respectively.

Figure 1: Yearly Seasonality – Total

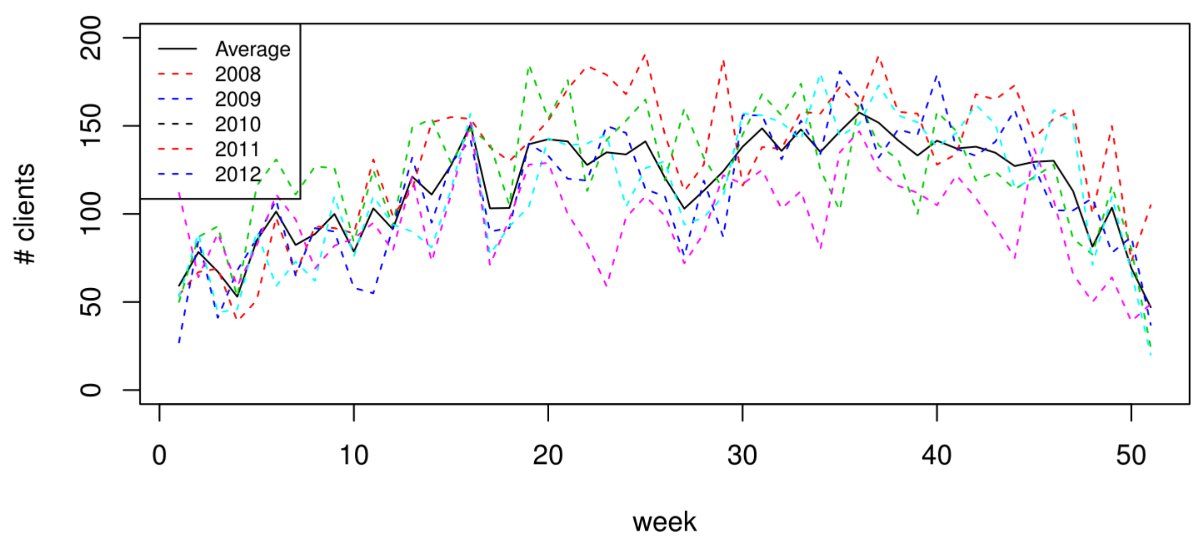


Figure 2: Yearly Seasonality – Business

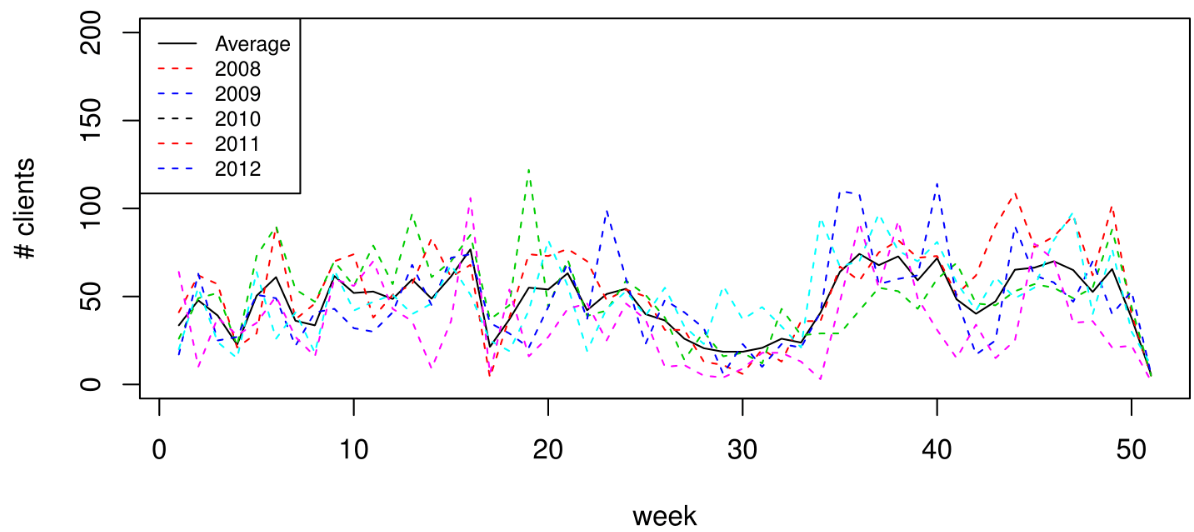
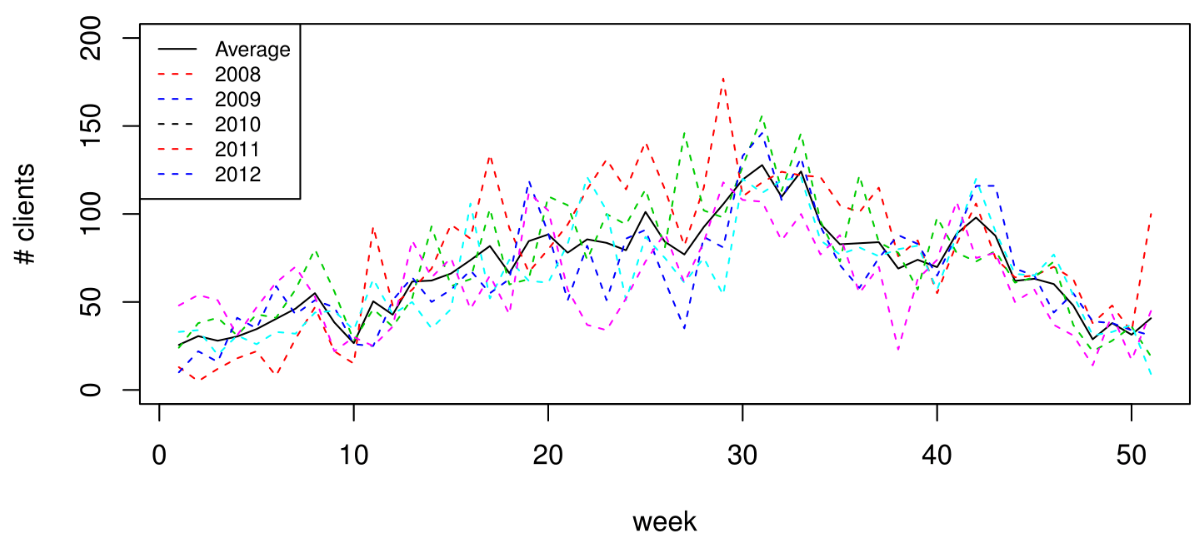


Figure 3: Yearly Seasonality – Leisure



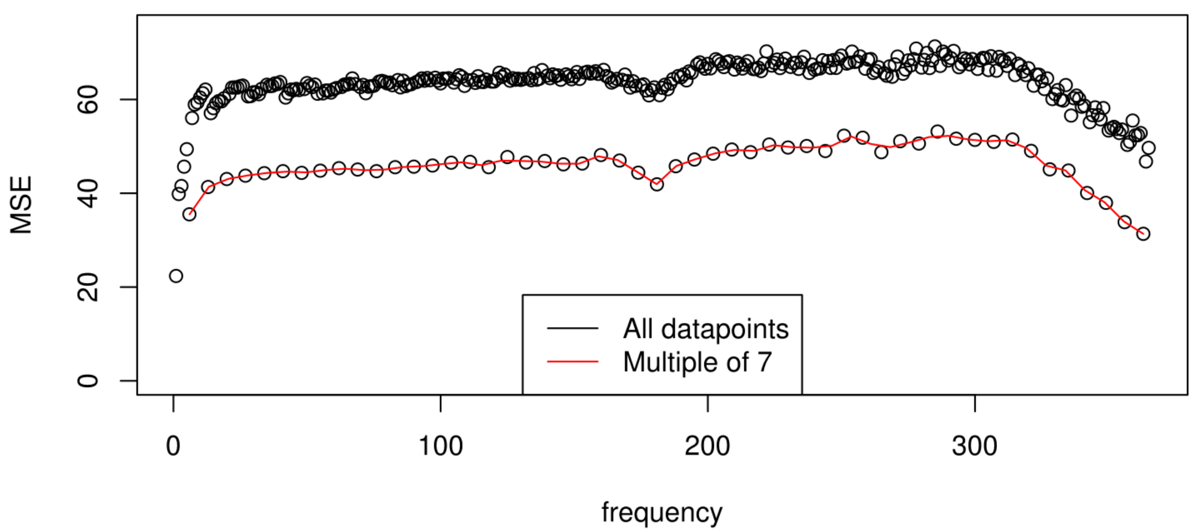
As Figure 1 illustrates from January to September total demand increased and from September to January total demand decreased. Figure 2 shows that the demand from business guests was quite stable during the whole year, except for a gap in July and August. Figure 3 shows that the demand from leisure guests was low in winter, and steadily rose until a peak in

July and August, in line with the Dutch summer holiday season. To examine whether leisure and business demand significantly differed a two-sided Kolmogorov-Smirnov test was applied to the time-series of 2008-2012. The null-hypothesis, stating that leisure and business demand were drawn from the same probability distribution, was rejected (D=0.2449, $p < 0.001$). The annual seasonality of leisure and business thus differed significantly.

4.2 Weekly Seasonality

The hotel manager claimed to observe demand similarities at the week level. This phenomenon is not uncommon in hotel revenue management practice. Using seasonal-trend decomposition analysis the presence of weekly seasonality was verified. Decompositions were calculated with a frequency varying between 1 (no seasonality) and 366 (a whole year). Then the corresponding mean squared errors (MSE) were compared. The results are presented in Figure 4.

Figure 4: Decomposition of Weekly Seasonality – Total



It can be observed that the MSE's for decompositions with frequency equal to a multiple of 7 are lower. This suggests that the observation of a weekly seasonality indeed is valid. Note that the

MSE for values lower than 7 are also low, but since multiples of these frequencies have high MSEs they are not true seasonality frequencies.

Figure 5: Average Number of Guests per Week – Total, Leisure, Business

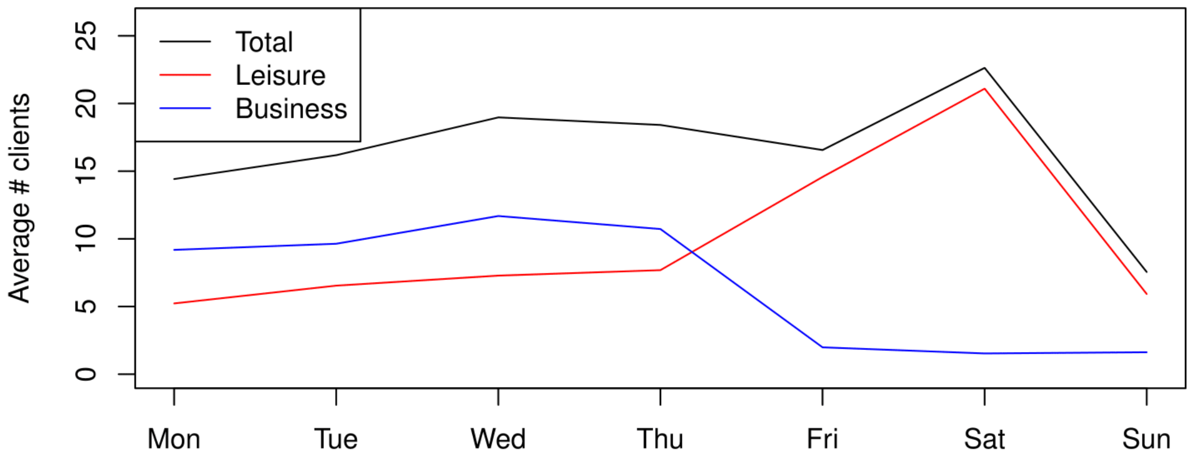


Figure 5 presents the average number of hotel guests per day of the week. To explore behavioral differences between business and leisure guests a distinction was made at the total, business and leisure level. A pattern was observed where leisure guests more frequently booked for Friday and Saturday, and business guests for Monday, Tuesday, Wednesday and Thursday, with a low occupancy on Sunday. Another observation was that the occupancy in the weekend was higher than on weekdays. This did not imply, however, that the hotel served more leisure than business guests.

4.3 Daily Seasonality

A crucial observation was made about the booking behavior at daily level. As Figures 6, 7, 8 and 9 illustrate, customer booking behavior depended on both the weekday *and the time of the day* on which an advanced booking was made.

Figure 6: Hourly Demand per Weekday – Total, Leisure, Business

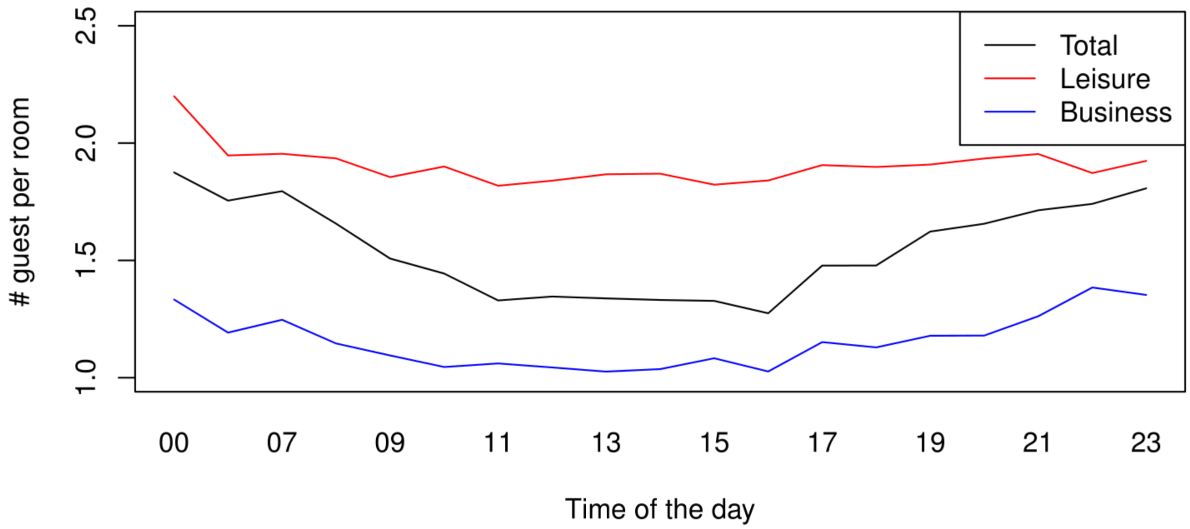
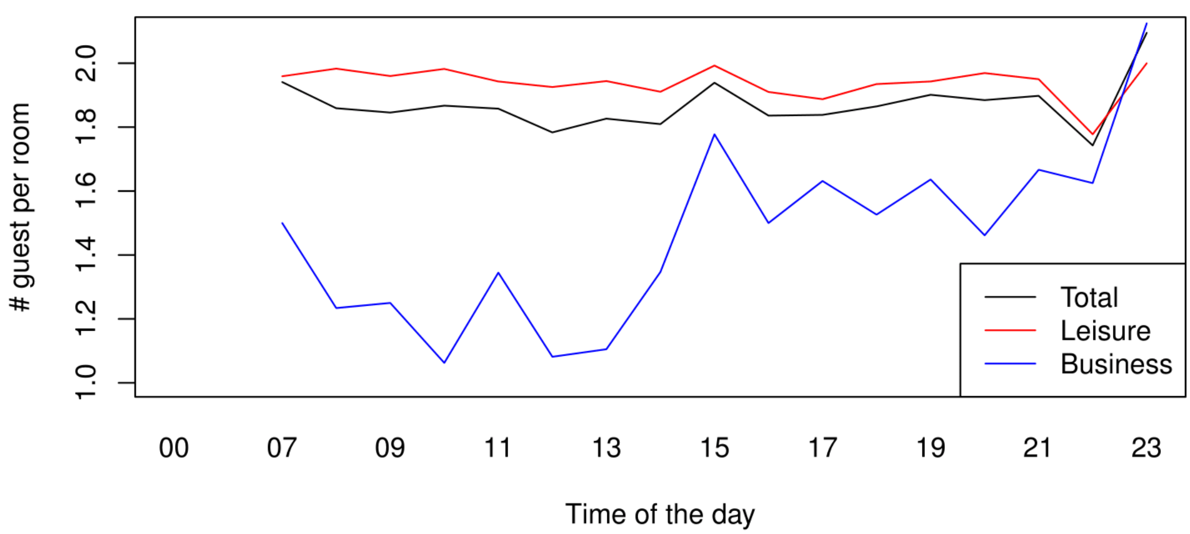


Figure 7: Hourly Demand per Weekend Day – Total, Leisure, Business



To examine whether the booking behavior of business and leisure guests significantly differed a Mann–Whitney U test was performed. Reservations that were made before 5pm on weekdays were more likely to have a higher price (M*dn* = 120.50, M = 132.97) than reservations made in the weekend and on weekdays after 5pm (M*dn* = 109.90, M = 120.01), U = 57031222, z = -21.505, $p < .000$, r = -.13. A chi-squared test confirmed a significant association between reservation moment and occupancy, $\chi^2(5, N = 25704) = 2497.756$, $p < .000$.

Figure 8: Hourly Price per Weekday – Total, Leisure, Business

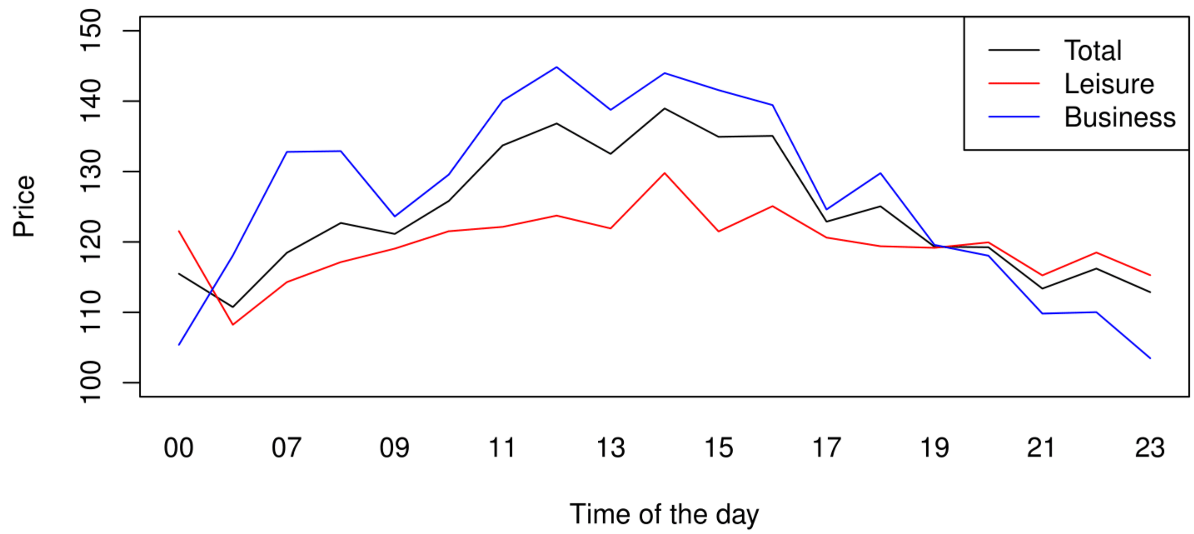
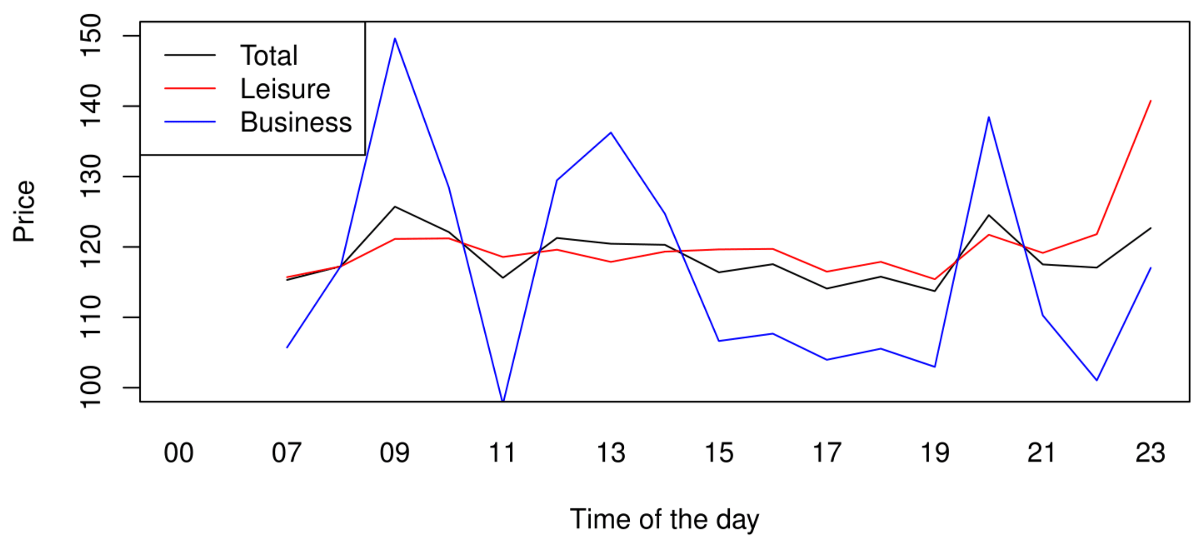


Figure 9: Hourly Price per Weekend Day – Total, Leisure, Business



This seems to represent the fact that before 5pm on weekdays rooms were 4.12 times more likely (based on the odds ratio) to be occupied by one person; in the weekend and on weekdays after 5pm it was more likely to be two persons or more. A second chi-squared test confirmed a significant association between reservation moment and segment (business/leisure), $\chi^2(1, N = 25704) = 1948.420$, $p < .000$. Before 5pm on weekdays rooms were 3.49 times more likely (based on the odds ratio) to be occupied by a business guest; in the weekend and on weekdays after 5pm it was more likely to be a leisure guest. The findings thus indicated that business guests, who tended to make purchases during working hours, were willing to pay a higher price than leisure travelers, commonly with an occupancy of more than one person per room, who tended to make purchases outside working hours.

Figure 10: Hourly Demand per Weekday – Total, Leisure, Business

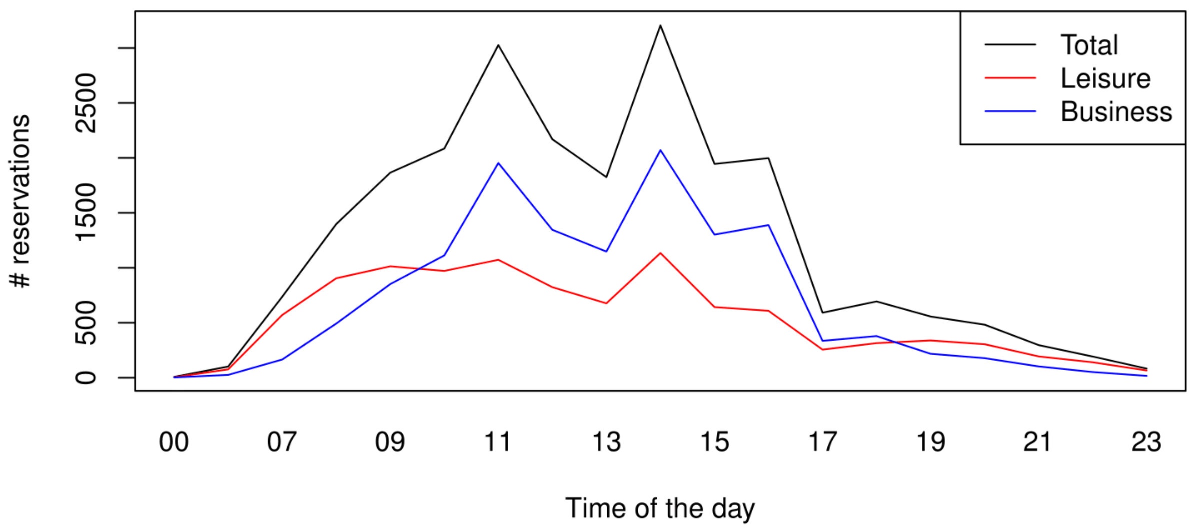
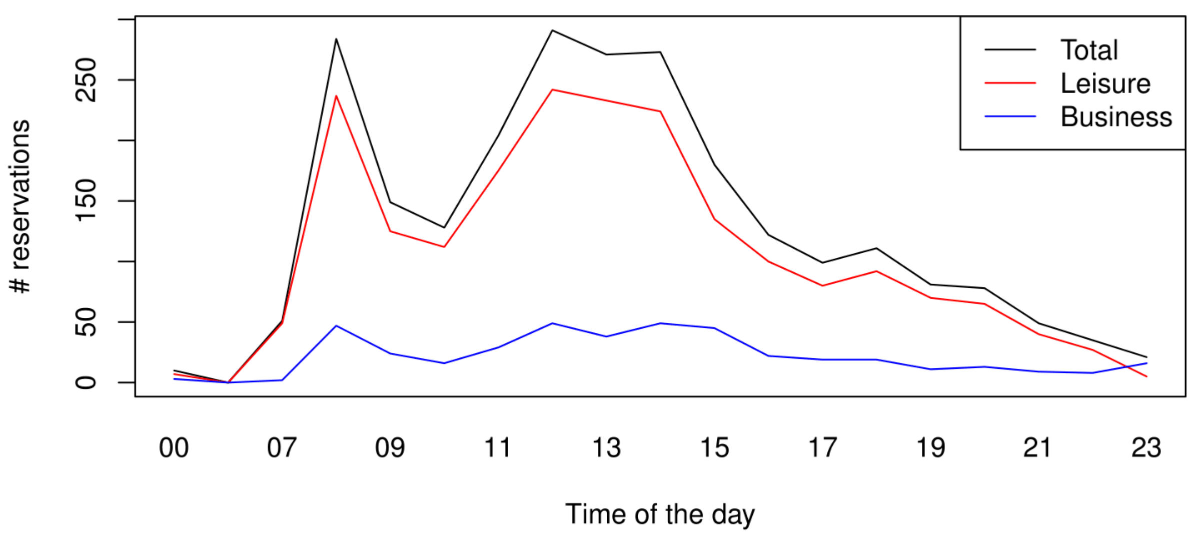


Figure 11: Hourly Demand per Weekend Day – Total, Leisure, Business



Figures 10 and 11 present business and leisure demand on an hourly basis. Purchases made in the weekend were dominated by leisure guests. Purchases made during the week consisted of a mix of business and leisure.

Table 4: Booking Statistics per Day and Hour

	Volume	**Leisure**	**Business**	**Guests**	**Price**
Total	25704	47.22%	52.78%	1.46	€ 129.15
Weekend *or* from 5pm.	7581	68.47%	31.53%	1.71	€ 120.01
Weekdays before 5pm.	18123	38.33%	61.67%	1.36	€ 132.97

Table 4 shows that more than one third of all reservations were made in the weekend or after 5pm. on weekdays. The average price of those reservations was lower (M=120.01) and the average number of guests per room was higher (M=1.71) than the reservations that were made during weekdays before 5pm. (M=132.97 resp. M=1.36). This suggests that the hotel can take advantage of the two discrete segments by dynamically changing prices both in the weekends *and* during the day, instead of maintaining the one-price policy per room regardless of day and time of the day, a practice that is commonly observed in small and independent hotels (See Table 1).

The revenue management forecasting literature does not take into account that demand can vary at certain hours of the day. Whereas it is complicated to develop tractable solution methods and accurate parameter estimation methods that perform well on computation time, taking a relevant model extension (based on exploratory data analysis) into account can have substantial impact on revenue, as is reported in literature. For example, the seminal model by Talluri and van Ryzin (2004b) increased revenue up to 12% compared to 1-2% differences in other literature at the time, by incorporating customer choice-behavior; and, Sierag et al. (2015) showed that by incorporating cancellations into the Talluri and van Ryzin (2004b) model, a substantial (additional) impact on revenue, up till 20%, could be achieved. The following theoretical proposition is, therefore, formulated:

> ***Proposition 1:*** there is room for optimization by bringing the revenue management strategy in line with demand for a specific month and day, *as well as* the observed booking behavior at the point of time during the day.

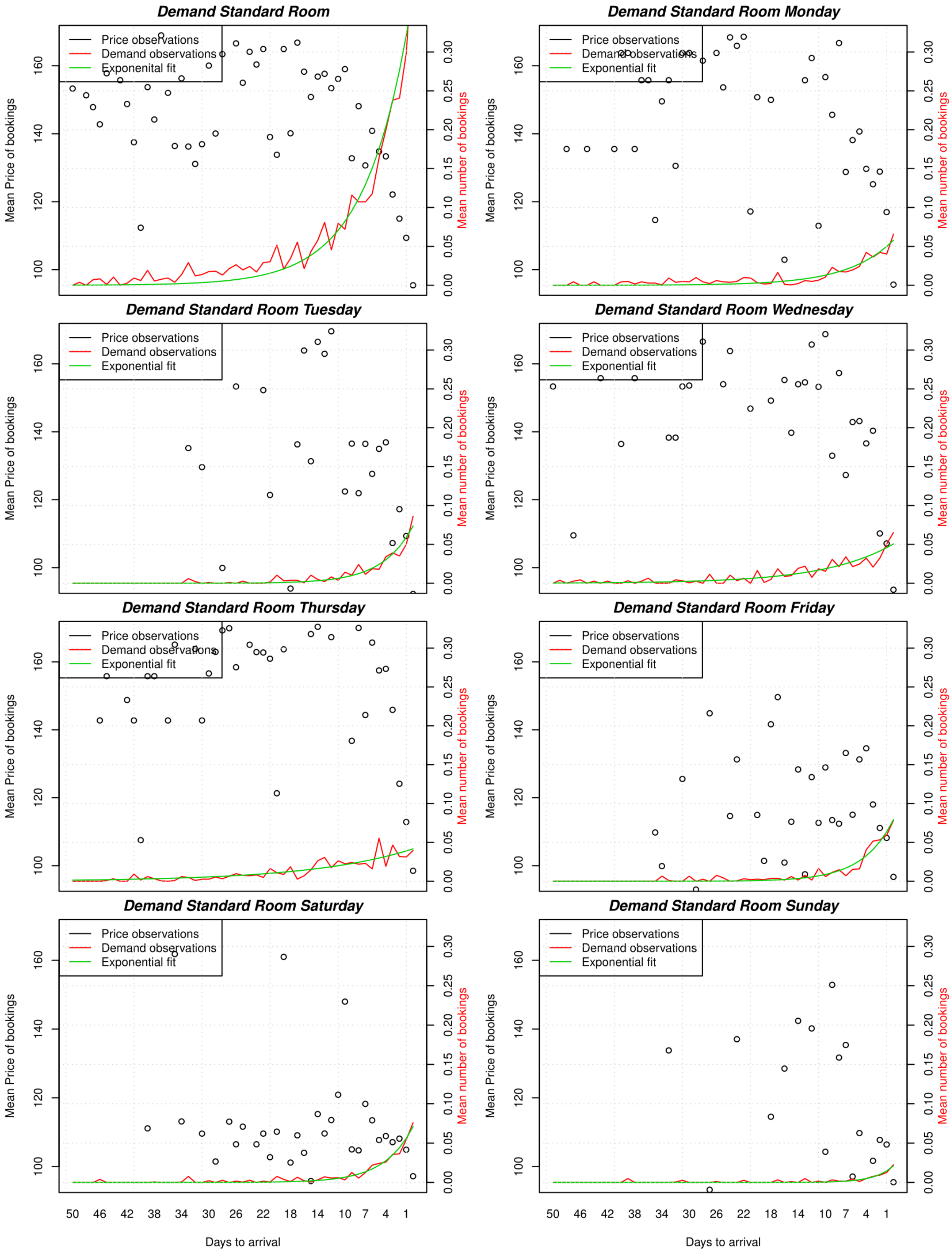
5. Exploring the Data: insight into group cancellation behavior

Hotels are vulnerable to demand and cancellation uncertainty (Chen, Schwartz & Vargas 2011). This can lead to sudden increases and decreases in pickup which is why hotel revenue managers – within the context of regular booking patterns – tend to closely monitor the booking pace at both total and segmented level. This section analyzes uncertainty in demand and cancellations.

5.1 Booking Pace

Figure 12 shows the relationship between demand and time until arrival for a standard room at total and weekday level. It was found that demand increased as the time until arrival decreased. Moreover, the majority of guests tended to plan not too far ahead. To test if demand increased exponentially as the time until arrival decreased, as the visual inspection suggested, an ordinary least squares regression was performed on the log of the mean number of bookings against the weeks before arrival. The results of the regression showed that the number of weeks before arrival significantly predicted mean number of bookings, β= -.908, $t(38) = 13.32$, $p < .001$. The number of weeks before arrival also explained a significant proportion of variance in mean number of bookings, $R^2 = .824$, $F(1, 38) = 177.42$, $p < .001$.

Figure 12: Advanced Bookings and Price of Standard Room – Total



Moreover, as price behavior was captured on the secondary axis, it was observed that during the last three months of the booking horizon average price decreased as the day of arrival came closer. The hotel thus dropped prices as the booking window shortened. To identify whether the increase in the number of bookings was also affected by a drop in price, with respect to the three month booking window, a multiple regression (with time and price as the predictors) was performed. The results show that the number of days before arrival (β= -.724, $t(88) = -10.94$, $p < .000$) and price (β= -.226, $t(88) = -3.42$, $p < .001$) both significantly predicted the mean number of bookings. Days before arrival and price also explained a significant proportion of variance in the mean number of bookings, $R^2 = .804$, $F(2, 88) = 180.1$, $p < .000$. Therefore, in addition to the shortening booking window, the drop in price affected demand.

The booking pace was different per weekday, suggesting that forecasting and pricing models should take this behavior into account. Arrivals on Monday through Thursday showed similar behavior patterns as well as the arrivals on Friday and Saturday. The arrivals on Sunday behaved differently. The curve for Monday through Thursday was more flat compared to Friday and Saturday. This implied that these reservations were made earlier in the booking horizon (M=37.06). On the other hand, reservations with arrival on Friday and Saturday tended to book closer to the day of arrival (M=29.10). For Sunday this was even closer (M=27.90).

5.2 Cancellations

Over five years of data on average about 21.71% of all reservations were cancelled. The number of cancellations varied per year and per month.

Table 5: Cancellation per Year per Month

Year	Month												Total
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	
2008	42.38%	26.69%	29.67%	33.99%	26.06%	23.13%	8.59%	13.62%	19.78%	32.85%	32.46%	28.47%	26.94%
2009	14.61%	18.22%	30.78%	32.46%	16.10%	38.41%	17.51%	23.01%	34.58%	18.90%	27.03%	13.86%	25.33%
2010	32.60%	16.79%	17.34%	15.33%	16.42%	22.32%	15.20%	18.89%	30.77%	16.35%	19.08%	15.29%	19.94%
2011	24.64%	14.23%	18.68%	6.91%	18.94%	12.59%	5.28%	9.50%	20.65%	10.43%	10.96%	16.60%	13.98%
2012	20.89%	22.88%	9.62%	29.89%	19.21%	31.82%	7.81%	13.66%	19.92%	16.42%	18.47%	20.50%	20.00%
Total	27.30%	20.02%	22.47%	25.94%	19.67%	26.51%	11.28%	16.21%	25.33%	19.93%	22.40%	19.61%	21.71%

As Table 5 illustrates, a higher cancellation rate was observed in 2008 (26.94%) than in 2011 (13.98%). Cancellation rates also varied per month. For example, in January the cancellation rate varied from 14.61% in 2009 to 42.38% in 2008.

Table 6: Cancellations and the Booking Window

Cancellations	Days Before Arrival				
	0-2	3-9	10-29	30-84	"85+"
2008	9,31%	11,96%	16,53%	30,40%	50,55%
2009	10,50%	8,29%	16,24%	28,42%	54,68%
2010	4,73%	10,13%	16,08%	31,66%	40,74%
2011	4,01%	7,73%	18,40%	24,35%	15,96%
2012	5,05%	6,13%	19,24%	34,66%	37,94%
Total	6,69%	8,90%	17,22%	29,75%	44,40%

Table 6 shows that overall, cancellation rates lowered when the booking window shrank. For example, 44.40% of all reservations that were made at least 85 days before arrival were cancelled eventually. Up to two days this was 6.69%. These rates varied per year. For example, 15.96% of the reservations were canceled in 2011 whereas this was 54.68% in 2009. For individual days this variation was even higher as only monthly averages were considered. Also, it was observed that 55% of the hotel's occupancy came from group bookings. These bookings showed higher cancellation rates.

Table 7: Cancellation Behavior per Segment

	Cancellations			Number of bookings		
	Transient	Group	Total	Transient	Group	Total
Business	8.26%	36.79%	30.78%	2857	10709	13566
Leisure	5.41%	26.72%	11.58%	8627	3511	12138
Total	6.12%	34.30%	21.71%	11484	14220	25704

As Table 7 presents, on average 34.30% of the group bookings were cancelled, as compared to 21.71% of all bookings and 6.12% for transient bookings. A chi-squared test confirmed a significant difference between group and transient, at the *total*, $\chi^2(1, N = 25704) = 2968.518$, $p < .000$, *leisure*, $\chi^2(1, N = 12138) = 1106.404$, $p < .000$, and *business* level, $\chi^2(1, N = 13566) = 861.630$, $p < .000$. Based on the odds ratio, it was found that groups were 8.00 times more likely to be cancelled than transient. For leisure and business groups this was respectively 6.37 times and 6.46 times more likely than transient. A similar pattern was observed for business versus leisure, with significant differences at *total*, $\chi^2(1, N = 25704) = 1390.407$, $p < .000$, *group*, $\chi^2(1, N = 14220) = 119.104$, $p < .000$, and *transient* level, $\chi^2(1, N = 11484) = 30.275$, $p < .000$. Using the odds ratio, it was found that business guests were 3.40 times more likely to cancel than leisure guests. Business groups were 1.60 times more likely to cancel than leisure groups. Business transient were 1.57 times more likely to cancel than leisure transient. Using Levene's test to identify differences in normalized variation between the segments ($p > .05$) it was found, with regard to group and transient cancellation behavior, that group business represents a relatively large proportion of the uncertainty in demand and cancellations.

Cancellations have received wide research attention in the hotel revenue management (e.g. Chen & Xie, 2013). The recent customer choice models in hotel revenue management forecasting literature, however, do not take cancellation into account, with the exception of Sierag et al.

(2015, p.3) who include cancellation but use three solution methods, each with different assumptions about cancellation, as the “problems are too large to solve exactly because of the curse of dimensionality”. Also, group cancellations (and lost/ turndown information) are not included in customer choice modelling. Group business, which – as one of four major areas – was identified “as having the greatest growth potential in hotel RM” (Milla & Shoemaker, 2008, p. 110), has properties that make modeling very complex. In addition, “transaction data, especially for the largest groups and smallest hotels, generally are sparse” (Hormby, Morrison, Dave, Meyers & Tenca, 2010, p. 49). Based on the exploratory data analysis the following theoretical proposition is, therefore, formulated:

> ***Proposition 2:*** there is room for a model extension in the (e.g., customer-choice based) forecasting literature by bringing the revenue management strategy in line with the more variable and statistically uncertain nature of group cancellations.

6. Exploratory Data Analysis: insight into demand uncertainty

The maximum capacity of the hotel was frequently reached in spring and autumn (for nearly all room types), but almost never in summer or winter. However, for an accurate forecasting and pricing model knowledge of the true unconstrained demand was necessary.

6.1 Probability distribution function

One of the most crucial assumptions in any revenue management model is the probability distribution function that demand follows. As was found in the analysis, on average the closer to the day of arrival, the more clients booked, but this finding did not reveal the nature of the demand distribution.

Revenue management literature generally assumes a (non-homogenous) Poisson distribution (e.g. McGill & van Ryzin, 1999; Bitran & Caldentey, 2003; Talluri & van Ryzin, 2004b; Sierag et al., 2014). That is, demand per time period is modeled as a homogeneous Poisson process. With the use of a likelihood ratio test as well as a chi-squared test, it was tested whether the data was Poisson distributed. All time periods (for the booking of a standard room) had *p*-values smaller than 0.001 so that the null hypothesis that the data was not Poisson distributed was rejected. Tests on the other six room types confirmed this finding. The finding that demand followed an non-homogeneous Poisson process was in line with earlier work by Haensel and Koole (2011b) who found that airline data was Poisson distributed. Assuming a Poisson distribution in forecasting modeling has the advantage of containing the Markov (memory-less) property (i.e. future demand does not depend on the guests who booked a room for the same arrival day in the past). This is in accordance with reality, since it is reasonable to assume that hotel guests arrive independent from each other.

6.2 Implication: logical inferences about hotel size

When demand follows a Poisson process different consequences can be inferred for smaller and larger hotels. Suppose demand is Poisson distributed with parameter λ, i.e. the expected number of guests who book a room. The standard deviation is then equal to $\sqrt{\lambda}$ such that the 95% confidence interval of the actual demand (D) is given by:

$$D = \in \left(\lambda - 2\sqrt{\lambda}, \lambda + 2\sqrt{\lambda}\right). \qquad (1)$$

The square root in formula (1) implies that the coefficient of variation decreases as λ increases, such that for smaller hotels the coefficient of variation in demand is higher than for large hotels.

Figure 13: Different consequences for smaller and larger hotels when demand follows a Poisson process

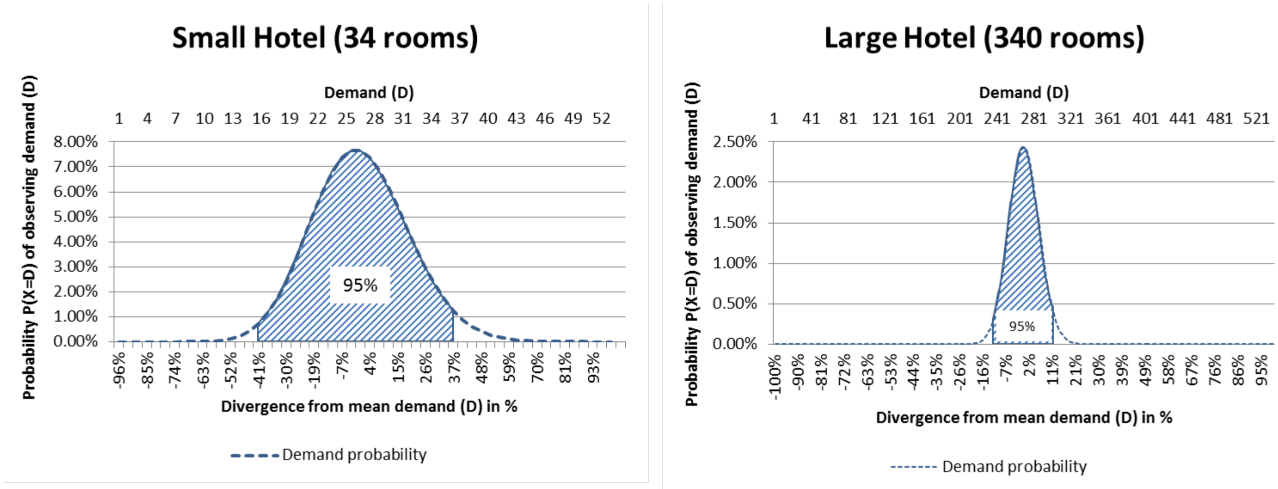


Figure 13 illustrates this size-implication inference for the case hotel (with 34 rooms) compared to a 10 times larger hotel (with 340 rooms). For illustrative purposes, market demand is assumed at 297 rooms from with each hotel gets its fair market share (27 respectively 270 rooms). Then, in 95% of the cases the larger hotel would have a demand between 237 and 302 rooms while the small case hotel would have a demand between 16 and 37 in 95% of the cases. In the worst case, for the small hotel this leads to 38% less demand than the average case, while in the worst case for the large hotel this leads to 12% less demand than the average case. The finding that demand is Poisson distributed thus implies that a small hotel is more vulnerable for demand uncertainty than a large hotel (cp.).

As (simple) forecasting and pricing models only consider average behavior, they provide an optimal strategy *on average*. However, when demand is volatile, as was found in this study, forecasts will be inaccurate, the errors being worse for smaller than larger hotels. To reduce forecasting error, forecasting and pricing models can take uncertainty into account, for example by considering the distribution of demand or by applying robust optimization techniques which take into account worst-case scenarios. This case study indicates that these methods are especially preferable for small hotels because demand variation is higher. Therefore, the following proposition is formulated:

Proposition 3: hotel demand follows a Poisson distribution. As a consequence, demand is more uncertain for smaller than for larger hotels.

7. Discussion

The purpose of this empirical study is to draw attention to the importance of preliminary and exploratory data analysis in hotel revenue management forecasting. Preliminary data analysis is key to the selection of the class of forecasting models, whereas exploratory data analysis is essential to evaluate whether a chosen model still is appropriate to capture changes that occur in the environment. As a whole, data analysis allows to determine whether a revenue management strategy is still optimal and to explore new opportunities for revenue optimization. In this context, the study identified three overlooked or ill-researched aspects of data analysis in hotel revenue management forecasting, each with different theoretical implications for demand modelling, forecasting and revenue optimization.

First of all, it provides empirical results on an inhomogeneous Poisson nature of the probability distribution function that demand follows. There is little evidence of this crucial and commonly assumed demand characteristic in the hotel forecasting literature, especially for small and independent hotels. This implies that especially for small hotels forecasting methods should be developed that take into account the uncertainty that comes with the Poisson distribution, for example by using robust optimization methods. Secondly, the study presents results on the random nature of group cancellations, an important but ill-researched segment in hotel revenue management. Optimization methods should take these cancellations into account. It is, however, unclear how such model would look like and therefore more research is needed. Thirdly, it finds that in a local market context business and leisure booking behavior significantly differ per point of time during the day. As the study shows, forecasting models that take this behavior into account can create a revenue increase. A further study that models this behavior could reveal the extent of this potential.

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. It presents three theoretical propositions that answer to Bodea et al. (2009) who call for more work in forecasting based on real-world data. In this context, as hotel revenue management forecasting can be perceived as 'a big data problem', the study also supports Xiang, Schwartz, Gerdes and Uysal (2015, p. 120) who observe that 'big data analytics approach in hospitality is yet to be well developed and established', and who reveal the potential of big data analytics to generate new insights.

The study is, however, not without limitations. One concern relates to the data which was collected from a single hotel. While forecasting research tends to rely on simulated data, and in this respect this exploratory study is a positive exception, its contribution is case-based. It thus is not possible to generalize the results; for *moderatum* generalizations (Williams, 2000, p. 215) additional studies first need to be performed to determine whether the findings of this study 'can be seen to be instances of a broader set of recognizable features'. As small and independent hotels generally collect incompletely and save poorly their data, it may be relatively time consuming to find hotels for these replication studies. Another limitation refers to the small and independent nature of the hotel and the specific local environment it operates in. The size and business mix in this study is specific to the context of the hotel. A comparative study, where hotels are grouped according to their location and business mix, could identify characteristics that can be generalized or are specific to a certain category of hotels. Such study is quite involved, since it requires the cooperation of a lot of hotels and the collection and cleansing of their data. A final limitation is that the study did not include competitive data. The effect of competitive prices on demand was not taken into account as sales and pricing data of the hotels in the competitive set were unavailable.

There are various implications for practice. Data analysis provides important insights in the booking and cancellation behavior of hotel guests. When analyzing at the segment level, data analysis can provide insights that are essential to maintain an optimal revenue management strategy and to explore new revenue opportunities. Data analysis also aids the process of evaluating the revenue management model as it tells how forecasting performs with respect to changes in demand. In this way, data analysis is vital for any hotel that seeks to stay competitive in a changing environment. In the case of the small and independent hotel cancellation was found to be much more severe than the hotel anticipated. Moreover, a daily booking pattern was identified. A rationale was thus provided for adjusting the revenue management strategy. Data analysis is however a laborious process (see also Bodea et al., 2009). Especially for small and independent hotels, such investment in time and analytical skills is often perceived as not worthwhile.

The study suggests three directions for future research. First of all, the findings indicate that there is room for an extension to the customer choice modeling literature in forecasting. An existing attempt of such extension is the research of Sierag et al. (2015) where the authors show that taking into account cancellation can impact revenue up till 20%. It would be interesting to examine whether their analysis holds for group bookings as well, and also whether their model can be extended to include differences in demand per point of time during the day. Secondly, the issue of variation in demand uncertainty as a result of differences in hotel size, and its implications for forecasting, can be further investigated. If demand and cancellations have a high variance, then conventional revenue management models are not appropriate. Empirical work could establish whether this variance indeed is higher for small hotels than for large hotels, as this exploratory study suggests. Finally, through systematical application of big data analytics techniques new sources of data could be analyzed to learn which customer behavior (such as day patterns) could be incorporated in forecasting modeling to further improve hotel revenue management performance.

References

- Aghazadeh, S-M. 2007. "Revenue Forecasting Models for Hotel Management." *Journal of Business Forecasting* 26(3): 33-37.
- Aydin, N., Ş. I. Birbil, J.B.D. Frenk and N. Noyan. 2012. "Single-leg airline revenue management with overbooking." *Transportation Science*, 47(4): 560-583.
- Baker, T., N. N. Murthy and V. Jayaraman. 2002. "Service Package Switching in Hotel Revenue Management Systems." *Decision Sciences* 33(1): 109-132.
- 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(1/2): 62-71.
- Bermúdez, J.D., A. Corberán-Vallet and E. Vercher. 2009. "Multivariate exponential smoothing: A Bayesian forecast approach based on simulation." *Mathematics and Computers in Simulation* 79(5): 1761-1769.

- Bitran, G. and Caldentey, R. 2003. “An overview of pricing models for revenue management.” *Manufacturing & Service Operations Management* 5(3): 203-229.
- Bodea, T., M. Ferguson and L. Garrow. 2009. “Data Set Choice-Based Revenue Management: Data from a Major Hotel Chain.” *Manufacturing & Service Operations Management* 11(2): 356-361.
- Brännäs, K., J. Hellström and J. Nordström. 2002. “A new approach to modelling and forecasting monthly guest nights in hotels.” *International Journal of Forecasting* 18(1): 19-30.
- Chen, C. and S. Kachani. 2007. “Forecasting and optimisation for hotel revenue management.” *Journal of Revenue and Pricing Management* 6(3): 163-174.
- Chen, C-C., Z. Schwartz and P. Vargas. 2011. “The search for the best deal: How hotel cancellation policies affect the search and booking decisions of deal-seeking customers.” *International Journal of Hospitality Management* 30(1): 129-135.
- Chen, C-C and K. L. Xie. 2013. “Differentiation of cancellation policies in the U.S. hotel industry.” *International Journal of Hospitality Management* 34(September): 66-72.
- Cranage, D. 2003. “Practical time series forecasting for the hospitality manager.” *International Journal of Contemporary Hospitality Management* 15(2): 86-93.
- El Gayar, N.F., M. Saleh, A. Atiya, H. El-Shishiny, A. Zakhary and H. Habib. 2011. “An integrated framework for advanced hotel revenue management.” *International Journal of Contemporary Hospitality Management* 23(1): 84-98.
- Erdelyi, A. and H. Topaloglu. 2010. “A dynamic programming decomposition method for making overbooking decisions over an airline network.” *INFORMS Journal on Computing* 22(3): 443-456.

- Ferguson, M., and S. Smith. 2014. “The changing landscape of hotel revenue management and the role of the hotel revenue manager.” *Journal of Revenue and Pricing Management* 13(3): 224-232.
- Fildes, R., K. Nikolopoulos, S. F. Crone and A. A. Syntetos. 2008. “Forecasting and operational research: a review.” *Journal of Operational Research Society* 59(9): 1150-1172.
- Fildes, R., P. Goodwin, M. Lawrence, and K. Nikolopoulos. 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.
- Fritz, R. G., C. Brandon and J. Xander. 1984. “Combining time-series and econometric forecast of tourism activity.” *Annals of Tourism Research* 11(2): 219-229.
- Guadix, J., P. Cortés, L. Onieva and J. Muñuzuri. 2010. “Technology revenue management system for customer groups in hotels.” *Journal of Business Research* 63(5): 519-527.
- Haensel, A. and G. Koole. 2011a. “Booking horizon forecasting with dynamic updating: A case study of hotel reservation data.” *International Journal of Forecasting* 27(3): 942-960.
- Haensel, A. and G. Koole. 2011b. “Estimating unconstrained demand rate functions using customer choice-sets.” *Journal of Revenue and Pricing Management* 10(5): 438-454.
- Holverson, S. and F. Revaz. 2006. “Perceptions of European independent hoteliers: hard and soft branding choices.” *International Journal of Contemporary Hospitality Management* 18(5): 398-413.
- Hormby, S., J. Morrison, P. Dave, M. Meyers and T. Tenca. 2010. “Marriott International Increases Revenue by Implementing a Group Pricing Optimizer.” *Interfaces* 40(1): 47-57.
- Kimes, S.E. 1999. “Group forecasting accuracy in hotels.” *Journal of the Operational Research Society* 50(11): 1104-1110.

- Koupriouchina, L., J.I. Van der Rest, and Z. Schwartz. (2014). “On Revenue Management and the Use of Occupancy Forecasting Error Measures.” *International Journal of Hospitality Management* 41(August): 104-114.
- 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 and Tourism Marketing* 16(2-3): 71-77.
- 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 Annual Meeting* (in Informs, Vol. 3), San Francisco, CA, November 17, 2012.
- Lee-Ross, D. and N. Johns. 1997. “Yield management in hospitality SMEs.” International Journal of Contemporary Hospitality Management 9(2): 66-69.
- Lemke, C., S. Riedel, and B. Gabrys. 2013. “Evolving forecast combination structures for airline revenue management.” *Journal of Revenue and Pricing Management* 12(3): 221-234.
- Li, G., H. Song and S. F. Witt. 2005. “Recent developments in econometric modeling and forecasting.” *Journal of Travel Research*, 44(1): 82-99.
- Li, G., H. Song and S. F. Witt. 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., K. Wong, H. Song and S. F. Witt. 2006. “Tourism demand forecasting: a time varying parameter error correction model.” *Journal of Travel Research* 45(2), 175-185.
- Lim, C., C. Chang and M. McAleer. 2009. “Forecasting h(m)otel guest nights in New

Zealand." *International Journal of Hospitality Management* 28(2): 228-235.

- De Livera, A. M., Hyndman, R. J. and R.D. Snyder (2011). "Forecasting time series with complex seasonal patterns using exponential smoothing." *Journal of the American Statistical Association* 106(496): 1513-1527.
- Liu, Q. and G. van Ryzin. 2008. "On the choice-based linear programming model for network revenue management." *Manufacturing & Service Operations Management* 10(2): 288-310.
- Luciani, S. 1999. "Implementing yield management in small and medium sized hotels: an investigation of obstacles and success factors in Florence hotels." *International Journal of Hospitality Management* 18(2): 129-142.
- Makridakis, M. H. S., S. C. Wheelwright, and R. Hyndman, 1998. *Forecasting Methods and Applications*. 3rd ed. New York: Wiley.
- McGill, J.I. and van Ryzin, G.J. 1999. "Revenue management: Research overview and prospects." *Transportation Science* 33(2): 233-256.
- Meissner, J. and Strauss, A. 2012. "Network revenue management with inventory-sensitive bid prices and customer choice." *European Journal of Operational Research* 216(2): 459-468.
- Milla, S. and S. Shoemaker. 2008. "Three decades of revenue management: What's next?" *Journal of Revenue and Pricing Management* 7(1): 110-114.
- Newman, J. P., L. A. Garrow, M. E. Ferguson and T. L. Jacobs. 2014. "Estimation of choice-based models using sales data from a single firm." *Manufacturing & Service Operations Management* 16(2): 184-197.
- Phillips, R. L. 2005. *Pricing and Revenue Optimization*. Stanford: Stanford Business Books.
- Rajopadhye, M., M. Ben Ghalia, P. P. Wang, T. Baker, C. V. Eister. 2001. "Forecasting

uncertain hotel room demand." *Information Sciences* 132(1-4): 1-11.

- Schwartz, Z. 2003. "Subjective estimates of occupancy forecast uncertainty by hotel revenue managers." *Journal of Travel and Tourism Marketing* 16(4): 59-66.
- Schwartz, Z. and E. Cohen. 2004. "Hotel Revenue-management Forecasting." *Cornell Hotel and Restaurant Administration Quarterly* 45(1): 85-98.
- Schwartz, Z. and S. Hiemstra. 1997. "Improving the accuracy of hotel reservations forecasting: curves similarity approach." *Journal of Travel Research* 36(1): 3-14.
- Sierag, D.D., G.M. Koole, R.D. Van der Mei, J.I. Van der Rest and B. Zwart. (2015). "Revenue management under customer choice behaviour with cancellations and overbooking." *European Journal of Operational Research* 246(1): 170-185.
- Song, H., G. Li. 2008. "Tourism demand modelling and forecasting – a review of recent research." *Tourism Management* 29(2): 203-220.
- Song, H., Z. Gao and S. Lin. 2013. "Combining statistical and judgmental forecasts via a web-based tourism demand forecasting system." *International Journal of Forecasting* 29(2): 295-310.
- Talluri, K. T. and G. van Ryzin. 2004a. "*Theory and Practice of Revenue Management*." Springer, New York.
- Talluri, K. T. and G. van Ryzin. 2004b. "Revenue management under a general choice model of consumer behavior." *Management Science* 50(1): 15-33.
- van Ryzin, G. 2005. "Models of demand." *Journal of Revenue and Pricing Management* 4(2): 204–210.
- van Ryzin, G. and G. Vulcano. 2013. "An expectation-maximization method to estimate a rank-based choice model of demand." *Technical Note*, November 2014, Columbia Business

School, http://pages.stern.nyu.edu/~gvulcano/EMPrefListsRev1.pdf (accessed April 22, 2015).

- Weatherford, L. R. and S. E. Kimes, S.E. 2003. “A comparison of forecasting methods for hotel revenue management.” *International Journal of Forecasting* 19(3): 401-415.
- Weatherford, L. R., S. E. Kimes and D. A. Scott. 2001. “Forecasting for hotel revenue management: testing aggregation against disaggregation.” *Cornell Hotel and Restaurant Administration Quarterly* 42(4): 53-64.
- Williams, M. 2000. “Interpretivism and Generalization” *Sociology* 34(2): 209-224.
- Xiang, Z., Schwartz, Z., Gerdes, J.H. and Uysal, M. (2015). What can big data and text analytics tell us about hotel guest experience and satisfaction? *International Journal of Hospitality Management* 44(January): 120-130.
- Yang, Y., B. Pan and H. Song. (2014). “Predicting Hotel Demand Using Destination Marketing Organization’s Web Traffic Data.” *Journal of Travel Research* 53(4): 433-447.
- Yüksel, S. 2007. “An integrated forecasting approach to hotel demand.” *Mathematical and Computer Modelling* 46(7-8): 1063-1070.
- Zakhary, A., A. F. Atiya, H. El-Shishiny and N. E. El Gayar. 2011. “Forecasting hotel arrivals and occupancy using Monte Carlo simulation.” *Journal of Revenue and Pricing Management* 10(4): 344-366.