STUDY OF DEMAND MODELS AND PRICE OPTIMIZATION PERFORMANCE

A Thesis
Presented to
The Academic Faculty

by

Seonah Lee

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2011

STUDY OF DEMAND MODELS AND PRICE OPTIMIZATION PERFORMANCE

Approved by:

Professor Pinar Keskinocak, Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Laurie A. Garrow, Co-Advisor
School of Civil and Environmental Engineering
Georgia Institute of Technology

Professor Ellis L. Johnson
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Paul M. Griffin
Harold and Inge Marcus Department of Industrial and Manufacturing Engineering
Pennsylvania State University

Doctor Jon A. Higbie
Revenue Analytics, Inc.

Date Approved: 26 October 2011

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Accurately representing the price-demand relationship is critical for the success of a price optimization system. This research first uses booking data from 28 U.S. hotels to investigate the validity of two key assumptions in hotel revenue management. The assumptions are: 1) customers who book later are willing to pay higher rates than customers who book earlier; and, 2) demand is stronger during the week than on the weekend. Empirical results based on an analysis of booking curves, average paid rates, and occupancy rates for group, restricted retail, unrestricted retail, and negotiated demand segments challenge the validity of these assumptions. The combination of lower utilization rates and greater product differentiation suggests that hotels should apply different approaches than simply matching competitor rates to avoid losing market share. On days when inventory is near capacity, traditional yield management tactics deliver tremendous value, but these should be augmented by incorporating price response of demand and competition effects. On days when demand is soft and occupancy is projected to be low, price and competition based strategies should dominate.

The hotel price optimization problem with linear demand model is a quadratic programming problem with prices of products that utilize multiple staynight rooms as the decision variable. The optimal solution of the hotel price optimization problems has unique properties that enables us to develop an alternative optimization algorithm that does not require solving quadratic optimization problem. Using the well known least norm problem as a subroutine, the optimization problem can be solved as

finding a minimum distance between a polyhedron defined by non-negative demand and capacity constraints. This algorithm is efficient when only a few of the staynights are highly constrained.

In practice, the choice of a demand model is largely driven by the ease of estimation and model fit statistics such as R^2 and mean absolute percentage error (MAPE). These metrics provide measures of statistical validity of the model, however, they do not measure how well the price optimization will perform which is the ultimate interest of the practitioners. In order to measure the impact of demand models on price optimization performance, we first investigate the goodness of fit of linear demand models with different driver variables using actual data from 23 U.S. hotels representing multiple brands and location types. We find that hotels within the same location types (such as urban, suburban, airport) share similar driver variables. Airport and suburban hotels have simpler model specifications with less drivers compared to the urban hotels. The airport hotel demand models are different from other location hotels in that the airport hotel demand level does not differ by day of week.

We then measure the impact of demand model misrepresentation on the performance of price optimization through simulation experiments, which are performed for different levels of demand and forecast accuracy to represent various market environments that hotels operate in. We find that using models with missing driver variables can reduce the potential revenue by 13%∼53% and using the wrong functional form 5%∼43% under our simulation environment. The findings from our research imply that correctly representing the demand model in price optimization is crucial to its success. In order for hotels to realize the maximum potential revenue through pricing, efforts should be focused on identifying the major driver variables influencing demand including the ones that we found to be significant.

CHAPTER I

INTRODUCTION

The revenue management discipline has traditionally developed around managing inventory assigned to a limited number of classes with pre-defined set of prices (for example, see Belobaba [8] and Talluri and Van Ryzin [53] for description of traditional revenue management methods). This approach is reasonable when the market environment is relatively static and frequently updated market response pricing is not necessary. On the other hand, for businesses that operate under highly competitive market with dynamic competitive environment shifts, ability to optimize price in response of market change in short time frame is crucial. In the hotel industry, where the market is quickly evolving to the previously described condition, price optimization has gained much interest over the recent few years [3]. This study investigates the commonly used demand models and their impacts on price optimization.

Demand forecasting is one of the most critical components in hotel revenue management and especially in price optimization. Most popular price-dependent demand models are univariate linear and exponential models (see Weatherford and Kimes [59]). In practice, the choice of demand model formulation is largely driven by the ease of estimation and optimization along with model fit statistics, but not by the impact of the different model specification to optimization performance. In this study we investigate the goodness of fit for different demand models using the data gathered from hotels and measures the impact of misrepresenting the demand model on price optimization. Econometric models with drivers such as price, time of booking, day of week and annual seasonality and autoregressive variable are tested to determine the best fit demand model. Also, the impacts of demand model specification on price optimization performance are measured.

The primary goal of this thesis is first to develop a demand model with both good explanatory and predicting power and then test the impact of demand models on price optimization performance. In order to achieve this goal, data gathered from U.S. domestic hotels are used. In Chapter 2, the characteristics of hotel demand are examined and validated against major revenue management assumptions originating from the airline industry. In Chapter 3, price-demand model for hotel demand is developed using both cross-sectional and temporal explanatory variables. Some of the explanatory variables tested include month, day of week, location, and competitor prices. After developing a valid demand model, sensitivity of the optimal solution in price optimization to different specifications of demand model is tested through a simulation study. Chapter 4 reports the results of simulation experiments comparing the performance of demand models and pricing methods. Chapter 6 contains summary of this thesis and future research directions.

CHAPTER II

STUDY OF HOTEL DEMAND

2.1 Introduction

Within the airline industry, revenue management (RM) has a well-established track-record of increasing profitability and has played an integral role in strategic and tactical managerial decision making. Over time, the utilization of analytics has evolved from using simple descriptive analysis to manage inventory to solving complex optimization problems that automatically set rate availability and other inventory controls. Based on the initial success of RM within the airline industry, it was not long before other industries began to adopt these practices. There are numerous industries using RM and many others considering using RM, including airlines, hotels, car rentals, casinos, restaurants, grocery chains, golf courses, cruise lines, apartment rentals, sports, performing arts, media, etc. (e.g., Garrow and Ferguson [25], Garrow *et al* [26], Gu [28], Hawtin [29], Heching *et al* [30], Kimes [37], Kimes and Schruben [38], Kimes *et al* [39], Kuyumucu [40], Lieberman and Dieck [43], Lippman [45], Vinod [58]). While the growth in revenue management across industries is impressive, one may, nonetheless question the wisdom of applying RM techniques originally developed for the airline industry to other industries without consideration of unique market characteristics.

The objective of this chapter is to investigate whether fundamental assumptions related to customer demand patterns typically observed in the airline industry (that are critical for successful RM implementation) also hold for the U.S. retail hotel sector. Empirical results, based on a study of 28 U.S. hotels representing five different brands and complete booking histories for 420 arrival dates (60 weeks), suggest that hotel retail demand has significant differences from airline demand. Specifically, this study challenges two classic assumptions used for the majority of hotel RM applications, namely: 1) late booking customers are willing to pay higher rates than early booking customers; and, 2) weekday demand is higher than weekend demand. The hotel demand is classified into three segments for the analysis - group, negotiated and retail. Readers not familiar with these terminologies can refer to 2.2 for the detailed description of these segments.

The expectation that late booking customers tend to be willing to pay higher rates is shared across the airline, hotel, and car rental industries (e.g., Alstrup *et al* [5] and Belobaba [9] for airline applications; Baker and Collier [6], Ben Ghalia and Wang [10], Schwarz [51] for hotel applications; and Carroll and Grimes [15] for car rental applications).

In general, the second assumption also appears reasonable, i.e., that hotel demand will be stronger on weekdays versus weekends, particularly for business-oriented properties that comprise the majority of hotels for large hospitality enterprises (and that also forms the basis for this analysis) . This assumption commonly appears in hotel literature (Choi and Kimes [16], Jeffrey and Barden [32]) and has been validated by several empirical studies; Rushmore [50] empirically observed that the transient demand is weaker on weekends and Jeffrey *et al* [33] examined 15 years of hotel data from England. It is important to note, that these observations are on the total demand and not focused on the true retail demand, which is the only segment that will be sensitive to price changes. While the negotiated segment, mainly comprised of mid-week business customers, can be strong during the weekdays, the price for

Table 1: Anticipated relationships between demand assumptions and observable data

Assumption	Expected observations
1. Late booking customers are willing to pay higher rates than early booking customers	• Higher-valued, unrestricted booking classes book later than lower-valued, restricted booking classes • Average paid rates increase as the arrival date approaches
2. Weekday demand is higher than weekend demand	• Occupancy rates are higher during weekdays • Booking rates are higher during weekdays

this segment is fixed by a pre-determined contract. If the assumption of strong midweek demand is mainly based on the patterns from the negotiated demand (or total demand which is largely impacted by the negotiated demand) and not the pure retail demand, it might end up misleading important retail pricing decisions.

In order to investigate the validity of these assumptions, we undertook extensive statistical analysis on booking curves, average paid rates, and occupancy rates. While it is not possible to directly observe the validity of these two assumptions using only actual booking data, it is possible to observe whether the expected relationships between these assumptions and booking curves, prices, and occupancy rates hold. As shown in Table 1, one would expect to observe the following relationships if the assumptions were valid.

One factor not considered in Table 1 is the possibility that the availability of products skewing the resulting price paid by customers. For example, if the discounted products are only available close to arrival dates, the resulting price paid by customers may decrease as it gets close to the arrival date, regardless of willingness to pay of customers. In order to verify that this unusual scenario is playing out, the availability information of products throughout the booking horizon is needed, but this information is not available. Instead the booking data can be used as a proxy for the availability. If a product was booked at a certain booking date, than it is reasonable to assume that the product was available at that date, though the inverse is not necessarily true. On the aggregate level, it is observed throughout this study that the discounted products (in restricted retail segment) are sold throughout the entire booking horizon (for example, see Figure 1). Given that about 50% of the bookings for discounted retail products occur within 7 days prior to departure, the sales of discounted products prior to and after 7 days to departure can be compared as an indicator of discounted products being available closer to departure only. Out of all possible hotel/arrival combinations, only 3.6% was recorded to have no restricted retail product sales further out than 7 days and have sales within 7 days. Since this is a booking statistic and not availability, it can be inferred that there is even less cases where the restricted product is not available before 7 days to departure. If the cut off is defined at 14 days to departure rather than 7, the percentage increases to 7.9%, which is still too small to impact the aggregate level average rates paid trend.

From the booking patterns of the restricted products, it can be concluded that the availability of restricted retail products are relatively stable over the booking horizon and is not skewing the average price or booking trends in some part of the booking window.

Empirical results based on an analysis of the expected outcomes suggest both of these assumptions may not be valid for U.S. hotel retail customers. Consequently, new recommendations for how to apply RM to transient hotel customers and how to price weekday versus weekend rates are presented. To this extent, it is the authors' hope that this study will serve as a broader warning of applying model assumptions developed for the airline industry to other industries without considering the market context. An example of such case would be hotels without the structural product differentiation where there is less differentiation of products and most of the products are highly substitutable with each other. In this case the virtual bucket or class based revenue management algorithms will not work. Another case of mis-adopted revenue management would be the price discrimination experiment by Amazon.com (see Talluri and Van Ryzin [53] for details of the incident); the experiment to offer different prices to different customers created a significantly negative perception for Amazon.com, unlike in the airlines where it has become acceptable to the customers to explicitly price discriminate.

2.2 Data

The data for this analysis is based on 60 weeks of booking data from March 2006 to April 2007. The data set represents 28 different hotels in the U.S. that span five different brands ranging from limited to premium full service. The hotels used in this analysis ranges widely in their variety - 5 luxury, 8 premium full-service, 6 full-service business, 6 limited service and 3 extended stay hotels. These hotels are located in urban (12), suburban (11), airport (5) and highway (1) locations. Of these hotels, only one property is located in a purely leisure destination, others are either heavily business oriented or business-leisure mixed destinations. Competitive unrestricted retail rates are available from two to seven competitors associated with each hotel property. The competitor rate data was obtained through a private company (Rubicon's Market Vision product) that collects shopping data through various channels including GDS and Internet.

In compiling the data for this analysis, demand was classified according to macro-channel as well as the level of restrictions typically associated with the demand classes. Four distinct segments are used for the analysis. Group demand refers to bookings that are associated with an allocated block of rooms, as would be the case for a conference or a corporate event. Negotiated demand refers to bookings that are associated with a corporate customer or large booking agency. Rates for this segment generally do not vary, and are available only to corporate employees or customers that book through travel agencies. The final two segments used for the analysis fall under the category of retail demand. Retail demand refers to all demand that is not group or negotiated. Retail customers book through channels that are available to the general public. In general, retail demand can be classified as unrestricted retail demand or restricted retail demand. For this study, unrestricted retail demand refers to bookings that have no advance purchase requirements and no cancellation fee (in this analysis rates that require cancellations before 5 PM on the day of check-in are classified as having no cancellation fee). Restricted retail demand refers to bookings that have restrictions associated with them, specifically advance purchase requirements, cancellation fees, and/or customer qualifications (e.g., requires AARP or AAA membership). Unrestricted retail demand is generally considered to be more valuable, since in theory these customers would be willing to pay more to have liberal cancellation policies and book closer to arrival dates.

Considering the group and transient segments (the latter of which includes negotiated demand and retail demand), the only transient demand that RM systems are able to influence is retail demand (since negotiated demand rates are fixed).

2.3 Assumption 1: Are late booking customers willing to pay more?

Different statistical analysis can be used to investigate the assumption that late booking customers are willing to pay higher rates than early booking customers. Specifically, if the assumption is true, one would expect that higher-valued, unrestricted classes book later than lower-valued, restricted classes and that the average rate paid by customers increases as the arrival date approaches.

Table 2: Summary statistics of booking days for unrestricted and restricted retail segments

(a) Overall

Demand	Median	Mean	Std. Dev.
Unrestricted Retail	5	17.0	32.5
Restricted Retail	7	20.0	34.4

(b) Property averaged

Demand	Median	Mean	Std. Dev.
Unrestricted Retail	4.5	13.6	26.0
Restricted Retail	7.4	18.5	30.8

2.3.1 Comparison of Booking Profiles

If higher valued customers tend to book later than lower valued customers, one would expect to see the distribution of unrestricted bookings more concentrated towards the day of arrival relative to the distribution of restricted customers. However, this relationship is only weakly observed in the data. As shown in Table 2(a), the median booking days from arrival are also similar (5 vs. 7 days, respectively). On average, restricted retail bookings occur three days earlier than unrestricted retail bookings (17 vs. 20 days, respectively).

The property average statistics in Table 2(b) show a little more gap between the unrestricted and restricted retail segment, implying that the larger hotels have less distinction between the timing of bookings between the two segments.

In addition to comparing descriptive statistics for the unrestricted and restricted retail segments, one can examine their booking profiles. Figure 1(a), which portrays the log of bookings for the restricted and unrestricted retail segments by days prior to arrival indicates that statistically, there is no discernible difference between the slopes of the booking profiles for restricted and unrestricted retail segments. The slopes of linear approximations for log unrestricted and log restricted bookings are displayed

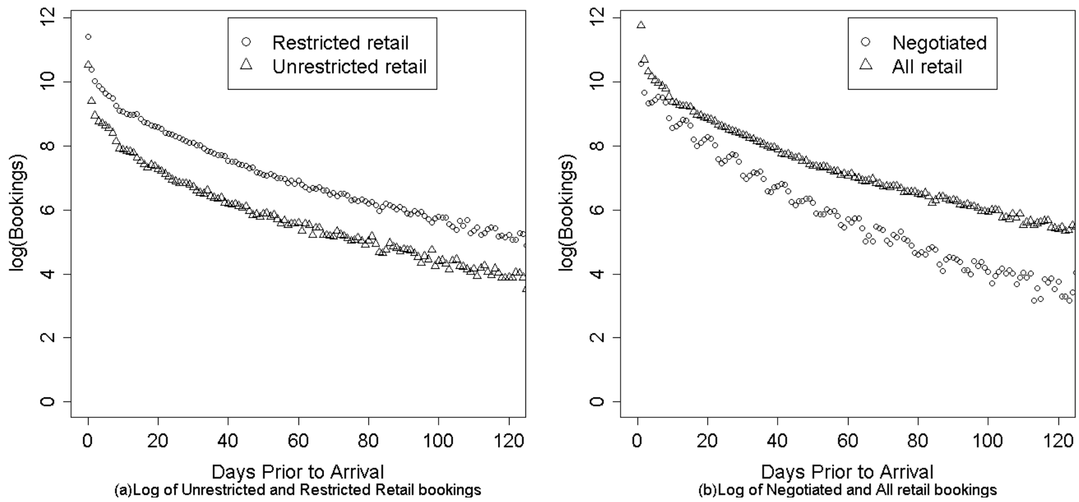


Figure 1: Log of arrivals for restricted and unrestricted retail segments

Table 3: Slope of linear approximation for log(unrestricted bookings) and log(restricted bookings)

Demand	# Obs.	Slope	Std. Err.	P value	R^2
Unrestricted Retail	348	-0.0194	3.9E-5	<< 0.001	0.88
Restricted Retail	365	-0.0191	3.2E-5	<< 0.001	0.68
Negotiated	328	-0.0237	5.2E-5	<< 0.001	0.87

in Table 3. A two sample t-test cannot reject the null hypothesis that the two slopes are the same (p=0.55). For comparison, the log of bookings for negotiated and all retail bookings are shown in Figure 1(b).

Finally, booking profiles can also be examined in terms of cumulative frequencies. If the assumption that higher valued customers tend to book later than lower valued customers is true, one would observe that the distribution for the unrestricted segment is more concentrated towards the arrival date than the restricted retail segment and that the cumulative distribution of unrestricted demand lies above the restricted retail frequency distribution. As seen in Figure 2, the cumulative distribution associated with the unrestricted retail bookings is slightly more concentrated towards arrival date. This is statistically confirmed via the Kolmogorov-Smirnov test (bootstrap

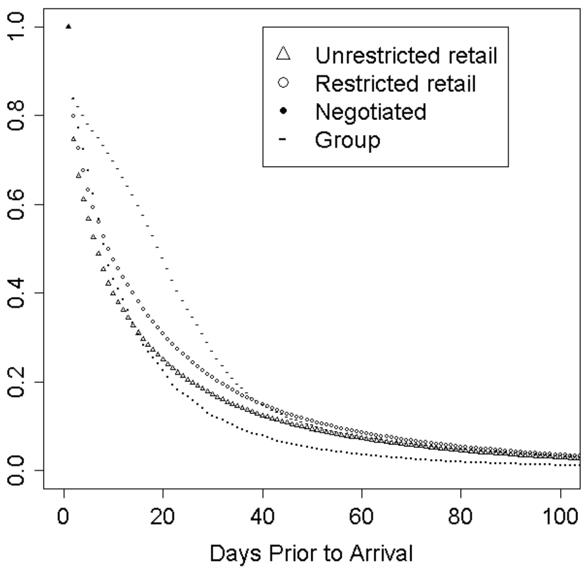


Figure 2: Cumulative frequency distributions by demand segment and days prior to arrival

version[1]) which rejects that null hypothesis that the unrestricted retail booking curve does not lie above the restricted booking curve (p=0.076).

To summarize, the statistical analysis provides only weak evidence that the distribution of unrestricted bookings is more concentrated towards the day of arrival relative to the distribution of restricted bookings. From a practical perspective, this implies that the assumption that higher value customers book closer to arrival date may not be entirely appropriate for hotel revenue management applications.

2.3.2 Comparison of Average Paid Rates

The assumption that higher valued customers tend to book later than lower valued customers can also be investigated by comparing average paid rates. Consistent with expectation, Figure 3 illustrates that unrestricted retail rates are on average 35.6%

[1]The Kolmogorov-Smirnov (K-S) test is not exact when the underlying distribution is discrete. Since the days prior distribution in nature is discrete, the bootstrap version of the K-S test was used (Abadie [4]).

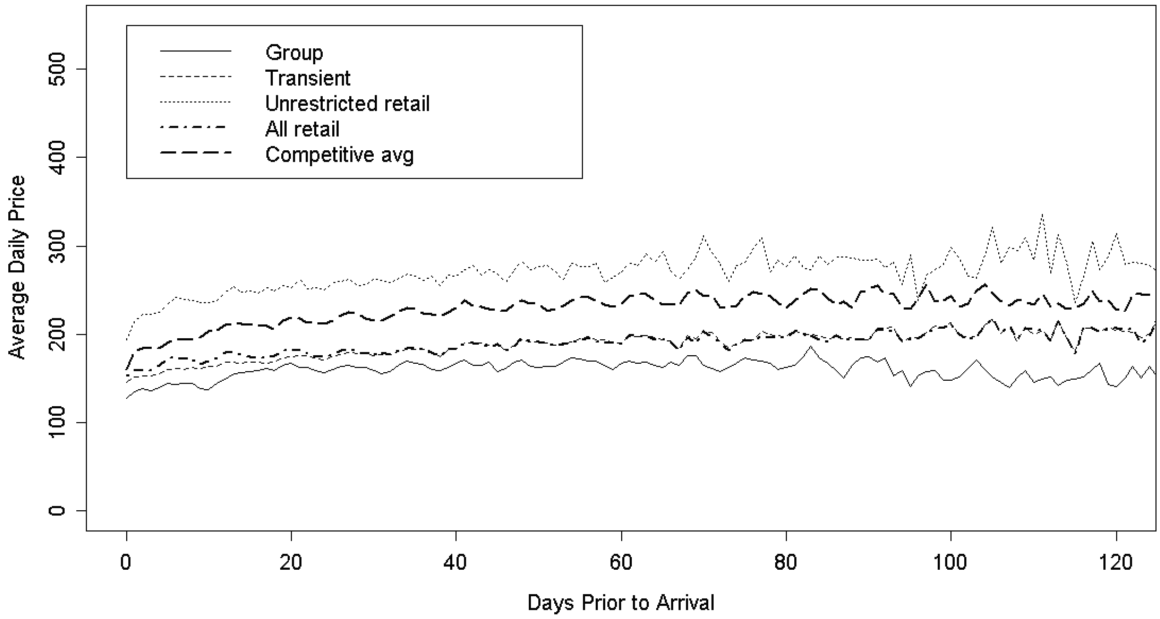


Figure 3: Average daily prices by days prior to arrival

higher than overall retail rates. However, the average rates paid (defined using arrival date as the unit of analysis) for hotel rooms declines as the day of arrival approaches, across both retail and negotiated demand segments. This result is also observed by competitor hotels. Table 4 summarizes the results of linear regression models associated with the slope of the average daily rate profiles. Note that the decline in average daily rates is steepest for the unrestricted retail segment (i.e., a decline of $0.59 per day)[2]. This is somewhat counterintuitive given that unrestricted retail products are generally designed for customers who book closer to arrival date (and in theory are willing to pay more for the flexibility of booking later and being able to change plans without cancellation fees).

One note of caution applies to the above result. Specifically, the result in Figure 3 may be influenced by "soft" demand days in which a large number of bookings at deeply discounted rates occur close to the arrival date and/or in which a few bookings

[2]Note that since "days" decreases as one nears the check-in date, a positive coefficient associated with days from arrival implies that the average daily price decreases.

Table 4: Regression results for average daily price as a function of days prior

Demand	# Obs.	Slope	Std. Err.	P value
Unrestricted Retail	95,515	0.589	0.01300	<< 0.001
All Retail	314,666	0.269	0.00526	<< 0.001
Group	134,509	0.109	0.00658	<< 0.001
Transient	405,219	0.347	0.00450	<< 0.001
Competitor Avg.	501,182	0.373	0.00383	<< 0.001

at very high rates occur far from the arrival date. In order to control for this potential effect, average booked rates were normalized to the average rate for each arrival date. Figures 4(a) and 4(b) portray the normalized curves for restricted and unrestricted retail bookings, respectively.

The light solid curves in Figure 4 represent the number of bookings for the segment by days prior and provide information on the number bookings at each price point during the booking cycle. The top and bottom edges in rectangles for each price point represents 25th and 75th percentiles associated with the normalized average rates. The dark solid line inside the box is the median rate at each day prior to arrival. There is also a light straight line at the normalized daily rate of 1.0 in Figure 4(a) that is used to highlight the decline in average rate for restricted retail bookings. This line is omitted in Figure 4(b) as the normalized rate is relatively flat.

The normalized average daily rate curves paint a slightly different picture than the non-normalized average daily rate curves. Table 5 summarizes the results of linear regression models associated with the slope of the normalized average daily rate profiles. Consistent with the non-normalized data, the restricted retail rates tend to decrease as the arrival date approaches; however, in contrast to the non-normalized data, the unrestricted retail rate slightly rises. Nonetheless, from a practical perspective the estimated increase in the normalized unrestricted retail rate, while statistically significant, will have little to no financial implications (the increase is only 0.01% of the

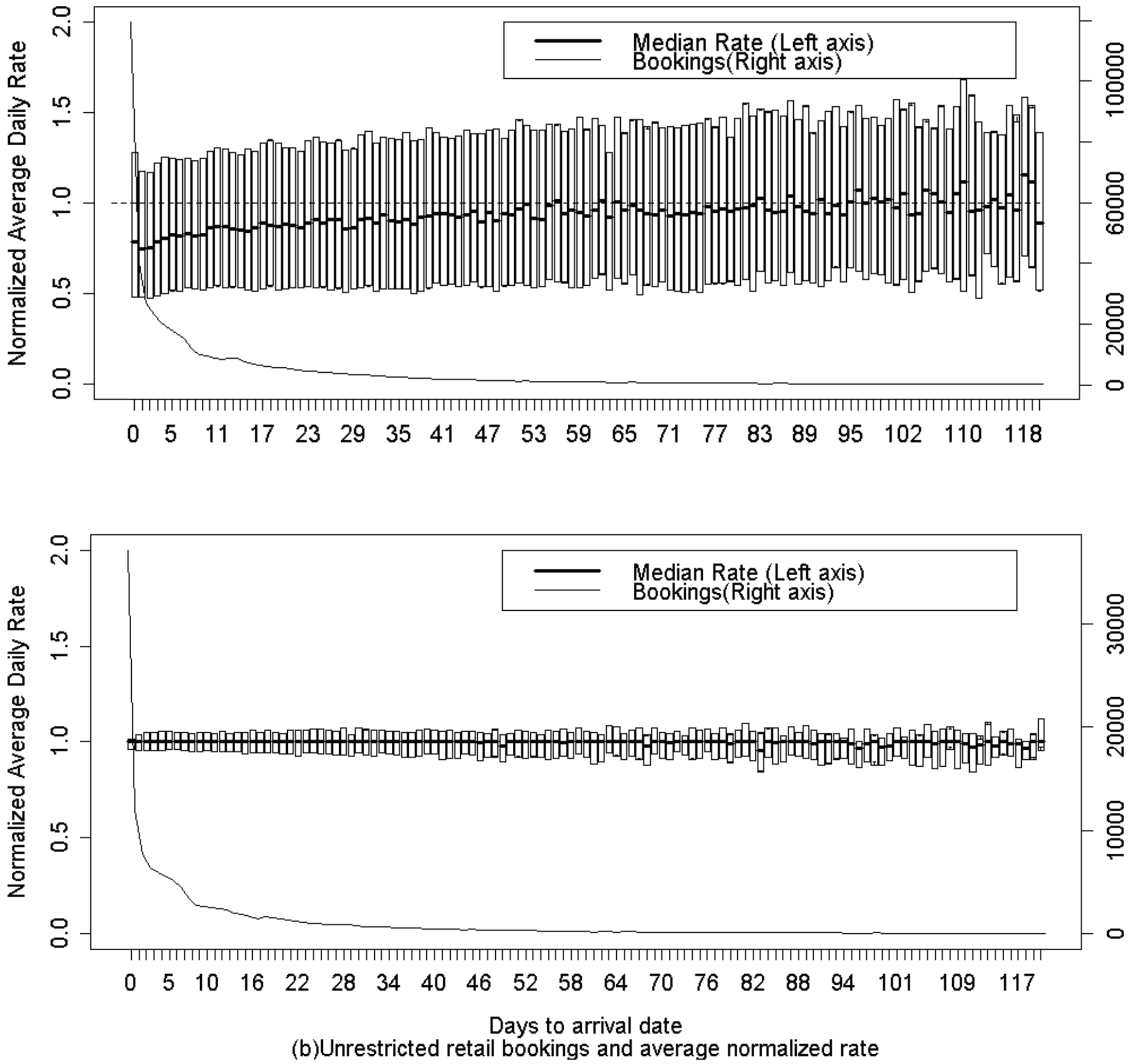


Figure 4: Restricted and unrestricted retail bookings and normalized rates by days prior to arrival

Table 5: Regression results for normalized average daily price as a function of days prior

Demand	# Obs.	Slope	Std. Err.	P value
Unrestricted Retail	93,096	-0.00010	-2.5E-6	<< 0.001
Restricted Retail	245,296	0.00190	5.6E-6	<< 0.001

average daily rate).

To summarize, both the analysis of booking curves for unrestricted and restricted classes as well as average daily rates paid by customers the statistical analysis provides only weak evidence in support of the assumption that late booking customers are willing to pay higher rates than early booking customers.

2.4 Assumption 2: Is weekday retail demand higher than weekend demand?

Statistical analysis can also be used to investigate the assumption that weekday demand is higher than weekend demand. Specifically, if the assumption is true, one would expect that occupancy rates and booking rates are higher during the weekdays. However, when looking for these expected relationships in the data, it is also important to recognize that the belief that retail demand is strongest midweek will lead to a general pricing strategy of charging higher rates during the week and lower rates during the weekend. The strategy is based upon the rationale that if occupancies are lower on the weekend, then prices should be lowered in order to stimulate more demand. Figure 5 clearly demonstrates the presence of such a pricing strategy both at the properties used in this study set as well as for their competitors.

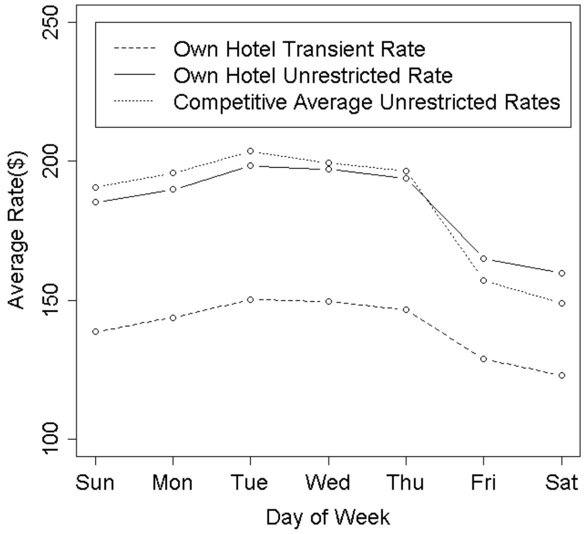


Figure 5: Transient, unrestricted retail, and unrestricted competitor rates by day of week

2.4.1 Comparison of Occupancy Rates and Demand

Latent (or unconstrained) demand is difficult to measure objectively. However, constrained demand is strongly correlated with unconstrained demand, and can be objectively measured. Occupancy is simply the constrained demand divided by the capacity. An analysis of occupancy rates for total demand reveals a pattern that is consistent with current industry intuition, namely that hotels tend to be busier during the weekdays versus weekends (see Figure 6(a)). Specifically, defining weekends as Friday and Saturday and weekdays as Sunday through Thursday, with the exception of Sunday, weekday occupancy rates are consistently higher than weekend demands. Richer insights can be gleaned by examining occupancy rates by demand segments (see Figure 6(b)) which reveals, counter to current business intuition, that retail demand is actually stronger on the weekends (Friday and Saturday) than for weekdays. This observation is confirmed via an ANOVA model of retail, transient and total rooms sold (Table 6). The coefficients capture how much the weekend retail and total rooms sold differ from the weekday. In the case of retail occupancy, the estimate for the weekend coefficient is positive and statistically significant, indicating

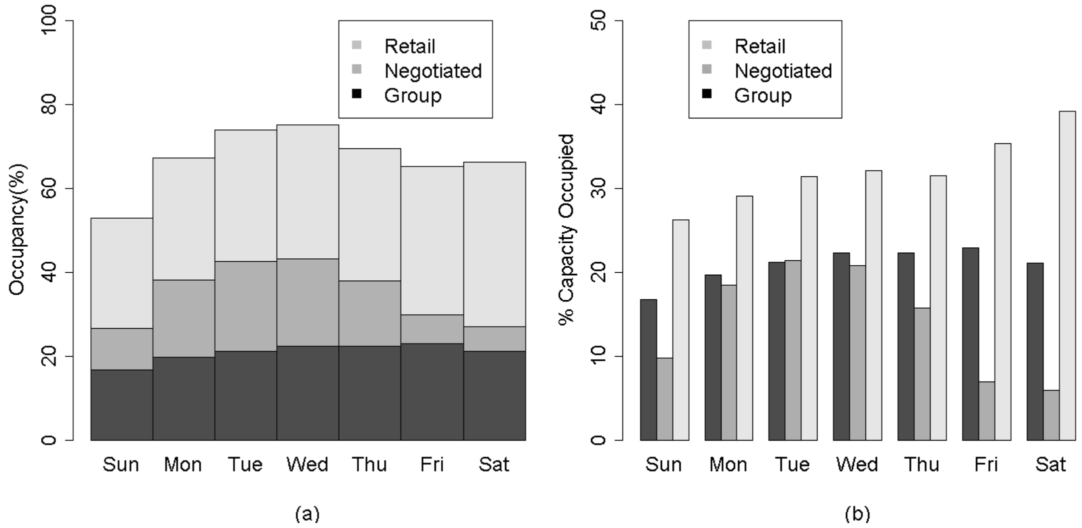


Figure 6: Average occupancy by day of week and demand segment

Table 6: T test from ANOVA model of weekend retail, transient and total rooms

Demand	Coeff.	T-stat	P value	Interpretation
Retail Rooms Sold	1.161	33.3	$<< 0.001$	Stronger retail demand on weekends
Transient Rooms Sold	-0.304	-6.36	$<< 0.001$	Stronger transient demand on weekdays
Total Rooms Sold	-0.122	-1.46	0.145	Stronger demand on weekdays vs. weekends

that the weekend retail demand is significantly larger than the weekday rooms sold.

This is a particularly interesting finding since the transient demand (defined as the combination of negotiated and retail demand) is the intended target of most current applications of hotel revenue management today. Negotiated rates are generally fixed, thus the price can not be adjusted up or down. Moreover, most corporate negotiated rates have a last room availability (LRA) clause. The hotels used for this study have 71.7% of the negotiated rooms sold under the LRA accounts. Rates with LRA can not be yielded out of the hotel even by length of stay controls. Even when transient demand is strongest during the week, which is the case on average for this study, a strategy that seeks to set length of stay controls or optimize rates based on a transient forecast would lead to suboptimal decisions. We submit that it is the retail segment in isolation that should be the focus of 'individual demand' revenue management, i.e., dynamic pricing (rate optimization) and inventory control actions.

It is also important to note that because retail demand is really strongest during the weekend and softest during the week, one may raise questions about the appropriateness of having lower retail rates on the weekends. Further study is required before one can conclude that having lower rates on the weekend is an incorrect (or correct) strategy. There are many factors that likely impact customers' willingness to pay higher retail rates on the weekend, including price elasticity and competitive rates.

2.5 Discussion

Using data from 28 different hotels, this study investigates two common assumptions common in the application of hotel pricing and revenue management: 1) customers who book later are willing to pay higher rates than customers who book earlier; and, 2) demand is stronger during the week than on the weekend. Empirical analysis indicates that rates, particularly retail rates, do not increase as the day of arrival approaches. Assumption two, while seemingly true in the aggregate, does not apply to the retail demand segment, yet the retail demand segment is the only segment impacted by traditional RM and dynamic pricing strategies. These findings challenge the current pricing and revenue management practices of most hotel companies. Table 7 provides a summary of the findings from this study.

One possible explanation for why hotel rates do not increase as the booking date approaches the arrival date - as observed in the airline industry - could be due to the different capacity constraints between the two industries. In general, airlines are more capacity constrained than the hotels. The International Air Transport Association

Table 7: Summary of analysis, findings and implications in Chapter 2

Assumption	Expected Outcome	Observed Outcome	Implication
1. Late booking customers are willing to pay higher rates	Higher valued customers book later than lower valued customers	Higher valued retail customers book at the same pace or only slightly later than lower valued retail customers.	Raises serious doubts regarding the assumption that late booking customers are willing to pay higher rates. If they are willing to pay higher rates, there is no evidence that they are being charged higher rates.
	Average rates paid increase as the booking date approaches the arrival date	Average rates for restricted retail demand decrease as day of arrival approaches. Average rates for unrestricted retail demand are flat or slightly increasing.	
2. Weekdays have higher demand than weekends	Occupancy is higher during the week than on the weekend	Total hotel occupancy is not significantly greater during the week than on the weekend. While Transient demand is stronger during the week, retail demand is in fact highest on the weekend.	Based on occupancy, weekday demand is not much stronger than weekend demand. Moreover, the key retail segment experiences peak demand on the weekend. At the same time rates, including retail rates are significantly lower on the weekend. One must ask the question - is it really necessary to lower rates so much on the weekend?
	Weekday rates are higher than weekend rates	Own property and the competitive set consistently price lower on the weekend.	

reports a North America average utilization (load factor) of 80.9% in 2006 [2], where U.S. hotels had an average occupancy rate of 63.4% [1]. In the classic microeconomic theory, positive shift of supply in a competitive market with other factors equal results in lower equilibrium price (see Varian [56] or any major economics textbook).

Differentiation is another key factor that must be considered when translating traditional airline yield management to the hotel industry. It is hard to differentiate one airline seat from another for a specific itinerary. Both leave from the same (or one of a few) origin airports and arrive at the same (or one of a few) destination airports. Hotels, on the other hand, are strongly differentiated by their location. Only one hotel can be the closest hotel to a traveler's intended destination. If the traveler has a business meeting downtown at 8 AM, a suburban hotel 25 miles from the city center may not be an acceptable alternative. However, a price conscious customer, particularly one with an automobile, can increase their alternatives by choosing a hotel further from their intended destination, but at a lower rate. Finally, hotels have a tremendous advantage over airlines in their ability to differentiate the customer experience through amenities and quality of service. Hotels seemingly have many opportunities to differentiate by both price and product attributes.

The combination of lower utilization rates, greater product differentiation, but more alternatives, suggests that hotels must apply different approaches than those learned from traditional yield management (as practiced by the airline industry). Simply matching competitor rates to avoid losing market share is not necessarily a profit optimal strategy. On days when inventory is tight, traditional yield management tactics deliver tremendous value, but these should be augmented by incorporating price response and competition. On days when demand is soft and occupancy is projected to be low, price and competition based strategies should dominate. In this study, very little difference in booking patterns between high and low valued customers was observed. The fact that high and low valued customers tend to book at the same time raises serious questions about appropriateness of applying traditional yield management methods that seek to protect rooms for late booking, high value customers. Cooper *et al* [18] show that applying the assumption "high valued demand books later" can lead to a downward spiral of rates in situations where demand is soft. When hotel demand is high, this is less of a concern. However, most in the hospitality industry saw their rates plummet during the post 9-11 travel recession. Leading hotel companies should revisit the assumptions inherent in their revenue management and pricing strategies before the onset of the next travel recession.

The general pricing strategy for properties in this study was to have lower rates on the weekend. This strategy is common - if not dominant - in the hotel industry as a whole, yet the retail segment is strongest on the weekend, suggesting that retail customers might be willing to accept higher rates. Retail customers booking over the weekend are likely to be leisure customers who only have leisure time on the weekend. Would they still travel and book a hotel room if the rates were a bit higher? In some cases, probably so. An understanding of the retail customer's response to own and competitive rates would be required to determine if retail rates could be increased over the weekend.

Hotels must reevaluate their pricing strategies and revenue management programs. Central to this reevaluation is to move from the traditional group-transient segmentation, to further differentiate true retail demand from negotiated demand. Negotiated demand can only be priced at the time of contract negotiation. Once set, these rates are not (typically) flexed as retail rates are changed. It is mainly the retail segment that is subject to the full array of both inventory controls and pricing actions. While restricting the focus of existing revenue management models to the retail segment, hotels must develop new approaches to ensure that revenue is also maximized for the group and negotiated segments and, in turn, for the total hotel.

Before optimizing the rate structure, revenue mangers need to thoroughly explore the data and truly understand the retail response to price, competition and other non price factors, particularly day of week. Modeling demand response to price with any degree of precision is not straightforward; yet developing such models will be a critical determiner of success for revenue management going forward. With a clear understanding of how demand will change under different market conditions and pricing structures, yield management and pricing models can be enhanced to incorporate the true nature of hotel demand. With this understanding, we will be able to use pricing as a powerful tool for maximizing profit.

CHAPTER III

DEMAND MODEL ESTIMATION USING HOTEL DATA

3.1 Introduction

Revenue management refers to the practice of applying analytics to predict consumer behaviors and optimizing the product availability and/or pricing decisions in order to maximize revenue. The traditional revenue management discipline was initially developed with its focus on controlling the availability of products (commonly referred as the yield management; see Talluri and Van Ryzin [53] and Cross *et al* [19] for a comprehensive review of the revenue management discipline). The traditional yield management has it's limitations, mainly due to the fact that the demand is assumed to be independent of prices. The objective of the price optimization is to maximize the revenue under the assumption that the demand is a function of the price.

In the hotel context, price optimization provides optimal itinerary (combination of arrival date and the length of stay) prices to maximize revenue where the expected number of itinerary bookings is a function of the itinerary price. One of the early adopters of hotel price optimization reported 2.7% additional revenue attributable to this capability [3].

In price optimization, the demand model determines the shape of the objective function and therefore is one of the most important inputs. Demand models are typically estimated using historical observations and are chosen based on various model fit statistics such as R^2, mean absolute percentage error (MAPE), Bayesian information criterion, etc. Besbes *et al* [11] reviews the various model testing methodologies commonly utilized in the revenue management field. The focus of demand modeling in revenue management is to achieve a good model fit, however, the impact of demand models on price optimization performance has not been actively explored and utilized in the practical context. Since efforts to achieve the best model fit requires extensive data collection and analysis, the benefits of using a better demand model needs to be justified. Also by identifying aspects of demand models more critical to revenue, the users can be better informed to make decisions such as whether to include certain driver variables and whether to choose a different functional form at the cost of increasing the complexity of the optimization problem.

In this chapter we investigate the goodness of fit for different demand models using data gathered from hotels. We evaluate various linear models using driver variables such as price, number of days prior to the check in date the booking was made, length of stay, day of week, demand from previous booking dates and the interaction of these variables. Booking and pricing data from U.S. domestic hotels are used to develop a comprehensive demand model that gives a good representation of the actual hotel demand.

3.2 Literature Review

Tourism demand literature focuses on finding driver variables and comparing the goodness of fit of different models for tourism related products such as transportation, accommodation and attractions. In these studies, econometric models with price as an explanatory variable have been popular for forecasting market level demand. Witt and Witt [63] give an overview of various demand models used in the tourism literature and compare their forecasting accuracy. One of the major findings in Witt and Witt [63] is that there is no single model superior in all markets, which suggests different models need to be considered for different market environments. Song *et al* [52] further introduces advanced econometric demand models to compare their forecasting accuracy with exponential price response models and time series models. The authors find that Time Varying Parameter model (allowing different parameters for different time lag) achieves the highest accuracy (MAPE, RMSPE (Root Mean Squared Percentage Error)) among the six econometric models they tested including error correction models, autoregressive distributed lag model, unrestricted vector autoregressive model, and ARIMA models. Lim [44] reviews existing literature on tourism demand forecasting.

According to Lim [44], the log linear model is the most frequently used model form in the existing tourism demand forecast studies followed by the linear model. The explanatory variables most often found in these studies include customer income, price, price of substitutables, trend and autoregressive terms of these variables collected from surveys or transactional data. In our research, we examine these variables as candidate driver variables in linear form to explain hotel demand.

3.3 Estimating the Hotel Demand

In this section we first review the summary statistics of hotel demand using historical booking data from actual hotels. Using the summary statistics, we explore aggregation schemes to be used in the regression analysis. In Section 3.3.4 we present the framework for estimating the linear demand models for each hotel in our dataset along with the analysis of model estimation results.

Table 8: Summary demand characteristics of hotel bookings

Parameter	Capacity	Occupancy	Daily booking	LOS	Days to arrival	c.v.
Minimum	81	39.0%	16.5	1.26	5.54	0.20
25 percentile	208	63.3%	28.5	1.60	8.50	0.32
Average	322	69.9%	56.6	1.87	14.0	0.40
75 percentile	423	77.4%	71.7	2.08	18.5	0.45
Maximum	807	89.4%	162.4	2.60	31.5	0.72

3.3.1 Data

We first take a look at the demand characteristics of different types of hotels based on historic booking data from a global hotel chain. This data is an extension of data used in Chapter 2, collected over 23 months from March 2006 to January 2008 from 23 U.S. non-extended stay hotels, consisting of 5 luxury, 7 premium full-service, 6 full-service business, and 5 limited service hotels. They are located in city center (11), suburban (9), and airport (3) locations. Only the *retail segment* (the non-group, non-negotiated segment of hotel demand) was analyzed for this analysis since we focus on optimization of retail rates. Retail segment typically occupies 25%∼45% of total capacity for the test hotels. Negotiated and group segments were included for calculating occupancy and coefficient of variation (c.v.).

Table 8 summarizes the key parameters of the hotel booking data. The capacity, which is the physical rooms available in a hotel, ranges from 81 to 807. The occupancy is calculated as the average of total bookings for a staynight divided by the hotel capacity, and ranges from 39% to 90%. Booking is the average number of the retail segment bookings by arrival date, and the length of stay (LOS) is the average number of staynights for a typical retail booking for the hotel. Finally, c.v. shows the average coefficient of variation for the total staynight bookings for a hotel.

Table 9: Length of stay distribution by location

Location	Length of stay 1	2	3	4	5+
Airport	76.8%	13.4%	5.7%	2.7%	1.4%
Suburban	60.0%	23.5%	9.7%	4.7%	2.1%
Urban	47.9%	27.5%	14.1%	7.2%	3.3%

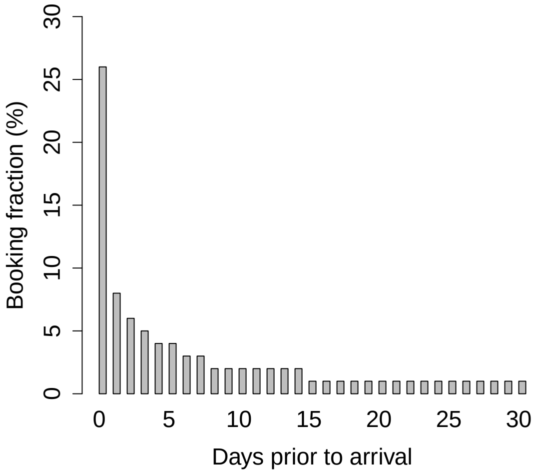


Figure 7: Booking fraction by days prior to arrival

3.3.2 Hotel Demand Characteristics

Table 9 shows the length of stay distribution of bookings for the three location types in the data. Location is an important factor influencing the demand characteristic and impacting the revenue management performance. We observe that more than 76% of airport bookings are for a single night stay. On the contrary, only 48% of urban hotel bookings are for a single night stay.

Figure 7 shows the average booking percentages by the number of days prior to arrival date (booking curves) for all hotels. Bookings are highly concentrated on the day of the arrival and the last week of the booking window.

Figure 8 plots the coefficient of variation of staynight demand against price levels.

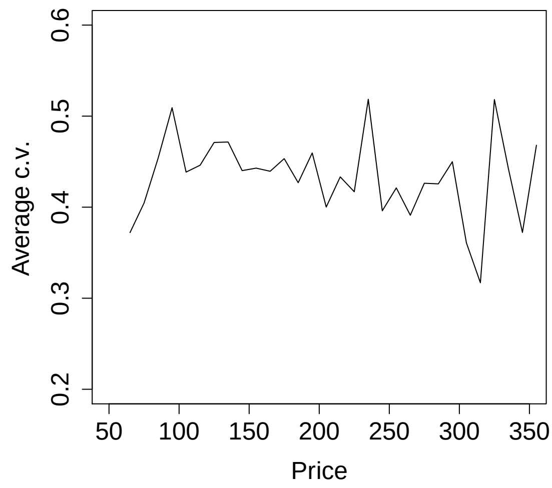


Figure 8: Coefficient of variation of staynight demand

We find that there is no significant trend between the coefficient of variation and the price level, as the p value of the slope coefficient in Figure 8 is greater than 0.1.

3.3.3 Aggregation of Data

On average, most of the hotel demand is booked on the last month of the booking window and is very sparse further out (average of 80%∼92% booked within 30 days prior to arrival, depending on the location). Also, majority of the bookings have one or two nights of stay and there is only a small percentage of bookings for longer stays (see Table 9). If we do not aggregate the higher days prior and length of stay levels, demand model coefficients are likely to be highly volatile or insignificant due to the data sparsity in these levels.

Table 9 shows LOS 4 segment has on average 3%∼7% of bookings, depending on the location. With average daily retail bookings of 57 (Table 8), any LOS higher than four would have less than 2∼4 bookings on average. Such a small segment, possibly with a large number of zero demand observations, could lead to unstable model estimation and therefore we group the lengths of stay 4 and above into a single category. For the days prior variable, we group the values ranging from 0 to 364 into a few number of categories. For this purpose, we employ the discrete wavelet transform method (DWT).

Wavelet transform is a method to decompose a function into mixtures of wavelets. DWT, which approximates a limited number of discretely sampled wavelets from a distribution, has been applied to nonparametric density estimation with success (Vidakovic [57]). Chui [17] provides a general exposition on wavelet theory. Recently, Popescu *et al* [49] used wavelet method to estimate cargo demand distribution as a function of show up rates.

DWT assumes the given distribution is composed of discrete number of signals (represented by the wavelet coefficients) and noise. Given the initial number of bins and the basis wavelet type assumption, the method estimates the wavelet coefficients for the given number of bins by ignoring the small details which it considers as noise. Once the noise is eliminated, adjacent bins with same frequency are grouped together resulting in variable sized bins.

Discrete wavelet basis used to decompose a distribution is a group of functions represented by scaling function ϕ and translations and dilations of the mother wavelet ψ, which define the basis that spans space of integrable functions, $L^2(R)$. In our implementation we use the Haar wavelet as the mother wavelet and the unity function as the scaling function.

$$\phi(x) = 1 \quad (0 \leq x < 1) \tag{1}$$

$$\psi(x) = \begin{cases} 1 & 0 \leq x < \frac{1}{2} \\ -1 & \frac{1}{2} \leq x < 1 \end{cases} \tag{2}$$

In our implementation we assume this distribution is composed of 2^d bins, with d

as log of Birgé-Rozenholc estimate of number of bins [12] rounded to the nearest integer. With 2^d initial bins, a quadratic variance-stabilizing transformation is used to estimate the optimal variable bin sizes as described in Vidakovic [57]. After the transform, we perform wavelet shrinkage to smooth out the noise in the original signal or the wavelet coefficients. The shrinkage procedure first sets coefficients below a universal threshold of $\sqrt{2log2^d}\sigma$ to zero (where σ^2 is the noise variance), and then shrinks the nonzero coefficients. The shrinkage method (known as soft thresholding) and the universal threshold value used in our implementation follow those from Popescu *et al* [49]. The steps of our DWT implementation is outlined as follows.

1. Estimate the initial number of number of bins, 2^d, where $d = logD$ rounded to the nearest integer and D is the recommended number of bins from the Birgé-Rozenholc procedure.

2. Perform variance-stabilizing transform to number of observations per bin, $2\sqrt{f_i + \frac{3}{8}}$, where f_i is number of observations for bin i.

3. Perform wavelet shrinkage and only choose coefficients passing the threshold criteria.

4. Construct the smoothed signal of f.

5. Calculate the variable size bins based on the smoothed signal.

Initial number of bins in DWT can be any power of two. In our case we followed the Birgé-Rozenholc procedure [12] to estimate the ideal number of bins and rounded it to the closest power of two. The Birgé-Rozenholc procedure and the DWT process was applied to only days prior 0 to 45 in order to avoid too many days prior groupings. The implementation of the DWT in this study closely follow Popescu *et al* [49].

Many statistical packages offer DWT functionality. In our study, *dwt* function from

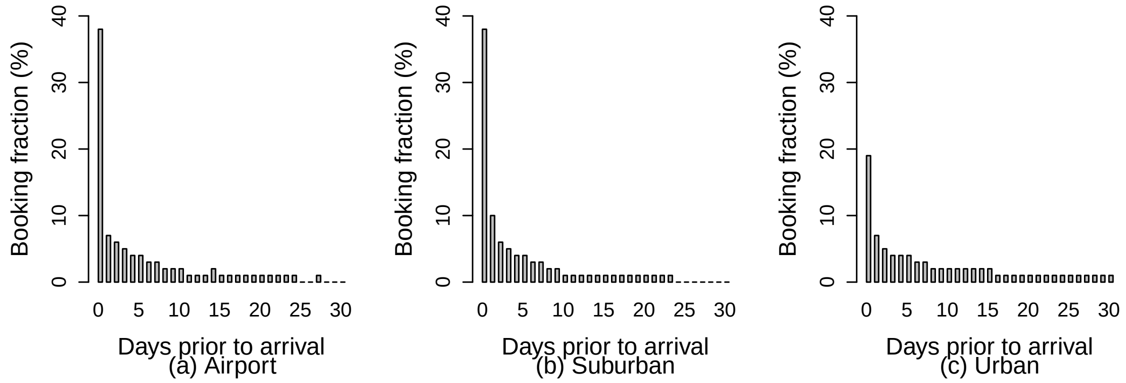


Figure 9: Distribution of booking density by DWT estimated days prior groupings

waveslim package in statistical software R (Whitcher [61]) was used to estimate the optimal days prior groupings with the Haar wavelet, which is a step function with value 1 on $[0, \frac{1}{2})$ and -1 on $[\frac{1}{2}, 1)$, as the basis wavelet.

In our analysis, DWT was performed separately for each location. The bin definition estimated by DWT is identical for all locations while the density of each bin significantly differed by location. Figure 9 shows the booking density distribution for days prior 0 to 45 for each location, which are grouped into the following categories by DWT: 0, 1$\sim$6, 7$\sim$11, 12$\sim$22, and 23$\sim$45.

Airport hotels with high percentage of transit demand have highly concentrated demand on the day of arrival (days prior 0). On the other hand, urban hotels bookings are less concentrated closer to the arrival date and more evenly distributed throughout the booking window as represented by the heavier tail in the booking density distribution.

3.3.4 Estimating the Linear Demand Model

In this section, we estimate hotel demand models using the explanatory variables shown in Table 10. The model estimated in this section has a linear price-demand

Table 10: List of potential explanatory variables

Variable	Type	Description
Price	Continuous	Average retail price for products sold
Average competitor price	Continuous	Average price of 4 competitors
Days prior	Ordinal	Number of days between the booking date and the arrival date
Length of stay	Ordinal	Number of nights
Day of week	Categorical	Day of week of arrival date
Month	Categorical	Month of arrival date
Saturday night stay	Categorical	Saturday night included in the stay (Y/N)

relationship and are referred as the linear model. Competitive unrestricted retail rates used in our study are based on the averages from four competitors designated by the hotel revenue managers.

Among the variables presented in Table 10, only the ones that have significant impact on the drivers are chosen to represent the base model (Table 11). In addition to the base model, Table 11 also identifies additional terms that could be included in the demand model to significantly improve the adjusted R^2. All of the base model and additional driver variables are chosen if adjusted R^2 is improved by more than 1% by including the variable and if the estimated coefficient for the variable is statistically significant (p value$<$0.05). It should be noted that instrumented variable regression was performed to test for price endogeniety but residual coefficients were insignificant (p$>$0.05) for the majority of the properites. This results indicates that price endogeneity is not significant, meaning that the price variable is not significantly correlated with the error term in the regression.

For each of the 23 hotels, the best fit model is determined by the following process.

Step 1 Estimate the following models

Model 1 (Base Model)

$$d = a + b_P P + \sum_{i=2}^{5} b^i_{DP} 1_{DP=i} + \sum_{i=2}^{4} b^i_{LOS} 1_{LOS=i} + \sum_{i=Mon}^{Sat} b^i_{DOW} 1_{DOW=i}$$
$$+ b_{DLAG1} DLAG1$$

Model 2

$$d = BaseModel + \sum_{i=2}^{5} b^i_{DP\times P} 1_{DP=i}$$

Model 3

$$d = BaseModel + \sum_{i=2}^{4} \sum_{j=2}^{5} b^{ij}_{LOS\times DP} 1_{LOS=i} 1_{DP=j}$$

Model 4

$$d = BaseModel + \sum_{i=Mon}^{Sat} \sum_{j=2}^{4} b^{ij}_{DOW\times LOS} 1_{DOW=i} 1_{LOS=j}$$

Model 5

$$d = BaseModel + \sum_{i=2}^{5} b^i_{DP\times DLAG1} 1_{DP=i}$$

Step 2 Among the additional terms in Models 2∼5, choose those that are significant and contributes more than 1% to adjusted R squared

Step 3 Estimate a model with all the terms chosen in Step 2 (Model 6)

Step 4 Among the parameters estimated for Model 6, identify coefficients which are not significantly different (two sample t-test)

Step 5 Merge the coefficients found in Step 4, drop insignificant coefficients

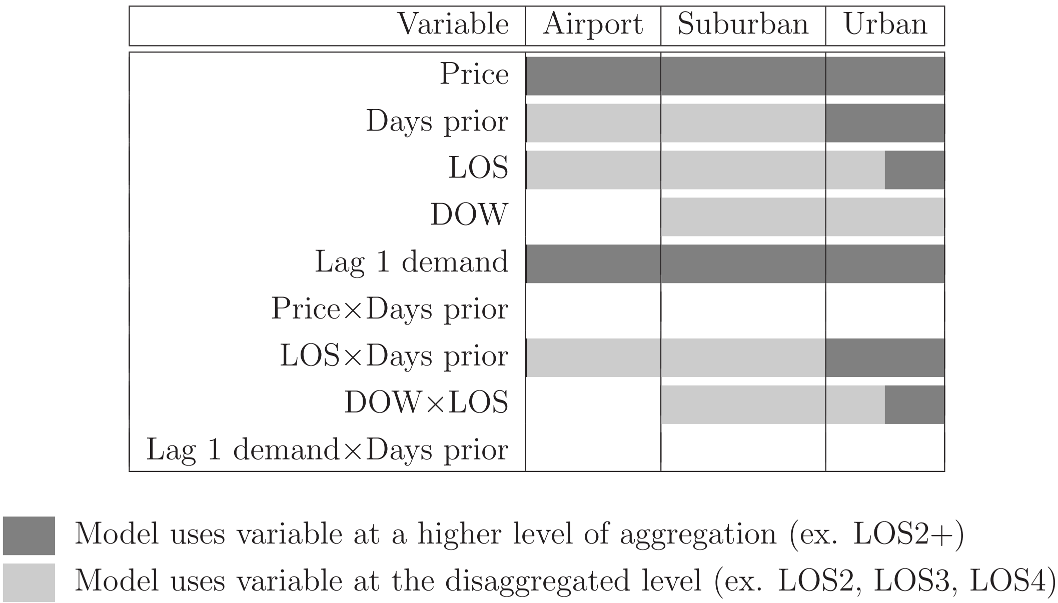
Figure 10 describes the best fit models for the 23 hotels and shows that hotels within the same location share similar model specifications.

Table 11: List of base model and candidate explanatory variables

	Variable	Category/Range
Base model variables	Price (P)	Non-negative
	Days prior (DP)	0,1∼6,7∼11,12∼22,23+
	Length of stay	1,2,3,4+
	Day of week (DOW)	Sun∼Sat
	Days prior lag demand (DLAG1)*	Non-negative
Candidate variables	P×DP	
	LOS×DP	
	DOW×LOS	
	DLAG1×DP	

*Note: *Average daily bookings for immediately previous days prior group*

Table 12: Summary of estimated demand models by location



Among the estimated demand models, we observe that the airport and suburban hotels have a simpler model specification with less drivers compared to the urban hotels. For example, the airport and suburban hotel demand do not have significantly different sensitivity to length of stay 2,3 and 4+, whereas the urban hotel demand has significantly different sensitivity to each LOS (i.e, the coefficients for LOS 2,3 and 4+ are significantly different). A similar pattern is observed for the days prior grouping coefficients. While the airport and suburban demand models have similar coefficients for days prior 7 and above, the urban hotel demand model has statistically distinct coefficients for all four levels of days prior groupings.

The airport hotel demand models are different from other location hotels in that the airport hotel demand level does not differ by day of week. The suburban and urban hotel demand exhibit different demand patterns for weekends (Friday and Saturday) compared to Sunday or Monday to Thursday. This observation highlights the distinctive behavior of airport hotel customers who mainly stay at these hotels for transit purposes. Other location hotels, which are mostly located near customers' final destinations, have clearly different demand patterns by day of week contrary to the airport hotels.

The demand models for urban properties can be classified into two types, one with more (the right side of the urban hotel column in Table 12) and another type with less detailed model specifications. The hotels identified as the first type are located in large metropolitan areas such as New York City, San Fransisco, Chicago and Washington DC, indicating that the large city hotel demand are influenced by more factors and are more sensitive to different levels of driver variables than the small city or suburban area hotel demand.

Overall, the results in Figure 10 suggest that the hotels within the same location types share similar demand patterns. Contrarily, different locations have distinctive set of explanatory variables at different levels, although a few sets of variables (days prior, length of stay, etc.) are common drivers for all hotel demands.

	Airport			Suburban								
	1	2	3	4	5	6	7	8	9	10	11	12
Pr Elas (at mean)	-0.190	-0.073	-0.092	-0.116	-0.380	-0.061	-0.107	-0.071	-0.102	-0.076	-0.067	-0.132
Pr Slope	-0.460	-0.086	-0.207	-0.097	0.285	-0.048	-0.099	-0.067	-0.075	-0.093	-0.097	-0.164
R sq	59.99%	45.24%	63.34%	65.15%	46.11%	26.65%	56.17%	41.84%	50.27%	57.12%	62.96%	57.56%
Intercept	16.47	6.17	20.96	8.48	4.37	3.16	3.50	3.17	2.09	6.71	7.14	8.52
Price/100	-0.46	-0.09	-0.21	-0.10	-0.28	-0.05	-0.10	-0.07	-0.08	-0.09	-0.10	-0.16
Days prior 1~6	-5.83	-0.98	-2.55	2.10	-0.67	-0.39	1.13	-1.52	0.41**	-4.62	-4.52	-3.19
Days prior 7~11						-2.45						
Days prior 12~22						-2.22						
Days prior 23~45	-11.67	-4.34	-14.70	-6.32	-3.11	-1.66	-4.88	-3.76	-3.01	-7.66	-6.88	-7.91
LOS2												
LOS3												
LOS4	-14.72	-4.40	-18.52	-6.11	-2.32	-1.64	-2.31	-1.67	-0.86	-5.03	-6.55	-7.49
Monday								1.56				
Tuesday								3.18				
Wednesday								2.67				
Thursday						0.73		1.32	1.27	2.14		
Friday			0.87	4.91	1.42	1.16	2.98	1.45	3.33	3.28		
Saturday			1.29	8.90	2.69	1.68	5.69	2.52	4.91	3.78	1.18	1.11
lag1	4.27	1.61	1.55	1.41**	3.16	0.88	2.80	1.92	2.56	1.59	2.88	2.50
Price*dp1~6					0.20							
Price*dp7~11												
Price*dp12~22												
Price*dp23~45					0.26							
LOS2*dp1~6	8.36	1.41	4.46	-1.08	1.02	0.71	0.74	2.01	0.96	5.24	5.01	4.43
LOS2*dp7~11												
LOS2*dp12~22												
LOS2*dp23~45	13.14	4.44	14.45	6.80	2.83	2.40	5.39	3.51	3.09	7.54	7.09	7.84
LOS3*dp1~6												
LOS3*dp7~11												
LOS3*dp12~22												
LOS3*dp23~45												
LOS4*dp1~6												
LOS4*dp7~11	merged	merged	merged	merged	merged	merged	merged	merged	merged	merged	merged	merged
LOS4*dp12~22	with	with	with	with	with	with	with	with	with	with	with	with
LOS4*dp23~45	LOS2	LOS2	LOS2	LOS2	LOS2	LOS2	LOS2	LOS2	LOS2	LOS2	LOS2	LOS2
Mon*LOS2												
Mon*LOS3												
Mon*LOS4								-1.18				
Tue*LOS2												
Tue*LOS3												
Tue*LOS4								-2.89				
Wed*LOS2												
Wed*LOS3												
Wed*LOS4								-2.54				
Thu*LOS2												
Thu*LOS3												
Thu*LOS4						-0.65		-1.47	-1.15	-1.87		
Fri*LOS2				1.56		0.31'			-1.26	-1.78		
Fri*LOS3												
Fri*LOS4				-5.00	-0.39*	-1.17	-2.43	-1.26	-3.38	-3.27		
Sat*LOS2												
Sat*LOS3												
Sat*LOS4				-8.87	-3.04	-1.69	-5.98	-2.65	-5.06	-3.92		
lag1*dp1~6				5.63							-2.03**	
lag1*dp7~11												
lag1*dp12~22												
lag1*dp23~45				1.44								

	Urban										
	13	14	15	16	17	18	19	20	21	22	23
Pr Elas (at mean)	-0.093	-0.107	-0.082	-0.136	-0.115	-0.084	-0.169	-0.112	-0.107	-0.078	-0.117
Pr Slope	-0.152	-0.077	-0.093	-0.069	-0.067	-0.045	-0.144	-0.206	-0.043	-0.050	-0.031
R sq	63.39%	32.09%	58.24%	45.84%	40.02%	33.34%	45.05%	48.48%	29.02%	33.33%	30.82%
Intercept	7.79	2.39	4.33	5.57	4.97	3.76	6.13	3.36	3.28	2.37	2.79
Price/100	-0.15	-0.08	-0.09	-0.07	-0.07	-0.04	-0.14	-0.21	-0.04	-0.05	-0.03
Days prior 1~6	-1.01	1.80	4.80	10.20	4.80	4.30	6.48	2.10	2.15	2.39	2.62
Days prior 7~11				-4.59	-5.77	-4.33	-4.09		-2.66	-1.94	-1.68
Days prior 12~22				-5.61	-6.51	-4.84	-4.97		-2.82	-2.18	-1.97
Days prior 23~45	-7.58	-1.56	-4.84	-1.45	*0.24*	-2.52	-1.11**	-3.12	*-0.32*	-0.55**	*-0.29*
LOS2			-3.25					-1.80			
LOS3											
LOS4	-6.26	-0.95	-2.43	-3.72	-2.99	-2.51	-4.05	-1.36	-1.53	-0.72**	-1.19
Monday											
Tuesday											
Wednesday											
Thursday	0.47**	0.77		2.72	1.95	2.94	-0.60	1.80	1.10**	0.61	
Friday	1.01	1.19	2.70	2.33	3.55	2.38	2.35	2.57	1.06*	2.11	0.94
Saturday	1.00	2.85	5.15	4.78	8.62	4.63	7.56	3.89	3.52	3.64	1.92
lag1	4.26	2.08	2.18	0.79	0.78	1.40	0.66	3.00	0.52	1.36	0.47
Price*dp1~6											
Price*dp7~11											
Price*dp12~22											
Price*dp23~45											
LOS2*dp1~6	3.87	-0.61	-0.76*	-4.59	-2.05	0.56	-3.42	-0.91**	*-0.01*	-0.41'	-1.06
LOS2*dp7~11											
LOS2*dp12~22											
LOS2*dp23~45	7.66	2.11	5.25	5.67	6.26	4.91	4.78	3.63	2.49	2.31	1.86
LOS3*dp1~6	1.74		-3.46	-8.21	-4.39	-1.92	-5.37	-1.63	-1.48	-1.63	-2.15
LOS3*dp7~11	merged		merged	merged	merged	merged		merged	merged	merged	merged
LOS3*dp12~22	with		with	with	with	with	merged	with	with	with	with
LOS3*dp23~45	LOS2		LOS2	LOS2	LOS2	LOS2	with LOS2	LOS2	LOS2	LOS2	LOS2
LOS4*dp1~6	mrg/LOS3		mrg/LOS3	mrg/LOS3	mrg/LOS3	mrg/LOS3	mrg/LOS3	mrg/LOS3	mrg/LOS3	mrg/LOS3	mrg/LOS3
LOS4*dp7~11	merged	merged	merged	merged	merged	merged		merged	merged	merged	merged
LOS4*dp12~22	with	with	with	with	with	with	merged	with	with	with	with
LOS4*dp23~45	LOS2	LOS2	LOS2	LOS2	LOS2	LOS2	with LOS2	LOS2	LOS2	LOS2	LOS2
Mon*LOS2			-2.14					-1.52			
Mon*LOS3											
Mon*LOS4			-3.00					-2.04			
Tue*LOS2											
Tue*LOS3											
Tue*LOS4											
Wed*LOS2											
Wed*LOS3											
Wed*LOS4											
Thu*LOS2			*merged					merged			
Thu*LOS3			with					with			
Thu*LOS4		-0.62*	Mon	-0.44*	-1.20**	-2.46		Mon	-0.90*	-0.69	-0.71
Fri*LOS2	2.10	*-0.52*	*-0.54*	3.49	1.76**	1.80**	5.30	-1.19**	1.77	-0.64*	0.74**
Fri*LOS3			*merged								
Fri*LOS4	-0.62'	-1.50	w Mon	-4.63	-2.10	-1.21'	-1.66	-2.96	*-0.65*	-1.94	-0.79
Sat*LOS2											
Sat*LOS3											
Sat*LOS4	-5.48	-3.28	-5.49	-4.68	-7.84	-4.55	-7.26	-4.37	-3.33	-3.76	-2.17
lag1*dp1~6			4.51	5.40	6.95		9.07		3.15	2.34	2.46
lag1*dp7~11											
lag1*dp12~22											
lag1*dp23~45											

** p<0.01, * p<0.1, ' p<0.5, *p>0.5*, otherwise p<0.001

Figure 10: Best Fit Demand Models for the 23 Sample Hotels

CHAPTER IV

HOTEL PRICE OPTIMIZATION PROBLEM

In this chapter we formulate the hotel price optimization problem in the general form. We also explore the analytical properties of hotel price optimization under linear demand and further use these properties to develop an alternate optimization algorithm to solve the hotel price optimization problem. The performance of the alternate algorithm is compared to the existing Quadratic Programming algorithms.

4.1 General Problem Formulation

In hotel price optimization the price of a product is the decision variable and the demand for a product is a function of the price. A hotel product is typically an itinerary, defined by the combination of an arrival date and length of stay. Depending on the business practice, it is possible to have multiple products within a single itinerary differentiated by factors such as customer type (corporate, individual, AARP), sales channel (global distribution system, Internet travel agency, etc.) and product restrictions (advance purchase, Saturday night stay, etc.). In our work we define an itinerary as the hotel product, but our framework can easily be extended to the case where multiple products exist within an itinerary. An itinerary is composed of multiple staynights, for example, an itinerary with arrival date of 1/10/2011 and length of stay 2 utilizes staynights January 10, 2011 and January 11, 2011. Available inventory is by staynights, for example, a hotel has 97 rooms available for the staynight of January 10, 2011 and 100 rooms for the staynight of January 11, 2011.

The objective function for our price optimization problem is the total revenue over a given period of time (decision horizon), which is the sum of itinerary revenues (Equation (3)). Total demand accommodated for a staynight is the sum of all itinerary demand that utilizes the staynight and is restricted by the total rooms available (capacity) for the given staynight (Equation (4)). The hotel price optimization problem can be expressed as following.

$$(P) \qquad \max_{r} \sum_{(a,l)\in A\times L} r_{a,l} d_{a,l}(r_{a,l}) \tag{3}$$

$$s.t. \qquad \sum_{(a,l):a\leq s\leq a+l-1} d_{a,l}(r_{a,l}) \leq c_s \qquad \forall s \in S \tag{4}$$

where the input parameters and the decision variables are

A	:	Set of arrival dates
L	:	Set of possible lengths of stay (duration)
(a,l)	:	An itinerary with arrival date a and duration of stay l
S	:	Set of staynights occupied by itineraries in $A \times L$
c_s	:	Remaining hotel capacity at staynight s
$r_{a,l}$	:	Price for itinerary (a,l)
$d_{a,l}(r_{a,l})$	:	Demand for itinerary (a,l) if price $r_{a,l}$ is charged

Optimization model *(P)* determines itinerary prices that maximize the expected revenue for the decision horizon at a given point of time. By solving *(P)* frequently over the decision horizon, the itinerary prices can be dynamically optimized to maximize the remaining revenue opportunities. When demand function is nonnegative and monotonic decreasing function of price, the optimal solution to (P) is always nonnegative.

4.2 Alternate Price Optimization Algorithm for Price Optimization with Linear Demand

In this section we explore the analytical properties of the price optimization solution and deduct an alternate formulation to the problem (P). When the demand function is linear, (P) becomes a quadratic optimization problem and the solutions have nice properties we can use to solve the problem efficiently. First, we introduce the following notations to express (P) with linear demand in vector/matrix form.

m	:	Number of itineraries
n	:	Number of staynights
α	:	Vector of intercepts for $d_{a,l}$ $(\alpha \in \mathbb{R}^m, \alpha > 0)$
β	:	Vector of absolute value of slopes for $d_{a,l}$ $(\beta \in \mathbb{R}^m, \beta > 0)$
B	:	Diagonal matrix with $\beta_{a,l}$ as diagonal elements $(B \in \mathbb{R}^{m \times m})$
T	:	$m \times n$ matrix with (i, j) element=1 if itinerary i utilizes staynight j, 0 otherwise
r	:	Vector of itinerary prices $(r \in \mathbb{R}^m)$
c	:	Vector of staynight capacity $(c \in \mathbb{R}^n)$
I^s	:	Set of itineraries that utilize staynight s
S^i	:	Set of staynights that itinerary i utilizes

(P) with linear demand in matrix form can be expressed as following.

$$(P2) \qquad \max_r \quad \alpha' r - r' B r \tag{5}$$

$$s.t. \quad \alpha - Br \geq 0 \tag{6}$$

$$T'(\alpha - Br) \leq c \tag{7}$$

The following property hold for the optimal solution of $(P2)$.

P2.1 $r^* \geq \frac{1}{2}B^{-1}\alpha$

Proof. Assume that r^* is an optimal solution of $(P2)$ with $r_i^* < \frac{\alpha_i}{2\beta_i}$. Let r^0 be same as r^* except for the i^{th} element, with $r_i^0 = \frac{\alpha_i}{2\beta_i}$. Then $\alpha_i - \beta_i r_i^* > \alpha_i - \beta_i r_i^0 = \frac{\alpha_i}{2} > 0$ and $\alpha_j - \beta_j r_j^* = \alpha_j - \beta_j r_j^0, \forall i \neq j$ and hence r_0 is a feasible solution for $(P2)$. Now $r_i^*(\alpha_i - \beta_i r_i^*) < r_i^0(\alpha_i - \beta_i r_i^0)$ and $r_j^*(\alpha_j - \beta_j r_j^*) = r_j^0(\alpha_j - \beta_j r_j^0), \forall i \neq j$ and r^* cannot be an optimal solution, thus $r_i^* \geq \frac{\alpha_i}{2\beta_i}, \forall i \in I$. □

P2.2 If $r_i^* > \frac{\alpha_i}{2\beta_i}$ then $c_s = \sum_{k \in I^s} \alpha_k - \beta_k r_k^*$ for at least one element in S^i

Proof. Assume r^* is an optimal solution of $(P2)$ with $r_i^* > \frac{\alpha_i}{2\beta_i}$ and $c_s > \sum_{k \in I^s} \alpha_k - \beta_k r_k^*, \forall s \in S^i$. Let $\delta = \min(\min_{s \in S^i} c_s - (\sum_{k \in I^s} \alpha_k - \beta_k r_k^*), \beta_i r_i^* - \frac{\alpha_i}{2})$, which is a positive number. Let r^1 be same as r^* except for the i^{th} element, with $r_i^1 = r_i^* - \frac{\delta}{\beta_i}$. Then $r_i^* > r_i^1 \geq \frac{\alpha_i}{2\beta_i}$ so $\alpha_i - \beta_i r_i^1 \geq \alpha_i - \beta_i r_i^* \geq 0$. Also, by definition of δ, $c_s \geq \sum_{k \in I^s} \alpha_k - \beta_k r_k^1, \forall s \in S^i$ hence r^1 is a feasible solution. Since revenue for itinerary i, $R_i(r_i) = r_i(\alpha_i - \beta_i r_i)$, decreases as r_i increases for $r_i \geq \frac{\alpha_i}{2\beta_i}$, $R_i(r_i^*) < R_i(r_i^1)$ and the optimality of r^* is violated. □

By substituting r with $\frac{1}{2}B^{-1}\alpha + B^{-\frac{1}{2}}x$, where $x \in \mathbb{R}^m$, we can create an equivalent problem with x as the decision variable. With this substitution, the objective function becomes $\frac{1}{4}\alpha' B^{-1}\alpha - x'x$. Since $\frac{1}{4}\alpha' B^{-1}\alpha$ is constant, solving the following problem is equivalent to solving $(P2)$.

$$
\begin{aligned}
(P3) \qquad & \min_x \quad x'x && (8)\\
& s.t. \quad K_1 x - l_1 \leq 0 && (9)\\
& \qquad\quad K_2 x - l_2 \leq 0 && (10)\\
& where \quad K_1 = B^{\frac{1}{2}}, K_2 = -T'B^{\frac{1}{2}}, l_1 = \frac{\alpha}{2}, l_2 = c - \frac{1}{2}T'\alpha && (11)
\end{aligned}
$$

Solving $(P3)$ can be interpreted as finding the minimum distance between origin and the polyhedron defined by constraints (9) and (10) (call this polyhedron H) . We can identify the following properties about $(P3)$ and its optimal solution.

P3.1 From property P2.1, the optimal solution of $(P3)$, x^*, is nonnegative.

P3.2 If the s^{th} element of l_2 is nonnegative, then the s^{th} element of $K_2x - l_2$ is always less or equal to zero and the s^{th} row of the inequality $K_2x - l_2 \leq 0$ is redundant.

P3.3 When all elements of l_2 are nonnegative, then constraint (9) is not required and the optimal solution is trivial, $x^* = 0$.

P3.4 If capacity for c_s is zero, $d_i(x_i^*) = 0, \forall i \in I^s$.

P3.5 From property P2.2, we know that if $x^* \neq 0$, at least one of the constraints in (10) is binding.

If we know the set of active constraints (let A) at the optimal solution for $(P3)$, we can show that solving the following problem is equivalent to solving $(P2)$:

$$(P4) \qquad \min_x \quad ||x|| \tag{12}$$

$$s.t. \quad K_i x - l_i = 0 \quad \forall i \in A \tag{13}$$

$$where \quad K = \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} \tag{14}$$

Proposition 1. *Solving $(P3)$ is equivalent to solving $(P4)$.*

Proof. Let H_S be the hyperplane defined by set of active constraints at the optimal solution (set A), $H_S = \{x | K_i x = l_i, \forall i \in A\}$. Let the optimal solution of $(P3)$ be x^*. Then $K_i x^* = l_i, \forall i \in A$ and $K_j x^* < l_j, \forall j \notin A$. Since x^* is in the interior of the polyhedron defined in H_S, $P_S = \{x \in H_S | K_j x^* \leq l_j, \forall j \notin A\}$, we can define an open ball $\epsilon_{x^*} \in H_S$ with x^* as the center, which all of its elements are elements of P_S. Then x^* has the minimum norm among elements in ϵ_{x^*} and is a local optimum for $(P4)$. Since the objective function of $(P4)$ is convex, x^* is also the global optimal solution of $(P4)$. By similar argument the optimal solution for $(P4)$ is also optimal for $(P3)$. □

Analytical solution to the least norm problem $(P4)$ is well known and can be calculated easily using matrix factorization methods such as Cholesky or QR factorization (for example, see Trefethen and Bau [54]). Using the least norm solutions and the properties above, we suggest an alternative algorithm for solving $(P3)$, the Minimum Norm Algorithm (MNA).

The primary idea of MNA is to utilize the least norm problem $(P4)$ as a subroutine to obtain the optimal solution for $(P3)$. Solving $(P4)$ only requires matrix algebra and therefore eliminates the need for Nonlinear Programming methods or solvers. Also, this algorithm is efficient when only a few of the staynights are highly constrained. This approach is similar to dual algorithms in the sense that it chooses infeasible solutions and add constraints until feasibility is obtained, and when feasibility is obtained for all constraints then we have an optimal solution.

One caveat to the least norm problem $(P4)$ is that the constraint matrix needs to be full rank. It is not a problem when the binding constraints consist of only capacity constraints (10), since no two staynights are utilized by exactly the same set of itineraries. When the binding constraints consist of both (9) and (10) constraints, the constraint matrix may not be of full rank. However this problem can be resolved by using row echelon reduction of the augmented constraint matrix. This algorithm needs to loop only through constraints in (10) according to property P3.5.

MNA algorithm using the least norm subproblem is outlined as following.

1. For all staynight s with zero capacity, eliminate all x_i from $(P3)$ for all $i \in I^s$ (P3.4).

2. For all rows in (10) such that corresponding element in l_2 is nonnegative, eliminate the constraint from (10) (P3.5) and let the remaining set be R. If all constraints in (10) are eliminated then $x^* = 0$ (P3.3), otherwise proceed to the next step.

3. For each constraint in R, $k_s x - l_s \leq 0, s \in R$, solve the following least norm problem

$$\min_x \quad ||x|| \tag{15}$$

$$s.t. \quad k_s x = l_s \tag{16}$$

   to obtain the least norm solution $\bar{x}^s = k_s'(k_s k_s')^{-1} l_s$.

4. For all $s \in R$, check the feasibility of $\bar{x}^s$ to $(P3)$. Let R_1 be set of s where $\bar{x}^s$ is feasible and R_0 be set of s where $\bar{x}^s$ is not feasible for $(P3)$.

5. Let $z = \min_{s \in R_1} ||\bar{x}^s||$ and $\bar{s}$ be $s \in R_1$ where the minimum is attained. z is the initial upper bound of the optimal value. If $R_0 = \emptyset$ then we have the minimum distance to all inequality constraints, and the optimal value of $(P3)$ is the minimum of all minimum distances, i.e., $x^* = \bar{x}^{\bar{s}}$. Otherwise proceed to the next step.

6. For all $s \in R_0$, the minimum distance to the feasible area of hyperplane $\{x | k_s x = l_s\}$ is attained at the intersection with other constraints. To compute the minimum distance, initialize $R_s = \{s\}$ and U as the set of all constraints and take the following steps:

(a) Compute $l - K\bar{x}^s$ and find the most negative value among constraints in U. Let the corresponding constraint index be i.

(b) If i is in (9), perform a row echelon reduction to the augmented matrix

$$\begin{bmatrix} K_{R_s} & l_{R_s} \\ K_i & l_i \end{bmatrix} \tag{17}$$

to 1) assure the current set of constraints are linearly independent (row echelon reduction returns full rank matrix) and proceed to the next step or 2) eliminate a redundant constraint (row echelon reduction returns rows with only zero as elements, eliminate these rows) and proceed to the next step or 3) find that current set of constraints are infeasible (row echelon reduction returns rows with all zeros except for the last element), update $U = U \setminus \{i\}$ and repeat 6a.

(c) Update $R_s = R_s \cup \{i\}$. Solve $(P4)$ again with constraints in R_s. Let the optimal solution be $\bar{x}^{s_i}$. If $\bar{x}^{s_i}$ is feasible for $(P3)$, check if $||\bar{x}^{s_i}|| < z$, if both conditions are satisfied then update z with $||\bar{x}^{s_i}||$ and R_0 with $R_0 \setminus s$, and let $x_z = \bar{x}^{s_i}$. If $\bar{x}^{s_i}$ is feasible for $(P3)$ but $||\bar{x}^{s_i}|| \geq z$, then update R_0 with $R_0 \backslash s$. If $R_0 = \emptyset$ then the optimum is z attained at $x^* = \bar{x}^{\bar{s}}$. If $\bar{x}^{s_i}$ is infeasible, update $U = U \setminus \{i\}$.

(d) Repeat 6a. z is the optimal solution when $R_0 = \emptyset$.

This algorithm iterates through each eligable constraints until it finds a feasible minimum norm point. In the best case this algorithm requires iterations equal to number of remaining constraints after eliminating the trivial constraints ($\leq m + n$). In the worst cast, it iterates through n^{m+n} iterations of $(P4)$, when the feasible region within each constraint hyperplane is constrained by all of the remaining constraints. Typically the iterations require algebra of very low rank matrices, from which the algorithm achieves its efficiency.

4.3 Computational Study of the Minimum Norm Algorithm

In order to compare the performance of the Minimum Norm Algorithm to existing QP algorithms, we performed a computation study comparing the run time and number of iterations that each algorithm performed to solve a price optimization problem. Six hotel demand profiles estimated in Chapter 3 were chosen, with two hotel profiles from each of the three locations. Three different occupancy levels were tested, with low/medium/high occupancy level designed to average 50%/70%/90% occupancy at average price. MNA was compared to three different QP algorithms in ILOG CPLEX, primal simplex, barrier algorithm and dual simplex. MNA was performed with statistical software R. All algorithms were run on the single core of the Intel(R) Core(TM)2 CPU T9400 2.53GHz Windows XP machine with 3GB RAM. In each scenario the optimizer solved a 112 day horizon problem with 112 itineraries and 28 staynights. The results of the computational study are reported in Table 13.

Table 13 reports the CPU time and number of iterations performed for each algorithm. Additionally for MNA the average number of constraint rows in subproblem $(P4)$ is reported. The results indicate that while MNA requires more iterations than other algorithms, the number of rows in the matrix to solve for each iteration is much smaller than typical QP algorithms require. Simplex and barrier algorithms typically solve matrices with row number in the order of number of constraints (in the above example, 140) and the dual simplex algorithm typically solves matrices with number of rows equal to the number of variables (in the above example, 112), whereas the average number of rows in the MNA constraint matrix is 7.3 in our scenarios. As a result, the CPU time elapsed for MNA is on average 38% of the barrier algorithm and 56% of the simplex algorithm. MNA run time is comparable to the dual simplex

Table 13: Performance of MNA compared to QP algorithms

	MNA			Simplex		Barrier		Dual Simplex	
	Time	Iter.	Rows	Time	Iter.	Time	Iter.	Time	Iter.
AP1(L)	0.15	38	3.26	1.23	144	1.93	9	0.23	5
AP1(M)	0.17	37	3.16	0.86	146	1.45	9	0.29	5
AP1(H)	1.02	507	11.52	1.36	134	1.84	8	0.78	22
AP2(L)	1.10	553	11.56	1.34	136	2.03	8	0.23	22
AP2(M)	1.58	703	13.52	1.81	128	1.36	7	0.75	26
AP2(H)	1.50	678	13.06	1.45	137	1.96	8	0.70	25
SUB1(L)	0.13	0	0.00	0.90	125	1.89	8	0.93	0
SUB1(M)	0.13	0	0.00	1.42	125	1.90	8	0.76	0
SUB1(H)	0.14	25	3.00	1.51	138	2.50	8	0.84	5
SUB2(L)	0.13	13	2.15	1.17	144	0.89	9	0.78	3
SUB2(M)	0.33	193	6.19	1.46	135	1.42	9	1.29	11
SUB2(H)	0.33	181	6.17	0.84	138	2.07	7	0.75	11
URB1(L)	1.17	571	10.71	1.43	117	1.90	8	0.62	23
URB1(M)	1.05	523	10.20	1.36	111	1.56	8	0.21	21
URB1(H)	0.68	360	8.69	1.48	119	2.60	9	0.20	16
URB2(L)	0.54	300	7.65	0.12	121	1.32	9	0.73	14
URB2(M)	1.15	571	11.50	0.73	117	1.92	8	0.71	23
URB2(H)	1.02	523	10.20	1.45	111	1.40	8	0.76	21
Avg	0.68	320.89	7.36	1.22	129.22	1.77	8.22	0.64	14.06

Table 14: Performance of MNA compared to QP algorithms in low demand scenarios

Algorithm	Avg. Time	% Difference with MNA time
MNA	0.52	0%
Simplex	0.92	44%
Barrier	1.17	56%
Dual Simplex	0.68	25%

algorithm since it uses an approach similar to the dual simplex, but is faster when the problem is "easier" to solve, with less number of constrained staynights. To further test the performance of MNA at low demand level where it is expected to outperform the existing QP algorithms, 60 instances of expected occupancy level of 50% were randomly generated. Table 14 illustrates the average run time of the four algorithms in the 60 trials. MNA algorithm was the most efficient algorithm among the tested algorithms. On average MNA performed 25% faster than the dual simplex and 44%, 56% faster than the primal simplex and the barrier algorithm respectively. Two tailed T test for the average values of MNA run times against the algorithms confirmed that the MNA algorithm was significantly faster than the simplex and barrier algorithms at 99% confidence level and 95% confidence level against the dual simplex algorithm. Note that the naive implementation of the MNA algorithm is comparable to the highly optimized CPLEX algorithms and additional performance improvements of the MNA algorithm is expected with further tuning. Overall, the Minimum Norm Algorithm, which can be easily implemented in general purpose high level languages such as C++ or R and does not require an optimization solver, performs comaparable to the existing QP algorithms and can be more effective than other algorithms when the hotel capacity is not too restricted relative to the demand level.

CHAPTER V

THE IMPACT OF THE DEMAND MODEL ON HOTEL PRICE OPTIMIZATION

5.1 Introduction

In this chapter we address the research question of how much revenue is lost by using an incorrect demand model. We explore the two most common cases of demand model misrepresentations: (1) Linear demand model with limited driver variables is used when the real demand model has more driver variables and (2) Incorrect functional form is used to represent the demand. More specifically, we test the performance of price optimization when a linear model is used when the true demand is an exponential function of price, and the opposite case of when an exponential model is used when the true demand is linear in price. These questions are answered by using price optimization on simulated demand where the true demand function is generated according to the results in Chapter 3.3. Revenues earned from using different models are compared to the revenues earned using the true demand model to measure the impact of different demand models.

5.2 Literature Review

The literature in revenue management with price sensitive demand can be classified into two categories: single period problems and multi-period problems. Table 15 summarizes the modeling literature reviewed in this section.

The single period price optimization problems are variations of the newsvendor problem with price-demand relationship added to the classic settings where price is originally assumed to be a fixed value. Whitin [62] is among the first papers in the single period revenue management literature to include price as a decision variable. The author shows how to optimize price and order quantity where demand is a deterministic linear function of the price. Mills [47] studies price optimization with demand uncertainty when the demand has additive random error, $D(p, \epsilon) = y(p) + \epsilon$, where $D(p, \epsilon)$ denotes demand at price p with random error ϵ and $y(p)$ is a decreasing function of price. Karlin and Carr [34] present a case where the demand has multiplicative random variation, $D(p, \epsilon) = y(p)\epsilon$.

More recent single period studies include Urban and Baker [55] and Weng [60] who examine price and markdown/discount optimization under exponential demand models for general products. Petruzzi and Dada [48] apply both the log linear and linear demand models to the newsvendor problem. Emmons and Gilbert [23] and Khouja [36] use linear demand model for catalogue goods price optimization and the newsvendor problem, respectively. Lau and Lau [42] compare the optimal solution for a multi echelon price optimization involving the manufacturer, wholesaler, and the retailer using exponential, linear and log linear demand models. Lau and Lau [42] and Petruzzi and Dada [48] both compare the theoretical optimal solutions under different demand models for a single period optimization problem, however they do not extend their work beyond a single period. In our study, we compare the price optimization performance simulated over multiple time periods under realistic hotel business environments.

The studies of single period pricing problem reviewed above are focused on analyzing the static optimal solution within the single period. Contrarily, multiple period problems in general are studied under the dynamic pricing setting where the price is updated over time as demand drivers and capacity fluctuate. Among the multiple period price optimization problems, we focus our review to those for fixed capacity and perishable products, as the hotel revenue management problem belongs in this category.

Pioneering work of price optimization within this category was done by Gallego and Van Ryzin [24]. The authors model the pricing problem as a stochastic dynamic program with demand arriving according to the Poisson distribution. They develop a deterministic heuristic which is found to be asymptotically optimal as the expected sales tend to infinity. Bitran and Caldentey [13] and Elmaghraby and Keskinocak [22] give overview of pricing models used in revenue management, a large part stemming from approaches in Gallego and Van Ryzin [24].

The work of Besbes *et al* [11] stems from similar research questions raised in our study. The authors observe that not enough attention has been given to the performance of an optimal decision using the estimated models compared to the attention given on the statistical validity of the demand model. They take the consumer choice model and an unlimited capacity profit optimization problem to develop a performance based metric to measure the impact on the optimization. They apply this metric to an auto lender sales data and show that a simple logit model with "bad" model fit for four out of eight instances does well under the performance based test for seven out of eight instances. While Besbes *et al* [11] focused their efforts on developing a performance based metric, our research emphasizes measuring the revenue impact of commonly used demand models in the hotel business (see Khouja [36], Lus and Muriel [46], Witt and Witt [63], etc.) and providing insights on effective demand models for price optimization.

While the dynamic programming approach is the most commonly researched in academia (such as in Bitran and Caldentey [13], Elmaghraby and Keskinocak [22],

Table 15: Summary of demand model literature

Literature	Model type	Dependent variable	Explanatory variables
Urban and Baker [55]	Log linear	Period demand	Price Inventory level Time trend
Weng [60]	General	Annual demand	Price
Petruzzi and Dada [48]	Linear Log linear	Period demand	Price
Emmons and Gilbert [23]	Linear	Period demand	Price
Khouja [36]	Linear	Period demand	Price
Lau and Lau [42]	Linear Log linear Exponential	Period demand	Price
Gallego and Van Ryzin [24]	General Log linear	Demand intensity	Price Time
Besbes *et al* [11]	Logit	Purchase probability	Price

Gallego and Van Ryzin [24]), applying dynamic programming to a large scale network problem is known to be a challenging (the challenge described as "curse of dimensionality" by Bellman [7]). In our work we take the quadratic programming framework for multiple period price optimization using the linear demand function. With this approach we are able to solve a dynamic network optimization problem close to the real life hotel price optimization problem. While all of the literature that examine the different demand models reviewed in this section either ignores capacity constraints (see Bitran and Caldentey [13], Gallego and Van Ryzin [24], Urban and Baker [55], etc.) and/or focus only on theoretical differences of the optimal solutions (for example, see Lai [41]), our research estimates demand models using actual hotel data and measures the performances of dynamic price optimization in the simulated environment close to the actual hotel business.

5.3 Impact of Demand Model on Price Optimization

In this section we measure the impact of demand model specification on the price optimization performance. We start by stating the specific research questions followed by the experimental design. In Section 5.3.5, we present the results and findings of the experiments.

5.3.1 Research Questions

In this section we address the following research questions:

1. How much revenue is lost by using a simple linear demand model when the true demand model is more complicated?

   Demand models used in price optimization in practice are often simpler, with fewer drover or segmentation variables, than the models estimated in Chapter 3.3 (see Lim [44] for typical demand models used). This is partly to ensure a more robust estimation, and partly due to the unavailability of certain driver variable information. Hence, measuring the impact of using simple models versus the "true" demand model can provide the hotels a valuable insight on what the benefits of demand model analysis.

2. How much revenue is lost by using the wrong functional form of demand model?

   (a) When linear demand model is used while the true demand model is exponential.

   (b) When exponential demand model is used while the true demand model is linear.

   Linear and exponential models are most widely used models in practice (Lim

[44]), but there can be cases where one model is used when the other is a better representation of actual demand. We measure the lost revenue due to misrepresentation of the functional form as well as the lost revenue due to using simple demand models through simulation experiments.

5.3.2 Overview of Simulation Process

Figure 11 illustrates the simulation process at a high level. One hotel for each of the three locations is chosen from the 23 sample hotels. For these 3 hotels, the best fit models estimated through the process described in Chapter 3 are assumed to be the true demand model. Note that the true demand model has different intercept terms by days prior group, length of stay, and day of week according to the corresponding coefficients. Using the historical demand data, the coefficients for test models are estimated with regression analysis.

The simulator generates demand and optimizes revenues for a moving 28-day horizon for 168 days, assuming that the maximum length of stay is four nights (following our observation in Section 3.3.3). For linear demand models, we solve a quadratic programming problem (($P2$) in Chapter 4) to optimize itinerary prices for the optimization horizon. For exponential demand models, we use the *alabama* package in statistical software R to solve a constrained convex nonlinear optimization problem form ((P) in Chapter 4 with exponential demand). Demand and capacity in these simulations are for retail segments only, since we focus on optimizing the retail itinerary rates. Optimal prices are calculated for each itinerary (arrival date, length of stay) based on the test demand model. Once the optimal price is calculated, we can use the true demand model to simulate the demand at that price, calculate demand that the hotel can accommodate under the hotel's capacity, and compute the corresponding revenue.

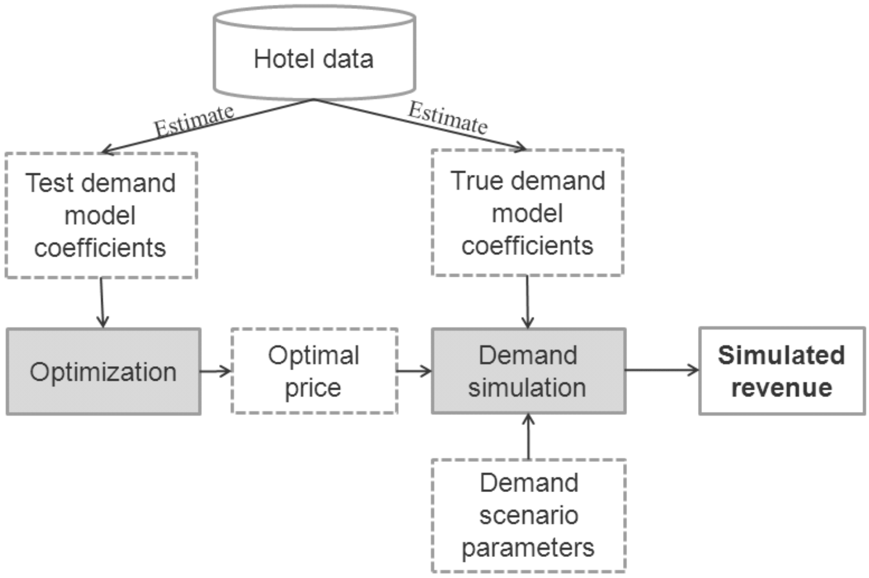


Figure 11: Simulation process flow diagram

The simulation study is performed as following for each scenario. First the "true" and "test" models are defined according to the simulation scenario. Also the capacity is set according to the demand level in each scenario, specifically, the average occupancy at average price according to the true model should be equal to the base occupancy for each scenario. For all itineraries within the simulation horizon, demand is generated according to the defined true demand model under randomness respective to the demand scenario as shown in the example in Section 5.3.3 where how the demand is generated to simulate what happens in the real world is described in detail.

In our simulation, the price recommendation is updated at every reading date. However, in cases where optimization horizon is longer with sparser demand towards the end of horizon, weekly update of prices can also be exercised to improve performance.

5.3.3 Design of Experiments and Demand Generation

Functional forms of demand models tested in this study are linear and exponential. Demand models compared in this experiment is summarized in Table 16. L2 model

Table 16: Demand models compared

Model	Functional Form	Explanatory Variables
L1	Linear	Price
L2	Linear	Price, DOW
L3	Linear	Price, DOW, Days prior group, LOS group
EXP*	Exponential	Price, DOW, Days prior group, LOS group

is not tested for the airport hotel since airport hotels do not have day of week explanatory variables.

Demand scenarios for numerical experiments are designed based on the real hotel demand characteristics described in Table 8. Different levels of occupancy and coefficient of variation of staynight demand are combined to create the demand scenarios. Experiments are performed at combinations of three occupancy levels (50%, 70%, and 90% approximating the minimum, average, and maximum values in Table 8). Coefficient of variation of simulated staynight demand are 0.25 and 0.40, close to the 25 and 75 percentile c.v. value of the real data. The coefficient of variation is assumed to be constant at all price levels. This assumption is based on our empirical observation on hotel booking data shown in Section 3.3.2. Individual arrival date and LOS distribution are simulated based on the true demand model with random errors.

Simulated demand is generated based on the following assumptions and inputs.

1. Average itinerary demand (deterministic portion of demand) is a function of the itinerary price:

$$\bar{d}(r) = \alpha - \beta r \qquad \text{for linear true demand} \tag{18}$$

$$\bar{d}(r) = \alpha \cdot e^{\beta r} \qquad \text{for exponential true demand} \tag{19}$$

   where α, β are positive constants.

2. Each demand scenario is built using the two coefficients of variation (0.25, 0.4)

of staynight demand and the average occupancy rate.

3. Generated demand has multiplicative errors independent of the price (Karlin and Carr [34], Petruzzi and Dada [48]). Errors are normally distributed with constant variation parameter σ^2 and are designed to achieve staynight c.v. of 0.25 and 0.40 for low and high demand variation scenarios respectively.

$$d(r) = \bar{d}(r)(1+\epsilon) \tag{20}$$

$$\epsilon \sim N(0, \sigma^2) \tag{21}$$

When creating random variation of demand, both additive error and multiplicative errors are viable options and are commonly used in stochastic price optimization problems similar to ours (Petruzzi and Dada [48]). However, additive errors are independent of demand level, and hence the coefficient of variation varies at different price levels. Since we assume the coefficient of variation is constant at all price levels (the standard deviation is proportional to demand), multiplicative errors are more appropriate in our case.

After models are defined and the capacity and demand coefficients are initialized, steps performed for each day in the simulation horizon are outlined below. Assume the following demand model is used for both test and true demand:

$$
\begin{aligned}
&D(dp, DOW, LOS, price) = \\
&\quad 7.1 \\
&\quad + (1 \leq dp \leq 6) \times 10.95 - (7 \leq dp \leq 11) \times 6.03 - (12 \leq dp \leq 22) \times 7.12 \\
&\quad - (23 \leq dp) \times 2.97 - (LOS \geq 2) \times 3.72 \\
&\quad + (DOW = Mon, Tue, Wed, Thu) \times 2.72 + (DOW = Fri) \times 2.33 \\
&\quad + (DOW = Sat) \times 4.78 \\
&\quad - (LOS = 2 \quad \& \quad 1 \leq dp \leq 6) \times 4.59 - (LOS \geq 3 \quad \& \quad 1 \leq dp \leq 6) \times 8.2 \\
&\quad - (LOS \geq 3 \quad \& \quad 1 \leq dp \leq 6) \times 5.67 \\
&\quad - (DOW = Mon, Tue, Wed, Thu \quad \& \quad LOS \geq 2) \times 0.44 \\
&\quad + (DOW = Fri \quad \& \quad LOS = 2) \times 3.45 - (DOW = Sat \quad \& \quad LOS \geq 2) \times 4.68 \\
&\quad - 0.7 \times \frac{price}{100}
\end{aligned}
$$

where dp is booking days prior to arrival date, LOS is length of stay and DOW is day of week.

1. For all itineraries in the optimization horizon, remaining demand model is calculated as sum of all remaining days prior demand models. For example, for booking date of Day 128, itinerary (arrival date=128(Mon), length of stay 1) has the following remaining demand:

   $$D(dp = 0, DOW = Mon, LOS = 1, price) = 9.82 - 0.7 \times \frac{price}{100}$$

2. Optimal prices are generated by optimizing the remaining revenue using the test demand model for the optimization horizon under capacity constraints. Assume the optimal price calculated be \$516.58.

3. Assuming the optimal prices based on the test demand model are implemented, true demand is generated for each itinerary in the optimization horizon for the simulating day using the true demand model (for respective day of week, length of stay, days prior and input price). In our example,

$$D(dp = 0, DOW = Mon, LOS = 1, \$516.58) = 6.2$$

4. Random number generated according to the coefficient of variation is multiplied to the true demand. In our example, assume the random number 0.93 is generated from normal distribution of mean equal to 1 and coefficient of variation of 0.25.

$$SimulatedDemand = 6.2 \times 0.93 = 5.8$$

5. For any staynight in the optimization horizon, if the generated demand for the staynight exceeds the remaining capacity, demand is constrained by taking shorter length of stay of earlier arrival dates first. For example, assume that in our example we have remaining capacity of 9 and there are following itinerary demands competing for this capacity on given booking date:

Itin 1	(arrival date=3, LOS=4) has demand of 1
Itin 2	(arrival date=4, LOS=3) has demand of 0
Itin 3	(arrival date=4, LOS=4) has demand of 1
Itin 4	(arrival date=5, LOS=2) has demand of 3
Itin 5	(arrival date=5, LOS=3) has demand of 0
Itin 6	(arrival date=5, LOS=4) has demand of 1
Itin 7	(arrival date=6, LOS=1) has demand of 4
Itin 8	(arrival date=5, LOS=2) has demand of 3
Itin 9	(arrival date=5, LOS=3) has demand of 1
Itin 10	(arrival date=5, LOS=3) has demand of 0

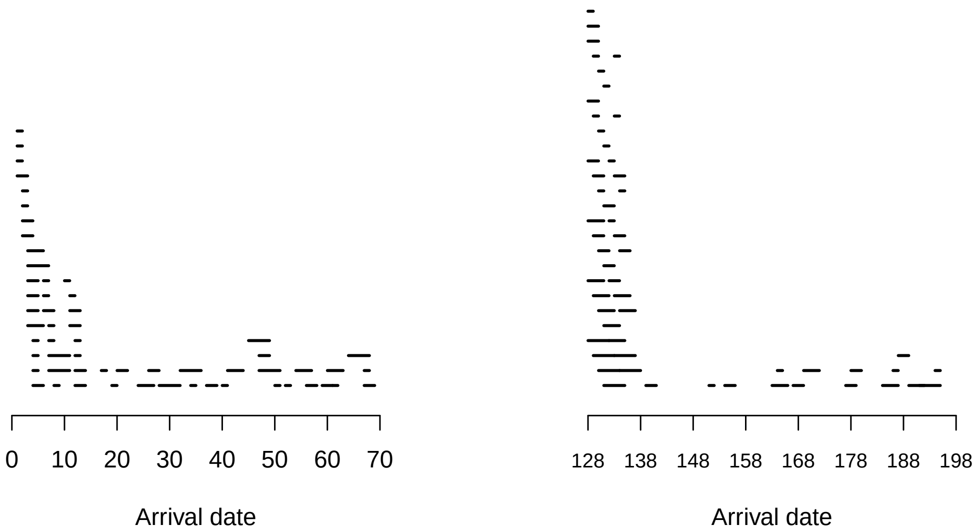
then itinerary 1$\sim$6 demand is fully accommodated and only 3 out of 4 of itinerary 7 demand is accommodated. Itinerary 8$\sim$10 demand is rejected.

6. Simulated revenue for the day is sum of all itinerary revenue in the optimization horizon, which is constrained itinerary demand multiplied by the optimal price.

7. Capacity for each staynight is updated as previous capacity less the constrained itinerary demand.

The simulation process mimics what happens in the real world, where a portion of total demand books for each days prior (with more demand booking closer to arrival date) and price optimization recommends price given the current remaining capacity and projected remaining demand. Figure 12 illustrates the booking distribution at a typical booking date in the actual data of the hotel simulated in the example above (Figure 12(a)) and the simulated demand at day=128 (Figure 12 (b)) from the steps outlined above. Each line segment in Figure 12 represents a simulated booking, with starting point at the arrival date and ending point at the last stay date of the itinerary. The length of the line segment represents the length of stay. The number of total lines corresponding to arrival date=128 and LOS=1 is 6, indicating 6 bookings arrived at day=128 for the itinerary of arrival date=128 and LOS=1 as illustrated in the example above (rounded from the simulated demand of 5.8). The distribution of demand shown in Figure 12(b) is similar to what is observed in practice as shown in Figure 12 (a), with most of the demand concentrated in itineraries within 7 days out to arrival date.

The detailed steps simulations process are outlined as follows. We assume that the demand starts to arrive at days prior 28, and the demand at that time is sum of all expected demand for days prior 28 or higher. This allows us to apply the initial optimal rate (optimal rate at the beginning of the booking horizon when there are



(a) Visualization of actual demand (b) Visualization of simulated demand

Figure 12: Visualization of simulated demand at day 128

no bookings on hand) to all the bookings booked at days prior 28 or higher when the arrival date hasn't rolled into the simulation horizon.

1. Set length of optimization and simulation horizon.

2. Read in the coefficients for the true and test demand model as well as capacity for the hotel and scenario simulated.

3. For each simulated day in the simulation horizon, for each itinerary in the optimization horizon, for each days prior less or equal to the current days prior (arrival date - simulated day), calculate the true and test demand model coefficients corresponding to the day of week, days prior and length of stay as shown in the demand generation example above. For itineraries with arrival date 28 or more days after the simulated day, the demand model should be sum of all demand models for days prior 28 and beyond (see explanation above).

4. For each simulated day in the simulation horizon do the following.

(a) For each itinerary in the optimization horizon, calculate the remaining demand model by adding all future days prior test demand models.

(b) Solve for optimal rates for each itinerary which maximized the total revenue in the optimization horizon.

(c) Apply the optimal rate to the true demand model of itineraries within the optimization horizon at corresponding day of week, days prior and length of stay.

(d) For each staynight in the optimization horizon, check if there are remaining capacity. If there are, accept generated demand following the order specified in the demand generation example above. Update the capacity by subtracting number of accepted bookings from the current capacity.

The total revenue gained at the end of the simulation horizon is compared to measure the impact of demand model performance for different scenarios.

5.3.4 Simulation Parameter Determination

For a simulation study, sufficient number of repetition and warm up period to ensure the statistical significance of the results need to be determined. We perform 10 repetition for each scenario, which results less than 5% standard deviation of revenue and average staynight occupancy for all scenarios and provide tight confidence intervals.

For first 28 days in the simulation horizon, the demand is not fully realized since the demand is generated beginning 28 days prior to the arrival date. Warm up period should be larger than 28 days for this reason. To determined the length of the warm up period, we follow the standard techniques in Kelton *et al* [35]. Figure 13 displays occupancy across simulation horizon for suburban location/base occupancy

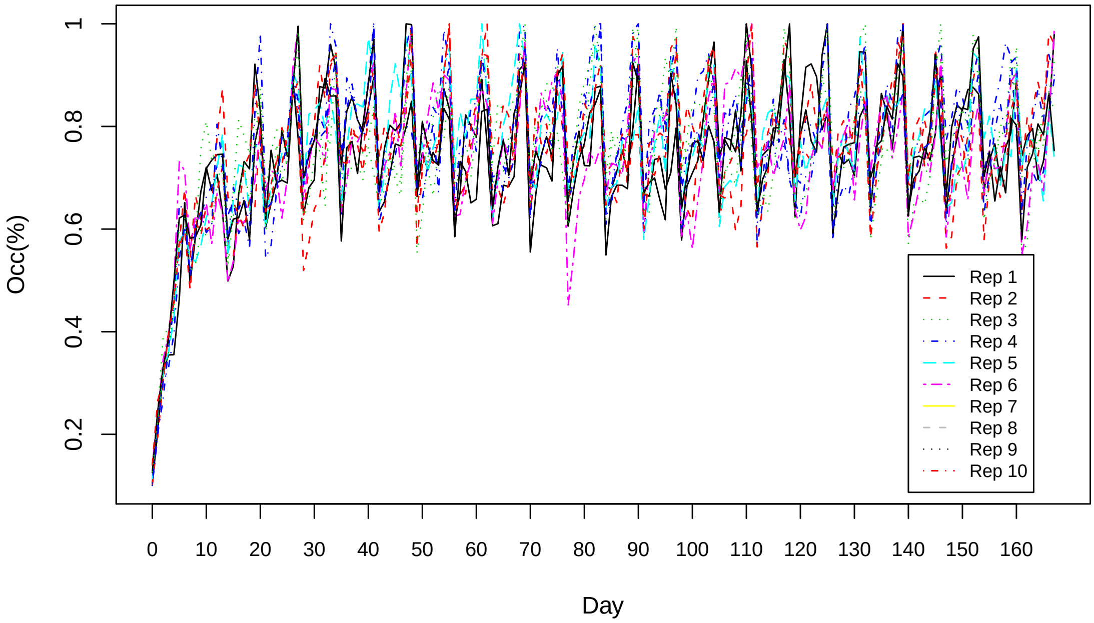


Figure 13: Occupancy across simulation horizon for suburban location, base occ.=0.7, c.v.=0.25 scenario

0.7/coefficient of variation 0.25 with true model L3 and test model L3. Other scenarios follow similar trend.

In Figure 13, we observe the occupancy stabilizing between day 20 and 30. Following Kelton *et al* [35], we can statistically determine if warm up period of 4 weeks (28 days) is sufficient by performing a t-test between occupancies between week 5 and 6 (day 29∼42) and week 7 and 8 (day 43∼56). The t test could not reject the null hypothesis that the two periods have same average occupancies at 90% confidence, which indicates the warm up period of 4 weeks is sufficient. Same test was performed for revenue and the same conclusion was reached. Following these results we employ the warm up period of 4 weeks.

Table 17: Ratio of simulated revenues between price optimization of simple demand models to true demand models (L3)

Base occ.	c.v.	Test model	Revenue ratio Airport	Suburban	Urban
50%	0.25	L1	67.6%	83.7%	60.2%
		L2		87.2%	86.4%
	0.40	L1	67.4%	83.2%	60.1%
		L2		87.2%	85.9%
70%	0.25	L1	67.6%	80.7%	53.1%
		L2		85.0%	76.0%
	0.40	L1	67.8%	80.2%	53.7%
		L2		85.2%	76.3%
90%	0.25	L1	67.6%	79.3%	47.0%
		L2		84.6%	67.3%
	0.40	L1	68.0%	78.8%	47.9%
		L2		85.1%	68.4%

5.3.5 Simulation Results

Table 17 shows the percentage of the true demand model (L3) revenue achieved when using test demand models L1 and L2. The values displayed in Table 17 are averages of 10 independent replications.

Using the simplest linear model, L1, our simulation results in 47%∼87% of the revenue earned with the true demand model. These percentages vary widely, with the urban hotel with more complex true demand model having the highest revenue loss from using the L1 model. L2 model with day of week variables achieves about 22%∼30% more for the urban hotel and about 4%∼6% more for the suburban hotel compared to when L1 model is used. Performance of simpler demand models L1 and L2 are worse for high occupancy scenarios in urban and suburban hotels, whereas the performance of the airport hotel is not sensitive to the occupancy level. For the urban hotel, L2 model performs better than the L1 model more for the high occupancy scenarios. The variability of demand (c.v.) does not affect the revenue performaces for all scenarios.

Table 18: Ratio of simulated revenues between price optimization of linear and exponential demand models to true demand models

True model	Test model	Base occ.	c.v.	Revenue ratio Airport	Suburban	Urban
L3	EXP	50%	0.25	61.8%	56.5%	85.5%
			0.40	61.5%	56.6%	85.4%
		70%	0.25	64.7%	57.2%	86.2%
			0.40	64.8%	57.6%	86.2%
		90%	0.25	67.7%	59.9%	89.6%
			0.40	68.0%	60.6%	90.1%
EXP	L3	50%	0.25	84.4%	82.0%	87.5%
			0.40	84.4%	82.0%	87.5%
		70%	0.25	89.5%	85.2%	87.7%
			0.40	89.0%	84.4%	87.7%
		90%	0.25	87.9%	86.%8	92.1%
			0.40	89.6%	88.3%	95.1%

For the impact of functional forms of demand models, Table 18 reports the simulated revenue when using linear (L3) and exponential (EXP) models as a percentage of revenue achieved with the true model.

When the true demand is generated by the linear model L3, price optimization with exponential demand model yields 57%∼90% of the revenue compared to using the true demand model. In general, urban hotel achieves the highest revenue percentages, contrary to the results in Table 17. When the true demand is generated by the exponential model, price optimization with linear demand model yields 82%∼95% of the potential revenue, having less revenue impact than when the true model is linear and exponential model is used. The revenue performance of linear model is similar across all location types. In general, the revenue impact was larger under low demand level scenarios.

CHAPTER VI

CONCLUSION AND FUTURE RESEARCH DIRECTIONS

In this thesis we first investigated the two common assumptions of the hotel revenue management: 1) late booking customers are willing to pay higher rates and 2) weekday demand higher than the weekend demand. Empirical analysis presented in Chapter 2 showed that hotel rates retail customers pay do not increase as the day of arrival approaches. We also showed that weekday retail demand is not really stronger than the weekend demand. Both these findings suggest that dynamic pricing customized to retail segment based on dynamically estimated price demand relationship is necessary in order to achieve the benefits through revenue management.

In Chapter 3, we identified the major demand drivers of hotel demand which are days prior, day of week, length of stay, previous days prior demand and interactions of these variables. We also observed that the airport hotel demand shows a similar pattern across all day of week unlike the demand for other locations. Also, the hotels in same locations shared similar demand model structures.

One noticeable observation from Chapter 3 is the inelastic response of demand to price variations. The estimated elasticity from this analysis was between -0.06 and -0.38, which is comparable to the values reported in the hotel literature: -0.35$\sim$-0.57 (Hiemstra and Ismail [31], 310 properties), -0.8$\sim$-1.8 (Damonte *et al* [20], Columbia County, South Carolina (1992$\sim$1995)), -0.1$\sim$-0.3 ([20], Charleston County, South Carolina (1992$\sim$1995)), and -0.12$\sim$-0.13 (Canina and Carvell [14], 480 urban hotels in metropolitan areas). The low price elasticity of hotel demand compared to related products like air travel (-0.5∼-1.5, [27]) suggests that market factors such as own and competitor hotel availability, overall market capacity, price of substitutable hotels and air fares should be considered in hotel demand prediction. Hence the relationship between air fares and hotel demand can be a promising area of research, as price of complementary products can have an impact on the demand according to Duvvuri *et al* [21].

In Chapter 4, we provided the general formulation for the hotel price optimization problem. We examined the analytical properties of the price optimization solution given the linear demand and proposed the Minimum Norm Algorithm which requires solving iterations of subproblems solvable with matrix algebra. We compared the computational results of MNA with existing Quadratic Programming algorithms and showed that MNA shows comparable performance to existing algorithms. In the case where the capacity constraints are not very tight, MNA algorithm showed better performance than even the dual simplex algorithm, which performed best among the existing algorithms for the tested hotel price optimization problems.

In Chapter 5 we identified the impact of driver variables on price optimization performance to be significant, with the revenue lost due to using simpler demand models being as high as 53%. Using simpler demand models lead to substantial revenue loss, especially for the hotels that have demand influenced by more variety of driver variables and interactions of those driver variables. This impact is even greater when the demand level is high and hotels frequently face capacity limitation, indicating that having an accurate demand forecast and pricing is crucial in realizing the revenue potential especially in high demand markets.

Impact of the functional form was less drastic compared to the impact from driver variables, but still considerable. Estimating demand model with a different functional form than the true demand resulted on average of 24% revenue loss for linear and exponential demand, ranging from 5% to 43% depending on the simulation environment. In real life, the true demand model is almost never linear or exponential, and the relative importance of identifying the correct driver variables becomes greater compared to the importance of representing the demand with the right funtional form.

The findings from our research imply that correctly representing the demand model in price optimization is crucial to its success. In order for hotels to realize the maximum potential revenue through pricing, efforts should be focused on identifying the major driver variables influencing demand including the ones that we found to be significant.

APPENDIX A

STAYNIGHT PRICE OPTIMIZATION

Pricing structure also has high impact on revenue management. Currently the hotels use two kinds of pricing structure, staynight pricing or length-of-stay (LOS) pricing. Staynight pricing in hotel revenue management refers to a pricing practice where pricing is done on an individual staynight level. On the other hand, in LOS pricing, there are no staynight prices and each itinerary (arrival date and length of stay combination) is independently priced. With LOS pricing hotels have more flexibility in pricing different itineraries, whereas with staynight pricing itinerary prices are restricted to be the sum amount of individual staynights. Therefore price optimization using LOS pricing is expected to achieve higher revenue when compared to staynight pricing optimization.

Hotels have traditionally used staynight pricing and revenue management until recently. Figure 14 shows how a typical hotel product is sold - itineraries priced as sum of staynight prices. While the LOS pricing is desirable in terms of flexibility in pricing, it has been challenging to hotels to make a transition from staynight pricing to the full LOS pricing. Transition to LOS pricing requires substantial changes in organization, infrastructure, and marketing practices not to mention the cost of implementation. More importantly, the benefits of implementing LOS pricing has not been validated with quantitative analysis in the hotel industry. In this paper, we simulate the earned revenue of staynight vs. LOS pricing to quantify the revenue impact of the pricing structure. Also we develop a heuristic for staynight pricing and compared the performance of heuristic optimization vs. exact staynight optimization.

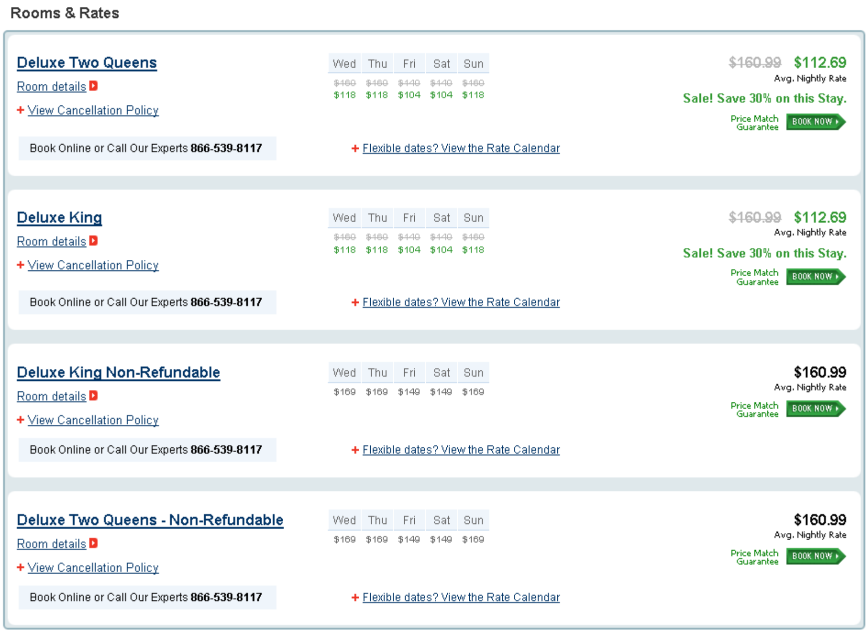


Figure 14: Hotel staynight pricing example from hotels.com

In this chapter, we research the question of how much revenue is lost by using the more restrictive staynight pricing instead of LOS pricing. We answer these questions by establishing a simulation environment where we can test the impact of demand models and pricing structure in hotel price optimization. Specifically, we compare the maximized revenue when both the price and quantity of sales are decision variables in optimization under various demand scenarios extracted from historical booking data.

When we use staynight price instead of LOS price in optimization, the demand function can no longer span the whole co-domain since the range of LOS prices are restricted to be sum of staynight prices. In this case we no longer have the guarantee that optimal sales quantity equals demand function at optimal price and we need to solve the bilinear optimization problem. To our best knowledge, there are no literature reporting implementation of bilinear optimization in the area of revenue management. In this paper, we do not attempt to develop a bilinear optimization algorithm but instead use discretized demand points to address this issue. Since the hotel sales quantity is always integer in real life, discretization using integer sales quantity gives an exact solution to the real world problem. The formulation of price optimization using discretized quantity is given below. In this formulation, price is the continuous decision variable. We limit the problem to the case where demand is a linear decreasing function of price. Following notations are used in addition to notations defined in Chapter 4.

$I_{a,d}$	:	Set of indices for discrete accepted demand at arrival date a, duration d. All integer values with maximum at $\lfloor \alpha_{a,d} \rfloor$.
$\alpha_{a,d}, \beta_{a,d}$	:	Intercept and slope for demand function at arrival date a, duration d
$r_{a,d,i}$	:	Price for arrival date a, duration d, when accepted demand is i
p_s	:	Price for staynight s
$y_{a,d,i}$	:	Indicator for accepted demand at arrival date a, duration d being i

The optimization problem (SN-OPT):

$$\max_{r} \sum_{(a,d)\in A\times D} \sum_{i\in I_{a,d}} i r_{a,d,i}$$

$$s.t. \sum_{\substack{(a,d): \\ a\le s\le a+d-1}} \sum_{i\in I_{a,d}} i y_{a,d,i} \le c_s \qquad \forall s \in S \tag{22}$$

$$r_{a,d,i} \le \frac{\alpha_{a,d} - i}{\beta_{a,d}} y_{a,d,i} \qquad \forall (a,d) \in A \times D \tag{23}$$

$$\sum_{i\in I_{a,d}} y_{a,d,i} = 1 \qquad \forall (a,d) \in A \times D, \forall i \in I \tag{24}$$

$$\sum_{i\in I_{a,d}} r_{a,d,i} = \sum_{s\in S: a\le s\le a+d-1} p_s \qquad \forall (a,d) \in A \times D \tag{25}$$

$$r_{a,d,i}, p_s \ge 0, \quad y_{a,d,i} \in \{0,1\} \tag{26}$$

In this formulation, the potential demand for an itinerary with arrival date a, duration of stay d is a linear function of the itinerary price, $r_{a,d}$. Specifically, the demand generated for this itinerary at price $r_{a,d}$ equals $\alpha_{a,d} - \beta_{a,d} r_{a,d}$. The coefficient i in the objective function is the discrete sales quantity point for arrival date a, duration of stay d. i ranges from 0 to maximum value $\lfloor \alpha_{a,d} \rfloor$. Discrete sales quantity i is multiplied by $r_{a,d}$ in the objective function to calculate maximum revenue. $r_{a,d}$ is the price for itinerary (arrival date a, duration d) and is structured to have positive value only when the corresponding discrete sales quantity i is chosen (constraint (23)). $y_{a,d,i}$ is an indicator, which equals 1 only when discrete sale quantity i is chosen, and constraint (24) enforces that only one discrete sale quantity is chosen per itinerary. Hence constraints (23) and (24) together enforces that only one discrete sales quantity per itinerary contributes to the objective function. Constraint (22) ensures that the sum of chosen sales quantity that share a staynight does not exceed hotel capacity. Constraint (23) also enforces that when discrete sales quantity i is chosen for itinerary (a, d), i.e. $y_{a,d,i} = 1$, the sales quantity i does not exceed the potential demand generated with the linear demand function. Finally, constraint (25) restricts the itinerary prices to be sum of individual staynight prices. Formulated this way, the bilinear price optimization can be solved as a discrete optimization problem.

We also considered discretizing price points and relaxing quantity to be continuous decision variable. However, discretized price formulation not only has more complicated constraints but also require more integer variables in order to represent the full feasible region and hence is more difficult to solve. For this reason, we use only the discretized quantity method for the simulation experiments in following sections.

Table 19 illustrates preliminary results from comparing the revenue earned by optimizing staynight prices for selected demand models with itinerary price optimization *(P)* presented in Chapter 4, as a percentage of earned revenue from *(P)*. Optimality gap for the staynight optimization problem *(SN-OPT)* is also presented. Initial results show that revenue difference between the two pricing methods are typically less than 10% when the optimality gap is small (less than 5%), far less significant compared to the revenue differences in Chapter 4, indicating that the pricing method may

Table 19: Revenue differences between itinerary price optimization and staynight price optimization

Hotel	Location	Occ.	Rev. diff.	Opt. gap
		50%	6.9%	0.1%
		100%	6.9%	0.1%
A	Airport	150%	5.3%	2.9%
		50%	13.9%	9.3%
		100%	13.8%	10.2%
B	Airport	150%	10.7%	9.8%
		50%	17.2%	9.2%
		100%	17.1%	13.3%
C	Suburban	150%	20.1%	26.0%
		50%	21.3%	13.0%
		100%	21.3%	17.0%
D	Suburban	150%	18.8%	2.9%
		50%	0.5%	0.2%
		100%	0.4%	0.2%
E	Urban	150%	0.5%	0.2%
		50%	7.0%	1.6%
		100%	7.2%	5.3%
F	Urban	150%	9.1%	10.8%

have less impact on the revenue earned compared to the impact of demand models on price optimization.

Future research directions on the staynight optimization performance may be focused on developing algorithms/formulations to enhance performance of the optimization and developing heuristic to solve the staynight and produce price recommendations close to the optimal solution.

REFERENCES

[1] "2006 U.S. Lodging Industry Results," tech. rep., Smith Travel Research, 2007.

[2] "Monthly traffic analysis, January 2008," tech. rep., International Air Transport Association, 2008.

[3] "IHG launches price optimization module," *Hotels*, September 2009.

[4] Abadie, A., "Bootstrap tests for distributional treatment effects in instrumental variable models," *Journal of the American Statistical Association*, vol. 97, pp. 284–292, 2002. [1].

[5] Alstrup, J., Boas, S., Madsen, O. B. G., and Vidal, R. V., "Booking policy for flights with two types of passengers," *European Journal of Operational Research*, vol. 27, no. 3, pp. 274–288, 1986.

[6] Baker, T. K. and Collier, D. A., "The benefits of optimizing prices to manage demand in hotel revenue management systems," *Production and Operations Management*, vol. 12, no. 4, pp. 502–518, 2003.

[7] Bellman, R. E., *Dynamic programming*. Dover Publications, Incorporated, 2003. 862270.

[8] Belobaba, P. P., "Airline yield management: An overview of seat inventory control," *Transportation Science*, vol. 21, no. 2, pp. 63–73, 1987.

[9] Belobaba, P. P., "Application of a probabilistic decision model to airline seat inventory control," *Operations Research*, vol. 37, no. 2, pp. 183–196, 1989.

[10] Ben Ghalia, M. and Wang, P. P., "Intelligent system to support judgmental business forecasting: the case of estimating hotel room demand," *Fuzzy Systems, IEEE Transactionsion*, vol. 8, no. 4, pp. 380–397, 2000.

[11] Besbes, O., Phillips, R., and Zeevi, A., "Testing the validity of a demand model: An operations perspective," *Manufacturing Service Operations Management*, vol. 12, no. 1, pp. 162–183, 2010.

[12] Birgé, L. and Rozenholc, Y., "How many bins should be put in a regular histogram." Working paper, Universite Paris VI, 2002.

[13] Bitran, G. and Caldentey, R., "An overview of pricing models for revenue management," *Manufacturing Service Operations Management*, vol. 5, no. 3, pp. 203–229, 2003.

[14] Canina, L. and Carvell, S., "Lodging demand for urban hotels in major metropolitan markets," *Journal of Hospitality and Tourism Research*, vol. 29, no. 3, pp. 291–311, 2005.

[15] Carroll, W. J. and Grimes, R. C., "Evolutionary change in product management: experiences in the car rental industry," *Interfaces*, vol. 25, no. 5, pp. 84–104, 1995.

[16] Choi, S. and Kimes, S. E., "Electronic distribution channels' effect on hotel revenue management," *The Cornell hotel and restaurant administration quarterly*, vol. 43, no. 3, pp. 23–31, 2002.

[17] Chui, C. K., *An introduction to wavelets*. Wavelet analysis and its applications, v. 1, Boston: Academic Press, 1992.

[18] Cooper, W. L., de Mello, T. H., and Kleywegt, A. J., "Models of the spiral-down effect in revenue management," *Operations Research*, vol. 54, no. 5, pp. 968–987, 2006.

[19] Cross, R. G., Higbie, J. A., and Cross, Z. N., "Milestones in the application of analytical pricing and revenue management," *Journal of Revenue and Pricing Management*, vol. Advanced online publication, no. doi: 10.1057/rpm.2010.39, 2010.

[20] Damonte, L. T., Domke-Damonte, D. J., and Morse, S. P., "The case for using destination-level price elasticity of demand for lodging services," *Asia Pacific Journal of Tourism Research*, vol. 3, no. 1, pp. 19–26, 1998.

[21] Duvvuri, S. D., Ansari, A., and Gupta, S., "Consumers' price sensitivities across complementary categories," *Management Science*, vol. 53, no. 12, pp. 1933–1945, 2007.

[22] Elmaghraby, W. and Keskinocak, P., "Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions," *Management Science*, vol. 49, no. 10, pp. 1287–1309, 2003.

[23] Emmons, H. and Gilbert, S. M., "The role of returns policies in pricing and inventory decisions for catalogue goods," *Management Science*, vol. 44, no. 2, pp. 276–283, 1998. 288758.

[24] Gallego, G. and Ryzin, G. J. V., "A multiproduct dynamic pricing problem and its applications to network yield management," *Operations Research*, vol. 45, no. 1, pp. 24–41, 1997.

[25] Garrow, L. A. and Ferguson, M. A., "Revenue management and the analytics explosion: perspectives from industry experts," *Journal of Revenue and Pricing Management*, vol. 7, no. 2, pp. 219–229, 2008.

[26] Garrow, L. A., Ferguson, M. A., Keskinocak, P., and Swann, J., "Expert opinions: Current pricing and revenue management practice across U.S. industries," *Journal of Revenue and Pricing Management*, vol. 5, no. 3, pp. 237–247, 2006.

[27] Gillen, D. W., Morrison, W. G., and Stewart, C., "Air travel demand elasticities: Concepts, issues and measurement," tech. rep., Department of Finance Canada, 2004.

[28] Gu, Z., "Product differentiation: Key to macaus gaming revenue growth," *Journal of Revenue and Pricing Management*, vol. 4, no. 4, pp. 382–388, 2006.

[29] Hawtin, M., "The practicalities and benefits of applying revenue management to grocery retailing, and the need for effective business rule management," *Journal of Revenue and Pricing Management*, vol. 2, no. 1, pp. 61–67, 2003.

[30] Heching, A., Garrow, L. A., and Ferguson, M. A., "Mark-down pricing: An empirical analysis of policies and revenue potential at one apparel retailer," *Journal of Revenue and Pricing Management*, vol. 1, no. 2, pp. 37–38, 2002.

[31] Hiemstra, S. J. and Ismail, J. A., "Analysis of room taxes levied on the lodging industry," *Journal of Travel Research*, vol. 31, no. 1, pp. 41–49, 1992.

[32] Jeffrey, D. and Barden, R. R. D., "Monitoring hotel performance using occupancy time-series analysis: the concept of occupancy performance space," *International Journal of Tourism Research*, vol. 2, no. 6, pp. 383–402, 2000.

[33] Jeffrey, D., Barden, R. R. D., Buckley, P. J., and Hubbard, N. J., "What makes for a successful hotel? Insights on hotel management following 15 years of hotel occupancy analysis in England," *Service Industries Journal*, vol. 22, no. 2, pp. 73–88, 2002.

[34] Karlin, S. and Carr, C. R., "Prices and optimal inventory policies," in *Studies in Applied Probability and Management Science* (Karlin, S., Arrow, K. J., and Scarf, H., eds.), Stanford, CA: Standford University Press, 1962.

[35] Kelton, W. D., Sadowski, R. P., and Sturrock, D. T., *Simulation with Arena, 4th edition*. McGraw-Hill, 2007.

[36] Khouja, M. J., "Optimal ordering, discounting, and pricing in the single-period problem," *International Journal of Production Economics*, vol. 65, no. 2, pp. 201–216, 2000.

[37] Kimes, S. E., "Restaurant revenue management: Could it work?," *Journal of Revenue and Pricing Management*, vol. 4, no. 1, pp. 95–97, 2005.

[38] Kimes, S. E. and Schruben, L. W., "Golf course revenue management: A study of tee time intervals," *Journal of Revenue and Pricing Management*, vol. 1, no. 3, pp. 220–233, 2002.

[39] Kimes, S. E., Wirtz, J., and Noone, B. M., "How long should dinner take? measuring expected meal duration for restaurant revenue management," *Journal of Revenue and Pricing Management*, vol. 1, no. 3, pp. 220–233, 2002.

[40] Kuyumucu, A. H., "Gaming twist in hotel revenue management," *Journal of Revenue and Pricing Management*, vol. 1, no. 2, pp. 161–167, 2002.

[41] Lai, K. S., "Price smoothing under capacity constraints," *Southern Economic Journal*, vol. 57, no. 1, pp. 150–159, 1990.

[42] Lau, A. H. L. and Lau, H., "Effects of a demand-curve's shape on the optimal solutions of a multi-echelon inventory/pricing model," *European Journal of Operational Research*, vol. 147, no. 3, pp. 530–548, 2003.

[43] Lieberman, W. H. and Dieck, T., "Expanding the revenue management frontier: Optimal air planning in the cruise industry," *Journal of Revenue and Pricing Management*, vol. 1, no. 2, pp. 7–24, 2002.

[44] Lim, C., "Review of international tourism demand models," *Annals of Tourism Research*, vol. 24, no. 4, pp. 835–849, 1997.

[45] Lippman, B. W., "Retail revenue management? competitive strategy for grocery retailers," *Journal of Revenue and Pricing Management*, vol. 2, no. 3, pp. 229–233, 2003.

[46] Lus, B. and Muriel, A., "Measuring the impact of increased product substitution on pricing and capacity decisions under linear demand models," 2007.

[47] Mills, E. S., "Uncertainty and price theory," *The Quarterly Journal of Economics*, vol. 73, no. 1, pp. 116–130, 1959.

[48] Petruzzi, N. C. and Dada, M., "Pricing and the news vendor problem: A review with extensions," *Operations Research*, vol. 47, no. 2, pp. 183–194, 1999.

[49] Popescu, A., Keskinocak, P., Johnson, E., LaDue, M., and Kasilingam, R., "Estimating air-cargo overbooking based on a discrete show-up-rate distribution," *Interfaces*, vol. 36, no. 3, pp. 248–258, 2006.

[50] Rushmore, S., "Mid-rate extended-stay provides best return," *Hotels*, vol. 34, no. 5, p. 42, 2000.

[51] Schwarz, Z., "Changes in hotel guests willingness to pay as the date of stay draws closer," *Journal of Hospitality and Tourism Research*, vol. 24, no. 2, pp. 180–198, 2000.

[52] Song, H., Witt, S. F., and Jensen, T. C., "Tourism forecasting: accuracy of alternative econometric models," *International Journal of Forecasting*, vol. 19, no. 1, pp. 123–141, 2003.

[53] Talluri, K. T. and Ryzin, G. J. V., *The theory and practice of revenue management*. New York: Springer, 2005.

[54] Trefethen, L. N. and Bau, D., *Numerical linear algebra*. SIAM, 1997.

[55] Urban, T. L. and Baker, R. C., "Optimal ordering and pricing policies in a single-period environment with multivariate demand and markdowns," *European Journal of Operational Research*, vol. 103, no. 3, pp. 573–583, 1997.

[56] Varian, H. R., *Microeconomic analysis*. New York: W. W. Norton, 3rd ed., 1992.

[57] Vidakovic, B., *Statistical modeling by wavelets*. New York: Wiley, 1999.

[58] Vinod, B., "Unlocking the value of revenue management in the hotel industry," *Journal of Revenue and Pricing Management*, vol. 3, no. 2, pp. 178–190, 2004.

[59] Weatherford, L. R. and Kimes, S. E., "A comparison of forecasting methods for hotel revenue management," *International Journal of Forecasting*, vol. 19, no. 3, pp. 401–415, 2003.

[60] Weng, Z. K., "Modeling quantity discounts under general price-sensitive demand functions: Optimal policies and relationships," *European Journal of Operational Research*, vol. 86, no. 2, pp. 300–314, 1996.

[61] Whitcher, B., *waveslim: Basic wavelet routines for one-, two- and three-dimensional signal processing*, 2007. R package version 1.6.1.

[62] Whitin, T. M., "Inventory control and price theory," *Management Science*, vol. 2, no. 1, pp. 61–68, 1955.

[63] WITT, S. F. and WITT, C. A., "Forecasting tourism demand: A review of empirical research," *International Journal of Forecasting*, vol. 11, no. 3, pp. 447–475, 1995.

VITA

Seonah Lee received her Bachelor's degree in Aerospace Engineering from Seoul National University in Seoul, Korea. She received her Master's Degree in Aeronautical and Astronautical Engineering from the Massachusetts Institute of Technology in Cambridge, Massachusetts. After graduating with the Master's degree she joined Northwest Airlines and then Korea Telecom and worked before she joined the Ph.D. program in Operations Research at the H. Milton Stewart School of Industrial Engineering at the Georgia Institute of Technology in Atlanta, GA. She is currently working at Revenue Analytics, Inc. as a Director of Operations Research.