The Dynamics of Hotel Pricing

Sungjin Cho, Seoul National University

Gong Lee, Georgetown University

John Rust, Georgetown University

Mengkai Yu, Georgetown University

March 10, 2018

Abstract

We use a unique data set from the reservation database at a luxury hotel based in Washington DC over a 37 month period to study how the hotel sets prices for various classes of customers and how its prices vary over time. Hotel pricing is a challenging, high-dimensional problem since hotels must not only set prices for each current date, but they must also quote prices for a range of future dates due to the prevalence of advance booking of hotel rooms. We are able to observe the path of room rates quoted for different classes of rooms quoted in advance of the date of occupancy. We find large within and between week variability in reservation rates, as well as huge seasonal variations in average daily rates and occupancy rates, not only for the hotel we study but also for its direct competitors. We formulate and estimate a dynamic structural model of expected revenue maximization by the Method of Simulated Moments (MSM). The estimated model provides accurate predictions of the actual prices set by this firm and the implied path of bookings and cancellations. Prices quoted for bookings in advance of occupancy generally decline as the date of occupancy arrives for non-busy days, but can increase dramatically in the final days before occupancy on busy days when management forecasts a high probability of sell-out. Prices of the hotel we study and its competitors are highly correlated with each other and with the aggregate hotel occupancy rates in the central DC luxury hotel market. This suggests that on busy days, when occupancy of all of the hotels is close to 100% prices rise in a way that resembles a sequential auction of scarce capacity to the highest bidder. However on non-busy days, prices can be much lower, and our findings suggest that the hotel starts by setting higher prices to try to attract less elastic business travelers, but as the date of occupancy approaches, it reduces rates in order to attract additional last minute customers whose demand may be more price elastic.

Keywords  price discrimination

---

*Department of Economics, Seoul National University e-mail: sungcho@snu.ac.kr
†Department of Economics, Georgetown University, e-mail: gl430@georgetown.edu
‡Department of Economics, Georgetown University, e-mail: jr1393@georgetown.edu
§Department of Economics, Georgetown University, e-mail: my262@georgetown.edu
1 Introduction

We analyze a unique new micro panel dataset of daily observations of reservations, prices, and occupancy of a luxury hotel based in major US city. Due to the confidential nature of this data set we are unable to reveal the name of the hotel or the city where it is located, and hereafter we refer to it as “hotel H”. We formulate and estimate a dynamic model of optimal pricing and show that our model provides surprisingly accurately predictions of the prices set by this hotel. Optimal pricing depends critically on having a an accurate model of customer demand. We introduce a dynamic model of hotel demand, where customers arrive stochastically and choose to stay at hotel H or one of six competing hotels that constitute a local luxury hotel market in this city. Though we have daily observations of the best available rate (BAR) for comparable rooms quoted by the six competing hotels, and we observe the number of new reservations (and cancellations) are hotel H, we do not observe the number of customers who are making reservations at hotel H’s competitors. Thus, we face a problem of censoring that makes it very challenging to estimate customer demand, and without good estimates of demand it is not clear how to set prices.

Via a matched dataset provided by Smith Travel Research, we are able to observe the total occupancy and average daily rate (ADR) on a daily basis for the seven hotels in this market. The ADR is an average of different prices paid by different customers at different dates prior to occupancy, and so at the very least we face a problem of errors in variables if we use ADR in place of BAR to try to estimate the demand function. But there is a more serious problem of endogeneity in hotel prices due to the strong co-movement of prices of all six hotels in response to variations in demand for rooms. Prices tend to peak on days where demand is high and occupancy is close to 100%, and prices fall on days when demand is low and there is significant excess room capacity. Regressions of hotel occupancy on hotel prices therefore produce spurious positively sloped demand functions due to the effect of demand shocks on endogenously determined prices. There are no relevant instrumental variables that can be used to deal with this endogeneity problem and in any case our model of hotel demand is not a simple linear demand equation but rather a more complicated nonlinear model that is derived from a stochastic point process for the number of customers wishing to reserve rooms on various future arrival dates combined with an individual level discrete choice model of which hotel to choose given the BAR quoted by hotel H and its competitors. It is not obvious how to control for endogeneity in a nonlinear dynamic model such as this even if we did have good instruments.

We show how the censoring, errors-in-variables and endogeneity problems can be solved using structural econometric methods. We provide credible structural estimates of the stochastic arrival process of customers and their preferences for the competing hotels using the method of simulated moments under the assumption that Hotel H is an expected profit maximizer. In essence, our structural estimation can be
regarded as process for inferring the hotel manager’s beliefs about customer demand that are implicit in the
array of prices the hotel sets on a daily basis. As such, our structural estimation method can be regarded as
a procedure for inferring the hotel manager’s revealed beliefs about customer demand from observations
of the prices they set, similar to the way that structural estimation is used to infer the revealed preferences
of consumers from observations of their choices.

The most important information needed to set hotel prices is the nature of customer demand,

Hotel pricing is a challenging problem since beside setting different prices for various room categories
(standard rooms, deluxe rooms, penthouse suites, etc) and customer categories (tourist versus business
guests, group discounts for corporations, governments, etc.) a hotel manager is continually required to
quote future prices since most customers book rooms well in advance of their intended arrival date.

a local market of seven luxury hotels in Washington DC over a 37 month period.

Phillips (2005)

**Hotel pricing system** Revenue Management System (RMS) provides optimal recommendations on price
and most major hotels today use the tool to maximize their revenue. Sophisticated systems analyzes his-
torical as well as current booking information on room availability, booking pace, group booking patterns,
competitors rates, and reputation index. Moreover, a forecast of future occupancy also serve as an input to
the pricing system. Meanwhile, hotels revenue per available room (RevPAR) is also influenced by several
factors affecting consumers’ hotel choices such as hotel loyalty programs, the quality of hotel services and
facilities, and convenience of the location Sturman, Corgel, and Verma (2011). As such, pricing strategies
for RMS needs to incorporate the price elasticity of demand and various components affecting the demand
while each component has different weight in the pricing system.

This paper develops a dynamic programming model for the hotel pricing system. Assuming the dy-
namic pricing, a hotels current price strategy has influence not only on current sales of the hotel, but also
on future sales. Current price setting system depends on expectation about booking inflow as well as
their price sensitivity in the future period. While those system is more static, our study pursuits the profit
maximization of the hotel considering all period between any booking day and actual occupancy day.

The reason this study is possible with dynamic programming comes from the characteristics/nature of
hotel business. Each hotel has fixed number of rooms to be provided to customers. It is necessary to set
the price efficiently for profit maximization given the number of rooms. The basic economic principle is
applicable to this hotel business. Once the hotel set the relatively lower price than its competitors, it would
result in sellout before the occupancy date. Since the hotel missed some customer who are willing to pay
more than customer who made a reservation early, it has negative effect on the revenue. On the contrary to
this, setting the high price would bring about empty rooms at the occupancy day. It negatively influences on hotel revenue for that occupancy day.

In a static market condition, previous records of price and demand determine the optimal future price. However, the reality a hotel business faces is far more complicated. For the specific occupancy day, the booking inflow trend depends on how far in advance to occupancy day. This trend has different aspect by the type of customer such as business trip and leisure trip. In addition, all of our competitor hotels have dynamic price trend which also influence on customer inflow of our hotel. Though our competitors have differentiated products, as the demand elasticity varies by leading day, type of customer and season. As gathering hotels of the same tiers, we are able to assume that those hotels including our hotel has mutual effect on demand and price. With these same tier hotels, we find the price index affecting on customers hotel choice.

In our study, we attempt to find dynamic optimal price of hotel taking under consideration competitors price index and booking inflow sensitivity in accordance with the price index. We believe this way of optimal price facilitate to find more realistic price than previous pricing system. Using dynamic programming of discrete choice, this study is able to go up a step in the oligopoly market study.

**DC area hotel demand** The presence of a number of major government agencies and thousands of its employees has made the Washington D.C. Metropolitan regions lodging market highly unique. The federal government has a significant influence on Washington D.C. hotel market with a large share of the demand comes from the government and its contractors in the private market. DC-based educational institutions, think tanks, international organizations such as the International Monetary Fund and the World Bank, and non-profit entities are a few others that drive demand for hotel rooms in the District area.

With the high volume of demand related to the federal government, the Washington D.C. lodging market has been kept stable in terms of the occupancy as well as the room rate. The government-related demand is tied to the federal per-diem rate which is set annually. The per-diem rate also becomes an important benchmark for corporate and group contracts. Corporate and group booking rates are often negotiated and anchored around the federal per-diem rate. The rate varies by season, but relatively stable throughout the year. As such, while the Great Recession (2007-2009) had damped the demand for hotels and cut hotels’ average daily rate greatly, Washington D.C. hotels could remain relatively intact from the fluctuation: the average rate in D.C. fell only by 5.5%; compared to 22% drop in New York and 15% drop in Chicago (HVS, 2017).
Washington D.C. area ranked as the eighth most visited city in US (2016) and also boasts steady tourism demand supported by its wealth of monuments, museums, memorials and other tourist attractions. Meanwhile, a strong seasonality is observed for the tourism demand in the area: demand peak for the cherry blossom in April; the commencement in May; and the fall foliage in October; whereas January and July are the slow season of the year.

More than 800 hotels provide 111,500 rooms in Washington D.C. Metro area. If we confine Washington D.C. area itself, 130 hotels supply 30,200 rooms in 2017 (HVS, 2017). With the limited land and development constraint by limit of building height, increased hotel demand is not able to catch up the room supply. So, more accurate price strategy is necessary to grab high demand of hotel room in DC area.

**Washington DC luxury hotel** If we attempt to study the whole Washington D.C. lodging market, it is closed to a perfect competition market. So, it is more reasonable to focus on comparable hotels as narrowing the number of hotels with the criteria. Our basic assumption is that potential hotel customers in DC debate with themselves between similar hotel options. This assumption become a basis of choice probability model.

The hotel which provided the data is located in Dupont circle, which is historic district in Northwest Washington D.C.. As a famous busy street, it is crowded with many restaurants and bars. And it is convenient for transportation as the subway station is handy. Numerous embassies are placed within walking distance and several well-known think tank such as Brookings institution, Peterson Institute, Paul H.Nitze school of advanced international Studies are in a neighborhood. Thus, the Dupont circle has great location advantage to business travelers and leisure travelers both.

The seven hotels used in this study including our main hotel are located around Dupont circle. Table 1 explains that they are 4-star hotels classified as upscale class or luxury class with total 2,153 rooms provided. While non-upscale class hotels are numerous and have alternative accommodation such as Airbnb, those upscale hotels are limited in supply and compose an oligopoly market.

**IDEAs revenue management system** In order to set the price, the hotel uses IDEaS price management system made by SAS software company. The hotel revenue manager uses her own discretion and selects 20 different BAR prices as possible choices (effectively, she discretizes the pricing space) and enters the
Table 1: Hotel List

<table>
<thead>
<tr>
<th>Property</th>
<th>Avg. BAR</th>
<th>Star</th>
<th>Class</th>
<th>Chained Brand</th>
<th>Rate</th>
<th>Rooms</th>
<th>Distance from Metro</th>
<th>Cancel Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>hotel 0</td>
<td>$293.26</td>
<td>4</td>
<td>Luxury</td>
<td>No</td>
<td>4.4</td>
<td>327</td>
<td>3 min walk</td>
<td>1 day before</td>
</tr>
<tr>
<td>hotel 1</td>
<td>$338.29</td>
<td>4</td>
<td>Upper Up</td>
<td>Yes</td>
<td>4.2</td>
<td>410</td>
<td>8 min walk</td>
<td>2 day before</td>
</tr>
<tr>
<td>hotel 2</td>
<td>$253.51</td>
<td>4</td>
<td>Upper Up</td>
<td>No</td>
<td>4.2</td>
<td>193</td>
<td>8 min walk</td>
<td>3 day before</td>
</tr>
<tr>
<td>hotel 3</td>
<td>$285.16</td>
<td>4</td>
<td>Upper Up</td>
<td>No</td>
<td>4.4</td>
<td>259</td>
<td>3 min walk</td>
<td>1 day before</td>
</tr>
<tr>
<td>hotel 4</td>
<td>$454.30</td>
<td>5</td>
<td>Luxury</td>
<td>Yes</td>
<td>4.7</td>
<td>216</td>
<td>10 min walk</td>
<td>1 day before</td>
</tr>
<tr>
<td>hotel 5</td>
<td>$397.09</td>
<td>4</td>
<td>Luxury</td>
<td>No</td>
<td>4.6</td>
<td>413</td>
<td>10 min walk</td>
<td>Strict</td>
</tr>
<tr>
<td>hotel 6</td>
<td>$282.64</td>
<td>4.5</td>
<td>Upper Up</td>
<td>No</td>
<td>4.4</td>
<td>335</td>
<td>5 min walk</td>
<td>3 day before</td>
</tr>
</tbody>
</table>

choice set into the system. Considering remaining availability, seasonal effects, cancellation ratio and competitors prices, the system picks one optimal BAR out of the 20 for each reservation. The price varies a lot, so there is huge fluctuation of price even within the same room type. Sometimes, the price offered by IDeaS system does not seem to be reasonable to hotel revenue manager. Then, the manager modifies the BAR by manually. Specially, this happens to hotels newly launched or hotels they has changed their RMS system recently because the management system does not have enough data for the past. But it still happens to the old-established hotel. Therefore, the hotel would like to hire a hotel revenue manager with a wide range of experience. Actually, it happens very often. According to a hotel revenue manager, they pursuit to maximize the hotel profit while RMS system is likely to focus on filling up a hotel with bookings as many as possible. Once the optimal BAR is provided, this price is distributed through the online channels to be searched by customers. And BAR can be set up several times a day depending on how often a hotel revenue manager want to change them.

**Literature review**  There is an abundance of operations research on the area of revenue management. Many of these focus on stylized fact of price Baum and Mudambi (1995); Hung, Shang, and Wang (2010); S. L. Shapiro and Drayer (2014) or descriptions of how tools developed from other science areas have been implemented into marketing science Anderson and Xie (2012); Chatwin (2000); Corstjens and Gautschi (1983); D. Shapiro and Shi (2008). Despite these achievements, there exists disparity between the mar-
Marketing science model and their practice in reality. This is because most marketing science models rely on unrealistic formidable assumptions of consumer behavior. It was expected that a well-developed theory model would arise in the real world. There have been many attempts to overcome this inconsistency, but the presence of barriers has not yet vanished. Therefore, our study is an attempt to put another economics tool, namely dynamic programming and structural econometrics, into operation research. If this is successful, it will bring new insights, understanding, and methodologies to the field of revenue management.

Basically, a hotel product has many of the properties of a perishable good. There have been plenty of studies on price discrimination of perishable goods such as flight tickets Escobari (2012); Lazarev (2013); Williams (2017), concert tickets Courty and Pagliero (2011); Leslie (2004), and food products Bhattacharjee and Ramesh (2000); Liu, Tang, and Huang (2008). Hotel products, like other perishable goods, have a finite time opportunity for sales and stochastic demand during this period, so if any products are not sold in a given period, they would literally perish after the last minute. Many literatures focus on the way of revenue maximization under the particular condition of perishable good. It implies the need of dynamic programming. The object value function starts at the last period and counts backward since products would be of no value after that. It takes into account the demand pattern from the last time of sales to any previous time available for sales. Considering whole possibility of occupancy status and other market status, the value functions enable us to set the optimal price. There are several essential conditions in dynamic programming of perishable goods and other literature handle some parts of those conditions. Our study attempts to aggregate all conditions which we are able to bring from within our data set. That is a notable contribution to this study.

First, intertemporal price discrimination has to be applicable to set the optimal price each time. Puller, Sengupta, and Wiggins (2012) and Gale and Holmes (1993) analyze the simple intertemporal price discrimination with an advance purchase. But if the price itself can be adjusted with lapse of time, the advance purchase discount would partly lose the meaning of intertemporal price discrimination. The advance purchase is still valuable strategy to keep more demands as it gives restriction on cancel policy. But in terms of setting optimal price every time period, the advance purchase is out of one we are looking for. What we pursuit is more unfettered, so can be used versatily. The advance purchase discount can be added to a dynamic optimal pricing later on. Instead of price discrimination by discount rate, we focused on the regular price of a hotel; the so called best available rate (rack rate). There are already several studies on this regular price of items in other industries such as stadium tickets and flight tickets. Although these
industries shared some properties, they have a different pricing strategy which stem from the difference of their industry structural. Our study tries to find the optimal price of hotels at each time carrying hotel industry attributes.

Second, another important feature when it comes to dynamic pricing of hotel industry is capacity allocation Dana (1999). Just as concert halls and sports stadiums have a seating capacity, each hotel also has their own limit on the number of available units. Therefore, depending on the status of the number of occupied units, how to keep or sell the remaining rooms is crucial for revenue maximization. For example, if there are few rooms left, a hotel revenue manager does not need to drop the price. On the other hand, if there are plenty of rooms left, they are highly motivated to drop the price. This example describes a kind of capacity allocation. Since we do not observe the actual capacity allocation by the hotel, we assume that the hotel keeps the remaining rooms which can be sold to urgent business customers and putting a premium on all the remaining rooms. This is the way of capacity allocation. This kind of price decision is achievable when we can see the demand by time and consumer type of demand. Favorably, we are able to monitor these information from our data. This is a critical benefit compared to other literatures. There is also plenty of literature on demand patterns and information on consumer types. But the data set we have provides the most information about the consumer.

Another good thing about this data set is the type of perishable good. Seat sections of concert halls and sports stadiums have a wide range of pricing within the same event. Even if the price is the same, each ticket has a different value, therefore making it difficult to measure demand patterns. But for hotel products, most rooms are equivalent and there are only a few number of luxury rooms available. Thus, within the same type of rooms, it is possible to analyze the capacity allocation. The information we gather from this enables us to figure out consumer behavior.

Finally, dynamic pricing depends on demand pattern, which is influenced by the market. The previous literatures mainly focus on monopoly markets or perfectly competitive markets (Monopoly : Gallego and van Ryzin (1994); Zhao and Zheng (2000), Competitive market :Enz, Canina, and Lomanno (2009); Sweeting (2012)). There are only a few that revolve around an oligopoly market. This is because it is required to know the competitors outcome, i.e. revenue or occupancy, as well as their competing reaction, and price. Since we obtained two points of information from the data set, it leads us to study dynamic pricing on the oligopoly market. Unless the market is oligopoly, the knowledge on cross price-elasticity would
be limited. As examining price and occupancy level at all arrival days, the price elasticity of oligopoly markets can be determined. However, there is one factor that hinders our ability to measure the degree of sensitivity. We are able to track the past reservations for the hotel back to one year before the arrival day. But it is meaningless to compute the price elasticity if only a few reservations have been made. So, we had to reduce the period of leading days to 45 days. Furthermore, the price elasticity can be computed by consumer type using the consumer code. Generally, it is believed that business customers are less elastic to price. This is consistent with our estimation result. The leisure customer seems to be very sensitive to price in our results and it gets more sensitive as it closes on the arrival day. Moreover, the degree of elasticity varies by seasonal movement.

In summary, the facts we have to consider in our dynamic hotel pricing are consumer price elasticity to competing hotels price, consumer type, time varying demand probability, and number of remaining units in the hotel. Along with these factors, seasonal demand effect is also crucial to generating an optimal price strategy. The key point of the hotel dynamic pricing model is how to combine all these factors into one model. We have yet to see this kind of attempt and it sheds light on a dynamic pricing model in an oligopoly market.

Moreover, there is one possible to be added to the model. As the hotel sets the optimal price, consumers might act strategically as well by waiting until the price drops. This guesswork is not allowed in our model. First of all, we do not observe consumers strategic action for the hotel reservation. Another thing is that it is extremely difficult to take action optimally following a price trend. For each arrival day, hotel regular price, best available rate, in process of time diverges greatly. Sometimes, it just goes up or down when it is close to the arrival day. But sometimes, it fluctuates between some range in unpredictable ways. The unstable price pattern hinders consumer from doing optimally. Our study excludes the chance of consumer optimal action, so we assume that all consumers would make a hotel reservation as soon as they are aware of his/her travel schedule. The consumer side optimal problem can only be applied when there are strong assumptions on the consumer action Gale and Holmes (1993); Horner and Samuelson (2011); Lazarev (2013); Su (2007). This exclusion also leads to preclude the hotels reputation for future consumers Selnes (1993).
Table 2: Data description

<table>
<thead>
<tr>
<th>Data</th>
<th>The first day of occupancy</th>
<th>The last day of occupancy</th>
<th>Observations</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>market vision</td>
<td>2010-09-21</td>
<td>2014-08-13</td>
<td>609,181</td>
<td>competitors’ price</td>
</tr>
<tr>
<td>reservation raw</td>
<td>2009-09-01</td>
<td>2013-10-31</td>
<td>201,176</td>
<td>reservations detail information</td>
</tr>
<tr>
<td>cancellation raw</td>
<td>2009-09-01</td>
<td>2013-10-31</td>
<td>29,241</td>
<td>cancel detail information</td>
</tr>
<tr>
<td>daily pick-up report</td>
<td>2010-09-16</td>
<td>2014-05-21</td>
<td>475,187</td>
<td>daily revenue report</td>
</tr>
<tr>
<td>STR market data</td>
<td>2010-01-01</td>
<td>2014-12-31</td>
<td>1,731</td>
<td>competitors’ occupancy</td>
</tr>
</tbody>
</table>

| Data range            | 2010-10-01                  | 2013-10-31                | 37 months    |                           |

2 Data

One hotel in the Dupont district area provided their detailed booking records, aggregate daily reports and their competitive daily rate records between September 2010 and October 2013, a total of 37 months. From STR, which handles hotel market data and benchmarking technique, we get the competitions aggregate occupancy information within the same time period. The summary of data set we used is shown in Table 2.

2.1 Market vision

One key component of hotel pricing is the pricing of other competitors. After considering the price of other hotels, the hotel needs to modify their own prices not to lose prospective consumers. Market vision data is collected for this reason. The Rubicon group establishes this market vision data by monitoring the competition and put their client hotel in a competitive position. Market vision data contains all channels such as GDS (global distribution system), travel websites and hotel websites. Although it collects only the lowest priced rooms for each hotel, they handle many variations of room products such as AAA, Adv purch, Any Non-qual, Gov, Unrest/ No Merch, and Unrestricted. The detailed descriptions for the above room products is as followed,
AAA  Products specifically identified as AAA or CAA, using keyword variations of AAA, CAA, auto club, and so on.

Adv Purch  Products requiring advance purchase or purchase at time of booking, using keywords such as deposit required, pre-payment, full payment due, and so on. Booking before 21-7 days and 10-15% off from original one.

Any Non-Qual  Lowest of Unrestricted, Adv Purch and Merchant Model combined. Excludes qualified rates that require membership, association or identification. Also excludes government. Sometimes it just indicates the Advance purchase.

Gov  Products specifically identified as government rates, using keyword variations of government and military. Includes but does not differentiate among federal, state and local rates.

Unrest/ No Merch  Products offered for general availability, without qualification or advance purchase requirements. Merchant Models are excluded.

Unrestricted  Products offered for general availability, without qualification or advance purchase requirements. Includes Merchant Model rates as applicable.

The market vision data includes the lowest room price for each room product and searching on either channels. According to the hotel manager, the hotel has the same room rate as the GDS rates regardless of channel. We should therefore focus on the room product only. In the middle of room product, Unrestricted indicates BAR (Best Available rate). When one is looking for the best hotel price, which varies at each hotels discretion, BAR is the only necessary data point. Other market segments, like CORN or GOV, are controlled by contracts and there is little chance on other market segments to be changed by the hotel pricing strategy.

Market vision data has the lowest room rate on 45 or 90 days in advance of the arrival date for 6 competitor hotels. For Unrestricted or BAR prices, they usually have a 24-hour advance cancellation policy. This means that any customer is able to cancel their hotel reservation without penalty as long as they have cancelled it a day in advance prior to the check-in time of the arrival date. When a hotel does not have this cancellation policy in place, the market vision data collects separate room pricing and labels it Unrest/Open cxl.
In most cases, the lowest priced room indicates a standard room. But when all standard rooms are booked or the remaining standard rooms are unserviceable due to other reasons, the market vision data takes the lowest priced room out of the remaining pool such as a suite room or beyond which causes some of this data to have an unusually high price level.

### 2.2 reservation and cancellation

A large strength of our data set is the detailed information on reservations and cancellation records. This data set contains all the information about each booking that the hotel knows. A reservation identification number is created when the reservation is made and becomes the permanent id for that reservation which is in chronological order. Besides those primary date information, when the booking is made and when it is for, other significant information included is the room type, booking channel, rate code and share amount (price). This hotel used to have 9 types of rooms before 2011-11-27 (Table 3). For some reasons we are not aware of, the hotel reformed its room type and currently has 11 types of rooms.

Typically, BAR is the rate for B1K and B2D (Table 3). However, when both are sold out, BAR changes their focus to the next level room such as A1K or A2D. We categorize these rooms into regular rooms and luxury rooms. The number of regular rooms is 312 out of 327.

There are around 200 rate codes which can be broken into 14 categories (Table 4). To simplify the analysis, we divided the codes into two; transient and group bookings. Transient signifies an individual traveler whose rate is based on BAR. Although their price differs depending on which channel is used for the booking, it can still be considered as the same group because a contract or deal between hotel and customer does not exist. Group bookings may also be based on BAR, however it will vary by rate code. When it is difficult to define their rate rule, it is associated with group reservations which we cannot anticipate the price and demand. So, it is likely that the price of group reservations is independent from transient reservations.

Share amount in reservation raw data indicates how much the guest has paid for the room per night excluding tax. In other words, this is revenue to the hotel for each room. When a guest has booked through a travel agency, the share amount still indicates the revenue to the hotel. But once the hotel takes this revenue, they would pay a commission fee (10%) to the travel agency at later date. This case differs from
Table 3: Room type and quantity

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th># of rooms (before 2011-11-27)</th>
<th># of rooms (since 2011-11-27)</th>
<th>Rack Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>B1K</td>
<td>Superior, 1 King</td>
<td>186</td>
<td>141</td>
<td>$203.15</td>
</tr>
<tr>
<td>B2D</td>
<td>Superior, 2 double beds</td>
<td>109</td>
<td>63</td>
<td>$203.15</td>
</tr>
<tr>
<td>A1K</td>
<td>Deluxe, 1 King</td>
<td>12</td>
<td>46</td>
<td>$253.15</td>
</tr>
<tr>
<td>A2D</td>
<td>Deluxe, 2 double beds</td>
<td>5</td>
<td>47</td>
<td>$253.15</td>
</tr>
<tr>
<td>GD1K</td>
<td>Grand Deluxe, 1 King</td>
<td>10</td>
<td></td>
<td>$303.15</td>
</tr>
<tr>
<td>GD2D</td>
<td>Grand Deluxe, 2 double beds</td>
<td>5</td>
<td></td>
<td>$303.15</td>
</tr>
<tr>
<td>U1K</td>
<td>Studio Suite, 1 King</td>
<td>4</td>
<td>4</td>
<td>$643.15</td>
</tr>
<tr>
<td>U1KB</td>
<td>Studio Suite, 1 King with balcony</td>
<td>5</td>
<td>5</td>
<td>$693.15</td>
</tr>
<tr>
<td>S1K</td>
<td>Luxury Suite, 1 King</td>
<td>2</td>
<td>2</td>
<td>$963.15</td>
</tr>
<tr>
<td>S1KB</td>
<td>Luxury Suite, 1 King with balcony</td>
<td>3</td>
<td>3</td>
<td>$1,013.15</td>
</tr>
<tr>
<td>R1K</td>
<td>Pent house Suite</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Category</td>
<td>Market Segment</td>
<td>Title</td>
<td>Description</td>
<td>Booking Share</td>
</tr>
<tr>
<td>----------</td>
<td>----------------</td>
<td>------------------------</td>
<td>-----------------------------------------------------------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Transient</td>
<td></td>
<td>BAR</td>
<td>Best Available Rate Best available rates that have hotel house cancellation policy, rate codes BAR only applicable in this segment</td>
<td>68.4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>CON</td>
<td>Consortia/TMC Consortia, Travel Management Companies bookings</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RESW</td>
<td>Restricted-Web collection web site with restrictions such as pre-paid/non-refundable i.e. 10% off 7 day advance purchase, 2mlos at 20% off, or limited time offer</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CORL</td>
<td>Corporate LRA Corporate/local negotiated rates with last room availability</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CORN</td>
<td>Corporate NLRA Corporate/local negotiated rates with Non-last room availability</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GOV</td>
<td>Government per diem equivalent rates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>PAK</td>
<td>Package Room package</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>FIT</td>
<td>Wholesale Locally negotiated wholesale accounts and Third party vacation package</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>DIS</td>
<td>Qualified Discount AAA, AARP, Employee rate or any qualified discounted rates</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>RESO</td>
<td>Restricted-OTAs Same rates as restricted segment available in OTA merchant sites</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>OPQ</td>
<td>Opaque Hotwire/ Priceline</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>ASS</td>
<td>Association convention group</td>
<td>31.6%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>TOT</td>
<td>Tour &amp; Travel tour group</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>group</td>
<td>uncategorized group</td>
<td></td>
</tr>
</tbody>
</table>
the case of OTA. When a guest has booked through an online travel agency (OTA), the OTA takes their margin (22%) and gives the remaining revenue to the hotel. Therefore, when analyzing the reservation records, the original price needs to be recovered when it comes through the OTA.

The cancellation raw has almost the exact same information as the reservations raw. Due to the cancellation policy of this hotel, the day before the occupancy has the highest cancellation rate.

2.3 Daily pace report

The hotel records their daily booking records and offers more reliability. However, it offers no detail, including cancellation information. Therefore, it is only used to check that the reservation in our data collection is valid. In our experience, they have only 2-3% of error. It does contain information on total revenue, occupancy, and average room rates.

2.4 STR

Although the hotel provided plenty of data, our model requires more data about other hotels. Since our study is based on the price comparison between our hotel and competitors, it is critical that we obtain price information on other hotels and their booking records. Acquiring reservation raw data from all competitors is crucial and it is possible to get this data, but very hard. Instead, we got a data of final occupancy rate of competitors from Smith Travel Agency (STR) that includes occupancy rate, average daily rate, revenue and so on. As using STR data set, we can find the price difference effect on the occupancy of hotel 0 and others. Without it, we have difficulty deriving outcome of price discrimination even if we have affluent data set of one hotel only. Though STR data set does not provide detail information of competing hotels, it facilitates further step on our analysis.

3 Stylized facts

3.1 Type of DC hotel customer

Figure 2 is a histogram of market codes for reservations, which is derived from reservation records. For group reservations, price decision is determined case by case. For simplicity, we assume that group reservations rate is a discounted rate of BAR. Furthermore, group reservation is less sensitive compared to transient reservation. These two assumptions enable group reservation to be distinguished from transient
Figure 1: Histogram of bookings by market code

Figure 2: Histogram of bookings by market code
reservation and prevent optimal price from being a corner solution.

There are several market codes in transient reservation. The most common market code is RESO (21.8%), which represents the reservation via an online travel agency such as hotel.com or agoda.com. The price level of RESO is close to BAR, but OTA charges about a 20% commission to the hotel, so revenue from this code is lower than the one from BAR. Another market code we need to focus on is GOV, which indicates reservations by a government sector. Any booking with this market code pays a per diem rate determined by the government and varies by month. Therefore, per diem rate increases in peak-season of Washington D.C. but lower in the off peak-season. Lastly, BAR takes up only 7.2%. But many of the market codes in transient reservation are discounted rates of BAR. Therefore, we regard all transient reservations as a discounted best available rate and simplify the problem by searching for the optimal BAR instead of surveying the whole market code rate.

3.2 Seasonality and weekday feature

Figure 3 illustrates the seasonality of Best available rate (BAR) for whole hotels including our hotel. We have plotted weekly average BAR from October 2010 to October 2013 which was drawn from Market vision data. The BARs we use are actually for same day reservations. Although the Walks-in rate is not factored into the average payment by hotel customers, it is necessary to use this rate for the same day to have them comparable to other hotels.

The bold line plots the BAR of our hotel while the other lines indicate BAR of six competitor hotels. All hotels have similar prices trend each year. The highest peak season is the middle of May; commencement ceremonies in the DC area, and tourist season. This is followed by the middle of April and the fall season. During this peak season, the BAR of our hotel is over $350, which is almost twice as high as the lowest weekly BAR in a year. The competitors also have similar peak season prices in late spring and fall.

Off-peak season is also apparent in all hotels we are looking at. August is the quietest season for DC area hotels along with some holidays, such as Christmas, Thanksgiving and New Years Day, leaving most DC hotels vacant. In order to match such low hotel demand, each hotel keeps the BAR relatively low. In fact, these low prices are not only for same day reservations, but also in advance of occupancy as we can see in the next section.
Figure 3: Price Seasonality for DC Luxury hotels
All of these characteristics of peak season and off-peak season appear in our competitor hotels. For hotel 1 and 2, they have an analogy in seasonality of price to our hotel and even the level of price is close to each other. Thus, we regard hotel 1 and 2 to be the actual competitors of hotel 0. Hotel 3,4 and 6 have a higher price level while hotel 5 is the only one which has a clearly lower price level than hotel 0. Since the average weekly BAR of hotel 4 and hotel 6 are over $100 higher than that of hotel 0, it is likely that these two hotels are not real competitors of hotel 0. But their activity, including their price setting practices can be a benchmark for the DC luxury hotel market. Thus, it is worth considering them as competitors of hotel 0 once we find that their price seasonality is alike.

Overall, the main highlights from Figure 3 can be summarized as follows. First of all, all seven DC luxury hotels have the same seasonality. They reach the summit price level during the same week and their price movement around the base price also has a similar pattern. Along with this movement, all prices of hotels tend to maintain the same interval throughout the year. Thus, it is rare to see the plotted lines intersect each other. Second, the seasonality is intense regarding price levels. Most hotels charge over twice as much as the price during their most quiet season. These price variations are not so stunning. It can often be seen that the variation of price level for seasonal tour areas in the US is much larger. Nonetheless, DC luxury hotels have a distinctive price difference depending on the season and those differences are sufficient to set the cause for this study since the optimization price range is plenty wide.

Figure 4 shows the distribution of number of bookings and average BAR of our hotel, i.e. hotel 0, based on the day of the week. The data set used here comes from reservation raw data, so these numbers of average bookings and price contain all reservation records regardless of their reservation type and market code. The feature of this figure is that it goes against common belief regarding hotel prices. The highest average price during the week appears on Monday and Tuesday, which goes against common sense. This is one of the unique features of the Washington D.C. area hotel market. As seen in Figure 2, many of the reservations are consisted of larger groups or people on business trips. These customers are relatively less sensitive to price according to the observation at the current price system. On the other hand, the hotel charges a lower price for weekend customers despite the considerable demand. Since tourists are sensitive to the price points, these price patterns are the result of revenue maximization. According to Figure 4, we conclude that we need to distinguish weekday and weekend bookings to find a better way of optimal pricing.
Figure 4: Bookings and Price by a weekday
3.3 Arrivals trends by leading day

Figure 5 plots the inflows and occupancy distribution by days before arrival (DBA), which were drawn from actual reservation records, i.e. reservation raw data. We classify them with quintile groups of demand to see their common patterns and particular patterns all at once. The first quintile group (demand rank larger than 80%) has 326 rooms of occupancy as an average, which looks dubious since our hotel capacity is 327. But this is caused by that “no show record is counted as an actual reservation in reservations raw data. So, if we exclude the no show record from Figure 5, it would be a more realistic value of inflow and occupancy. However, we will go with the reservation data including no show since we are not convinced of how the hotel marks no show records in their data set. The second quintile group shows an average of 305 rooms sold in a 60-80% demand rank. Hotel 0 has 312 of regular rooms, so this means that sales of hotel rooms are successful even in the second quintile group. The average room sold in the third quintile group should be close to the mean or median of whole occupancy set. Yet, 273, the average room sold in the third quintile is larger than the mean of total sample, 223.4. This reveals that occupancy distribution is concentrated in the high occupancy group. In other words, the frequency of occupancy value is denser as occupancy increases in the third quintile group. Lastly, two remaining quintile group are placed around 200 rooms as their final occupancy. All occupancy quintile groups entail a consistent increasing plot which depicts a near-linear relation between leading day and occupancy.

The left-side panel of Figure 5 shows inflows booking patterns of hotel 0. It is also classified with five quintile groups and is rather crowded compared to the right-side panel of occupancy. We discover an interesting pattern in inflows series in two peak points in the leading day. The first one is, as expected, is just before the occupancy day. The inflow tendency rapidly increases near this day. The second peak appears around 20 day in advance to occupancy day. The exact day varies by quintile group, but all groups have an apparent peak point at some point before the 20th leading day. In order to study this more in depth, it is necessary to divide the inflow data into two groups such as group reservations and transient reservations. This separate analysis facilitates to find a clear tendency in inflow pattern. We will dive deeper into the detail in the next section.
Figure 5: Inflow and Occupancy distribution by leading days
3.4 cancellation rate and policy on it

There are four panels in Figure 6 related to cancellation distribution.

**Upper-left** Frequency of cancellation is shown in upper left panel of Figure 6. The inflection point exists around two or three day before the occupancy. Since hotel 0 allows free cancellation one day prior to occupancy, the number of cancellations abruptly decrease as it approaches the occupancy date. Similar inflection points can be found in the other three panels as well. The average of daily cancellation is 0.97 and the maximum cancellations is 3.24. The cancellation pattern by the time is a increasing function before the inflection points located around two or three day before the occupancy.

**Upper-right** The next panel, upper-right, indicates the cancellation rate. Basically, this panel has similar shape of the frequency of cancellation panel, but having thick tails at the beginning of period. The cancellation rate moves around 0.5% until 10 days prior to occupancy, so the average is 0.62 and the median is 0.49.
For the simplicity, we assume that cancellations are lumped with the probability of cancellation existence. Without cancellation dynamics, our computation burden can be reduced much. For this purpose, two bottom panels in Figure 6 shows the statistics of cancellation for the estimation. The left panel indicates the probability that there exists any cancellation case. It is obvious that this probability is increasing as it approaches the occupancy date since the number of occupancy is also increasing. The average of cancel existence is almost 40% and this probability is peaked around a day prior to occupancy.

The lower-right panel shows the cancellation rate conditional on that any cancellation occurs. The average is 1.72% cancellation a day and it has different time trending compared to the upper-right panel for cancellation rate. Instead of using upper right panel of Figure 6, we use panel 3 (lower-left) and panel 4 (lower-right) in the model estimation.

### 3.5 Revenue vs Price vs Booking

We conclude our review of the hotel data set by showing Figure 7, which plots the summary of weekly hotel management. While Figure 3 is brought from market vision data which shows market price of hotel rooms on the online search channels, this data comes from actual reservation records. So the average price is significantly lower than the one found in Figure 3. The variability of hotel 0 revenue (dashed line) is large, compared to booking (bar plot) and price (solid line). This implies that off-peak season is a huge obstacle to the hotel manager. Even if the hotel restores the number of booking by lowering the price during off-peak season, it has to give up a huge loss in revenue. This graph launches our main goal to determine which price is optimal in this condition.

### 4 Preliminary Data Analysis

#### 4.1 $\rho'$ and $\rho$

\[
\log \rho_{t+1} = \alpha_t + \beta_t \log \rho_t + e_t
\]

We attempt to estimate the transition matrix for competitor hotels price ($\rho$) with AR(1) model. And the result is presented in Figure 8. As you see, $\beta_t$, the slope of AR(1) model, has slight variation with 1 as its center. And the magnitude of variance of $e_t$ has increased as it goes to the occupancy day. It denotes
Figure 7: Weekly Summary
Figure 8: parameters of price transition
that the accuracy of price($\rho$) transition matrix is reduced just before the arrival day.

4.2 BAR example of IDEAs pricing system

Figure 9 shows four typical observations of busy day when the hotel rooms are almost sold out.

**Upper-left**: April 18, 2013 is expected to be busy with cherry blossom and other spring festival in DC. The pricing system hotel 0 used forecasts high demand of accommodation based on historical records and suggests a severe price, $419, even at 45 days in advance of occupancy day. Keeping this price in succession, the occupancy rate has been increased consistently and reaches almost 98% on 20 days before the occupancy as 320 rooms occupied. But the hotel price starts increasing behind this 98% occupancy rate was attained and rises suddenly to $1050 on 15 day before the occupancy when it is likely that all of regular rooms are sold out. This striking price does not seem to be for a regular room since BAR term comes from hotel rooms which is currently available for reservation as mentioned before. So such a high
Another interesting fact we could observe is that hotel 0 is over-booked between 15 days and 6 days before the occupancy. At 13 days before the occupancy, the number of room reserved in hotel 0 is 345 which is 18 reservation rooms are over hotel 0 capacity. These excess reservations are reduced by cancellation until the occupancy day, so the occupancy of hotel reaches 324 in 3 days before and 319 in 2 days before. Since several vacant rooms are emerged by cancellation, the price of hotel 0 shows a sharp fall of $399 which is even lower than one of 45 days before the occupancy, initial price. Finally, the occupancy of hotel 0 re-increases 322 and BAR also rises to $599 around the last day. As April 18 is predictable busy day, it starts with high price and keep slight increasing the price. But the hotel rooms are almost sold out 20 days before the occupancy day. It yields that there is valid motivation to increase the price more than current level before 20 day in advance of occupancy. As a result, the hotel allows the overbooking and accept the risk of it with their tactical position.

Upper-right : July ordinary has average demand level of hotel in DC (Figure 2). So, BAR of July 14, 2012 starts with a normal price level, $209. Although the occupancy of hotel 0 is quite high (137) for the 45 days before the occupancy, the hotel decreases the price to $169 after 40 days in advance of occupancy. These lower prices catch the considerable demand class which has looking for the hotel room 20 days before the occupancy, so hotel 0 achieves the occupancy of 253, i.e. 77.4% of occupancy rate, when it comes to 15 days before the occupancy. Eventually, it results in high occupancy (323) of final day by keeping this price. In terms of occupancy, their price strategy was successful, but for the revenue of hotel, we need to examine it to see success or failure since high occupancy does not always guarantee high revenue.

Lower-left November 18, 2010 is the end of peak period. Since then and before Thanksgiving day, it becomes to a slack period of DC hotel. Thanksgiving day of 2010, November 25, is relatively late, compared November 24 of 2011 and November 22 of 2012. Thus, November 18, 2010 is classified as a peak period. The staring price of hotel 0 in 45 days before the occupancy is also as high as $319 which is a rather high and their occupancy of that day is also high as 120 rooms are already booked. There is rapid increase in occupany until 40 days in advance of occupany and it results increase in price to $339 after 37 days before the occupancy. Such an updraft in price continues, so it reaches $379 of price and 298 of occupany around 9 days before the occupany. But since then, it suddenly drops the price to $319 and it may cause the overbooking of rooms, 332 rooms are booked in 3 days before, and provides a cause of price lifting after then. Actually, it is reported that the number of booking at the occupancy day is over-
booking, 329, but it shows that final occupancy is 324 in the report issued after the occupancy day. The discrepancy between 329 and 324 rooms may be caused by no show though we do not have exact reason of it. Anyway, final occupancy, 324, means that the rooms of hotel 0 are almost sold-out, so it cannot be better in terms of occupancy. On the contrary, it has high risk of overbooking which was able to embarrass the hotel manager, so can be a huge failure in management.

**Lower-right** : March 22, 2013 has a demand above the average in a year. There is also high occupancy on the same day of previous year, March 22, 2012: both days are weekday. Reflecting this, the initial price of March 22, 2013 is not low, $239 since high demand is expected. However, due to the low flow of arrival, it reduces the price to $185 when it is 40 days before the occupancy. And since then there is a slight fluctuation around that price until the end of arrival day. It has quite consistent increase in occupancy and is concluded with occupancy of 310. It shows the accuracy of pricing system in practice.

Figure 10 illustrates four observations of slack day which has extremely low occupancy level. Considering 327 rooms of hotel capacity, all cases in Figure 10 have less than 30% occupancy rate. And their occupancy and price patterns are quite simple.

**Upper-left** : Hotel 0 keeps $279 as its BAR entire time. The demand of hotel can be expected to be a level of average or above it. But since the occupancy of 45 days before the occupancy is only 8, it should have decreased the price at some point before the arrival day. It implies that the pricing system can be improved as using more aggressive pricing strategy in this situation. At the end, the final occupancy is 73 which is lowest level in every occupancy records.

**Upper-right** : The price of room starts with $259 at the beginning. It is followed by lower price $239 in 28 days before the occupancy. But in spite of continuing low inflow, the hotel increases the price as $279 and it results the very low level of occupancy, 77, on the arrival day. It might be since the hotel focuses only customers whose willingness to pay is high. So, it is worth looking at how the optimal price from dynamic programming is changing this condition in comparison with actual price of the hotel.

**Lower-left** : December 20, 2012 is very quiet season in Washington D.C.. The initial price of $169 is maintained continuously with a slight drop later the period. Since the hotel charges a low price at the
Figure 10: Non Busy day example
beginning, they do not have many price options to take.

**Lower-right**: Usually middle of January has low demand (Figure 7). Despite it, hotel 0 sets the price high as $259 or $270, and it lasts until 6 days before the occupancy. In spite of low occupancy rate, they did not lower the price earlier than 6 days before the occupancy. We will attempt to find a better way of pricing while finding the reasons of pricing strategy of this hotel.

4.3 Arrival mean and sd

For the inflow process estimation, we have tried to apply Poisson distribution and Negative Binomial distribution both. Poisson distribution has restriction that their average and variance have to be the same. Under the result of observation, we find that the variance of arrival inflow has been increasing as it approaches the occupancy day. So, we examine Negative binomial distribution for inflow process instead of Poisson distribution.

As seen in Figure 5, the overall arrival pattern has a feature that they have rising curve around 20 days before the occupancy and changed to downward curve after that. The inflow process is calm for a while, and then it is increased rapidly again just before the occupancy date. We are able to succeed to explain the feature more accurately once we divide the data into two parts.

The upper two graphs in Figure 11 are inflow pattern of transient reservation and group reservation respectively. It presents the average inflow level by leading day. It shows that two type of reservations has totally different inflow pattern. First, transient reservations, which are occupied 6-70% of whole (Table 4), illustrates an increasing slope of inflow process (dashed line in the upper-left panel of Figure 11). Whereas, group reservations has rising cure in the middle of period which implies the feature we found in Figure 5 as well. The peak point is placed near 25 days in advance of occupancy and it decreases after then. If we combine these two graphs, it will show up as Figure 5.

The solid lines in Figure 11 are the result of polynomial fitting estimation based on shape of actual inflow trend. We use 3rd degree polynomial and result is fit as seen in upper graphs of Figure 11. The lower graphs are also the result of 3rd degree polynomial fitting, but for the negative binomial parameter $r$. As we see here, it should be a good estimation if we reflect what we found here in Figure 11.
Figure 11: Negative binomial dist. on Inflow
5 A Simple Model of Hotel Room Pricing

But in the interest of “starting simple” using the data we have, we propose the following model that has 3 state variables: \((n, \overline{p}, \rho)\) where \(n\) is the number of rooms reserved for occupancy on a specific date, \(\overline{p}\) is the average price (a combination of best available rate, BAR, and contractual rates for reservations made under pre-negotiated corporate and government rates) for these rooms, and \(\rho\) is an index of the prices charged by competing hotels for a comparable room.

The number of days prior to occupancy, \(t\), is also an implicit state variable. We will assume that hotel 0’s pricing problem starts some fixed number of days, say 30 or 90, prior to a given occupancy date. Hotel 0’s problem is to choose its own pricing strategy to maximize its revenues on each occupancy day. Thus, the hotel is solving many such dynamic programs in “parallel” since for any given calendar day customers are “arriving” to stay in the hotel on the same day (via past reservations or as “walk-ins” on the same day without advance reservation) or to make reservation at some occupancy date in the future.

In our analysis we will distinguish between several types or categories of occupancy days such as weekdays versus weekends. The types of customers who stay in the hotel during a weekday could be different than a weekend: for example, weekday occupancies are more likely to be for business travellers whereas weekends are more likely to be booked by tourists or other types of non-business travellers who may be more price-sensitive than a business traveller (whose hotel bill may be reimbursed and paid by his/her firm). Also among weekend and weekday occupancies we might distinguish certain days where the hotel knows in advance that there is likely to be overall excess demand for rooms. This could occur due to large conferences or demonstrations/marches occurring in the city or due to other seasonal reasons (i.e. at peak season for Cherry blossoms) and so forth. Let there be \(K\) different types of occupancy days.

We will assume that the particular days of any given occupancy type are essentially generic: i.e. we assume that the stochastic process governing the pattern of arrivals and cancellations, the types of customers who make reservations, and their willingness to pay for hotel 0 relative to other hotels are the same for all occupancy days of a given type. This allows us to pool all days of a given occupancy date type for purposes of econometric analysis and for solving the optimal pricing strategy for the firm. Thus, if there are \(K\) types of occupancy days, we will need to solve \(K\) corresponding dynamic programs. Let \(\delta_{t,k}(n_t, \overline{p}_t, \rho_t)\) be the optimal price hotel 0 will charge as a BAR reservation at non-contract rates for an occupancy day of type \(k\) when it is currently \(t\) prior to that occupancy date. If we have solved all \(K\) dynamic programs, then will have \(K\) corresponding optimal pricing rules \(\{\delta_1, \ldots, \delta_K\}\) and we can then regard these as “dynamic price schedules” that the firm simply quotes from when customers contact the hotel to make reservations.

Note that hotel 0 does not sell blocks of rooms at wholesale rates to third party hotel/travel inter-
mediaries such as Orbis, Hotels.com, Expedia, Travelocity, etc. However it does pay a commission for reservations that are made via these intermediaries. Hotel 0 provides daily updates to all of these third party hotel/travel websites of the uniform price it charges to all customers who are “transients” i.e. customers who do not have pre-negotiated contract rates with the hotel. The revenue that the hotel actually received when a reservation is made through a 3rd party travel website is this uniform price less the commission. Though the company can choose not to provide prices or take reservations through one of more of these third party hotel/travel websites, in practice we observe that it does pay the commission and provide price updates to these sites. Thus, we treat hotel 0’s decision to not sell blocks of its rooms at wholesale rates (which then amounts to a partial delegation of its pricing decision to third parties), as well as its decision to maintain a uniform pricing strategy and pay commissions as exogenously given and beyond the scope of this analysis. Similarly, we take hotel 0’s decision about negotiating government and corporate discount rates for group reservations as given and focus only on the pricing of the residual set of rooms that remain unreserved after any reservations have been made for rooms at these prenegotiated group/corporate/government rates.

Before we write the Bellman equation for this problem and explain the nature of the problem further, how is decomposition being used to simplify this problem? The first key assumption is that the hotel solves a sequence of separate dynamic programming problems, one for each occupancy date, i.e. the date that the rooms are being reserved for. There are at most 365 possible occupancy dates, and hence at most 365 separate dynamic programming problems that would need to be solved to determine the hotel’s pricing problem in each day of the calendar year. However by grouping the 365 possible occupancy dates into \( K \) possible types, it follows we only need to solve \( K < 365 \) dynamic programs to be able to generate a feasible hotel pricing strategy that enables hotel 0 to operate on a daily basis through the year.

When a reservation comes in, it can be a reservation for just one room or multiple rooms, and it can be for just one day or multiple days. So one of the first restrictions or assumptions we need to make to employ our decomposition principle is to ignore the possibility of length of stay based discounts, i.e. a lower daily rate for customers who book a room for a longer period of time. Initial empirical analysis using our data seems to indicate that length of stay based discounts are not easy to see in the data, and this suggests that they do not play a major role in attracting customers to stay in this hotel. This assumption could be wrong, however, and we note it as an assumption we might want to relax in future work if we find from an analysis of consumer demand that length of stay based discounts could be effective in attracting more customers and generating more revenue. But initially we will treat a single customer who wishes to reserve a room for \( J \) successive days as equivalent to \( J \) individual customers making separate, independent
1 day reservations. In future work we will attempt to relax this assumption and allow different prices to be quoted for customers who make reservations for multi-day occupancies.

Due to lack of data on how individual customers choose when to stay in Washington DC and which hotel to stay at in DC, the model we outline below is rather silent at first on the underlying model of “consumer demand” and adopts a semi “reduced form approach” that we may want to extend and develop further as our work proceeds. The key assumption is that there is essentially exogenous arrival of customers who wish to stay in a luxury class hotel in downtown DC for a given period of time and this fundamental underlying demand for a hotel room is not significantly affected by the price the hotel charges: i.e. customers do not change their dates of stay significantly based on the hotel prices they are quoted. Instead, we assume that where we are likely to see substitution is a decision on the part of consumers of which hotel to choose to stay. We are fortunate to have prices on competing hotels and we expect these competing prices will play an important role in our analysis and this is why we include this (as a price index or average of competing prices quoted for the given occupancy date) as one of the key state variables in the DP problem.

So we imagine a customer wants to arrive, for example, on July 5th and stay for 3 days. Their demand to be in Washington DC for days from July 5-7th, checking out on the 8th, is something we treat as an exogenous demand for most customers. For example it might be a business customer who needs to come for a business trip for these 3 days, or a person attending a 3 day convention, or a family who is on vacation and had planned this visit during a vacation period that they cannot easily change. So we will assume that the date of occupancy and the duration of occupancy are exogenously determined random variables that are not responsive to changes in the price charged by the hotel.

However we do assume that customers are likely to “comparison shop” and they will switch to a competing hotel in the same overall “quality class” if the price is significantly lower. The set of competing hotels to hotel 0 are all luxury hotels in a relatively close proximity to each other, and have comparable amenities. So we do assume there is an “elasticity of substitution” across these hotels, but no substitution with respect to staying in lower quality hotels or in luxury hotels that are further outside DC, or near the airports, etc. Again these are assumptions we can try to relax or examine later, perhaps with better data, but our data on hotel customers is rather limited in that we are observing a choice-based sample of consumers who chose to stay at hotel 0 and not at one of its competitors (though we do have overall occupancy data for the competitors). In particular, we do not observe customers who called for a price quote (or looked up a price on the web) and chose to stay at another hotel, or not to stay in any of the hotels. Further, we do not observe directly whether individual customers are on business or pleasure visits, or their income and
socioeconomic status and so forth.

One thing that helps our analysis is that hotel 0 adopts a uniform pricing policy for each of its room classes. That is, on each day there is a given price for a given class of room, and while this price may vary from day to day (and some additional variation due to corporate and government pre-negotiated discount rates that we will discuss later) the “best available rate” charged to customers who are not reserving under one of these pre-negotiated rate classes will be the same whether they call the hotel reservation line, or look up the price on the hotel website, or obtain a quote on any one of the many hotel website that quote prices for hotel rooms such as Travelocity, Expedia, Hotels.com, Orbitz, Trivago, and so forth. Thus, there may be intertemporal variation in hotel room rates for a given class of hotel room, and there is variation between corporate/business pre-negotiated rates versus the best available rate, there is not customer-specific within day variation in the best available rate as one can observe for other hotels that decide to sell blocks of rooms to third party sites such as Travelocity or Hotels.com at “wholesale rates” and then allow these firms the ability to try to resell them at a profit.

When hotels sell blocks of rooms in the “wholesale market”, there can be large variations in the price of a hotel room depending on which website one consults: if a given website such as Expedia has a block of unsold rooms and the occupancy date is quickly approaching, it may decide to sell these rooms at a deep discount if it is pessimistic about the number of consumers likely to buy the remaining rooms, whereas another website may be more optimistic about its ability to sell these remaining rooms and charge a higher price. Websites such as Trivago have arisen to pool information on hotel prices across the many different hotel websites to enable customers to obtain the lowest possible price on any given occupancy date, and by doing this they may tend to “arbitrage away” large variations in hotel prices for a given room on a given occupancy date. As we noted, hotel 0 maintains a uniform price but does pay variable commission rates to different 3rd party websites if they initiate the reservation for one of its rooms.

Finally, as we noted, our model treats reservations for either multiple rooms for reservations for multiple day periods as a sequence of independent single room, single day reservations. Thus, for the example of a customer who reserves 2 rooms on July 5-7th, we treat this as 3 separate one day reservations for 2 rooms: one of the 5th, one on the 6th and the third on the 7th and we analyze the pricing decisions for the 5th, 6th, and 7th occupancy dates as separate dynamic programs.

Similarly we ignore substitution across the 9 different room classes in our date set, i.e. between standard rooms and “luxury” rooms (which are usually on higher floors), rooms with balconies, rooms that are suites with kitchenettes, or additional living areas, and so forth. Most rooms are allocated to the standard rooms at the lowest price tier, and thus we ignore the possibility that a customer might contact
the hotel reservation desk looking for a luxury room with a balcony but decide to choose a standard room instead because of the significantly lower price of a standard room. Similar to our exogenous assumption about occupancy dates and length of stay, we assume that the customers make a choice of room class in advance and may substitute between hotels in an attempt to, say, find a luxury room in one of the DC luxury hotels on a given occupancy date for the lowest possible price, but they will not be likely to substitute and take a standard room, or a significantly more expensive luxury room (e.g. a penthouse suite).

This assumption is another example of a decomposition assumption that allows us to analyze the room reservation and pricing decisions for each of the 9 different room classes as separate dynamic programs for each occupancy dates. Thus, we will have to solve a total of 9*K dynamic programs to determine the pricing strategy for all nine room classes for hotel 0 for the K different types of occupancy dates in a given calendar year. If, with more detailed analysis of our existing data or analysis of other data leads us to realize that there is significant substitution across room classes, then we can attempt to relax these assumptions. For example if we felt that there was significant substitution between the standard and luxury room classes (the lowest two price tiers among the 9 room classes), then we could solve the pricing problem jointly as a 6-dimensional dynamic programming problem with state variables \((n_1, \bar{p}_1, \rho_1, n_2, \bar{p}_2, \rho_2)\) where \(n_1\) is the number of reservations for a standard room with occupancy on a specific future date, \(n_2\) is the number of reservations for a luxury room on this same date, and \(\bar{p}_1\) and \(\bar{p}_2\) are the average room rates quoted for standard and luxury room reservations made so far, and \(\rho_1\) and \(\rho_2\) are the average prices of standard and luxury hotel rooms quoted by the hotel’s competitors for the same occupancy date.

Now a bit of further notation and we can state the hotel’s DP problem. The hotel has a variety of corporate and government contracts and these contracts allow the clients who sign these contracts to reserve any available rooms in the hotel at pre-negotiated prices, subject to room availability. Though there are different government and corporate rates, we will let \(\bar{p}_c\) denote the average of these corporate and government contractually pre-negotiated rates and initially we will assume these prices are time invariant. The actual corporate and government contracts may be more complicated and include “block out dates” such as holidays where the pre-negotiated rate is not applicable, and there may be different rates for weekend vs weekday, or the rates can vary over the season of the year. We believe we can take some of these details into account in subsequent work but for now we will try to keep things simple and consider that there is a single average, pre-negotiated contract rate \(\bar{p}_c\) and that demand for hotel reservations from these corporate and government contracts is just an exogenous stochastic process. Let \(\nabla n_c\) denote the number of new corporate reservations that “arrive” in given period. If \(\nabla n_c < 0\) we treat this as a cancellation of one or more reservations at the pre-negotiated rate. Since these reservations are assumed to be exogenous, we
let \( g_t(\nabla n_c) \) be the probability distribution for the number of new corporate/govt room reservations that are made \( t \) days before the occupancy date.

Our decision problem will focus on the sequential determination of the best available rate (BAR) charged to customers who make reservations that are not covered under the corporate/government pre-negotiated contract rate \( \overline{p}_c \). Let \( \nabla n_b \) be the number of new reservations that are made at the current BAR \( p_t \) that the firm sets each day, where \( t \) is again the number of days prior to the occupancy date. We assume that if \( \nabla n_b < 0 \) this represents the cancellation of one or more previously made reservations, whereas if \( \nabla n_b > 0 \) then this represents a number of new reservations made for occupancy \( t \) days in the future at the price per night of \( p_t \). Let \( f_t(\nabla n_b|n,p,\rho) \) denote the conditional probability distribution of the number of new net reservations made (or cancellations made if \( \nabla n_b < 0 \)) for occupancy \( t \) days in the future when the number of reservations for this occupancy date totals \( n \) and the hotel quotes a room rate of \( p \) and the price index for the hotel rates charged by competing hotels for the same room class is \( \rho \). Finally let \( h_t(\rho'|\rho) \) represent an (exogenous) transition probability for the price index of competing hotels for occupancy \( t \) days in the future. That is, \( h_t(\rho'|\rho) \) is the probability distribution over the price \( \rho' \) that the hotel’s competitors may charge tomorrow (\( t - 1 \) days prior to occupancy) given that the price they are charging today, \( t \) days prior to occupancy, is \( \rho \).

Let \( \overline{p}_{t-1} \) be the average expected room rate for reservations for this hotel after today’s reservations have been made. If \( \nabla n_b > 0 \) so there are net new reservations made today (\( t \) days before occupancy) at price \( p_t \) then we have

\[
\overline{p}_{t-1} = \frac{n_t \overline{p}_t + \nabla n_b p_t + \nabla n_c \overline{p}_c}{n_t + \nabla n_b + \nabla n_c} \quad \nabla n_b > 0
\]

so that \( \overline{p}_{t-1} \) is a moving average that accounts for new reservations at the best available rate \( p_t \) as well as new reservations at the corporate/government rate \( \overline{p}_c \). If there are net cancellations on a given day, we will assume that the cancellations are at the current average room rate \( \overline{p}_t \), so we have

\[
\overline{p}_{t-1} = \frac{n_t \overline{p}_t + \nabla n_b \overline{p}_t + \nabla n_c \overline{p}_c}{n_t + \nabla n_b + \nabla n_c} \quad \nabla n_b < 0
\]

(3)

Let \( \overline{p}_t = \lambda(n, \nabla n_b, \nabla n_c, \overline{p}_t, \overline{p}_c, p_t) \) denote the law of motion for the average reserved price in equations (2) and (3) above.

Now we have the notation we need to write down hotel 0’s dynamic program. Let \( V_0(n,p,\rho) \) be the hotel’s value on the occupancy date, i.e. \( t = 0 \) days before occupancy. Then this will simply be given by

\[
V_0(n,p,\rho) = n \overline{p}(1 - \kappa)
\]

i.e. the hotel receives the average reservation price of \( \overline{p} \) for the \( n \) reserved rooms in this rate class. The factor \( \kappa \) is the product of the share of reservations that arrive via third party websites times the average

38
commission rate charged by these sites. Thus, \( n\bar{p}(1 - \kappa) \) is total revenue from occupancy on a given date net commissions. We treat other costs such as electricity, heating/cooling, room cleaning, and use of water, internet, television/cable and other services in the hotel as fixed costs, but in future work with more data we will adjust the terminal payoff on the date of occupancy to deduct variable costs that are a function of occupancy \( n \).

Let \( T \) be the horizon in days prior to any given occupancy date where the first reservations start to be made. Typically we can assume \( T = 90 \) since reservations are rarely made more than 90 days in advance of occupancy, but this can easily be changed to be the largest integer that encompasses all advance bookings made for a given room class in this hotel on a given occupancy date. Then we have the Bellman recursion, ignoring discounting,

\[
V_t(n, \bar{p}, p) = \max_p \left[ \int \rho' \sum_{\n_b \n_c} V_{t-1}(n + \n_b + \n_c, \lambda(n, \n_b, \n_c, \bar{p}, \bar{p}_c, p), p') g_t(\n_c) f_t(\n_b | n, p, p) h_t(p' | p) \right] \quad (5)
\]

Thus, hotel 0 sequentially decides on the optimal price \( p_t \) to charge at the start of each day, \( t \) days before occupancy, in order to maximize the expected revenue it will earn on the final occupancy day, \( V_0(n, \bar{p}, p) \). The hotel pays attention to the price index of its competitors, \( \rho \), because if it raises its own price \( p_t \) too high, it can expect customers to substitute to a competing luxury hotel in DC.

Note that for notational simplicity we have omitted the index \( k, k \in \{1, \ldots, K\} \) indexing the type of occupancy date and an index \( i \) for room type, \( i \in \{1, \ldots, 9\} \). In principle each of the stochastic laws of motion in the Bellman equation (5), \( g_t, f_t \) and \( h_t \), also depend on \( k \) and \( i \) and thus \( V_t \) will also depend on \( k \) and \( i \), which by construction are time-invariant characteristics. Thus, we actually have to solve \( 9 \times K \) separate dynamic programs for each of the \( 9 \times K \) types of rooms and occupancy dates (e.g. weekend, weekday occupancies, etc).

Each of these dynamic programs results in an optimal decision rule \( p_{t,i,k} = \delta_{i,k}(n, \bar{p}, p) \) that specifies the hotel’s optimal price for a room in room class \( i \) \( t \) days prior to occupancy (where occupancy date type is \( k \)) when hotel 0 has \( n \) rooms reserved already for that occupancy date at an average price of \( \bar{p} \) per room, and the index of prices charged by hotel 0’s competitors for a similar class room on the given occupancy date is \( \rho \). Our interest is to see if we can solve this dynamic program numerically and determine how far the actual prices charged by this hotel are from the “optimal prices” from the solution to our dynamic programming problem.

In the next section we discuss our econometric approach to estimation of a behaviorally justified transition probability \( f_t(\n_b | n, p_t, p) \) that represents the number of new reservations that the hotel expects to
make \( t \) days prior to occupancy if it quotes a price of \( p_t \) when its competitor hotels are charging an average price of \( p \) for the same class of hotel room.

### 6 Model of Demand for Hotel Rooms

This note lays out the simplest dynamic model of pricing, intended to be a starting point for discussions, not necessarily a proposal for the model we will ultimately want to estimate. But to keep the dynamic programming problem tractable, we employ extensively the principle of decomposition — namely decomposing a difficult overall problem the hotel is facing to solve its revenue maximization and pricing problem into a number of more tractable subproblems. Of course, this decomposition approach relies on assumptions to simplify the overall problem, but the assumptions may not be good ones. So we fully expect that as we become aware of problematic aspects of the simplifying assumptions underlying this simplest first generation model, we will start to work on a succession of less restrictive, more realistic but computationally more demanding models. To respect the anonymity of the hotel we are studying, we refer to it as “hotel 0” below.

We will start with a very simple model for \( \nabla n_{b,t} \), the number of net new best available rate reservations, net of cancellations, where hotel cancellations are assumed to be “exogenous” and random (i.e. not dependent on the firm’s pricing or that of its competitors), and new reservations are a compounding of an exogenous stochastic arrival process (i.e. a stochastic process for the number of new customers making reservations at luxury hotels in the Washington DC market that does not depend on the firm’s pricing or that of its competitors). Thus we can write

\[
\nabla n_{b,t} = n_{b,t}^+ - n_{b,t}^-
\]

where \( n_{b,t}^+ \) is the number of new reservations for occupancy at date 0 that are made \( t \) days prior to occupancy and \( n_{b,t}^- \) is the number of cancellations of reservations for occupancy at date 0 made \( t \) days prior to occupancy. As above, the ‘b’ subscript refers to the “best available rate” (BAR). We could write a similar equation decomposing the number of new corporate-rate room reservations arriving \( t \) days prior to occupancy, \( \nabla n_{c,t} \), but in general we treat these reservations as an exogenous random process since they depend on a pre-negotiated corporate room rate \( \bar{p}_c \), which we assume is exogenous and pre-determined with respect to the day by day decisions the hotel makes over the best available rate \( p_t \) it sets for individual room reservations that are not covered under a government or corporate pre-negotiated room rate agreement. Though it is possible that \( p_t \) may be a covariate that could be useful in predicting \( \nabla n_{c,t} \), we will not adopt
a more detailed “structural interpretation” of this process as we do for \( \nabla n_{b,t} \), which has a choice-theoretic foundation that we discuss below. We will discuss the econometric model for the evolution of \( \nabla n_{c,t} \) below, after discussing our more structural econometric model for \( \nabla n_{b,t} \).

We assume that with \( n_t \) total reservations made prior to occupancy, that the number of cancellations \( \nabla n_{c,t} \) is a binomial random variable, \( \nabla n_{c,t} \sim \text{bin}(n_t, \pi^-) \) where \( \pi^- \) is the probability that any individual room of the \( n_t \) reservations is cancelled (and hence cancellations of individual rooms are treated as independent Bernoulli trials with probability \( \pi^- \)). We will adopt a reduced-form approach to room cancellations and we may allow \( \pi^- \) to depend on the prices \( (p_t, \rho_t) \) in a flexible way, though we will not have any particular structural interpretation of theory of why consumers choose to cancel a reservation that they previously made.

For the number of new incoming reservations, \( n_{b,t}^+ \) we adopt a more complex compound arrival/binomial model that includes a simple static discrete choice model as a submodel explaining how many new reservations are made \( t \) days prior to occupancy. We assume that reservations are made for only individual rooms in a given room class, and reservations are independent of each other. Assume that there are a total of \( k_t \) individuals seeking to reserve a single room \( t \) days prior their intended occupancy date. We do not have any complicated dynamic model for the optimal timing of a hotel reservation in mind in our first simple specification of the model. Instead we assume that consumers simply try to reserve a room as soon as they have confirmed their travel dates, or whenever they are in a position to focus on making a hotel reservation for their trip, and at that time they make a simple static choice about whether to stay at the hotel we are studying, or to stay at one of the competing luxury hotels in Washington DC.

We will shortly discuss the probability distribution that \( k_t \) is drawn from, but conditional on \( k_t \), the \( k_t \) individual reservations involve independent binomial choices of whether to reserve at room at the hotel we are studying at price \( p_t \), or to make the reservation at one of the competing hotels at price \( \rho_t \). Let the consumer “type” be indexed by \( \tau \) and assume that a consumer of type \( \tau \) chooses to reserve the room at the hotel that provides the highest utility, taking into account Type 1 extreme value distributed shocks that represent other idiosyncratic factors affecting their choice of which hotel to reserve at. We normalize the net utility of staying at a competing hotel to be 0 and the net utility of the hotel we are studying to be \( \alpha_{\tau} \), and let \( b_{\tau} > 0 \) denote consumer \( \tau \)'s degree of price-sensitivity. Then a consumer of type \( \tau \) chooses our hotel, which we denote by the choice \( d = 0 \) if

\[
a_{\tau} + b_{\tau} p_t + \varepsilon_0 \geq b_{\tau} \rho_t + \varepsilon_1
\]

where \( \varepsilon_0 \) and \( \varepsilon_1 \) are independent Type 1 extreme value distributed random variables with mean 0 and scale parameters normalized to 1. This implies that the probability that the consumer chooses to reserve at the
hotel we are studying, \( d = 0 \), is given by

\[
P_r \{ d = 0 | \tau, p_t, \rho_t \} = \frac{\exp \{ a_\tau + b_\tau p_t \}}{1 + \exp \{ -a_\tau - b_\tau (p_t - \rho_t) \}}.
\]  

(8)

Assume that \( f(\tau) \) is the probability that an individual consumer is of type \( \tau \) and assume there are a finite number \( K \) of types of different consumers, then we have

\[
P_t(p_t, \rho_t) = \sum_{k=1}^{K} P_r \{ d = 0 | \tau_k, p_t, \rho_t \} f(\tau_k)
\]

(9)

It follows that the number of new reservations at hotel 0 will be \( n_{b,t}^+ \sim \text{bin}(k_t, P_t(p_t, \rho_t)) \), i.e. a binomial distribution with parameters \( k_t \) and \( P_t(p_t, \rho_t) \).

Finally we assume that the arrivals of new reservations \( k_t \) is a realization from a distribution \( f_t(k) \) such as a Poisson distribution or a negative binomial distribution. For example, if we use the Possion distribution with parameters \( k_t \) and \( P_t(p_t, \rho_t) \),

\[
f_t(k|\mu_t) = \exp \{-\mu_t\} \frac{\mu_t^k}{k!},
\]

(10)

and as well known the Poisson distribution has the restriction that \( E\{k|t\} = \mu_t = \text{var}(k|t) \). The negative binomial distribution is a two parameter family given by

\[
f_t(k|r_t, \mu_t) = \binom{k + r_t - 1}{k} (1 - \mu_t)^{r_t} \mu_t^k
\]

(11)

and for this distribution we have \( E\{k|t\} = \mu_t r_t / (1 - \mu_t) \) and \( \text{var}(k|t) = \mu_t r_t / (1 - \mu_t)^2 \) where \( \mu_t \in (0, 1) \) and \( r_t \) is a positive real number.

Now consider the likelihood for observations of \( n_{b,t}^+ \), the number of new room reservations made at hotel \( 0 \) \( t \) days prior to occupancy. We do observe this quantity from our data and though we do not observe the total number of room reservations \( k_t \) made at competing DC luxury hotels, we can conclude that \( k_t \geq n_{b,t}^+ \). Thus let \( L(n_{b,t}^+|\theta) \) be the probability of observing \( n_{b,t}^+ \) new reservations \( t \) days prior to occupancy, where \( \theta \) is the vector of parameters of the choice model, \( (a_\tau, b_\tau, f(\tau)) \), \( k \in \{1, \ldots, K\} \) and any other parameters entering \( (\mu_t, r_t) \) for the negative binomial or Poisson model of new arrivals of new customers seeking room reservations at a luxury hotel in DC \( t \) days prior to occupancy. We have

\[
L(n_{b,t}^+|\theta) = \sum_{k \geq n_{b,t}^+} \binom{k}{n_{b,t}^+} P_t(p_t, \rho_t)^{n_{b,t}^+} [1 - P_t(p_t, \rho_t)]^{(k-n_{b,t}^+)} f_t(k|\mu_t, r_t).
\]

(12)

In principle the parameters \( \theta \) can be estimated by the method of maximum likelihood, by pooling observations of new reservations \( n_{b,t}^+ \) at the various different dates \( t \) prior to occupancy for different but “similar” occupancy days (i.e. separating week days and weekend days and treating “special dates” such as holidays
or other events where occupancy is likely to be unusually high or low separately from other more “normal” occupancy dates).

Identification of this model of demand may be problematic without independent observations on $k_t$, the number of total new reservations made at DC luxury hotels $t$ days prior to occupancy. The basic problem is clear: if we see, say, a total of 10 new reservations made on a given day, was this because hotel 0 managed to attract 50% of a total of 20 new reservations that were made that day at luxury hotels in DC, or only 10% of 100 new reservations? Since the total number of new reservations $k_t$ that are made on any given day at luxury hotels in DC $t$ days before occupancy is not a variable we observe, there is an issue of how to identify the $(a_\tau, b_\tau, f(\tau))$ parameters from the heterogeneous agent choice model from the parameters $(r_t, \mu_t)$ of the stochastic model of the total number of “arrivals” $k_t$ of new hotel reservations $t$ days prior to occupancy.

Part of the identification may come via functional form assumptions such as the “exclusion” restriction that $(p_t, \rho_t)$ only enter the average choice probability $P_t$ of the binomial model $\text{bin}(k_t, P_t(p_t, \rho_t))$ for the number of reservations that are made at hotel 0 conditional on the total number of arrivals equalling $k_t$, and not the parameters $(\mu_t, r_t)$. However it seems clear that we need other identifying information that can help determine overall “market share” such as aggregate information on the number of paid rooms at different luxury hotels in Washington DC. There is a hotel room tax and it may be possible to get this information from the Washington DC Department of Revenue, but in the absence of such data, more thought will have to go into how to identify the demand model. While it is possible to try to adopt a more “reduced form” approach of only attempting a distribution $f_t(\nabla n_{b,t} | p_t, \rho_t)$ that does not necessarily have any structural interpretation as given above, there is a possibility that such an approach will lead to “spurious causality” such as an implied upward sloping demand for rooms in the hotel’s price $p_t$. That is, if there are specific days for which occupancy at this hotel is particularly desirable (such as visiting Washington DC during peak period for the spring cherry blossoms) there could be surges in new room reservations $\nabla n_{b,t}$ that are matched by corresponding increases in room rates $p_t$. Thus, $p_t$ may be raised in response to surges in $\nabla n_{b,t}$ but there is a real danger that a reduced form model may interpret this correlation in the opposite direction, i.e. that increases in $p_t$ “cause” surges in $\nabla n_{b,t}$. Thus, it is important to search for “instruments” — i.e. variables that are likely to change the relative prices $p_t$ and $\rho_t$ that might be considered as “exogenous price shifters” that can lead to a correct inferences about the effect on changes in $p_t$ on the probability a typical consumer would choose to stay at hotel 0.
7 Update for new data from Smith Travel Research

Via a stroke of good luck, we found out that Georgetown has a contract with a company called Smith Travel Research that maintains daily occupancy and revenue data for individual hotels. They were able to provide us a spreadsheet with daily level total occupancy for hotel 0 and its six competitors for the period January 1, 2010 to October 31, 2016. Though we do not observe sequence of reservations made at these other hotels in advance of each occupancy date, this new data is a big boon to our research program as it provides key additional info we need to identify market share which is critical to the identification of the parameters underlying the individual customers’ choice of which hotel to stay at. For confidentiality reasons, STR did not provide us data on daily occupancy of each of the individual hotels, but only the total for all of them on a day by day basis. It also only provides the average daily rate for all of the hotels as a group, not individual hotel ADR’s. However this is not a problem since we can patch in the Market Vision data to get the ADR’s for the individual hotels that compete with hotel 0. Also our initial model really only has two prices: the price quoted by our hotel (on a day by day basis for reservations in advance of any given occupancy date) and the ADR of competing hotels from the Market Vision data. In this section we will continue to denote the price quoted by hotel 0 for a room \( t \) days prior to occupancy as \( p_t \) and use the notation \( \rho_t \) for the ADR quoted by the other competing hotels for a comparable room \( t \) days prior to the desired occupancy date.

The new data from STR provides us with \( n_t \), the total occupancy in the seven properties covered by the STR spreadsheet on calendar date \( t \). We do not observe the path of reservations and cancellation that lead to this total occupancy \( n_t \) on calendar day \( t \) but we do observe the path of reservations and cancellations for our particular hotel. We now describe how to form a likelihood function that uses the combined STR data and the data we already have on our particular hotel. We first discuss and then dismiss maximum likelihood due to the high dimensional numerical integrations necessary to write down the likelihood function for the combined set of data, though Monte Carlo integration might be a feasible way to estimate the parameters of the demand/arrival/cancellation model by simulated maximum likelihood.

Instead we will recommend an alternative estimation strategy, simulated minimum distance, that is much easier to implement. Intuitively it is based on combining data we have on paths of advance reservations and cancellations of rooms for “similar” occupancy days at our hotel with the data from STR to construct a set of moments that we want our simulation model to match. Thus we will consider, for example, “regular weekdays” as a set of days that we can pool data over, and perhaps estimate a separate model for “regular weekends” where “regular” means that these are not holidays or times of special demand for rooms (tourist high times such as spring Cherry Blossom Festival, or Inauguration Week, or periods
where there are large professional conferences occurring (e.g. the ASSA meetings) that put heavy demand on available hotel space of the luxury hotel group we are analyzing. These “extraordinary” weekdays and weekend occupancy dates would be separately pooled and analyzed as the dynamics of reservations, cancellations and pricing are likely to be different for these types of days than for “regular” days.

By pooling over similar types of occupancy days we can obtain a large number of moments such as the average number of rooms reserved \( t \) days prior to occupancy, the average number of new reservations and cancellations on each day, as well as the total number of rooms sold in the competing hotels on the occupancy day from the STR data set. What we don’t observe is the path of reservations and cancellations leading up to the number of rooms actually sold on a given occupancy date at the competing hotels, but we do know that from our observation of the number of new reservations at our hotel, then the number of total new reservations (which we also term as “arrivals” of new reservations) and denote by \( a_t \) is at least as great as the number made at hotel 0.

To understand the data we have and the complexity of the likelihood, let’s focus on a particular calendar day as the fixed “occupancy date” that we are trying to analyze, such as July 12, 2012. Let \( n_0 \) be the total occupancy in the hotels we study on that date, which from the STR data was 1955 rooms sold. We also observe \( n^d_0 = 321 \) which is the number of rooms sold on that day at hotel 0. Let \( a^d_t \) and \( a_t \) be the number of new reservations made for occupancy on the given date (July 12, 2012) \( t \) days prior to occupancy at hotel 0 and at all of the competing hotels (including hotel 0), respectively. Clearly we must have \( a_t \geq a^d_t \). Let \( n^d_t \) and \( n_t \) be the total number of rooms reserved on this occupancy date at hotel 0 and at all hotels in this group, respectively. Finally let \( c^d_t \) and \( c_t \) be the number of cancellations of reserved rooms at hotel 0 and all hotels in the group, respectively. Then we have \( n_t \geq n^d_t \) and \( c_t \geq c^d_t \), and the following accounting identities

\[
\begin{align*}
    n^d_{t+1} &= n^d_t + a^d_t - c^d_t \\
    n_{t+1} &= n_t + a_t - c_t \quad (13)
\end{align*}
\]

We will assume that the arrivals to all of the hotels is an independent but not necessarily identically distributed arrival process: that is, we assume that \( a_t \) is a draw from a probability distribution \( f_t(a_t|x_t) \) that may depend on the number of days prior to occupancy as well as other seasonal characteristics captured in a vector of variables \( x_t \) that might include seasonal dummy variables and so forth. We thus assume that \( a_t \) and \( a_s \) are independently distributed if \( t \neq s \) but do not assume that \( f_t(a_t|x_t) \) and \( f_s(a_s|x_s) \) have the same probability distribution.

We also assume that cancellations are independent events, i.e. if there are \( n_t \) reserved rooms, each room reservation may be cancelled independently of other cancellations but the probability of a cancellation,
\( p_t^d(x_t) \) may depend on the number of days in advance of the occupancy the cancellation occurs, \( t \), as well as a vector of covariates \( x_t \) to also capture seasonal effects on cancellation rates. Thus we allow for potential time-varying cancellation rates, but we assume that the cancellation rates at all hotels in the group we are studying are the same. This implies that \( c_t \) is a binomial distribution with parameters \( n_t \) and \( p_t^d \), and similarly, the number of cancellations at hotel \( d \) is \( c_t^d \) which is binomial with parameters \( n_t^d \) and \( p_t^d \).

The probability distribution for new arrivals of reservations \( a_t \) may be a Poisson or Negative binomial distribution (to allow for potential “overdispersion”, i.e. where the variance of new reservations is larger than its mean, which are constrained to be equal under the Poisson specification), but the probability distribution for \( a_t^d \) at hotel \( d \) results from an additional choice process as described above. That is, if we let \( f_t(a_t|x_t) \) denote the probability distribution of total arrivals of new reservations for a given occupancy date \( t \) days prior to this date, then the distribution of \( a_t^d \), the number of these new arrivals choosing to reserve at hotel \( d \) when faced with price quotes \( (p_t, p_t) \) is denoted by \( f_t(a_t^d|x_t, p_t, p_t) \) and given by

\[
   f_t(a_t^d|x_t, p_t, p_t) = \sum_{k=0}^{a_t^d} \binom{k}{a_t^d} p_t(x_t)^{a_t^d} [1 - p_t(x_t)]^{(k-a_t^d)} f_t(k|x_t). \tag{14}
\]

The probability for the number of new reservations at hotel \( d \) in equation (14) is a mixture of binomial probabilities that a total of \( k \) new reservations arrived but only \( a_t^d \leq k \) of them chose to reserve at hotel \( d \) compared to one of the other competing hotels. We must sum over all \( k \geq a_t^d \) because we do not observe the number of new reservations made at all of the hotels, but only \( a_t^d \), the number of new reservations at hotel \( d \).

There are additional implicit constraints imposed by the fact that we do know that on the date of occupancy there were \( n_0 \) rooms sold (and occupied), so this places an implicit upper bound on the total number of new reservations since it cannot exceed the total capacity in all of the hotels. However even if the hotels are close to being sold out collectively (i.e. \( n_t \), the total number of reservations made \( t \) days prior to occupancy) there may be cancellations of rooms in the days \( t-1, t-2, \ldots, 0 \) before occupancy.

It is easy to write a computer simulation program to simulate a set of paths \( \{a_t^d, c_t^d, n_t^d, a_t, c_t, n_t\} \). For example if we start at \( t = 90 \) prior to occupancy (or some date such that the hotels do not take reservations further in advance than this), we set \( n_{90} = 0 \) and \( n_{90}^d = 0 \) and then simulate in order: 1) total arrivals \( a_t \), 2) using the binomial probability with parameters \( (P_t, a_t) \) we simulate the number of the \( a_t \) reservations that are made at hotel \( d \), 3) we simulate total cancellations \( c_t^d \) at all of the hotels as a draw from a binomial distribution with parameters \( (n_t, p_t^d) \) and then simulate independently the number of cancellations at hotel \( d \), \( c_t^d \) as a draw from a binomial with parameters \( (n_t^d, p_t^d) \), and 4) we update the total reservation counts using the identities (13) and then advance to the next time period (i.e. decrement the time counter \( t \) to \( t-1 \).

46
Suppose we have obtained a vector of moments $\overline{m}_N$ by pooling data from $N$ different “similar” occupancy days. This vector will contain the average values over these $N$ occupancy dates of the number of arrivals, cancellations and total occupancy at hotel $d$ ($n^d_t, a^d_t, c^d_t$) as well as the average occupancy on the occupancy date for all of the hotel over these $N$ days. These moments then represent a “typical path” of reservations, cancellations and room reservations for hotel $d$ as well as the typical total occupancy in all of the competing hotels. Other moments can also be included such as the variance or quantiles of the distributions of $n_0$ and $n^d_t$ for various $t$. Let $\hat{\Omega}_N$ be a $K \times K$ covariance matrix of the moments $\overline{m}_N$, which we assume is a $K \times 1$ vector with $K \geq J$ where $J$ is the number of unknown parameters in our probability models of customer choices, arrivals, and cancellations.

Let $\theta \in \mathbb{R}^J$ be a guess of the parameters of the model. Using $\theta$ we can simulate trajectories as described above. Using $S$ independent simulations, we can construct simulated moments $\overline{m}_S(\theta)$ that censor our simulated data in the same way they are censored in our actual data — i.e. we do not observe $\{n_t, a_t, c_t\}$ for $t > 0$, but only $n_0$. Then if $\theta$ represents the $J \times 1$ vector of parameters of the choice model, arrival model and cancellation model, the SMD estimator $\hat{\theta}$ is given by

$$
\hat{\theta} = \arg\min_{\theta} [\overline{m}_s(\theta) - \overline{m}_N]' \hat{\Omega}_N [\overline{m}_s(\theta) - \overline{m}_N].
$$

(15)

8 Estimation

8.1 Estimation Strategy

Simulated Methods of Moments :

Process of SMD estimation :

8.2 Identification

8.3 Estimation result and Model Fit

This section presents estimate results of the hotel model implemented by the simulated method of moments (SMD) estimation. As mentioned in the previous section, the aim of the SMD is to minimize the quadratic form in a distance vector between actual and simulated moments we are interested in the hotel model. Our main empirical task is to see how well the actual data and simulation data match up with each other. The process of estimation enables the predicted optimal price of our hotel to fit actual price of the
hotel we observe in the data when taking different parameters. We illustrate our principal empirical finding in Figure 12 before turning to the details of the estimation results. We classified the data set into eight groups according to weekend and seasonal hotel demands in the Washington D.C. market. The suggested plot is brought from one specific day of sample20 which is a weekday and the second highest demand sample (top 25-50%). Since the SMD basically uses moments of sample, the estimated parameters is the average value of the sample.

**Model fit example of sample20 day41**  The top left panel of Figure 12 plots the BAR of our hotel and the average BAR of competing hotels, while the top right panel shows the optimal price of our hotel generated through our simulation. In both panels, we use the same data for the average BAR of competing hotels. It shows that the computed optimal BAR price, the mean simulation on the top right panel, closely follows the competitors movements. Through this, we are able to find the effect of these price differences between the actual data and simulated data on occupancy. In the middle left panel, you find two different lines...
for occupancy. The dashed-line describes the simulated occupancy of our hotel while the star-solid line indicates the actual occupancy level of our hotel on the same day. Around 10 days before arrival, the simulated data (dashed-line) displays higher occupancy compared to the actual occupancy. As it approaches the arrival date, i.e. moving to the left end along the horizontal line, the actual occupancy level increases more than the simulated one. The higher occupancy in the actual data leads us to conjecture that the real hotel price is higher than the price created by the model. As expected, we are able to find a lower price in real data during the 0-5 days before arrival in the two top panels of Figure 12. This leads to a sensible suggested price by the model and reveals that occupancy maximization does not always lead to revenue maximization.

The remaining three panels of Figure 12 demonstrate daily inflow and outflow of our hotel reservation. The middle right panel called BAR reservations at H tells us how many new reservations have been made on that day. BAR reservations signify the aggregate reservations of both leisure and business customers. The actual inflow marked with the star-dash fluctuates widely, so it is hard for the model to capture the details of its daily movement. Despite this difficulty, our model succeeds in simulating the tendency of arrival inflow.

Two panels located in the bottom of Figure 12 represent daily inflow of group reservations and daily cancellation within overall bookings. Unlike the previous panels, they have distinct differences between the data and simulation. Since our model depends on a simple assumption regarding occurrence probability, it is unavoidable to allow these gaps. Although the simulation result is not always consistent with the data for a specific day, it is designed to be consistent when it comes to average rate of the sample. We will describe in more detail with estimation results.

Basically, the SMD uses the moments of the data set. So, the moments which represent the properties of the data set are expected to have similar values in both the actual data and the simulated data. As seen previously, we found the good-fit example of the individual day where the simulated data is consistent with the actual data. Considering the process of the SMD, the model fit has to be improved when we see a case of average factor of data set instead of any individual day. Indeed, the model closely matches the real data, especially the mean of BAR and the mean of occupancy in Figure 13 - 15.

Figure 13 and Figure 14 show the average BAR movement in each sample group. Figure 13 is com-
Figure 13: BAR trajectory for Sample20

puted in sample20, which is a weekday and the second highest demand sample (top 25-50%), while Figure 14 is derived from sample31 which represents a weekend and the highest demand sample (top 25%). Although Figure 14 speaks for most peak seasons for the weekend, their BARs of our hotel, marked po and shaped star-solid, revolve between $310 and $330 which is lower than BARs of Figure 13 which is around $350. The same pattern appears in competitors price marked as pc in the Figure 13 and Figure 14. This observed fact reflects the essential feature of the Washington D.C. hotel market. Since there is high demand on the hotel rooms during weekdays, it implies that business travelers are influential to the hotel market in this area. We can also compare price elasticities for business and leisure customers. As expected, business customers are less elastic to hotel room price. This enables the hotel manager to set a higher price during weekdays.

Another thing you need to pay attention to in Figure 13 and Figure 14 are the similarity of our hotel price, po. In the estimation, the model attempts to find parameters which make the simulated data closer
to the real data. Although we do not use the actual data of our hotel price when we compute the optimal price in the model, they are both close in terms of pattern and value. When the star-solid actual price goes down over time, the star-dashed simulated price also has a downward slope in Figure 13. Figure 14 shows a more dramatic result of estimation. Average BAR estimation works well even for the irregular pattern of actual BAR. Given the increasing price of competitors, our hotel price in the actual data moves up and down which the simulated data follows this movement very closely. This reveals that the estimation with the SMD is very successful for the BAR.

The well-simulated price affects the progression of occupancy. In Figure 15, we see the average occupancy flows for our hotel and competing hotels. Since we do not have reservations details of other hotels, the occupancy flow of others is not tractable. We only observe the final day occupancy for other hotels, so we have to rely on some assumptions related to flow of other hotels reservations. We surmise that other hotels booking flow has a similar shape to our hotel. Unfortunately, the bottom panel of Figure 15
only shows disparity in shape between ours and others. We know the insufficiency in our model on the competing hotel occupancy. However, for our hotels occupancy, the model fit is matched so well that we feel our model does an excellent job of capturing the feature of our hotel data. The top panel of Figure 15 plots the simulated occupancy and the actual occupancy. The two lines are so close that it is hard to distinguish from one another.

Parameter Estimates Overall, we are confident that the estimation works excellent and our model is able to mirror the real world hotel data. Now we discuss the parameters which constitute the main result of estimation. We use 46 parameters to estimate our model and we have those parameter sets for 8 different sample groups. The parameters we would like to mention first are related to choice probability. Equation (16) defines the parameter alpha and beta using Equation (8).

\[
Pr\{d = 0 | \tau, p_t, \rho_t\} = \frac{1}{1 + \exp\{-a_\tau - b_\tau(p_t - \rho_t)\}} = \frac{1}{1 + \exp\{\alpha_\tau + \beta_\tau(p_t - \rho_t)\}}.
\] (16)
From Equation (16) above, we derive the price elasticity of choice probability as follows,

$$e_p = \frac{dQ}{dP} \frac{P}{Q} = -\beta \tau \cdot p_t,$$

 prominently, we find that beta coefficients are always higher in leisure than that of business except in the case of a weekday during the highest demand season (Table 5). That season is an exceptional time for the Washington DC luxury hotel market. The minimum occupancy rate of seven hotels analyzed in this paper was higher than 92.0% during this peak period. Considering the fact that 95.4% rooms of our hotel are regular rooms, almost all of the regular rooms of the luxury hotels are literally sold out in this season. So, the highest demand season of weekday is inappropriate to use to generalize price elasticity. Our main interest focuses on other parameters in the remaining 7 seasons. For example, the sample20 we use for Figure 13 and Figure 15 has a 0.010 beta coefficient for the leisure customer while the beta coefficient for business customer is 0.006 as seen in Table 5.

Another thing we find is that all beta parameters have the same sign. With Equation (16), we know that a positive $\beta$ value implies that the choice probability of our hotel decreases when our hotel price increases, so the price elasticity is negative. It turns out that our estimation result is on the right way regarding demand. To take care of the endogeneity of demand function, we were planning to find an instrumental variable such as supply shock. Since the estimation results shows that the endogeneity problem is not severe, we decided to stick with our current demand model.

The remaining 39 parameters are in Table 6 and Table 7. Rows 1-4 and 25-28 of Table 6 show the 3rd degree polynomial coefficients for $\mu_t$ of negative binomial model which addresses leisure customers arrival probability of luxury hotel markets in this area. The coefficients of the leading term are all zero and the coefficients of the second term are also close to zero and concentrated around 0.002. It is hard to say that $\mu_t$ of negative binomial model is close to linear model. Instead, we say the coefficient of $t$ are most dominant during parameter terms.

Rows 5-8 and 29-32 of Table 6 show the 3rd degree polynomial coefficients for $r_t$ of negative binomial model which addresses leisure customers arrival probability of luxury hotel markets in this area. As same as $\mu_t$, it seems to be linear model. However, there is one thing different that the coefficients of the third term and the fourth term are not consistent each other. Compared to those of $\mu_t$ in leisure, these terms fluctuate more in $r_t$ parameter.

Rows 9-12 and 33-36 of Table 6 show the 3rd degree polynomial coefficients for $\mu_t$ of negative binomial model which addresses business customers arrival probability of luxury hotel markets in this area. The
Table 5: Choice probability parameters

<table>
<thead>
<tr>
<th>Segment</th>
<th>Parameter</th>
<th>Lowest Demand (0-25%)</th>
<th>Medium-Low Demand (25-50%)</th>
<th>Medium-high Demand (50-75%)</th>
<th>Highest Demand (75-100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leisure</td>
<td>alpha (α)</td>
<td>1.698 (0.384)</td>
<td>1.546 (0.338)</td>
<td>1.329 (0.174)</td>
<td>2.300 (2.798)</td>
</tr>
<tr>
<td></td>
<td>beta (β)</td>
<td>0.008 (0.001)</td>
<td>0.007 (0.001)</td>
<td>0.010 (0.001)</td>
<td>0.074 (0.036)</td>
</tr>
<tr>
<td>Weekday</td>
<td>Leisure</td>
<td>alpha (α)</td>
<td>1.618 (1.115)</td>
<td>1.904 (0.150)</td>
<td>1.047 (0.134)</td>
</tr>
<tr>
<td></td>
<td>business</td>
<td>beta (β)</td>
<td>0.006 (0.002)</td>
<td>0.006 (0.000)</td>
<td>0.006 (0.001)</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>alpha (α)</td>
<td>0.539 (1.115)</td>
<td>0.935 (0.152)</td>
<td>1.167 (0.360)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>beta (β)</td>
<td>0.012 (0.005)</td>
<td>0.011 (0.002)</td>
<td>0.012 (0.002)</td>
</tr>
<tr>
<td>Weekend</td>
<td>Leisure</td>
<td>alpha (α)</td>
<td>1.580 (0.091)</td>
<td>1.803 (0.328)</td>
<td>0.296 (0.515)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>beta (β)</td>
<td>0.008 (0.001)</td>
<td>0.009 (0.002)</td>
<td>0.035 (0.046)</td>
</tr>
<tr>
<td></td>
<td>Business</td>
<td>alpha (α)</td>
<td>1.358 (0.149)</td>
<td>1.262 (0.314)</td>
<td>2.203 (2.480)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>beta (β)</td>
<td>0.007 (0.001)</td>
<td>0.007 (0.003)</td>
<td>0.007 (0.010)</td>
</tr>
<tr>
<td></td>
<td>Group</td>
<td>alpha (α)</td>
<td>0.813 (0.076)</td>
<td>0.913 (0.217)</td>
<td>0.002 (0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>beta (β)</td>
<td>0.012 (0.003)</td>
<td>0.017 (0.002)</td>
<td>0.015 (0.010)</td>
</tr>
</tbody>
</table>

Note: t indicates days before occupancy and numbers in parentheses are variances.
<table>
<thead>
<tr>
<th>Segment</th>
<th>Parameter (Coefficient)</th>
<th>Lowest Demand (0-25%)</th>
<th>Medium-Low Demand (25-50%)</th>
<th>Medium-high Demand (50-75%)</th>
<th>Highest Demand (75-100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.002 (0.002)</td>
<td>0.001 (0.001)</td>
<td>0.002 (0.001)</td>
<td>0.002 (0.001)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.111 (0.050)</td>
<td>-0.110 (0.033)</td>
<td>-0.093 (0.031)</td>
<td>-0.088 (0.033)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.677 (2.931)</td>
<td>3.639 (1.269)</td>
<td>3.397 (0.563)</td>
<td>3.372 (0.579)</td>
</tr>
<tr>
<td>Weekday</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.003 (0.001)</td>
<td>-0.002 (0.001)</td>
<td>-0.002 (0.000)</td>
<td>-0.003 (0.003)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.048 (0.015)</td>
<td>0.029 (0.022)</td>
<td>0.034 (0.004)</td>
<td>0.036 (0.063)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.933 (1.310)</td>
<td>2.749 (0.554)</td>
<td>2.246 (0.639)</td>
<td>3.003 (2.184)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-0.003 (0.001)</td>
<td>-0.001 (0.000)</td>
<td>-0.002 (0.000)</td>
<td>0.001 (0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.012 (0.012)</td>
<td>-0.008 (0.010)</td>
<td>0.009 (0.001)</td>
<td>-0.022 (0.011)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1.096 (1.630)</td>
<td>-0.677 (0.541)</td>
<td>-0.378 (0.049)</td>
<td>-0.648 (0.338)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.4E-18 (0.000)</td>
<td>9.8E-19 (0.000)</td>
<td>3.58E-18 (0.000)</td>
<td>1.3E-18 (0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-1.08E-16 (0.000)</td>
<td>-2.11E-17 (0.000)</td>
<td>-1.49E-17 (0.000)</td>
<td>-4.32E-17 (0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.32E-15 (0.000)</td>
<td>8.97E-16 (0.000)</td>
<td>8.97E-16 (0.000)</td>
<td>7.42E-16 (0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18.796 (55.835)</td>
<td>26.195 (79.966)</td>
<td>32.996 (76.012)</td>
<td>19.045 (22.840)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.11E-18 (0.000)</td>
<td>1.19E-18 (0.000)</td>
<td>1.55E-18 (0.000)</td>
<td>7.56E-19 (0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-8.03E-17 (0.000)</td>
<td>-4.67E-17 (0.000)</td>
<td>-6.71E-17 (0.000)</td>
<td>-4.75E-17 (0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5E-15 (0.000)</td>
<td>1.07E-15 (0.000)</td>
<td>7.72E-16 (0.000)</td>
<td>7.13E-16 (0.000)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
</tbody>
</table>

Note: \( t \) indicates days before occupancy and numbers in parentheses are variances.
estimation results are partly similar to $\mu_t$ in leisure data. These parameters constitute a weak linear approximation and remaining terms such as the third and fourth terms are consistent each other in weekday demand. Yet, the parameters of $\mu_t$ vary in weekend demand.

Rows 13-16 and 37-40 of Table 6 show the 3rd degree polynomial coefficients for $r_t$ of negative binomial model which addresses business customers arrival probability of luxury hotel markets in this area. Still, the coefficients of the leading term are close zero and other terms are influential in movement of $r_t$.

Rows 17-24 and 41-48 of Table 6 show the 3rd degree polynomial coefficients for zero-inflated negative binomial model of leisure and business respectively. All coefficients, except for the constant coefficient, are close to zero. It implies that they are not time-varying parameter term. It seems that zero-inflated negative binomial model is still valid. However, the terms related to time $t$ are not necessary for estimation.

Rows 1-4 and 17-20 of Table 7 show the 3rd degree polynomial coefficients for probability that there is no group reservations t-days before arrival day. Although most parameters have the same sign and similar scale regardless of the sample, they undulate smoothly.

Rows 5-8 and 21-24 of Table 7 show the 3rd degree polynomial coefficients for average group reservation numbers given a case when group reservations exist. It is highly volatile, so it turns out how hard to expect the number of group reservations.

Rows 9-12 and 25-28 of Table 7 show the 3rd degree polynomial coefficients for probability that there exists any cancellation made by booking customers. Cancel probability are very stable, so it enable the model predict statistically accurate the number of cancel.

Rows 13-16 and 29-32 of Table 7 show the 3rd degree polynomial coefficients for average cancellation numbers given a case when there exists cancellation cases. Same with average cancel rate, the parameters are quite stable between samples. Rather than group reservations, Table 7 shows that cancel part is better to estimate.

All parameters above are derived from the SMD estimation. This method pursues minimization of moments between the actual data and the simulated data. The final distance between those moments are as following Table 8.

It can be also shown graphically as Figure 16. The round-solid line indicates the moments created by the actual data, while the star-solid line indicates the moments of simulation. We used over 450 moments for estimation and almost all moments look very close to each other for sample20. The other 7 samples also have identical results.
<table>
<thead>
<tr>
<th>Segment</th>
<th>Parameter (Coefficient)</th>
<th>Lowest Demand (0-25%)</th>
<th>Medium-Low Demand (25-50%)</th>
<th>Medium-high Demand (50-75%)</th>
<th>Highest Demand (75-100%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Group Probability</td>
<td>$t^3$</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td>$t^2$</td>
<td>0.003 (0.002)</td>
<td>0.006 (0.002)</td>
<td>0.008 (0.003)</td>
<td>0.012 (0.011)</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>-0.019 (0.010)</td>
<td>-0.054 (0.018)</td>
<td>-0.027 (0.037)</td>
<td>-0.048 (0.054)</td>
</tr>
<tr>
<td></td>
<td>$1$</td>
<td>-0.971 (0.960)</td>
<td>0.402 (0.401)</td>
<td>0.636 (0.333)</td>
<td>-0.714 (0.632)</td>
</tr>
<tr>
<td>Weekday Group Mean (if group &gt; 0)</td>
<td>$t^3$</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td>$t^2$</td>
<td>0.003 (0.000)</td>
<td>-0.004 (0.001)</td>
<td>-0.001 (0.001)</td>
<td>-0.006 (0.001)</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>0.058 (0.037)</td>
<td>0.151 (0.042)</td>
<td>0.076 (0.004)</td>
<td>0.201 (0.067)</td>
</tr>
<tr>
<td></td>
<td>$1$</td>
<td>0.457 (2.140)</td>
<td>0.304 (1.511)</td>
<td>0.792 (0.166)</td>
<td>0.423 (0.513)</td>
</tr>
<tr>
<td>Cancel Probability</td>
<td>$t^3$</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td>$t^2$</td>
<td>-0.003 (0.001)</td>
<td>-0.004 (0.001)</td>
<td>-0.004 (0.001)</td>
<td>-0.007 (0.002)</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>0.147 (0.031)</td>
<td>0.157 (0.019)</td>
<td>0.175 (0.021)</td>
<td>0.240 (0.055)</td>
</tr>
<tr>
<td></td>
<td>$1$</td>
<td>-1.028 (0.073)</td>
<td>-1.472 (0.153)</td>
<td>-1.989 (0.201)</td>
<td>-2.522 (0.418)</td>
</tr>
<tr>
<td>Cancel rate</td>
<td>$t^3$</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td>$t^2$</td>
<td>-0.003 (0.001)</td>
<td>-0.001 (0.000)</td>
<td>0.000 (0.000)</td>
<td>-0.002 (0.001)</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>0.060 (0.025)</td>
<td>0.061 (0.012)</td>
<td>0.025 (0.011)</td>
<td>0.047 (0.024)</td>
</tr>
<tr>
<td></td>
<td>$1$</td>
<td>3.884 (3.923)</td>
<td>3.206 (1.025)</td>
<td>3.432 (1.042)</td>
<td>4.043 (0.729)</td>
</tr>
<tr>
<td>No Group Probability</td>
<td>$t^3$</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td>$t^2$</td>
<td>0.005 (0.003)</td>
<td>0.007 (0.001)</td>
<td>0.007 (0.004)</td>
<td>0.008 (0.002)</td>
</tr>
<tr>
<td></td>
<td>$t$</td>
<td>-0.069 (0.032)</td>
<td>-0.057 (0.063)</td>
<td>-0.080 (0.070)</td>
<td>-0.106 (0.172)</td>
</tr>
<tr>
<td></td>
<td>$1$</td>
<td>0.158 (0.296)</td>
<td>-0.661 (1.238)</td>
<td>-0.281 (0.261)</td>
<td>-0.991 (3.938)</td>
</tr>
<tr>
<td>Weekday Group Mean (if group &gt; 0)</td>
<td>$t^3$</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
<td>0.000 (0.000)</td>
</tr>
<tr>
<td></td>
<td>$t^2$</td>
<td>0.003 (0.002)</td>
<td>0.001 (0.000)</td>
<td>-0.001 (0.001)</td>
<td>-0.006 (0.010)</td>
</tr>
</tbody>
</table>

Note: $t$ indicates days before occupancy and numbers in parentheses are variances.
Table 8: SMD Finalized Distance by Sample

<table>
<thead>
<tr>
<th>Sample (Weekday)</th>
<th>Final Distance (SMD)</th>
<th>Sample (Weekend)</th>
<th>Final Distance (SMD)</th>
<th>Demand Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>00</td>
<td>10,658.7</td>
<td>01</td>
<td>9,248.0</td>
<td>Lowest</td>
</tr>
<tr>
<td>10</td>
<td>12,482.4</td>
<td>11</td>
<td>11,605.6</td>
<td>Medium-Low</td>
</tr>
<tr>
<td>20</td>
<td>20,670.8</td>
<td>21</td>
<td>14,111.5</td>
<td>Medium-high</td>
</tr>
<tr>
<td>30</td>
<td>45,978.7</td>
<td>31</td>
<td>20,950.7</td>
<td>Highest</td>
</tr>
</tbody>
</table>

Figure 16: Estimated Moments
**Model Fit**: Lastly, we can compare the estimated results with the hotel daily record. Figure 17 shows the daily average rate of our hotel and competing hotels between the period of Feb 16, 2012 and May 26, 2012. In addition to this, Figure 18 shows the confirmed occupancy level of our hotel and competing hotels in the same period. For each arrival day, we can collect the reservation detail data of our hotel from reservations raw and cancel while we use STR data set to collect the equivalent information to plot the bottom of Figure 17 and Figure 18. for the competing hotels information,. The plot derived from those data is marked with a star-solid line under the name of data. Whereas, the dashed-line in both figures indicate the simulated ADR and occupancy level.

Average daily rate (ADR) is a distinguishable rate from the best available rate (BAR). BAR is the official price before any discount has been taken. Even if two reservations for the same arrival day are made simultaneously, their discount rates are different according to their market segments. In simulation, we need to know discount rate for each simulated reservation made. There is practical difficulty in finding the exact discount rate for each reservation. Instead of applying various discount rates to each booking, we use average discount rates for each sample and for each category such as group, leisure, and business. Once we get the average discount rate from the actual data, ideally, ADR should be consistent with the actual data and simulated data if our estimation is correct. However, as seen in Figure 17, the simulated ADR is often much higher than the data ADR, especially at peak-points.

There might be several reasons for these gaps. Since average discount rates for each day vary even within the same sample, the computed discount rate is not able to reflect the exact discount rate for each day. When there are ample bookings by a company or group, it cannot be expected that their discount rate is the same as the sample average rate. In this case, we need to split the whole data set into smaller samples. However, due to the shortage of observations for each sample, we did not employ this attempt with a narrower sample. It is also possible that this problem is related to the computation of optimal price, BAR. When the optimal price is computed higher than the actual BAR price, ADR with a fixed discount rate must also be a higher value in simulation results. The above reasons are not enough to explain the occurrence of Figure 17. Its possible that it is caused by a computation problem with our code or it may not be a soluble problem with our model. This is the limiting factor of our hotel model. Even so, we successfully accomplished the estimation in terms of average level of sample which, in itself is valuable work.
Figure 17: Time Series of Average Daily Rate (ADR)
Figure 18: Time Series of Occupancy
The gap between simulation and actual also occurs occasionally in time series of occupancy in Figure 18. However, the two data sets appear to match each other closely with the exception of a few dates such as April 1, 2012. The model-fit is more successful in the competing hotels plot which is the bottom panel of Figure 18. Unlike the ADR plot, the gaps in occupancy plot appear unevenly. This might be caused by unexpected reservation situations such as a dump sale by online travel agencies or group reservations. Indeed, many reservations for April 1, 2012 were made with the market code RESO or OPQ which represents a discounted price carried by a travel agency. Figure 18 shows that the model fits mostly well for the occupancy level in our hotel and competing hotels.

9 Counter-factual experiment

**single day example**  With the outcome of the model estimation, we can attempt several experiments by setting different price points for our hotel. Then, we can compare the results from these experiments from the point of occupancy, revenue, and so on. Our goal of this experiment is not to find a better way for hotel manager or customers. Instead, we are looking to see if these experiments are sensible and whether it is consistent with what hotel managers experience therefore, evaluating the practical value of our hotel model and estimation.

We have conducted three different experiments. The first experiment is to keep a fixed BAR instead of dynamic pricing where we use the average actual BAR of each day as a constant price. The next experiment is to put a 20% discount on the optimal price derived from our hotel dynamic program. Inversely, we set a 20% higher price in the third experiment. For the sake of convenience, we start this section with experiments for a single day. The experiment result contains a simulated occupancy, revenue, and arrival flow. Once we show the single day event, this section looks through the overall results, including the distribution of all the experiments.

Figure 19 plots the first experiment example with a constant price of actual data of April 19, 2012. This figure compares the experiment result with the actual data record. At the top panel of Figure 19, the constant price marked with a dashed line keeps it at $410 during the whole period while the actual BAR, the solid line, fluctuates between $280 and $550. The effect of price shifting is seen in the bottom panel of Figure 19. Our hotels occupancy rate is similar in both the actual data and simulated data of constant price until 38 days before arrival. But since then, the gap of occupancy rate increases up to the day of arrival. The actual occupancy rate is close to being overbooked, whereas occupancy rate of the constant price stays
Figure 19: Experiment example of Constant BAR
around 73%. The constant price keeps occupancy rate increasing smoothly. Since its final occupancy rate is much lower than the actual data, those gaps have an impact on revenue as shown in the bar plot of the top panel. The left bar plot indicates the total revenue of the actual data which comes from the reservation data and the right bar plot represents the total revenue of experiment under constant price. The hotel earned around $80,000 on April 19, 2012 which is higher than the $70,000 revenue from the constant price experiment. Compared to the constant price experiment, this example shows that the current hotel pricing system is efficient in terms of revenue maximization and occupancy maximization. In the next figure, we see how our model has improved the hotel pricing system regarding revenue maximization.

As mentioned at the beginning of this section, we examined cases where we increased and decreased the price by 20%. Figure 20 plots 4 panels related to those two experiments which are implemented for the same day as Figure 19. The top left panel introduces how the experiment process works. Keeping the optimal price until 15-days before the day of arrival, BAR is increased by 20% for the second experiment (dashed line) and decreased by 20% for the third experiment (dotted line). Once their BARs are changed, we assume that the hotel keeps these price points until the end of period, the arrival day. Therefore, the new BAR price on experiment fluctuates correspondingly with the optimal BAR (solid line).

With a 20% increase (dashed line), the maximum BAR is close to $520 while the maximum of optimal BAR is $435, as shown in the top left panel of Figure 20. The impact of high price appears in the occupancy level immediately. At the start of the experiment, the occupancy rate was 0.96, since then the occupancy rate has decreased because new reservations are no longer made while there is an outflow due to cancellations. This shows how huge the higher price influences the inflow of the hotel reservations. After the price hike, new reservations for the transient, which are comprised of business and leisure customers, drops sharply in the bottom left panel of Figure 20. The low reservations are recovered at the end of the period, yet that is not enough to cover up the loss by the higher price. The same pattern also shows up in group reservations at the bottom right panel of Figure 20. However, group reservations inflow has not recovered even at the end of period because there is only a small chance of new arrivals in the case of a group as it approaches the arrival day.

With a 20% lower price scenario (dotted line), the BAR drops to $300 which results in a shift up in occupancy level. Since the occupancy rate is already 0.96 just before the experiment price hike, the occupancy hits the hotel capacity in a few ways. The dramatic discount on the price affects the arrival pattern as well. In the bottom left of Figure 20, you see the dotted line (discounted price) has a spike right after the discount. Unusually, new arrivals go down quickly as well. As mentioned before, the hotel is already filled with reservations, so they are not able to accept any new reservations even if there is a high demand.
Figure 20: Experiment examples of BAR addition/deduction
under such a discounted price and there are bound to be cancellations. We also find a hike in arrival in group reservations as seen in the bottom right panel. Due to the low demand of group reservations on our example day, the plot of the third experiment does not have any notable distinction from the benchmark plot.

In summary, the change in the optimal price affects the new reservations and occupancy rate and the results of the experiment are consistent with our expectations. When we apply these experiments to another single day, it is clear that our experiment results are sensible, so our estimation of the hotel model has huge practical value. The hotel is most interested in maximizing revenue rather than maximizing occupancy. We compare the revenue of cases we mentioned such as the actual price data, the optimal data, and three scenarios of experiments. In the bar plots of the top left panel of Figure 19, the revenue of each case is around $88,000 at the optimal price, around $86,000 at the discounted price, and only $82,000 at the higher price point. All of them are still higher than revenue at the constant price ($70,000) or the revenue of the actual data ($80,000). In the view of revenue maximization, our model estimation is very successful for the suggested day, April 19, 2012.

**entire distribution**  Now, we turn to the collective experiments distribution to see if the previous observations on the single days can be accepted as a whole. As seen in Figure 21, experiment3 which enforces a 20% reduction in optimal price has the highest occupancy rate among the four cases. And its density is highly concentrated on the level of fully occupied. With a low price point, it is easy for the hotel to fill the hotel at capacity level. And the second-best case for occupancy maximization is our optimal pricing model. It is less than occupancy level of discounted price scenario, but still keeps around 0.8 rate as its average occupancy rate.

On the revenue side, Figure 22 plots the revenue distribution of all cases. As expected, the optimal price computed by the hotel model has the highest mean of revenue distribution. On average, the hotel can expect a revenue of about $60,000 per day. The second highest revenue occurs in the discounted price scenario. The expected revenue of this case is just below $58,000 and its revenue distribution is similar to the optimal price case. The increase by 20% scenario has the third highest revenue average which is just above the average revenue of the constant price case.
Figure 21: Distribution of Occupancy for Experiments
Figure 22: Distribution of Revenue for Experiments
10 Conclusion

This paper estimated a dynamic price of oligopoly hotel markets in Washington D.C. area.

References


A Introduction

This note discusses how it might be that ADR of past room reservations may affect the hotel manager’s decision to set BAR \( t \) days prior to occupancy, \( p_t \). We conclude that unless the number of cancellations are a function of ADR, then the optimal price decision rule for BAR should not depend on ADR. The note also lays out an alternative formulation of the Bellman equation for the hotel DP problem using number of rooms reserved and total revenues booked so far, \((n_t, R_t)\), as the state variables in addition to mean BAR of competing hotels, \( \rho_t \), as the state variables instead of using bookings so far and the ADR, \((n_t, \overline{p}_t)\), but concludes that this version is mathematically equivalent to the version we are using, except that our current formulation has the continuous state variable ADR \( \overline{p}_t \) that spans a more compact range of values than revenues booked so far, \( R_t \), and thus has numerical advantages even though the two formulations are mathematically equivalent (where issues of discretization numerical approximation are ignored).

This note was motivated by the finding that the optimal BAR rule seems to depend on the ADR in a monotonic fashion (i.e. higher ADR leads to higher BAR) and ADR has especially strong influence on BAR when hotel capacity is close to 100%. In this note we argue that unless cancellation rates are a function of ADR, then BAR should be independent of ADR.

B A Simple Model of Hotel Room Pricing

Recall that in the original version of the model, the value function had 3 state variables besides \( t \), the number of days before occupancy:

1. \( \overline{p}_t \): the average price of rooms booked so far (i.e. the ADR as of \( t \) before occupancy)
2. \( n_t \): the number of rooms booked so far
3. \( \rho_t \): the BAR of competing hotels for rooms on the occupancy date, i.e. what they are quoting \( t \) days before occupancy

In addition, we will let \( K \) denote the total capacity of the hotel and assume this is a known, time-invariant parameter. Now for the change in notation, previously I used \( \nabla n_t^b \) for the number of new reservations \( t \) days before occupancy, and \( \nabla n_t^c \) to be the number of cancellations. I have decided that it is more compact and easier to read to just use \( n_t^r \) for the number of new reservations and \( n_t^c \) to be the number of cancellations, so we have the law of motion

\[
n_{t-1} = n_t + n_t^r - n_t^c \tag{18}\]

72
I also want to introduce the marginal cost of cleaning/providing for an occupied room and will use the symbol $c$ for this and I will assume this is a time invariant parameter to be estimated. With this new notation, the value function on the occupancy day is

$$V_0(n_0, \bar{p}_0, \rho_0) = n_0(\bar{p}_0(1 - \kappa) - c)$$

(19)

where as before $\kappa$ is the average percentage commission rate that the hotel pays to third party websites and travel agents who book rooms at the hotel. The parameter $\kappa$ may also capture different discounts that the hotel offers to various customers, though we do treat group reservations separately with their own group discounts. We will continue this treatment here. Thus only difference between this new formulation and what we had previously is that we now account for the marginal cost of cleaning/servicing the room (including water, heating, electricity, providing soap, cleaning linens, etc).

Note that we deduct average commissions/discounts from revenues only at the “end” by multiplying $\bar{p}$ by $(1 - \kappa)$, so properly speaking, we should regard $\bar{p}_t(1 - \kappa)$ as the estimate of the ADR for rooms booked $t$ days before occupancy, and not $\bar{p}_t$. We can also add in separate treatment of group reservations with separate arrival process for groups, which we denote by $n^g_t$. But to keep things simpler, in this note we ignore the group reservation dynamics. Also in the cancellation dynamics, we assume that group cancellations are lumped together with cancellations by non-group transient customers, so $n^c_t$ represents the total number of cancellations of all reservations made before date $t$, which is $t$ days before occupancy.

Now consider the value function at $t = 1$, one day before occupancy. Then the firm, by quoting a BAR $\bar{p}$ will be able to get some new reservations on the last day before occupancy, and there may be cancellations. Recall we provided an updating formula for the ADR to reflect, correctly, the ADR on the next day given the BAR quoted on date $t$, $\bar{p}_t$:

$$\bar{p}_{t-1} = \lambda(p_t, \bar{p}_t, n_t, n^c_t, n^f_t) = \frac{(n_t - n^c_t)\bar{p}_t + n^f_t p_t}{n_t + n^c_t - n^f_t}.$$  

(20)

Note that in this equation, for simplicity, we have left out the group reservations but the equation can easily be modified to include group reservations as well. Notice that under this definition we have

$$\bar{p}_0 n_0 = (n_1 - n^c_1)\bar{p}_1 + n^f_1 p_1$$

(21)

so total (gross) revenue on the occupancy day is equal to the revenue earned for all $n_1$ consumers booked so far, $n_1\bar{p}_1$, less revenue lost due to last-day cancellations, $n^c_1\bar{p}_1$, plus new revenues from last day reservations at the BAR quoted on day $t = 1$, $n^f_1 p_1$. So the trick of carrying the ADR $\bar{p}_t$ as a state variable in the problem is a justifiable way to formulate the problem, since it does result in a correct law of motion for the evolution of total revenues at the hotel.
To understand this completely, we will carry out several backward induction steps and characterize the form of the value function $V_t(n_t, \overline{p}_t, \rho_t)$ and the optimal BAR decision rule, which is potentially a function of $t$ and all three state variables $(n_t, \overline{p}_t, \rho_t)$, which we denote by

$$p_t = \pi_t(n_t, \overline{p}_t, \rho_t)$$ (22)

We now show by induction that unless the random variables $n^c_t$ or $n^r_t$ depend on $\overline{p}_t$, then $p_t$ will not depend on $\overline{p}_t$. Instead it will only depend on $(n_t, \rho_t)$ as in

$$p_t = \pi_t(n_t, \rho_t)$$ (23)

reflecting the economic intuition that $\overline{p}_t$ captures “bygones” but hotel pricing should be an entirely forward looking calculation and as such should not depend this bygone information. On the other hand, BAR will depend on $n_t$ which is essential to determine remaining unsold capacity $K - n_t$.

Using the Bellman equation at $t = 1$ we have

$$V_1(n_1, \overline{p}_1, \rho_1) = \max_p E \{ V_0(n_1 - \tilde{n}^c_1 + \tilde{n}^r_1, \lambda(p, \overline{p}_1, n_1, \tilde{n}^c_1, \tilde{n}^r_1, \rho_0)|n_1, p, \overline{p}_1, \rho_1 \}$$ (24)

In Bellman equation (24) we use the random variable notation $\tilde{n}^c_t$ and $\tilde{n}^r_t$ to emphasize that cancellations and new reservations are random variables. We assume cancellations only depend on $(n_t, \overline{p}_t)$, the number of rooms booked so far and possibly on the average (pre-commission, pre-discount) price at which they were booked. So far, we have only considered models where $n^c_t$ depends on $n_t$ but not on $\overline{p}_t$ and our conclusion that the optimal decision rule for BAR should be independent of $\overline{p}_t$ depends on the assumption that cancellations are independent of $\overline{p}_t$. However if there are “strategic cancellations” $\tilde{n}^c_t$ could depend both on $\overline{p}_t$ and on BAR $p_t$. The logic is that if there is no cost to cancelling a booking prior to occupancy and if guests are monitoring hotel rates, they are more likely to cancel (and potentially rebook) if they see that $p_t$ is sufficiently lower than $\overline{p}_t$, which proxies for the rate that they reserved at. Conversely cancellation may fall if $p_t$ is sufficiently higher than $\overline{p}_t$. So it may be plausible to conjecture that cancellation rates may be a function of the price difference $\overline{p}_t - p_t$ and possibly on the BAR of competing hotels $\rho_t$ as well. But for our main result below, we will assume that the probability distribution for $\tilde{n}^c_t$ depends only on $n_t$ and we will write $E\{n^c_t|n_t\}$ to denote expected cancellations. For example if $\alpha_t$ represents a cancellation rate $t$ days before occupancy, we have $E\{n^c_t|n_t\} = \alpha_t n_t$.

The number of new reservations $\tilde{n}^r_t$ made $t$ days before occupancy is a draw from a probability distribution that depends on $(p_t, \rho_t)$, since presumably this is just a result of a discrete choice by customers arriving and making a hotel choice decision. However we also have to recognize that the number of rooms
reserved, \( \bar{n}^c_t \), must satisfy the hotel capacity constraint and thus the hotel has to turn away new reservation requests in situations where it is full. Let \( n^d_t \) denote the number of rooms demanded \( t \) days before occupancy. Assuming the hotel does not overbook, we have

\[
n^c_t = \min[K - n_t - \bar{n}^c_t, n^d_t].
\] (25)

In equation (25) we have assumed that the hotel can observe cancellations before it makes new reservations. If this is not the case, we can modify the upper bound on room demand \( n^d_t \) to equal \( K - n_t \), number of unsold rooms but not accounting for cancellations on day \( t \). Let \( E\{n^c_t|n_t, p_t, \bar{p}_t\} \) denote the conditional expectation of new reservations. It depends on \( \bar{p}_t \) only if a) the number of new reservations at \( t \) responds to cancellations \( \bar{n}^c_t \) as in equation (25) above, and b) if the probability distribution for \( \bar{n}^c_t \) depends on \( p_t \).

Since we assume initially that cancellations is independent of \( \bar{p}_t \) we can write \( E\{\bar{n}^c_t|n_t, p_t, \bar{p}_t\} \), which is independent of \( \bar{p}_t \).

Substituting the conditional expectation formulas for \( \bar{n}^c_t \) and \( n^c_t \) into the Bellman equation (24) using the formula for \( V_0(n_0, \bar{p}_0, \rho_0) \) in equation (19) we have

\[
V_1(n_1, \bar{p}_1, \rho_1) = \max_p E\{\bar{n}^c_t|n_1, p, \rho_1\}[p(1 - \kappa) - c] + (n_1 - E\{\bar{n}^c_t|n_1\})[\bar{p}_1(1 - \kappa) - c].
\] (26)

The Bellman equation (26) has a simple interpretation: \( V_1(n_1, \bar{p}_1, \rho_1) \) equals the sum of the expected profits from the “last day” reservations \( \bar{n}^c_t \) (first term on the right hand side) plus the expected profits from the already booked customers less lost profits from expected cancellations (second term on the right hand side). Note that \( \bar{p}_1 \) only enters the second term of (26) but the BAR decision only enters the first term, so it follows that BAR will only depend on \( (n_1, \rho_1) \) and not on \( \bar{p}_1 \), which we write as

\[
p_1 = \pi_1(n_1, \rho_1).
\] (27)

**Theorem 1** In the hotel pricing problem, if the probability distribution for the the number of cancellations 1 day prior to occupancy is independent of \( \bar{p}_1 \), then the optimal BAR at \( t = 1 \) is independent of \( \bar{p}_1 \).

Substituting the optimal BAR decision rule we obtain the following expression for the value function \( V_1(n_1, \bar{p}_1, \rho_1) \)

\[
V_1(n_1, \bar{p}_1, \rho_1) = E\{\bar{n}^c_t|n_1, \pi_1(n_1, \rho_1), \bar{p}_1\}][\pi_1(n_1, \rho_1)(1 - \kappa) - c] + (n_1 - E\{\bar{n}^c_t|n_1\})[\bar{p}_1(1 - \kappa) - c].
\] (28)

Continuing the backward induction to days \( t = 2, 3, \ldots \) prior to occupancy gets more complicated since we now see that \( n_1 \) enters both terms in the expression for \( V_1(n_1, \bar{p}_1, \rho_1) \) and thus the hotel manager needs to consider how setting BAR at \( t = 2 \) days prior to occupancy affects not only \( \bar{n}^c_2 \), the number of new
reservations she will make at \( t = 2 \) but she must also consider the effect of those reservations on filling up any remaining capacity which impinges on the net profits she can earn at \( t = 1 \) as well. However under an additional simplifying assumption on cancellations, we can show that at \( t = 2 \), the BAR will also be independent of \( \bar{p}_2 \).

**Assumption C** expected cancellations are a time-varying fraction of the number of rooms already reserved:

\[
E\{\tilde{n}_t^c|n_t\} = \alpha_t n_t, \quad \alpha_t \in (0,1).
\]  

Under this assumption it is easy to see that \( V_1(n_1, \bar{p}_1, \rho_1) \) depends on \( \bar{p}_1 \) only in the second term in (28) and then only via the product \( n_1 \bar{p}_1 \). This property will imply that at \( t = 2 \) the optimal BAR will also be independent of \( \bar{p}_2 \), i.e. \( p_2 = \pi_2(n_2, \rho_2) \), and continuing inductively we can show that in every period \( t \) the optimal BAR decision rule is independent of \( \bar{p}_t \). That is we will be able to prove by induction that

**Theorem 2** In the hotel pricing problem, if the probability distribution for the number of cancellations \( \tilde{n}_t^c \) is independent of \( \bar{p}_t \) and Assumption C holds, then the optimal BAR is independent of \( p_t \) for all \( t \), i.e.

\[
p_t = \pi_t(n_t, \rho_t), \quad t = 1, \ldots, T
\]

**Proof** We have already showed that the independence condition (27) holds at \( t = 1 \) (and for this last period, Assumption C was not required, only the weaker condition that expected cancellations do not depend on \( \bar{p}_1 \)). So the result holds in the next to last period, \( t = 1 \). Then to complete the proof we only need to show that in every period \( t \), \( V_t(n_t, \bar{p}_t, \rho_t) \) depends on \( \bar{p}_t \) only via the product \( n_t \bar{p}_t \). It is easy to see from equation (28) that this property holds at \( t = 1 \) when Assumption C holds. To complete the proof it is easy to show by induction that for each \( t = 2, 3, \ldots, T \) a) if \( V_t(n_t, \bar{p}_t, \rho_t) \) depends on \( \bar{p}_t \) via the product \( n_t \bar{p}_t \) then \( p_t = \pi_t(n_t, \rho_t) \), i.e. the optimal BAR decision rule does not depend on \( \bar{p}_t \), and b) then \( V_{t+1}(n_t, \bar{p}_{t+1}, \rho_{t+1}) \) depends on \( \bar{p}_{t+1} \) via the product \( n_{t+1} \bar{p}_{t+1} \).

For example we have, via the definition of the law of motion for number of rooms reserved \( n_t \) in equation (18) and average price in equation (20), we have

\[
\bar{p}_{t-1} n_{t-1} = (n_t - n_t^c) \bar{p}_t + n_t^c \rho_t
\]

Taking conditional expectations given \((n_t, \bar{p}_t, \rho_t)\) (where \( p \) is a possible BAR that the hotel manager might choose at time \( t \)) and invoking Assumption C we have

\[
E\{\bar{p}_{t-1} n_{t-1}|n_t, \bar{p}_t, \rho_t, \rho_t\} = n_t \bar{p}_t (1 - \alpha_t) + E\{\tilde{n}_t^c|n_t, \rho_t, \rho_t\} \rho_t
\]

so it follows that if \( V_{t-1}(n_{t-1}, \bar{p}_{t-1}, \rho_{t-1}) \) depends on \( \bar{p}_{t-1} \) only via the product \( n_{t-1} \bar{p}_{t-1} \) then \( V_t(n_t, \bar{p}_t, \rho_t) \) will depend on \( \bar{p}_t \) only via the product \( n_t \bar{p}_t \). Finally, it is easy to see that when this is true, the conditional
Thus, $V_{n-1}(n_{t-1}, \bar{p}_{t-1}, \rho_{t-1})$ is additively separable into terms that depend on $n_t \bar{p}_t$ but not on $p$ and other terms that depend on $(p, n_t, \rho_t)$ but not on $\bar{p}_t$. This implies that the optimal decision rule for BAR, $p_t = \pi_t(n_t, \rho_t)$, does not depend on $\bar{p}_t$.

**Comments** The insight from this “semi-analytic” characterization of the value functions to the hotel pricing problem is that it is fundamentally a two-dimensional dynamic programming rather than a three dimensional problem. That is, the insights above have revealed special structure of the value functions that they are additively separable into a component that depends on profits from customers who have already made reservations (less expected cancellations) plus the “forward looking” component of revenues of new reservations from the residual remaining capacity. The first component of expected profits depends on $(n_t, \bar{p}_t)$ only whereas the the latter component depends only on $(n_t, \rho_t, p)$, where $p$ is a possible BAR price to be set at time $t$. So knowing this special structure of the value function we can significantly speed up the dynamic programming recursion by focusing only on the latter “forward looking” component of the value function and just “adding in” the backward looking component of profits from already reserved customers to the forward looking component.

We will write down a decomposition formula $V_t(n_t, \bar{p}_t, \rho_t)$ more explicitly below but we can already see the nature of the decomposition in the formula for $V_t(n_t, \bar{p}_t, \rho_t)$ in equation (28). Under Assumption C, the backward-looking component of the value function is $n_t(1 - \alpha_t)[\bar{p}_t(1 - \kappa) - c] = n_t \bar{p}_t(1 - \alpha_t)(1 - \kappa) - n_t(1 - \alpha_t)c$, which is the component of $V_t(n_t, \bar{p}_t, \rho_t)$ that depends on $\bar{p}_t$ only via the product $n_t \bar{p}_t$.

The other term is the “forward looking component” of profits from new reservations on the last day, which depends on $(n_1, \rho_1, p)$ where $p$ is a possible BAR price. Optimizing this first component over $p$ results in decision rule $p_t = \pi_t(n_t, \rho_t)$ and the first component of the value function thus depends only on $(n_1, \rho_1)$. Since we have a closed form expression for the second component of the value function, we can subtract it off for purposes of optimizing the first component, and then add it back in later when needed to get the total value. This means that what might have initially appeared to be a three dimensional dynamic programming problem is really only a two dimensional problem.

So let us define the following backward induction recursion for the forward looking component $V_t^f(n_t, \rho_t)$ of the value function, which we note only depends on $(n_t, \rho_t)$ and not on $\bar{p}_t$. The optimal decision rule for BAR will be determined from the forward looking recursion and thus $p_t = \pi_t(n_t, \rho_t)$ will depend only on $(n_t, \rho_t)$ as well. We start the recursion by defining $V_0^f(n_0, \rho_0) = 0$ (i.e. there is no remaining forward looking component on the last, occupancy day), and we define $V_1^f(n_1, \rho_1)$ by

$$V_1^f(n_1, \rho_1) = \max_p E\{\hat{n}_1^f|n_1, p, \rho_1\} [p(1-\kappa) - c] \}. \quad (33)$$

Thus, $V_1^f(n_1, \rho_1)$ is the maximal profits that the hotel can earn from customers who arrive on the last
possible date, \( t = 1 \), prior to occupancy (e.g. by the evening of night of stay). Naturally we assume there are no cancellations by these customers who have arrived just in time to check in on their desired night of stay.

Now continue the backward induction as follows

\[
V_f^t(n_t, \rho_t) = \max_{p_t} \left[ \prod_{s=1}^{t-1} (1 - \alpha_s) E \{ \tilde{n}_t | n_t, p_t, \rho_t \} [p_t (1 - \kappa) - c] + E \{ V_{t-1}^f(n_t - \tilde{n}_t^{t-1} + \tilde{n}_t^{t-1}, \tilde{\rho}_{t-1}) | n_t, p_t, \rho_t \} \right].
\]

(34)

Thus, \( V_f^t(n_t, \rho_t) \) represents the hotel manager’s expectation of profits from new reservations made in the future, on days \( t, t - 1, \ldots, 1 \). Notice how this expectation adjusts for expected cancellations in the future. That is, for a reservation made \( t \) days before occupancy, there is a probability \( \alpha_{t-1} \) it will be cancelled on the next day, \( \alpha_{t-2} \) it will be cancelled \( t - 2 \) days before occupancy and so on, so the probability that the reservation will not be cancelled between the day it is made, \( t \), and occupancy at \( t = 0 \) is

\[
Pr \{ \text{no cancellation between day } t \text{ and day } 0 | \text{reservation made on day } t \} = \prod_{s=1}^{t-1} (1 - \alpha_s)
\]

(35)

Thus, we can see that equation (34) constitutes the “forward looking” component of the value function that determines the sequential determination of the BAR, where the hotel manager considers the tradeoff in the decision to charge a lower price to reserve more customers today and the effect this has on reduction in remaining capacity, which constrains her ability to earn more profits through reservations of new customers in the future.

The backward-looking component of the value function is given by

\[
V_b^t(n_t, \bar{\rho}_t) = \prod_{s=1}^{t} (1 - \alpha_s) n_t [p_t (1 - \kappa) - c].
\]

(36)

Thus, \( V_b^t(n_t, \bar{\rho}_t) \) represents the profits that are “locked in” from the \( n_t \) customers who are already booked, except that it adjusts for expected future cancellations of some of these customers using a similar cancellation factor as the one given in equation (35). Thus, any of the \( n_t \) customers who are reserved at the start of day \( t \) may cancel later on day \( t \) with probability \( \alpha_t \), or on day \( t - 1 \) with probability \( \alpha_{t-1} \), or on day \( t - 2 \) with probability \( \alpha_{t-2} \), and so forth. Thus the “backward looking” component of the Bellman equation (36) does look forward to adjust profits of already booked customers for expected future cancellations.

**Theorem 3** If the probability distribution of cancellations does not depend on \( \bar{\rho}_t \) and Assumption C holds, then the value function has the representation

\[
V_t(n_t, \bar{\rho}_t, \rho_t) = V_f^t(n_t, \rho_t) + V_b^t(n_t, \bar{\rho}_t), \quad t = 0, 1, \ldots, T.
\]

(37)

Note that the backward looking component of the value function \( V_b^t(n_t, \bar{\rho}_t) \) has a closed form expression given in equation (36), and thus it is only necessary to write a dynamic programming recursion to calculate
the forward looking component $V^f_t(n_t, \rho_t)$ in equation (34). This is only a two dimensional problem (since $V^f_t(n_t, \rho_t)$ depends only on the two variables $(n_t, \rho_t)$ and thus taking account of the “special structure” of the value function given in Theorem 3 not only explains why the BAR decision rule cannot depend on $\pi_t$, but it also enables us to develop significantly faster solution algorithms that exploit this special structure.