Estimating Price Discrimination in the Online Distribution Channel*

Babur De los Santos[†] Renáta Kosová[‡] Sergei Koulayev[§]

October 2011

Abstract

This article explores price discrimination strategies in the online sales channel. We combine a rich set of consumer search data for hotel rooms at a major online travel consolidator with data on hotels' daily revenue and performance and hotel census data providing a nearly complete picture of the competitive environment for 13 major U.S. hotel markets. Contrary to theoretical predictions on the unambiguous profitability from price discrimination, we find that across different quality segments a significant proportion of hotels forgo the option of price discrimination by not participating in the online distribution channel. We find that hotels with an online distribution channel differ in important ways, and online prices are consistently higher across quality segments and other hotel attributes. The different pricing strategies between online and offline channels across establishments of different quality suggest that the degree of price discrimination depends on the degree of product differentiation. The results suggest behavior considered by Corts (1998), namely that hotels have asymmetric assessments of consumers' demand elasticities across distribution channels, and adjust their pricing strategies accordingly. We correct for selection into the online channel and find that it noticeably changes the direction of the effect of online pricing across quality segments. These results suggest that online pricing strategies are opposite when comparing high vs. low quality segments.

Keywords: online distribution channel, hotel industry, market segmentation, online bookings.

*We would like to thank In Kyung Kim for excellent research assistance.

[†]Kelley School of Business, Indiana University, e-mail: babur@indiana.edu.

[‡]School of Hotel Administration, Cornell University, e-mail: rk373@cornell.edu.

[§]Keystone Strategy, E-mail: skoulaye@columbia.edu.

1 Introduction

Since Pigou's (1920) seminal work, price discrimination has been a centerpiece of economic analysis. The treatment of monopoly price discrimination in the literature is extensive (e.g. Schmalensee, 1981; Varian, 1985). The degree of price discrimination and its effects on prices, profitability, and welfare depends on firms' information about consumers, what instruments firms can use to leverage this information—such as two-part tariffs—to segment consumers according to their elasticities, and the firms' ability to commit to its pricing policies. Gertsner and Holthausen (1986) show that the profitability of price discrimination is directly related to the degree of consumer leakage between markets. The majority of this research shows that price discrimination leads to unambiguous increases in prices and profits.

The research on price discrimination under competition is less extensive than that under monopoly.[1] Holmes (1989) analyzes the case of a duopoly with product differentiation in which the effects on output depend not only on demand elasticities but also on competitive interactions through cross-price elasticities. Corts (1998) furthemore shows that the effects of price discrimination depend on the degree of product differentiation and firm's asymmetric assessment of consumers' elasticities across markets. A market might be strong for one firm but weak for the other. Armstrong and Vickers (2001) and Rochet and Stole (2002) provide similar useful frameworks in which firms compete to provide consumers utils and quality, respectively, in cases with consumer heterogeneity. If price discrimination is allowed, the effect on profits is unambiguous, and though in general consumer welfare decreases, it depends on the degree of market segmentation. Armstrong and Vickers show that if markets are perfectly segmented (no leakage) price increases in one market lead to price decreases in the weak market. Rochet and Stole show that in some cases the market will be segmented in two classes: a weak market where consumers have strong tastes and where competition drives prices to marginal cost, and a strong market where firms have a local monopoly.

[1] There are several excellent surveys on price discrimination, for example Anderson and Renault (2008) or Armstrong (2005).Varian (1989) surveys price discrimination literature under monopoly. Stole (2007) surveys recent advances of price discrimination with imperfect competition.

The common thread of theoretical predictions is that firms benefit if price discrimination is allowed. Recent empirical research brings evidence of similar kind. Besanko, Dube and Gupta (2003) estimate a model in which taste heterogeneity offers firms the opportunity to price discriminate profitably by charging different prices across segments. They have a vertical channel setting, in which manufacturers provide coupons directly to consumers and retailers engage in third-degree price discrimination across consumer segments defined by observable consumer characteristics. Zettelmeyer (2000) shows that the profitability of competing in new online channels enables firms to provide selected groups of consumers with different amounts of information and achieve a finer market segmentation that increases the firm's market power. The results suggest that a proliferation of channels will decrease competition among firms. In this case, firms always benefit from the new channel.

Despite the large amount of work showing that price discrimination is optimal, there is little research that reconciles the fact that firms often forgo the option to discriminate. The notable exception is the well-known result from Stokey (1979) in which inter-temporal price discrimination is allowed. In this setup firms never find it advantageous to discriminate and they commit to a fixed price to induce consumers to buy at the first opportunity.[2] Anderson and Dana (2009) show that profitability of price discrimination depends on the incremental gain in consumers' willingness to pay relative to the increase in costs.

In this paper we use novel data that combines three important sources of hotel information for aproximately 3,000 hotels situated in 13 major U.S. markets to explore various aspects of price discrimination across distributions channels. The hotel industry (explored in Section 2) provides an excellent case for the analysis of price discrimination since consumer segmentation is clearly linked to the distribution channel. According to Kalnins (2006) half of the room rentals corresponds to corporate accounts and conference customers, with rates often negotiated at the national level. The other half of rooms goes to transient demand and

[2]Salant (1989) shows that the viability of inter-temporal price discrimination is related to the cost structure of the monopolist. Armstrong (2005) shows that in the presence of naive consumers inter-temporal price discrimination is viable.

individual business. Our analyses show that the online distribution channel, which targets primarily transient customers, involves higher prices.

The data described in detail in Section 3 consists of a rich set of consumer search data for hotel rooms at a major online travel consolidator combined with a nearly complete picture of the competitive environment of each market: a comprehensive census of hotels in these markets and their main characteristics, such as hotel rooms, date of opening, quality segment, amenities, and location. We supplement this data with hotels' revenue and performance data that contain daily average prices, occupancy rates, and other high frequency revenue information. Our methodology exploits one important feature of the online travel aggregator—it consolidates hotel information from most of the major online booking sites—to investigate the effects of the online distribution channel on the firms' pricing behavior.

Strikingly, we observe that across different hotel quality segments there is a significant proportion of hotels which do not use the online channel. Overall, 41 percent of the hotels in our sample do not have an online booking presence. This is surprising as most theoretical research predicts increased profits if price discrimination is allowed as it permits a finer segmentation of consumers. In the first part of Section 4 we present estimates of the establishment's participation decision and price discrimination between channels that correct for establishment's decision to distribute rooms through the online channel or not. We find that hotels with an online distribution channel differ in important ways, such as hotel capacity, quality segment, and location.

In the second part of Section 4 we show that online prices are consistently higher across different hotel quality segments and other hotel attributes. In 97 percent of observations, online prices are higher, which indicates different pricing strategies between online and offline channels. As we only observe online prices if establishments expect positive profits from participating in that channel, we correct for selection in the pricing equation using a semiparametric approximation of the selection term. The results corrected for selection markedly change the direction of the effect of online pricing across all quality segments.

These results suggest that the online pricing strategies are opposite at the extremes of hotel quality rankings.

2 Background on the Hotel Industry

According to the American Hotel and Lodging Association, the industry had $99.3 billion in sales in 2010 (down from $140 billion in 2008) from more than 50,000 properties and more than 4.8 million rooms. An industry profile by Kalnins (2006) illustrates that even though the industry appears competitive with properties owned by more than 30,000 owners and proprietors, the 10 largest brands control 50 percent of the market. Kalnins also provides a breakdown of demand, where half of the room rentals corresponds to corporate accounts and conference customers with rates often negotiated at the national level. The other half of rooms goes to transient demand and individual business. Most of the revenues come from the transient consumers as corporate and conference customers often involve large discounts.

In recent years, the distribution of hotel rooms through the online channel—which targets primarily transient demand—has become commonplace and replaced traditional agency distribution models. Online bookings accounted for over half of all U.S. travel bookings in 2010, and up to 75 percent of bookings are influenced by the Internet.[3] Using the Forrester Technographics Technology Survey, Lieber and Syverson (2011) find that 15.1 percent of households in the U.S. reserve hotels online. According to an industry report relying on data from 130,000 visitors to online travel sites, 62 percent of online customers represent leisure travelers, compared to 29 percent who are business travelers (9 percent represent other travellers). This is consistent with the official figures for 2009 published in Hospitality and Tourism Industry Report (2011) that similarly points out that 60 percent of consumers traveled for leisure and 40 percent for business.[4]

Given the importance of transient demand, it is surprising that we find that a significant

[3] Market research report by PhoCusWright Inc.

[4] The report is prepared by iPerceptions and American and Hotel Lodging Association.

proportion of hotels across different hotel quality segments do not have an online distribution channel. In our data, described in detail in the next section, only 59 percent of hotel properties sell rooms online. A potential explanation is the cost associated with this channel. Online distribution channels involve commission-based online travel agencies (OTAs), which represent about 40 percent of the bookings and sales directly to consumers.[5] The top 30 hotel brands have a larger proportion of direct online bookings (72 percent).[6] Commission payments to OTAs represent between 15-35 percent of the hotel's booking revenues (Mourier, 2011). The distribution of fees to Expedia, the largest OTA with 50 percent of the market, amounted to $2.3 billion dollars from $9 billion in gross bookings.[7]

The airline industry provides a good example of firm's cost considerations with regard to online channel distribution. Airlines pay fees for online bookings by OTA's and travel information consolidators. In 2008, American Airlines, the third-largest carrier in the U.S. withdraw its flight listings from Orbitz, a major OTA, in order to bypass the fees of information consolidators such as Sabre. In the airline industry, these costs represent 3-4% of operating costs. Another example is Southwest Airlines, which does not offer flights through OTAs but sells directly to consumers. In sharp contrast to hotel room pricing, online flight prices typically involve significantly lower fares than those generated by offline travel agencies.

[5] The largest OTAs include Expedia, Orbitz and Travelocity. A subgroup of online merchants, which operate differently than the traditional travel agencies, include sites such as Hotwire or Priceline. These offer large discounts by shielding hotel names until a consumer has booked the room with a non-refundable fare. Contrary to traditional agencies, these merchants typically book at a price net of a contracted margin percentage or dollar amount. The figures were obtained from PhoCusWright, a Connecticut-based travel research firm. See Robinson-Jacobs (2010).

[6] The share of OTAs in online booking revenues grew from 1.34 percent in 2001 to about 7.35 percent in 2010 (Tourism Economics, a subsidiary of consulting company Oxford Economics). OTAs' bookings also represented 9.8 percent of all U.S. room-nights in 2010 (STR). HotelNewsNow.com (August 2011), "OTAs Cost US Hotels $2.5b in 2010".

[7] According to Expedia's SEC filing.

3 Data

3.1 Data Sources and Sample

To analyze price strategies across channels we combined data from three sources. The first source is the hotel census database from Smith Travel Research (STR).[8] The database contains information for approximately 98% of the hotel properties in the U.S. and offers a nearly complete picture of the competitive environment for each market. For each hotel it provides hotel identifier, date of opening, number of rooms (hotel capacity), whether a hotel offers extended stay rooms (upper or lower standard) or not, information about operation type (organizational form) of the hotel (franchised, company management or owner -operator) and various characteristics of the hotel location (e.g., is the hotel in an urban or suburban area; near an airport or highway and zip code). Based on characteristics of the major chain brand categories and differences in product and service quality, each hotel is also classified by STR into one of the six hotel quality segments: luxury, upper upscale, upscale, midscale with food & beverage (F&B), midscale without F&B, and economy. The scale segment classification, developed in 1996, is widely recognized by the industry.[9]

The second source consists of STR's hotel performance data, which contains information on daily revenue, number of rooms sold and available for each hotel, as well as a set of its competitors.[10] We use this data to construct average daily prices (revenue/rooms sold) and occupancy rates (i.e. hotel capacity utilization ratio) for each hotel as well as its competitors.

The third source consists of online consumer search data for hotels from a major online travel engine platform. The unique feature of this online platform is that it serves as a search aggregator, i.e. it consolidates hotel information from the online booking sites of individual

[8]STR is an independent research firm that collects information on hotel properties in the U.S. and internationally and represents one of the most comprehensive sources on the hotel industry available.

[9]See Canina et al. (2005) for further details about hotel segment classifications.

[10]STR defines a competitive set as a group of hotels by which a property can compare itself to the group's aggregate performance. There must be a minimum of three hotels in any competitive set excluding the subject hotel. To protect proprietary data, a single hotel or brand cannot exceed 40% of the competitive set. A single hotel company (i.e. Marriott brands, Choice brands, etc.) may only comprise 60 percent of the competitive set room supply.

hotels as well as most travel sites, such as Expedia and Orbitz. As most hotels that allow online booking use one or more of these services, the use of an aggregator allows us to identify those hotels that participate in the online channel. Consumers looking for a stay provide the platform with city, arrival and departure dates, and the number of people staying in the room. After the request is submitted, search results include prices (daily average for the duration of the stay) and other hotel characteristics. We have all the data from all the hotels in the search results for each request.[11] Unfortunately, we only observe search. Once a consumer selects a hotel, he or she is directed to the OTA or hotel website to complete the transaction. Based on the search and stay dates we observe advance purchase, length of stay, and seasonality demand shocks.

We combine these data sources in a sample of 2,940 hotels and daily performance information for the period May 1, 2007 through April 30, 2008 across 13 major U.S. markets. An advantage of using this time period is that it avoids the economic recession which has significantly impacted the hotel industry, leading to 7 percent decrease in demand in 2009. We use the hotel market definition constructed by STR that closely follows Metropolitan Statistical Areas. Our hotels represent nearly the entire population of branded hotels in these markets.[12] We focus on branded hotels for several reasons: first, these hotels are individual properties affiliated with major brands, often indistinguishable from the customer's perspective when it comes to quality. Second, as Rushmore and Baum (2001) point out, overall chain affiliation rose from 35% in 1970 to over 80% just 30 years later. Furthermore, according to the Economic Census, the top four lodging companies commanded 17.5% of overall sales in the industry in 2002, up from 16.3% just five years earlier. Across our 13 markets there were 2,067 unbranded hotels in the STR census database that were excluded from the sample as they did not have segment quality ranking or price information. These

[11] For hotels that use the online channel, we have additional information on hotel amenities for each searched hotel, namely: whether the hotel offers a fitness center, Internet, parking, pool, restaurant, and allows pets.

[12] Initially, we had 3,067 branded hotels, but 127 hotels had missing information on opening date and thus hotel age. These were dropped from the sample. Dropping these 127 hotels did not change in any important way the overall distributions and cross-sectional characteristics of our data.

represent much smaller properties, with an average of 99 rooms compared to 172 rooms for branded hotels.

3.2 Hotel Characteristics

Table 1 shows that hotel characteristics differ between online and offline channels in many dimensions. Only 59 percent (1,729 out of 2,940) hotels sell through the online channel and participation decreases with hotel quality. Specifically, among luxury hotels 84 percent of hotels are online, while when looking among economy hotels only 53 percent of hotels are online. Similar distribution across segments is apparent regardless of whether hotels appear in our online travel consolidator. On average all hotels represent medium size properties with about 172 rooms, but those that use online channel are on average larger by about 75 rooms than hotels that do not sell online. Moreover, hotels that do not use online channel are on average younger (16 years old), while hotels that rely on online channels are older (21 years). Type of hotel operation also differs between channels, with 33 percent of hotels with online channels operated under management contract and 59 percent via franchise, compared with 22 percent management and 70 percent franchise, respectively, among those hotels which do not participate in the online channel.[13]

In terms of hotel location characteristics, more than 55 percent of hotels in our data operate in suburban areas, about 17 percent in urban areas, and about 12 percent near the airport. Relatively, hotels in urban areas, resorts, and airport locations are overrepresented in the online channel. Hotels without online channels are more likely to appear in suburban areas and near highways. Oahu Island has the largest participation rate in the online channel (93 percent), followed by Phoenix (77 percent), Milwaukee (73 percent), and New York (71 percent). Surprisingly, Chicago on the other hand represents the market with the smaller

[13] Under management contract (see e.g. Kehoe, 1996) a third party (usually a local investor/developer) owns the physical property, but the hotel company (e.g. Marriott) hires managers to operate the hotel under its brand name. Under a franchise agreement the hotel company of a hotel brand (the franchisor, e.g. Hilton International) grants an owner or developer (the franchisee) the right to use and operate the hotel under its brand name (e.g. Doubletree).

proportion (24 percent) of hotels with online channel. There are no differences between channels when it comes to offering the extended stay rooms.

In order to explore the strength of competition for each hotels, we calculated the number of other hotels operating in the same segment, as well as segments above (higher quality) and below (lower quality) at various distances of each hotel. Table 1 shows that for every distance definition and across segments there are fewer competing hotels near hotels without an online distribution channel. For example, hotels selling online have on average 3 hotels in the same quality segment within one mile, compared to 1.8 hotels for hotels that do not participate in the online channel. The same pattern repeats for the number of hotels at further distances.

3.3 Online and Offline Prices and Stay Characteristics

We observe consumers' online search requests in the travel aggregator for hotel stays in 13 major U.S. markets. There were 122,772 search request for rooms in May 2007 for hotels in the sample for stays until the following year (through April 2008). On average there are 15.4 hotels per search resulting in 1,885,940 hotel price observations. Panel A of Table 2 presents summary statistics of the search requests. The average online price is \$212, the average stay is 3.5 days, and the most popular days of arrival are Friday (22 percent) and Saturday (17 percent). Consumers search for hotels 49 days before arrival on average. Even though we do not observe hotel bookings, only consumer search, we observe hotel's price offerings and hence price discounts for advance purchase.

We obtained average prices between distribution channels (i.e. including rooms sold online and offline) and occupancy rates for each hotel, as well as its competitive set, from the STR hotel performance database. Important to note is the fact that though our variable Average price includes rooms sold offline and online, in case of rooms sold online the prices represent the whole-sale prices for which a hotel sells the rooms to online sites (e.g. Expedia) not the prices that a customer pays when booking a hotel online. Since online price data is in the form of averages for a particular stay, we constructed the corresponding Average price from STR data by averaging the daily prices across the days of arrival and departure for each consumer request.[14] Panel B of Table 2 shows that the average prices are $47 lower ($164.5 vs. $211.5) than the online price for the same dates of stay.

Figure 1 presents percentage price differences between the online channel and the average price of a particular stay according to the number of days between the search and the arrival date. The figure tracks the most popular hotel per quality segment, according to number of searches, in Chicago and Boston. The majority of searches for a particular stay occur closer to the arrival date and decrease the longer the time until arrival. The figure shows that the majority of online prices are higher than the averages (97 percent of observations in the sample) and this relationship does not change over time.

Table 3 shows that price setting between online and offline channels cannot be entirely explained by quality differences across segments. Column 3 shows that online prices are 31 percent higher than the average room price and this relationship persists across all the segments. The absolute price difference decreases with segment quality, from $75 on average for luxury hotels to $22 for economy hotels. In percentage terms, Upper and Upper Upscale segments have the largest price differences: 37 percent, while other segments fluctuate between 26 and 32 percent higher online prices. Figure 2 shows the distribution of the percentage price difference by segment for two types of advance searches: within one month of stay and between 11 and 12 months before the stay. It shows that the distribution and the magnitude of prices above the average follows different patterns by segment. For the Luxury segment, price differences are higher the longer the time window between the date of search and stay, while the opposite is apparent among other segments, particularly Economy.

[14] The industry jargon for daily price is average daily rate (ADR) calculated as the total daily revenue of a given hotel divided by its number of room-nights sold.

4 Empirical Methodology and Results

4.1 Online Channel Participation

The participation in online channel of hotel j is determined by the expected change in profits, $\Delta^o \pi_j$. Profits are unobservable, thus we observe participation in the online channel only if $\Delta^o \pi_j \geq 0$. Let an indicator variable $I_j = 1 \left(\Delta^o \pi_j \geq 0 \right)$, relate latent profits to online participation. We observe hotel j participating in the online channel with probability

$$\rho_j = \Pr \left(I_j = 1 \mid X_j \right) \tag{1}$$

where X_j are hotel attributes. Consistently with the nature of our sample, hotel's participation decision does not change over time, $\rho_{jt} = \rho_j$ for $\forall t$.

Table 4 presents marginal effects evaluated at sample means from Probit model, estimating the probability of participation in the online channel on hotel characteristics. Due to missing observations in the STR performance data, 2,550 properties out of 2,940 branded hotels have revenue and performance data, we present results for three different samples. A key characteristic is the hotel's capacity, i.e. number of rooms. Larger hotels are more likely to participate in the online channel (extra 100 rooms increases online participation by 4 percent). Compared to Midscale hotels with F&B (omitted group), Luxury hotels are about 18 percent more likely to participate in the online channel. Similarly, hotels in Upper Upscale and Upscale segments are more likely to participate by 14 and 11 percent, respectively.

There are few differences between the estimates of all the hotels and hotels with revenue information. Hotels in the Florida Keys provide the extreme example of the need for breakdown on hotel's revenue information: 12 out of 26 hotels in the sample have revenue information, and 11 of those participate in the online channel. Of the 14 properties without revenue information, only 2 are online. Resort destinations such as the Florida Keys and Oahu Island, Hawaii are more likely to participate in the online channel (32 and 25 percent, respectively) than Phoenix, which has the second largest proportion of hotels that participate in the online channel. Comparable cities to Phoenix are Boston, Charlotte, and Milwaukee. And surprisingly hotels in Chicago, New York, and Los Angeles are less likely to participate in the online channel.

Table 4 also shows the effect of the number of competitors on the decision to participate in the online channel. Surprisingly the number of hotels within the same quality segment does not increase the likelihood of online participation. However the number of hotels in the quality segments above and below increases hotel's online participation.

4.2 Selection Correction

This section explores different price strategies between online and offline channels. Let p^o_{jst} be the average price quoted online to customers at hotel j for a particular stay s and time of request t. Let $\bar{p}_{js}$ be the average daily price between online and offline distribution channels of stay s.

$$\ln\left(p^o_{jst}/\bar{p}_{js}\right) = X_j^{'}\beta_1 + W_{js}^{'}\beta_2 + T_{jst}^{'}\beta_3 + month_s + \xi_{jts}, \quad \text{observed if } I_j = 1$$

where X_j are observable hotel attributes; W_{js} are observable attributes of hotel j that vary across different stays s such as date, day of arrival, length of stay, competitors' average prices for the stay, and occupancy rate of the focal hotel and its competitors, and T_{jst} represents observables related to the consumer request at time t, such as the number of days in advance of stay. Finally, $month_s$ stands for the month of a stay s. We include monthly dummies in order to control for unobserved time-varying shocks and monthly seasonality in demand, typical for hotel industry. Important to note for our analysis is the fact that in order to provide the price estimates outside our sample and control for hotel self-selection, rather than including hotel-fixed effects, we rely on time-invariant hotel characteristics in X_j, including hotel segment, market, location, and hotel operation type. ξ_{jts} is the disturbance term. As we only observe online prices if the firm has decided to participate in the online channel, in general $E\left(\xi_{jts} \mid I_j = 1, X_j, W_{js}, T_{jst},\right) \neq 0$. In the case that ξ_{jts} is distributed normally, this is the familiar setting from Heckman (1979) where the selection correction term is the Mills ratio. We correct for selection into the online channel with a semiparametric approximation of $E\left(\xi_{jts} \mid I_j = 1, X_j, W_{js}, T_{jst},\right)$.[15] Notice, due to relatively short time period of our sample, we rely on the assumption $\rho_{sjt} = \rho_j$ for $\forall t, s$ as hotel participation in the sample is decided at the beginning of our sample period. [16]

The two-stage estimation involves obtaining predicted probabilities $\hat{\rho}_j = \Pr\left(I_j = 1 \mid X_j\right)$ from equation (1).[17] Table 5 presents OLS regression estimates and the second stage of the selection correction model by approximating the control function with a polynomial on $\hat{\rho}_j$. Panel A presents estimates of searches and time covariates. The OLS and two-stage selection correction are similar in signs and magnitude across both specifications. The elasticity of the number of days consumers search in advance on relative online price is -0.003. This result is consistent with previous research on advance purchase discounts and price discrimination, in particular in industries where sellers commit with capacity in advance and face uncertain demand.[18] As expected, price differentials increase with hotel's occupancy rate (elasticity .029 in column (2)) and decrease with competitors' occupancy rates by a similar magnitude.

Panel B of Table 5 shows that correcting for selection is important in estimating the relationship between hotel characteristics and the price differentials between the channels. The coefficients across specifications differ noticeably. For example, the percentage online difference between Luxury and Midscale with F&B hotels (control group) is positive and

[15] See Newey (1999) for an excelent discussion of two-stage methods to correct for sample selection and the semiparameric approximations for selection correction that do not rely on the normality assumption of the Mills ratio in original Heckman's (1979) two-stage methology.

[16] If this assumption does not hold, the standard selection model does not provide consistent estimates and an alternative model of selection correction in panel data is required, such as Kyriazidou (1997), Wooldridge (1995), and Dustmann and Rochina-Barrachina (2007). See for example the survey on selection models in panel data by Vella (1998) and the treatment by Semykina and Wooldridge (2005).

[17] The marginal effects from the Probit used to predict $\hat{\rho}_j$ are presented in column (2) of Table 4.

[18] See for example, Gale and Holmes (1993); Dana (1998); Nocke and Peitz (2007); Puller, Sengupta and Wiggins (2008); and Nocke, Peitz, and Rosar (2011). For other work on advance purchase under monopolist capacity decisions see e.g., Gale and Holmes (1993) .

statistically significant (3.8 percent) in OLS, but the difference turns out to be negative (-1.4 percent) once we correct for selection in two-stage model. Upper Upscale and Upscale segments also present lower coefficients for the two-stage model. However, we find the opposite effect for Economy segment, the relative online price in the selection model is higher when correcting for selection (elasticity doubles from .016 to .032). These results suggest that online pricing strategies are opposite at the extremes of quality segments. We also find significant changes in signs and magnitudes when correcting for selection across locations, operation type, and several markets for hotels with a relatively high online participation. For instance, there are changes in the signs of coefficients of hotels in Urban and Airport settings, in Florida Keys and Oahu, HI resulting in statistically significant lower online prices. The opposite change in signs occurs in markets with relatively lower online participation, such as Los Angeles and Chicago (as per columns (2) and (3), Table 1), where online prices are 3.1 and 17 percent higher, respectively.

In order to control for the market structure, the specification in Table 5 includes the number of other hotels at different distances and across different quality segments. Figure 3 shows the total effects on relative prices for the number of other hotels in the same quality segment. The results indicate that Luxury hotels are more responsive to competition in the online channel. Moreover, online prices increase (by about 2 percent) with larger number of other hotels located at further distances, but decrease with number of other hotels located closer. This suggests that overall the higher number of other hotels strengthens competition and reduces online prices.

The different pricing strategies between online and offline channels across establishments of different quality suggest that the degree of price discrimination depends on the degree of product differentiation. In fact, as in Corts (1998) firms have different assessments of consumers' demand elasticities and adjust their pricing strategies accordingly. For instance, the online market might seem to be strong for Luxury hotels and weak for Economy hotels. But once selection into the market is accounted for, there is heightened competition in the online channel for higher quality segments.

5 Conclusions

This article explores price discrimination strategies in the online channel. Contrary to theoretical prediction, we find that a large proportion of hotels forgo the option of price discrimination by not participating in the online distribution channel. The different pricing strategies between online and offline channels across hotels of different quality suggest that the degree of price discrimination depends on the degree of product differentiation. In fact, as in Corts (1998), hotels seem to have asymmetric assessments of consumers' demand elasticities across distribution channels, and adjust their pricing strategies accordingly. For instance, the online market might seem to be strong for Luxury hotels and weak for Economy hotels. But once selection into the market is accounted for, there is heightened competition in the online channel for higher quality segments.

References

ANDERSON, E. T., AND J. D. DANA, JR (2009): "When is Price Discrimination Profitable?," *Management Science*, 55(6), 980–989.

ANDERSON, S. P., AND R. RENAULT (2009): "Comparative Advertising: Disclosing Horizontal Match Information," *RAND Journal of Economicss*, 40(3), 558–581.

ARMSTRONG, M. (2005): *Recent Developments in the Economics of Price Discrimination: Theory and Applications: Ninth World Congress*vol. 2. Cambridge University Press.

ARMSTRONG, M., AND J. VICKERS (2001): "Competitive Price Discrimination," *RAND Journal of Economics*, 32(4), 579–605.

BESANKO, D., J.-P. DUBE, AND S. GUPTA (2003): "Competitive Price Discrimination Strategies in a Vertical Channel Using Aggregate Retail Data," *Management Science*, 49(9), 1121–1138.

CANINA, L., C. ENZ, AND J. HARRISON (2005): "Agglomeration Effects and Strategic Orientations; Evidence from the U.S. Lodging Industry," *Academy of Management Journal*, 48, 565–581.

CORTS, K. S. (1998): "Third-Degree Price Discrimination in Oligopoly: All-Out Competition and Strategic Commitment," *RAND Journal of Economics*, 29(2), 306–323.

DANA, JR, J. D. (1998): "Advanced Purchase Discounts and Price Discrimination in Competitive Markets," *Journal of Political Economy*, 106(2), 395–422.

DUSTMANN, C., AND M. E. ROCHINA-BARRACHINA (2007): "Selection Correction in Panel Data Models: An Application to the Estimation of Females' Wage Equations," *Econometrics Journal*, 10(2), 263–293.

GALE, I. L., AND T. J. HOLMES (1993): "Advance-Purchase Discounts and Monopoly Allocation of Capacity," *American Economic Review*, 83(1), 135–146.

GERSTNER, E., AND D. M. HOLTHAUSEN (1986): "Profitable Pricing When Market Segments Overlap," *Marketing Science*, 5(1), 55–69.

HECKMAN, J. (1979): "Sample selection bias as a specification error," *Econometrica*, 47(1), 153–161.

HOLMES, T. J. (1989): "The Effects of Third-degree Price Discrimination in Oligopoly," *American Economic Review*, 79(1), 244–250.

KALNINS, A. (2006): "Markets: The U.S. Lodging Industry," *Journal of Economic Perspectives*, 20(4), 203–218.

KEHOE, M. R. (1996): "Franchising, Agency Problems, and the Cost of Capital," *Applied Economics*, 28, 1485–1493.

Kyriazidou, E. (1997): "Estimation of a Panel Data Sample Selection Model," *Econometrica*, 65(6), 1335–1364.

Lieber, E., and C. Syverson (2011): "Online vs. Offline Competition," Working paper: Prepared for the Oxford Handbook of the Digital Economy.

Mourier, J. F. (2011): "Are You Leaving Money on the Table?," Published online www.hospitalitynet.org.

Newey, W. K. (1999): "Two-Step Series Estimation of Sample Selection Models," Working Paper.

Nocke, V., and M. Peitz (2007): "A Theory of Clearance Sales," *Economic Journal*, 117, 964–990.

Nocke, V., M. Peitz, and F. Rosar (2011): "Advance-Purchase Discounts as a Price Discrimination Device," *Journal of Economic Theory*, 146(1), 141–162.

Pigou, A. C. (1920): *The Economics of Welfare*. London: Macmillan.

Puller, S. L., A. Sengupta, and S. N. Wiggins (2009): "Testing Theories of Scarcity Pricing in the Airline Industry," NBER Working Papers 15555, National Bureau of Economic Research, Inc.

Robinson-Jacobs, K. (2010): "Practice that Holds Rates steady Among Hotels, Travel Sites Coming under Fire," Dallas Business News.

Rochet, J.-C., and L. A. Stole (2002): "Nonlinear Pricing with Random Participation," *The Review of Economic Studies*, 69(1), 277–311.

Rushmore, S., and E. Baum (2001): *Hotels and Motels: Valuation and Market Studies*. Appraisal Institute.

Salant, S. W. (1989): "When is Inducing Self-Selection Suboptimal For a Monopolist?," *Quarterly Journal of Economics*, 104(2), 391–397.

Schmalensee, R. (1981): "Output and Welfare Implications of Monopolistic Third-Degree Price Discrimination," *American Economic Review*, 71(1), 242–247.

Semykina, A., and J. M. Wooldridge (2005): "Estimating Panel Data Models in the Presence of Endogeneity and Selection: Theory and Application," Working paper.

Stokey, N. L. (1979): "Intertemporal Price Discrimination," *Quarterly Journal of Economics*, 93(3), 355–371.

Stole, L. A. (2007): *Price Discrimination and Competition*vol. 3. North-Holland Press.

Varian, H. R. (1985): "Price Discrimination and Social Welfare," *American Economic Review*, 75(4), 870–875.

——— (1989): *Price discrimination*vol. 1. North-Holland Press.

Vella, F. (1998): "Estimating Models with Sample Selection Bias: A survey," *Journal of Human Resources*, 33(1), 127–169.

Wooldridge, J. M. (1995): "Selection Corrections for Panel Data Models under Conditional Mean Independence Assumptions," *Journal of Econometrics*, 68, 115–132.

Zettelmeyer, F. (2000): "Expanding to the internet: Pricing and Communications Strategies when Firms Compete on Multiple Channels," *Journal of Marketing Research*, 37(3), 292–308.

Table 1: Descriptive Statistics of Hotels between Channels

	All Hotels		Online Channel		Not Online Channel	
Variable	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Age (years)	18.800	15.605	20.513	17.067	16.355	12.859
Number of rooms (100's)	171.865	189.276	202.737	227.226	127.789	99.877
Segment						
Luxury	0.029	0.168	0.041	0.198	0.012	0.107
Upper upscale	0.123	0.328	0.154	0.361	0.078	0.269
Upscale	0.178	0.382	0.191	0.394	0.158	0.365
Midscale w/ F &B	0.156	0.363	0.156	0.363	0.157	0.364
Midscale w/o F &B	0.260	0.439	0.230	0.421	0.302	0.459
Economy	0.254	0.436	0.227	0.419	0.293	0.455
Location						
Urban	0.165	0.371	0.238	0.426	0.059	0.237
Suburban	0.563	0.496	0.463	0.499	0.704	0.457
Airport	0.117	0.321	0.176	0.381	0.032	0.177
Highway	0.047	0.212	0.001	0.024	0.114	0.318
Resort	0.096	0.294	0.121	0.327	0.059	0.235
Small Metro/Town	0.013	0.114	0.001	0.024	0.031	0.174
Operation Type						
Chain Management	0.285	0.451	0.333	0.471	0.216	0.411
Franchise	0.635	0.482	0.588	0.492	0.702	0.458
Owner-operator	0.080	0.272	0.079	0.269	0.083	0.275
Market						
Phoenix	0.096	0.295	0.127	0.333	0.053	0.224
Los Angeles	0.129	0.335	0.120	0.325	0.141	0.348
Orlando	0.097	0.296	0.113	0.317	0.073	0.261
Florida Keys	0.009	0.094	0.008	0.086	0.011	0.103
Oahu Island, HI	0.015	0.121	0.024	0.152	0.002	0.050
Chicago	0.157	0.364	0.064	0.244	0.291	0.454
Boston	0.072	0.258	0.072	0.258	0.072	0.258
New York	0.063	0.242	0.076	0.265	0.044	0.205
Charlotte, NC-SC	0.073	0.260	0.082	0.274	0.061	0.240
Raleigh, NC area	0.059	0.236	0.053	0.223	0.069	0.253
Dallas	0.136	0.343	0.150	0.357	0.116	0.320
Seattle	0.064	0.245	0.076	0.265	0.048	0.214
Milwaukee	0.030	0.171	0.038	0.190	0.020	0.139
Extended Stay						
Upper	0.055	0.228	0.056	0.229	0.055	0.227
Lower	0.087	0.282	0.088	0.284	0.085	0.279
No Extended Stay	0.858	0.349	0.856	0.351	0.860	0.347
Number of Other Hotels						
Within 1 mile						
Same segment	2.495	3.411	2.976	3.840	1.808	2.528
Segment above	4.066	7.572	5.247	9.085	2.380	4.069
Segment below	4.203	7.954	5.414	9.540	2.473	4.315
Between 1-3 miles						
Same segment	1.608	2.800	2.205	3.249	0.756	1.654
Segment above	3.206	6.839	4.481	8.061	1.386	3.895
Segment below	3.338	7.313	4.704	8.679	1.389	3.982
Between 3-5 miles						
Same segment	2.592	3.221	3.225	3.544	1.688	2.425
Segment above	4.814	7.748	6.081	8.648	3.006	5.785
Segment below**	5.068	7.447	6.464	8.388	3.074	5.239
Between 5 and 10 miles						
Same segment	9.605	8.439	11.282	8.838	7.211	7.190
Segment above	17.582	20.883	21.157	22.678	12.478	16.748
Segment below	18.524	22.879	23.054	25.104	12.057	17.328
Number of Hotels	2,940		1,729		1,211	

Table 2: Online Search for Hotels and Stay Characteristics

	Mean	Std. Dev.	Min	Max
	Panel A. Consumer Requests			
Online price	211.537	136.318	16.00	1500.00
Days advance search	48.64	53.439	0	360
Length of stay	3.461	4.26	1	337
Day of Arrival				
Sunday	0.107	0.309	0	1
Monday	0.098	0.297	0	1
Tuesday	0.104	0.305	0	1
Wednesday	0.129	0.335	0	1
Thursday	0.168	0.374	0	1
Friday	0.223	0.416	0	1
Saturday	0.172	0.377	0	1
	Panel B. STR performance			
Average price	164.464	107.34	8.21	1439.73
Competitors' average price	164.504	109.976	26.82	1270.46
Occupancy rate	0.792	0.185	0.01	1.23
Competitors' occupancy rate	0.790	0.148	0.08	1.1

Note: The table presents descriptive statistics of online searches for hotels and stay characteristics. There are 122,772 search requests and 1,885,940 performance observations.

Table 3: Prices between Channels by Hotel Segment

	Online Price		Avg. Price		Difference (%)		Difference ($)	
Segment	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Luxury	421.224	190.072	345.741	180.350	28.821	37.670	75.483	107.123
Upper upscale	272.521	125.970	207.515	94.603	34.495	32.922	65.006	66.637
Upscale	205.612	100.636	154.470	70.065	35.277	33.865	51.142	56.570
Midscale w/ F&B	156.887	79.436	126.224	59.360	26.124	32.400	30.664	41.901
Midscale w/o F&B	149.197	85.515	117.671	54.280	26.737	29.106	31.526	46.728
Economy	97.192	56.920	75.077	39.096	31.834	32.762	22.114	29.397
Overall	211.537	136.318	164.464	107.340	31.156	32.931	47.072	61.760

Note: The table presents price descriptive statistics of the online channel and the average of all the stays by hotel quality segment. There are 1,885,940 price observations.

Table 4: Average Marginal Effects of Online Channel Participation

	All Hotels		Hotels with Rev.		Hotels without Rev.	
Variable	Coeff.	Std. Err.	Coeff.	Std. Err.	Coeff.	Std. Err.
Age of hotel (years)	0.002	(0.001)*	0.002	(0.001)	0.005	(0.003)
Number of rooms (100's)	0.039	(0.012)***	0.040	(0.012)***	0.016	(0.069)
Segment						
Luxury	0.179	(0.069)**	0.195	(0.066)**	-0.452	(0.409)
Upper upscale	0.140	(0.052)**	0.151	(0.052)**	-0.183	(0.344)
Upscale	0.105	(0.045)*	0.120	(0.046)**	-0.182	(0.241)
Midscale w/o F&B	0.026	(0.045)	0.051	(0.047)	-0.446	(0.168)**
Economy	-0.050	(0.054)	-0.026	(0.059)	-0.072	(0.169)
Location						
Urban	0.264	(0.030)***	0.269	(0.030)***	0.051	(0.130)
Airport	0.273	(0.030)***	0.266	(0.029)***	0.334	(0.115)**
Highway	-0.590	(0.053)***	-0.602	(0.060)***		
Resort	-0.145	(0.064)*	-0.148	(0.067)*	-0.118	(0.276)
Small Metro/Town	-0.487	(0.107)***	-0.440	(0.153)**		
Operation Type						
Chain Management	0.028	(0.031)	0.022	(0.032)	0.209	(0.117)
Owner-operator	0.084	(0.045)	0.135	(0.056)*	0.027	(0.105)
Market						
Los Angeles	-0.446	(0.042)***	-0.480	(0.044)***	-0.269	(0.171)
Orlando	-0.187	(0.060)**	-0.225	(0.062)***	0.291	(0.127)*
Florida Keys	0.061	(0.111)	0.324	(0.049)***	0.214	(0.260)
Oahu Island, HI	0.297	(0.059)***	0.248	(0.072)***		
Chicago	-0.651	(0.026)***	-0.674	(0.027)***	-0.556	(0.105)***
Boston	-0.244	(0.058)***	-0.156	(0.067)*	-0.360	(0.201)
New York	-0.569	(0.035)***	-0.583	(0.038)***	-0.449	(0.233)
Charlotte, NC-SC	-0.008	(0.060)	-0.027	(0.064)	0.273	(0.118)*
Raleigh, NC area	-0.161	(0.060)**	-0.189	(0.063)**	0.120	(0.198)
Dallas	-0.275	(0.055)***	-0.317	(0.058)***	0.182	(0.150)
Seattle	-0.166	(0.062)**	-0.216	(0.064)***	0.420	(0.051)***
Milwaukee	0.107	(0.067)	0.116	(0.068)	0.163	(0.225)
Extended Stay						
Upper	-0.015	(0.074)	-0.024	(0.077)	0.001	(0.343)
No Extended Stay	0.020	(0.048)	0.001	(0.051)	0.143	(0.171)
Number of Other Hotels						
Within 1 mile						
Same segment	-0.009	(0.006)	-0.005	(0.006)	-0.027	(0.021)
Segment above	0.012	(0.003)***	0.011	(0.003)***	0.002	(0.009)
Segment below	0.006	(0.002)*	0.004	(0.002)*	0.009	(0.009)
Between 1-3 miles						
Same segment	0.011	(0.008)	0.010	(0.008)	0.014	(0.028)
Segment above	0.012	(0.003)***	0.010	(0.003)***	0.029	(0.010)**
Segment below	0.009	(0.003)**	0.009	(0.003)**	0.011	(0.014)
Between 3-5 miles						
Same segment	0.005	(0.005)	-0.000	(0.006)	0.034	(0.017)
Segment above	0.007	(0.002)**	0.009	(0.002)***	-0.006	(0.008)
Segment below	0.005	(0.003)*	0.006	(0.003)*	0.014	(0.013)
More than 5 miles						
Same segment	0.003	(0.002)	0.002	(0.003)	0.016	(0.008)*
Segment above	0.007	(0.001)***	0.007	(0.001)***	0.005	(0.004)
Segment below	0.002	(0.001)*	0.002	(0.001)*	0.007	(0.005)
Log-likelihood	-1218.6		-1043.4		-124.5	
Number of observations	2,940		2,550		346	

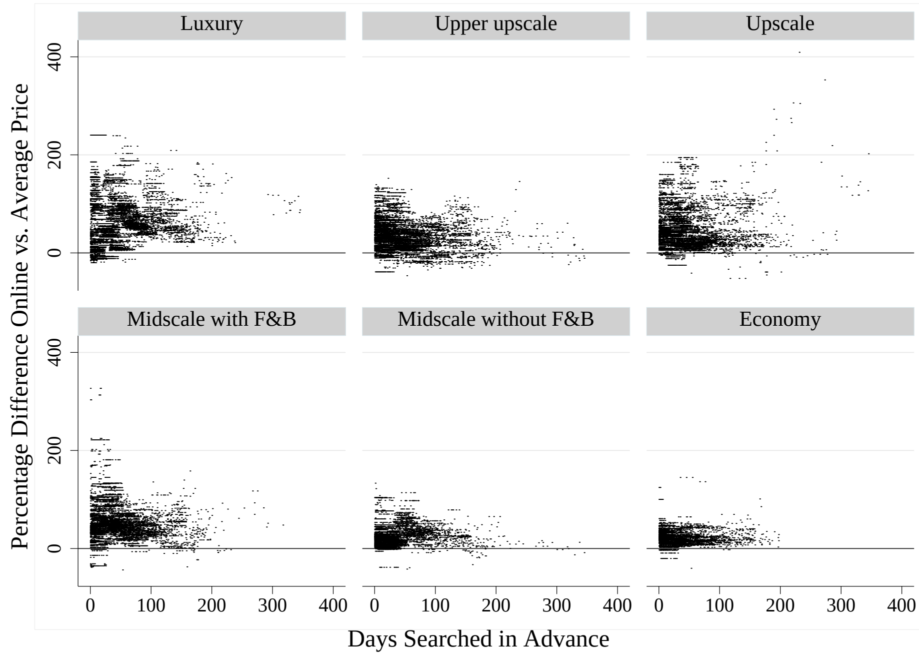
Note: The table presents average marginal effects at sample means for probit where the dependent variable is an indicator that equals one when hotel participates in the online channel and 0 when it does not. The omitted categories are: midscale hotels with F&B; suburban area, franchised operation type, Phoenix market, and the lower standard of extended stay. Standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Table 5: Estimates of Percentage Price Differences between Channels

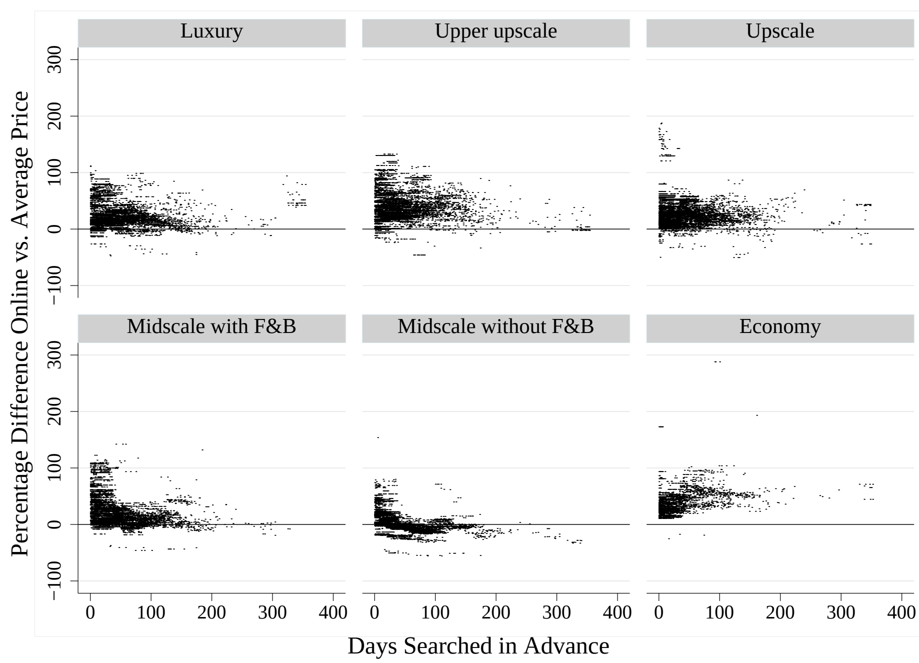
	OLS		Two-Step Polynomial	
Variable	Coeff.	Std. Err.	Coeff.	Std. Err.
	Panel A. Search and Stay			
Ln(days advance search)	-0.003	(0.000)***	-0.003	(0.000)***
Ln(length of stay)	0.010	(0.000)***	0.010	(0.000)***
Ln(competitors' average price)	0.017	(0.001)***	0.019	(0.001)***
Ln(occupancy rate)	0.028	(0.001)***	0.029	(0.001)***
Ln(competitors' occupancy rate)	-0.030	(0.001)***	-0.033	(0.001)***
Day of Arrival				
Sunday	0.003	(0.001)***	0.003	(0.001)***
Monday	0.003	(0.001)***	0.003	(0.001)***
Tuesday	0.007	(0.001)***	0.007	(0.001)***
Wednesday	0.018	(0.001)***	0.018	(0.001)***
Thursday	-0.004	(0.001)***	-0.004	(0.001)***
Friday	-0.010	(0.001)***	-0.010	(0.001)***
	Panel B. Hotel Chracteristics			
Age of hotel (years)	0.002	(0.000)***	0.002	(0.000)***
Number of rooms (100's)	0.046	(0.000)***	0.045	(0.000)***
Segment				
Luxury	0.038	(0.002)***	-0.014	(0.002)***
Upper upscale	-0.008	(0.002)***	-0.055	(0.002)***
Upscale	0.037	(0.001)***	0.013	(0.002)***
Midscale w/o F&B	-0.065	(0.001)***	-0.064	(0.001)***
Economy	0.016	(0.002)***	0.032	(0.002)***
Location				
Urban	0.037	(0.001)***	-0.042	(0.002)***
Airport	0.057	(0.001)***	-0.024	(0.002)***
Highway	-0.018	(0.003)***	0.191	(0.005)***
Resort	-0.019	(0.001)***	0.003	(0.001)***
Small Metro/Town	0.008	(0.009)	0.156	(0.009)***
Operation Type				
Chain Management	-0.011	(0.000)***	-0.022	(0.001)***
Owner-operator	0.006	(0.001)***	-0.009	(0.001)***
Market				
Los Angeles	-0.058	(0.001)***	0.031	(0.002)***
Orlando	0.024	(0.001)***	0.051	(0.001)***
Florida Keys	-0.007	(0.002)***	-0.018	(0.002)***
Oahu Island, HI	-0.084	(0.002)***	-0.203	(0.003)***
Chicago	-0.001	(0.001)	0.174	(0.004)***
Boston	-0.054	(0.001)***	-0.016	(0.001)***
New York	-0.150	(0.001)***	-0.005	(0.003)
Charlotte, NC-SC	0.035	(0.001)***	0.024	(0.002)***
Raleigh, NC area	0.009	(0.002)***	0.035	(0.002)***
Dallas	0.020	(0.001)***	0.070	(0.002)***
Seattle	-0.069	(0.001)***	-0.050	(0.001)***
Milwaukee	0.060	(0.002)***	0.019	(0.002)***
Extended Stay				
Upper	-0.108	(0.001)***	-0.100	(0.001)***
No Extended Stay	-0.099	(0.001)***	-0.107	(0.001)***
Hotel Amenities				
Internet	0.003	(0.000)***	0.003	(0.000)***
Fitness	-0.006	(0.000)***	-0.006	(0.000)***
Pets	0.009	(0.000)***	0.009	(0.000)***
Parking	-0.002	(0.000)***	-0.003	(0.000)***
Pool	0.019	(0.000)***	0.019	(0.000)***
Restaurant	-0.001	(0.000)***	-0.001	(0.000)***
	Panel C. Selection terms			
$\hat{\rho}_j$			1.034	(0.021)***
$\hat{\rho}_j^2$			-1.698	(0.037)***
$\hat{\rho}_j^3$			1.096	(0.024)***
R-squared	0.13		0.14	
Number of observations	1,845,498		1,845,498	

Note: $ln(p_j^o/\bar{p}_j)$ is the dependent variable in the two specifications. OLS regression and two-step selection correction with a polynomial approximation are presented. The omitted categories are: midscale hotels with F&B; suburban area,franchised operation type, Phoenix market, and the lower standard of extended stay. Not reported monthly dummies, number of other hotels by segment and distance with interactions, and constant included in all specifications. Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

Figure 1: Price Differences across Channels



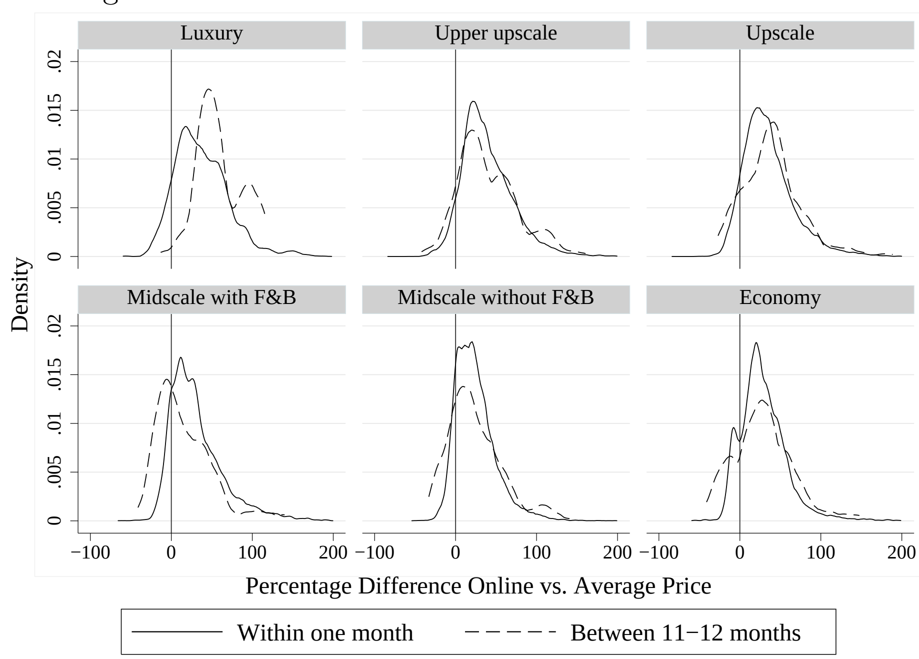
(a) Chicago



(b) Boston

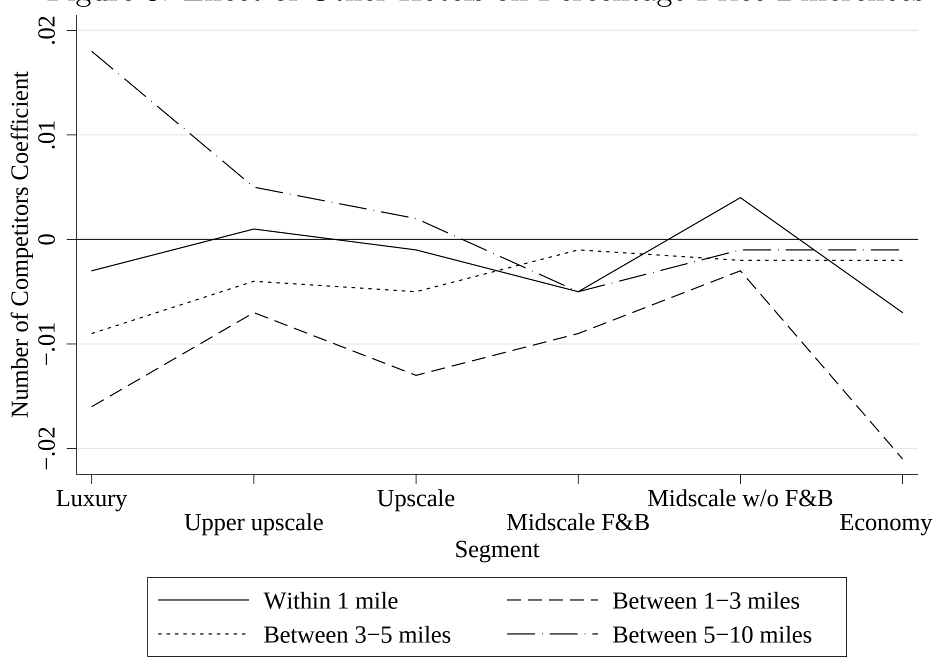
Note: The figure shows the percentage price difference between the online channel and the average price of a particular stay searched by a consumer according to the number of days between the search and the arrival date. The figure tracks the most popular hotel, defined by the total number of searches, for each hotel quality segment in Chicago and Boston.

Figure 2: Distribution of Price Differences between Channels



Note: The figure shows the distribution of the percentage price difference between the online channel and the average price of a particular stay searched by a consumer 30 days or less before the arrival date, as well as searched between 11 and 12 months before the stay.

Figure 3: Effect of Other Hotels on Percentage Price Differences



Note: The figure shows coefficients (baseline and interactions) for the number of hotels in the same segment as focal hotel by distance and hotel segment from the selection specification where the dependent variable is $ln(p^o_j/\bar{p}_j)$. Table 5 presents the estimates of the full specification.