# Price Discrimination 2.0: The Case of Opaque Bookings in the Hotel Industry*

Ricardo Cossa Mariano Tappata

This draft: October 3, 2013

**Abstract**

The emergence of opaque selling (i.e., when a product characteristic is only revealed to the customer after payment) has motivated a growing theoretical literature but little progress on the empirical front. This paper tries to fill this gap. We document the main features of opaque selling using a unique dataset from the lodging industry that includes bookings made in opaque platforms—Priceline and Hotwire—matched to their equivalent booking in the transparent market. The discounts across channels are significant and consistent with self-selection based on transaction costs and strong or weak preferences over transparent products. The average opaque discounts, relative to a transparent booking, are 40% for the posted-price platform and 49% in the auction-based platform. We find no evidence of price-skimming strategies traditionally associated to revenue management practices in the travel industry. Moreover, we reject the hypothesis that hotels use opaque selling as a last-minute resource to dispose unsold inventory. We argue that hotels achieve demand segmentation by offering customers a menu of selling channels instead of within-channel intertemporal price discrimination strategies.

**JEL Codes**: D43, D47, L11, M31.

**Keywords**: Two-sided markets, Opaque products, Price discrimination, Hotel Industry, Hotwire, Priceline.

*Correspondence to Mariano Tappata: Strategy and Business Economics Division, Sauder School of Business, University of British Columbia. 2053 Main Mall, Vancouver, BC; Canada, V6T–1Z2. E-mail: mariano.tappata@sauder.ubc.ca. Ricardo Cossa: Charles Rivers Associates, E-mail: rcossa@crai.com. We thank Mariana Marchionni for insightful comments and discussions. Thanks also to Pascal Courty, Vera Brenčič, Marco Celentani, Nathan Schiff and seminar participants at various institutions for helpful comments. Mariano Tappata is grateful for financial support from SSHRC.

# 1 Introduction

The diffusion of e-commerce has produced shakeouts in nearly all retail industries. As consumers shifted their offline purchases to the internet, markets experienced entry by new players simultaneously with exit and repositioning of incumbent firms. The travel agency industry is a case in point of this transformation. The rise in online bookings and the evolution in IT reshaped the sources of competitive advantages and led to the replacement of small, local, brick-and-mortar agencies by large, global, online travel agencies (OTAs).[1] More important to us, the change in the market structure of the travel industry allowed for disruptive innovations in business processes like the introduction of opaque selling by OTAs. Although still incipient, different opaque selling mechanisms are currently implemented by OTAs and accounts for a quarter of hotels' transient room-nights bookings (HSMAI Foundation, 2012). Moreover the large margins captured by opaque intermediaries suggests a promising future for this new selling channel in other industries.[2] The goal of this paper is to contribute to the understanding of this new selling mechanism using pricing data from the lodging industry.

Opaque platforms offer online travel booking services (hotel, rental car and airline bookings) concealing the providers' identity from the buyer until the transaction is completed.[3] Understandably opaque products are also called probabilistic goods since they represent a lottery over a set of traditional or transparent products. Thus, an opaque booking is viewed as inferior relative to a transparent booking by a risk-neutral or risk-averse consumer. It is for this reason that, all else equal, opaque bookings are cheaper than transparent bookings and appeal to buyers with no strong preference over a specific service provider. In addition, consumers can self-select among opaque platforms based on the nature of the transaction mechanism used by each platform (auction-based or posted prices) and the amount of information suppressed by the intermediary (opacity level). The magnitude of the price discounts required to compensate some—but not all—buyers of opaque products for bearing the transaction cost and risks associated with buying opaque is an empirical question that we address in this paper. Consumers in our sample require an unconditional 44% percent rate discount to make an opaque booking. In this paper we document how these large discounts vary across transaction mechanisms, opacity level, and other

[1] See Goldmanis et al. (2010) for a quantitative analysis of the 1994-2003 market structure evolution in this industry.

[2] For example, Scorebig.com uses opaque selling of seats at live concerts and sporting events.

[3] The nature of opaque products requires all bookings to be non-refundable, non-changeable and non-transferable.

explanatory variables relevant to the lodging industry.

Since its development, revenue and yield management practices in the travel industry have helped service providers to price discriminate consumers in a static and inter-temporarily fashion. Consumers booking a hotel room can differ on their willingness to pay for quality, hotel amenities, location or even have strong preferences over a particular hotel (e.g., due to frequent flyer programs). They can also differ in their marginal utility of income, demand realizations and associated uncertainty (e.g., a work-related meeting scheduled a month in advance can be canceled in the last minute). Absent the possibility of first-degree price discrimination, revenue management strategies exploit the correlation of consumers' preferences over these attributes and enable within-selling channel price discrimination. For example, leisure travelers have been traditionally characterized as price-sensitive, with low preference for special hotel amenities, and associated with early demand realization and flexible schedule. As a result, hotels and airlines have targeted this type of customer minimizing the cannibalization of sales to business travelers using early-booking discounts and lower weekend-rates or Saturday stay-over requirements (Talluri and Van Ryzin, 2005; Phillips, 2005).[4] One natural question we ask is whether hotels use opaque selling channels to substitute or complement these price discrimination strategies. In other words, do consumer types fully sort by choosing from the menu of selling channels or do hotels need to use within-platform pricing strategies to achieve finer demand segmentation? The answer to this depends on the underlying heterogeneity in consumers' preferences and therefore is an empirical question that we tackle in this paper.

OTAs offering opaque bookings are examples of two-sided markets. A service provider's willingness to participate in the platform depends on the number of travelers using the platform *and* their preferences. At the same time, buyers' expected utility is directly affected by the number of participating providers. An advantage of opaque platforms, highlighted in the industry, is that it allows hotels to sell to price-sensitive customers at lower prices avoiding the negative effect on reputation associated with traditional markdowns. This is so because the identity of the hotel is only revealed to the customer after the transaction is finalized.[5] A disadvantage

[4]These "rate feces" are based on the time of purchase and consumption. Hotels can also use physical rate fences according to room-type attributes (view, room size, bedding type), or group-based (senior and AAA member discounts) fences (Kimes, 2002). Aside from anecdotal evidence by industry experts and scholars, there is little work using pricing data that documents the use of these practices by hotels. An exception being Abrate et al. (2012) that finds evidence of inter-temporal price discrimination by hotels in Europe, a market where opaque selling is still in its infancy.

[5]Note however that even if a price-insensitive and high willingness-to-pay customer knows that

for hotels selling opaque is that, since product differentiation becomes immaterial, competition in the wholesale market is fierce and lowers their margins in the opaque *and possibly* transparent segments. This latter point has generated industry debate and a new theoretical literature that studies the strategic benefits of opaque selling (Fay, 2008; Shapiro and Shi, 2008; Jerath et al., 2010; Tappata, 2011).[6] These models are stylized and capture the basic logic underlying the emergence and adoption of opaque selling discussed above. They share the prediction that, conditional on adopting opaque selling, the price spread between the transparent and opaque prices increases with product differentiation. The main distinction among these models relates to whether opaque selling involves a static or inter-temporal price discrimination strategy. That is, do hotels use the opaque channel to sell distressed inventory in the last minute (Jerath et al., 2010) or as a static screening device?

We use a large dataset with opaque and transparent hotel-room bookings in eighteen cities in North America from October 2011 to June 2012. There are two salient features of our data that make the analysis unique. First, we observe the hotel identity in each opaque transaction and can therefore compare the opaque and transparent rates controlling for the exact same service provider.[7] Second, we use data from two opaque platforms—Priceline and Hotwire—that differ in their selling format.[8] More specifically, the data includes successful bids on Priceline's Name Your Own Price system (NYOP) and transactions on Hotwire's Secret Hot Rates (SHR).[9] The two platforms differ in the way buyers and the seller interact as well as the amount of information disclosed to the buyers by the platform. The following example helps to clarify these differences. Imagine a customer attempting to book a room on a three-star hotel on a given city and date. When using `Hotwire.com`, the consumer is presented with a list of hotels, one for each area–stars pair, displaying room rates and

her preferred hotel sells rooms in the opaque channel at large discounts, she would not want to take the risk associated with an opaque booking.

[6] As the CEO of Delta Airlines, Leo Mullin, said after Delta joined Priceline's platform: "There was great suspicion that this would lead to excessive discounting. How would it threaten Delta? We had very strong arguments. The general tenor was definitely against it." (*The E Gang*, Forbes; July $24^{th}$, 2000).

[7] Courty and Liu (2013) and Fay and Laran (2009) are the only two studies that employ opaque prices. In both cases, small samples of Hotwire's posted prices are used. Since these data do not contain the hotel identity associated to each opaque rate its use is limited to inferences on availability and price volatility over time in this specific opaque channel.

[8] Priceline and Hotwire are the opaque selling pioneers. They introduced their systems in 1999 and 2000 respectively. Although offerings by opaque competitors have increased in recent years (Expedia and Travelocity launched opaque selling channels in 2011), they are the clear dominant players in the industry ("Opaque Selling in 2011" Hotel Yearbook 2011, p.96-98).

[9] `Priceline.com` has traditionally offered users the option to book transparent and opaque rooms. More recently, Priceline added the "Express Deals" system that is similar to Hotwire's. We only have data on Priceline's NYOP transactions and refer to them loosely as rooms booked on Priceline.

hotel amenities (e.g. parking, fitness room, free–wifi, etc). The booking process and user interface is simple and mimics that of a transparent platform like Hotels.com. The key difference however is that hotels names are hidden from the buyer and only revealed after the booking has been paid.

On the other hand, a customer using Priceline.com is asked to select the desired room characteristics and enter her willingness to pay for it.[10] The room characteristics include check-in and check-out dates, city, hotel quality and opaque city area.[11] Once the customer commits to the booking (i.e., credit card information is entered), Priceline attempts to procure the room by running an auction among all participating hotels that match the previously specified characteristics.[12] Within less than a minute the customer receives a notification with the booking outcome. The hotel name is revealed if the transaction is successful and this happens when a secret threshold price, in principle the winning bid plus Priceline's commissions, is below the customer's named price. If the named price is below this threshold price the transaction is rejected and the customer is prompted to try with a new bid on the following day.[13]

It is important to highlight the differences in the booking mechanisms used by Hotwire and Priceline. First, while room rates are posted on Hotwire's website, buyers using Priceline actively bid for a room. In Priceline's system, the role of transaction costs, impatience, expectation formation about the platform's threshold price, and other frictions can affect consumer sorting and the equilibrium discount. Second, Hotwire discloses additional information regarding the amenities of the hotel underlying the opaque booking. In some cases, it is possible that this extra information reduces the level of opacity and therefore consumers' uncertainty. Booking opacity levels vary significantly across markets (e.g., a four-star booking is a different probabilistic good if the opaque area is Manhattan's downtown than if it is La Guardia airport). This allows us to exploit variation in booking discounts within platforms to estimate the effects of opacity on consumers' willingness to pay and relate the variation in discounts across platforms with the different features of the opaque selling format (posted price vs. auction).

[10] Figures 1 and 2 in Appendix A display snapshots of Hotwire and Pricelines's user interfaces.

[11] The opaque areas used by Hotwire and Priceline are indistinguishable from each other. Figures 11–13 in Appendix B show some of Hotwire's opaque areas.

[12] Strictly speaking, Priceline selects the winning hotel from the GDS (Global Distribution Systems) with hotels' schedules of rooms and rates available. Hotels can change their Priceline rates and inventory at any time making the process equivalent to a real-time auction.

[13] In many cases, the bidder is allowed to try again by redefining the product category selected (i.e., the area–quality pair).

Our findings suggest that hotels effectively use the opaque channels to price discriminate consumers based on price sensitivity and transaction costs. Predicted opaque rates in the auction-based platform are on average 49.3% lower than in the transparent market. Notably, the discounts when posted opaque rates are used drop to 39.5%. The fact that both systems have survived in equilibrium highlight the importance of selling format design and information suppression to further segment price-sensitive customers. In both systems, discounts increase monotonically with hotels' quality (star-rating). We relate this finding to a demand-side story since consumer heterogeneity is expected to be larger in the higher quality hotels. Interestingly, we find that opaque bookings are not used as a last-minute resource to dispose unsold inventory as suggested in Jerath et al. (2010). Instead, the data is consistent with the static price-discrimination models in the literature (Fay, 2008; Shapiro and Shi, 2008; Tappata, 2011). Moreover, the lack of evidence of within-platform price discrimination based on check-in dates or time of booking is consistent with the notion that opaque selling channels are a substitute rather than a complement to the traditional price discrimination strategies associated with the travel industry.

The rest of the paper is structured as follows. In Section 2 we discuss the economics of opaque selling and predictions from the existing models. The empirical specification, together with descriptions of the data and summary statistics are presented in Section 3. The econometric results are discussed in Section 4. Section 5 concludes.

# 2 Industry and theoretical background

Why do hotels engage in opaque selling? The short answer is that it allows them to segment demand and price discriminate business and leisure travelers similar to how airlines price discriminate using Saturday stay-over requirements or early-booking discounts. However, there are new issues associated with opaque selling that make the price discrimination logic more intertwined. Before discussing the literature and economics of opaque selling, it helps to begin describing the main features of the industry.

First, though hotels book transparent rooms directly or through "passive" intermediaries, they can only sell opaque through an intermediary. The reason is that a third-party is required to pool hotels and create the opaque good hence buyer uncertainty.[14] Second, transparent and opaque channels coexist (sometimes offered by

[14]In principle, a multiproduct firm (a hotel with different room-types) could sell opaque. We do

the same intermediary) and are easily accessed by customers. The choice between transparent and opaque bookings is mostly driven by heterogeneity in buyers' preferences than by search or switching cost frictions. Third, the opaque wholesale market operates as a "reverse auction" and it is specific to each platform. Hotels submit their secret supply schedules (rates and number of rooms available) and the platform chooses the winner for each marginal transaction. Market clearing in the wholesale market is critical in determining the shape of competition in the opaque channel, its interaction with the transparent market, and the final effect on hotels' profits. Last, advance purchasing is a common practice in this industry and therefore dynamic pricing or revenue management considerations can impact both, transparent and opaque pricing strategies.

Aside from quality differences, hotels differ in many horizontal dimensions and so the demand faced by each hotel is not perfectly elastic. Selling through an opaque channel can help a hotel to attract consumers from the lower part of the demand curve without cannibalizing too much the sales to high-valuation consumers in the transparent market. Consumers can vary in their intrinsic willingness to pay as well as flexibility regarding the room characteristic: hotel brand, room usage date and room booking date. But they can also vary in their marginal utility of income. For example, other things equal, business travelers are expected to be less price sensitive than leisure travelers from the basic fact that their companies pay for their trips. In addition, they present strong preferences regarding the hotel (closest to company headquarters) and room usage dates (weekdays). A conference or a meeting organized on a specific hotel and date make business travelers inflexible and willing to pay significantly more than a casual traveler that can reschedule her visit with little cost. In general, the literature refers to those price-insensitive customers as loyal customers. A non-loyal or shopper customer is price-sensitive and values most hotels similarly (which makes her even more price-sensitive).

Opaque products are inferior goods and can be thought of as a special case of damaged goods. Deneckere and McAfee (1996) and Shapiro and Varian (2000) showed that a monopolist can increase profits and welfare by selling a "damaged" version of the product to price-sensitive consumers. In a sense, the value destroyed by the producer of a damaged good might allow for the creation of even more value. This same logic underlies opaque selling. However, a key difference with the damaged good argument is that hotels selling opaque need to do so through the opaque platform and

not observe this practice in the market and therefore ignore it.

therefore compete aggressively for customers directly with other hotels.[15] Moreover, it is in the best interest of the platform to generate intense competition among hotels to capture a larger margin between the wholesale cost and customers' willingness to pay. Nevertheless, a successful platform is required to increase the utility of some type of customers *and* attract hotel participation. Low opaque prices impact hotels' profits directly and might also attract loyal customers to the opaque platform. In order to prevent cannibalization, price reductions in the transparent market might be needed, depressing profits even more. In that sense–and unlike in damaged goods–intense competition in the wholesale opaque market might make firms worse-off with opaque channels than without it. To summarize, opaque selling destroys value to create more value as suggested by the damaged good theory, but part of the value generated is captured by a new player: the opaque platform. Much of the current debate in the industry is on the impact of opaque selling on hotels' profits. Entry by the opaque intermediary can increase hotels' profits but it is also possible that it creates a prisoner's dilemma for hotels making profitable individual hotel participation but lowering their aggregated profits.

The current theoretical literature has focused on the conditions required for opaque selling to increase industry profits. Fay (2008) and Shapiro and Shi (2008) show that price discrimination through opaque selling can be a successful strategy as long as consumers are sufficiently heterogeneous or products sufficiently differentiated. The market is segmented in equilibrium when the opaque prices attract shoppers but are not low enough to attract the loyal customers.[16] In these cases, introducing opaque selling into fully-served markets increases industry profits at the expense of consumers' surplus. It is easy to see that opaque selling increases welfare if it also operates in the extensive margin selling to price-sensitive customers that would have otherwise not consumed. Tappata (2011) shows that this market expansion effect is larger in markets exposed to demand seasonality where the free-entry equilibrium leaves market gaps in the low-demand seasons.

It is important to note that the existing theories are mostly motivated to explain why we observe opaque pricing in a given industry and not in predicting discount or price patterns in the opaque and transparent market. The models are very stylized and allow for few dimensions of heterogeneity among consumers and firms. Moreover,

[15] By definition, hotels lose their market power in the opaque market since consumers self-selecting into the opaque channel are no longer interested in a differentiated product.

[16] Formally, Fay (2008) assumes exogenous customer loyalty in a Hotelling duopoly while Shapiro and Shi (2008) model two groups of buyers, one with low and one with high transportation costs, in a circular city.

firms are by assumption not able to discriminate consumers in the absence of opaque selling. Similarly, we do not have a clear understanding of the predictions regarding competition between opaque platforms. We do not attempt to take these models literally to the data. Instead, we use them as a basic framework to think about opaque selling. In general, the predictions from these theories is that i) significant consumer heterogeneity is required to justify the introduction of opaque selling, and ii) conditionally on selling opaque, the difference between transparent and opaque prices increases with product differentiation but not necessarily with consumer heterogeneity.[17] A common feature from general price discrimination models is that the price in the no-discrimination scenario falls between the highest and lowest of the prices when the firm price discriminates. This might not be the case with opaque selling since prices for opaque and transparent prices can be lower than before the entry by the opaque intermediary (Tappata, 2011).[18] We can not test this prediction since we do not observe hotel prices before and after the emergence of opaque selling.

The forces in favor of opaque selling provided above are present in static models and we have so far ignored dynamic elements associated with demand uncertainty and capacity constrained firms. The travel industry has traditionally been characterized by leisure travelers booking early in advance and business travelers on dates closer to the consumption date. Exploiting the correlation between room-booking date and willingness to pay has been the main reason for the use of revenue management programs in the travel industry. However, the incentive to use dynamic pricing might be attenuated for a hotel once it can use the opaque channel to cater price-sensitive consumers. For example, Jerath et al. (2010) present a two period model where firms can choose to use the opaque market only in the last-minute to sell excessive inventory. A hotel would prefer to use the competitive opaque market last-minute to discounts in the transparent market if product differentiation is large enough to keep loyal buyers away from leaking to the opaque market. Otherwise, hotels avoid the opaque market altogether and sell excess inventory (if any) at a discount in the transparent market. As we explain in the next section, we exploit the structure of our data to test whether opaque selling is a last-minute practice in this industry. We also relate our findings to the existing opaque selling theories and highlight unexplored dimensions for future research.

[17] In the duopoly model of Fay (2008) for example, a symmetric equilibrium predicts a larger gap between transparent and opaque prices when consumers face larger transportation costs and the proportion of loyal customers is larger. In Tappata (2011), the gap increases with transportation cost and the distance separating firms.

[18] More formally, different modeling choices (e.g. exogenous vs endogenous loyalty) can lead to transparent and opaque prices being strategic substitute or complements.

# 3 Data and Empirical Specification

## 3.1 The structure of the data

Our data contains opaque and transparent hotel room rates combined with hotel and market characteristics. In order to describe these data, it is helpful to define first the unit of analysis more precisely. A "booked room" $j$ is defined by the set of attributes $\{h, r, b, t\}$ where $h$ is the hotel identity, $r$ the room-type (non-refundable, standard, with view, bedding, etc), $b$ the booking date and $t$ the check-in date. In other words, we define a hotel room $j$ as the product observed by the buyer at the time of consumption (check-in date) instead of purchase (booking date). Note that, since opaque and transparent products only differ at the time of purchase, consumers' information about $h$ varies depending on whether the room was booked through the opaque or the transparent channels. Thus, the same booked room $j$ can be associated to three different rates based on the channel used. The opaque and transparent hotel room rates were obtained from different sources. We refer to them as the opaque and transparent data. A major challenge involves the matching of each price paid in the opaque platforms ($p_j^o$) to the relevant transparent posted price ($p_j^\tau$). We explain the main features of this process in this section and leave the details to Appendix B.

The opaque data includes hotel room bookings made on Priceline's NYOP system and Hotwire's SHR listings. The source is the website `Betterbidding.com`, a forum created to benefit users of these two popular platforms. The different sites in the forum are designed so that users discuss strategies, share experiences and exchange detailed information on their opaque bookings. Forum participants routinely post the identity of the hotel, rate paid and dates when booking through Hotwire or from a winning bid on Priceline. We have compiled and parsed these posts for the most popular destinations in North America. We emphasize that the nature of opaque bookings makes it impossible to obtain transaction-level information (rate paid and supplier identity) from sources other than the opaque platforms. Though price-protection is a feature that attracts hotels to the opaque platform, the secrecy goes beyond potential buyers and even hotels are not aware of the rate paid by consumers to the opaque platforms. Self-reported data has become available recently and this is, to the best of our knowledge, the first attempt to work with this kind of data. It must be noted that it is not clear whether forum participants represent the average user of opaque platforms. On the one hand, one expects forum participant to be frequent travelers willing to invest time and effort to obtain better-than-average deals. On the other hand, inexperienced users are the ones most likely to seek information and report

back in these forums. This is a common issue with user-reported data in the IO literature however we do not think it has major implications in our study.[19,20]

We paired each booked room $j$ in the opaque data to a room rate posted in the transparent channel. To do so, we first collected one-night transparent room rates from Hotwire.com for the largest cities in Canada and the US from October 1, 2011 to June 25, 2012.[21] For each booking date $b$, we recorded the website's posted rates offered by all hotels in the city allowing for check-in date $t = \{b, b+1, ...., b+90\}$. Note that this data with transparent room rates is very large and includes rates for all room types and booking/check-in date combinations quoted by more than 5,000 hotels (i.e., about 8.1 million transparent rates per day). Such detailed data allows us to do a proper match between opaque and transparent rates controlling for all the attributes that define a booked room $j$. The rooms booked through Priceline or Hotwire are non-refundable standard rooms. Thus, each opaque booking was matched to the price quoted by the same hotel for its cheapest room quoted on the same check-in date and, when possible, the same booking date. As explained in Appendix B, if the latter condition was not met, we selected the transparent rate with closest booking date that satisfied the other restrictions. There are 4,116 opaque transactions successfully matched. Table B.1 in Appendix A displays the distribution of matches across cities.

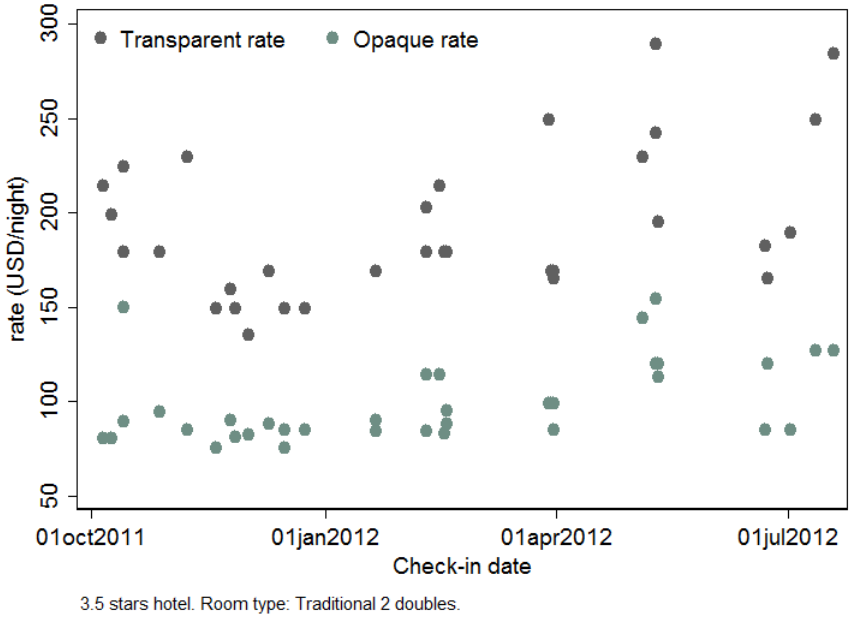
To illustrate the rate-pairing procedure, Figure 1 displays the check-in date and room rates (transparent and opaque) for the 37 matches made for a single hotel in Los Angeles (Westin Bonaventure Hotel). For example, the first pair of observations from the left in Figure 1 corresponds to a Priceline booking made on Oct 3 2011 (not shown) with check-in date Oct 6 for \$80/night. From the transparent data, we find that the rate for the "same" booked room is \$214/night (i.e., 62% discount). The figure shows that the transparent and opaque rates charged by the same hotel vary significantly as well as the implied opaque discounts. Not shown in the figure are other sources that affect such variation as the opaque platform used and booking date. To highlight the dimension of the transparent data required for the matchings, only 37 of the 37,564 transparent rates collected for the Westin Bonaventure Hotel in Los Angeles were used in this figure. We take advantage and use the full transparent data to create variables

[19] Note that sample bias, if any, would only affect the opaque market since semi-opaque offers are posted and do not depend on the NYOP abilities of the buyer.

[20] Examples of work that use self-reported data include Lewis (2008) and Dai et al. (2012). The former uses gasoline prices reported by drivers to GasBuddy.com and the latter uses retail store's ratings from users of Yelp.com.

[21] We have also included rates from a small sample of New York City hotels from March to October 2011.

Figure 1: Matched bookings. Westin Bonaventure Hotel, LA




Note: Each opaque booking in the Westin Bonaventure hotel is characterized by the triple (booking-date, check-in date and room-type). The figure displays the opaque rate paid and the transparent rate a customer would have paid if booked through the transparent channel (hotels.com) on the same booking-date.

that capture market and competitor's characteristics. We discuss these variables and present summary statistics after describing the econometric specifications.

## 3.2 Empirical specification

The main goal of this paper is to rigorously document opaque pricing and answer the following questions: What is the discount required by a consumer that books opaquely instead of using the transparent channel? Do consumers view posted-price and auction-based opaque selling systems as equivalent? What is the effect of the opacity level on the willingness to pay for opaque bookings? Are opaque bookings a last-minute practice in the industry? Do hotels price discriminate inter-temporarily within channel? Are opaque discounts related to quality? To investigate these questions we first estimate a general discount equation. We further explore the determinants of opaque and transparent rates by analyzing pricing equations for each channel

separately. The baseline specifications are as follows:

$$
\begin{aligned}
Disc_j &= \delta_0 + \delta_1 Platform + \delta_2 X_{h,r} + \delta_3 X_{b,t} + u_j && (1)\\
\ln p_j^{\tau} &= \beta_{\tau} + \gamma_{\tau} X_{h,r} + \alpha_{\tau} X_{b,t} + v_j^{\tau} && (2)\\
\ln p_j^{o} &= \beta_{o} + \lambda Platform + \gamma_{o} X_{h,r} + \alpha_{o} X_{b,t} + v_j^{o} && (3)
\end{aligned}
$$

with $Disc_j = 1 - p_j^o/p_j^\tau$. The vector $X_{h,r}$ contains variables related to the hotel room and market fixed characteristics. The hotel-room attributes include hotel quality (star rating), dummy variables to identify hotels that belong to large chains and refundable rooms sold in the transparent channel (only in equations 1 and 2). The static market characteristics are captured by city fixed-effects and local market metrics. Among them, we use the number of hotels within a certain walking distance, dummies for airport areas and other variables related to opacity level of a given opaque area-star market. The dummy variable *Platform* (=1 if Hotwire is used) captures the premium consumers place on posted-price instead of auction-based selling systems. Thus, the estimated coefficients $\delta_1$ and $\delta_2$ in equation (1) inform us of the extent to which hotels use selling channels to screen consumers based on preferences and transaction costs in a static sense.

Buyers only know the vector $X_{h,r}$ when booking in the transparent channel. Users of opaque platforms are uncertain about the identity of the hotel they are booking and therefore form expectations over the hotels in the population that match the user specified characteristics (opaque area and quality).[22] We control for the number of hotels, the opaque area size, and dispersion in hotels' locations. In terms of address models, these variables influence the expected "transportation cost" underlying an opaque booking and therefore the opaque discount expected by consumers. All else equal, a consumer faces more uncertainty and therefore a lower expected utility when booking opaquely in a market with a large opaque area. The expected effect of the number of hotels or their dispersion within the opaque area is ambiguous. Increasing the number of firms can decrease the expected "transaction cost" if consumers' utility is convex on distance. Similarly, more dispersed hotels might increase or decrease consumers utility. Consider first 2 markets with the same number of firms and opaque area size. The only difference being that in market 1 there are 5 equally distant hotels. In market 2, all firms are located on the same block. In this case, consumers

[22] Opaque intermediaries do not make public a list of hotels that participate in their platform and therefore it is unclear what consumers' priors are. We assume that consumers expect all hotels selling through a transparent intermediary to also participate in the opaque channel. In addition, to the best of our knowledge, hotels do not announce to their customers that they do not sell opaque.

are expected to benefit from the lower dispersion in market 2. Consider now the possibility of 2 of the firms in market 2 being isolated and far away from the other three. In this case, even if Dispersion in market 2 is larger than in market 1, it is not clear which market is preferred for consumers. In addition, we include airport area dummy variable to control for special market characteristics not picked up by the competition and opacity controls. We expect less product-differentiation in these markets since consumers booking hotel rooms near airports are less concerned with location and other hotel attributes than when booking rooms in other areas. Note also that the number of hotels influences competition in all channels, transparent and opaque. We instrument the number of firms using the office rent cost per area.[23] This instrument proxies the fixed cost faced by hotels and impacts firms' entry/exit decisions and therefore is expected to be correlated with the number of firms in the market. At the same time, fixed costs do not enter in hotels' pricing equation.

Our data allows us to detect price and discount variation between and within platforms. In principle, Hotels can segment their demand by setting different prices between selling channels as well as within a selling channel. In the latter case, pricing can differ based on the check-in date or advance booking variables. This has been assumed a common practice in the airline, hotel and rental car industries where leisure travelers (price-sensitive) are thought to purchase at earlier dates than business travelers and more likely to book on weekends than weekdays. The vector $X_{b,t}$ include dynamic variables that capture changes in demand composition (i.e., elasticity of the demand curve faced by each hotel in each channel) helping us to identify the existence of such practices. We use indicators for weekend/weekday check-in dates, advance bookings (days between booking and check-in dates) and a proxy for temporal demand shocks. Time-related demand shifts affect the likelihood of binding capacity constraints. A hotel's marginal cost at booking time $b$ is composed by the cost of servicing a room and the opportunity cost of not selling it at a higher price at a date closer to the check-in date $t$. While the former can be assumed constant, the latter depends on the probability, at time $b$, of reaching full capacity before date $t$. We do not observe hotels' marginal costs. Instead, we proxy demand fluctuations using Google Trend's weekly measures of search popularity (SPI) for each destination city. Unlike city fixed-effects, the SPI is a relative measure of search popularity that captures variation of popularity across cities and over time. We describe this variable in more detail in the Appendix B.4.

[23] Asking rent values were obtained from Jones Lang LaSalle's (www.joneslanglasalle.com). Details on the construction of our instrumental variable and summary statistics are provided in the Appendix B.3.

## 3.3 Descriptive statistics

Table 1 presents the summary statistics of the variables used in the estimation of equations 1–3.[24] On average, an opaque booking allows the buyer to obtain a 44% discount with respect to the transparent rate. The average transparent and opaque rates are $185 and $99 dollars per night respectively. Both, opaque and transparent, markets present significant rate dispersion. This variation is influenced by hotel quality rating and city effects (Figure 4 in Appendix A) although, as shown in Figure 1, the rate dispersion for a given hotel can be very large.[25]

Table 1: Summary Statistics

| Variable | Mean | SD | Med | Min | Max |
|---|---|---|---|---|---|
| Discount $(1 - p^o/p^\tau)$ | 0.44 | 0.17 | 0.45 | -0.28 | 0.90 |
| Transp Rate $(p^\tau)$ | 185 | 88.53 | 165 | 29 | 803 |
| Opaque Rate $(p^o)$ | 99 | 47.96 | 87 | 13 | 380 |
| Platform (Hotwire = 1) | 0.31 | 0.46 | 0 | 0 | 1 |
| Non-refund =1 | 0.31 | 0.46 | 0 | 0 | 1 |
| Chain = 1 | 0.68 | 0.47 | 1 | 0 | 1 |
| Stars | 3.70 | 0.66 | 4 | 2 | 5 |
| Weekend = 1 | 0.48 | 0.50 | 0 | 0 | 1 |
| Adv Booking (weeks) | 4.40 | 4.75 | 3 | 0 | 28 |
| SPI | 40.32 | 23.92 | 33 | 3 | 100 |
| Airport = 1 | 0.11 | 0.31 | 0 | 0 | 1 |
| Area size (sqmi) | 8.17 | 13.27 | 2.20 | 0.13 | 107.57 |
| Dispersion=Var(dist) | 0.38 | 1.03 | 0.06 | 0 | 19.08 |
| Comp_sim 0.5mi | 8.56 | 10.05 | 5 | 0 | 49 |
| Comp_same 0.5mi | 4.02 | 5.20 | 2 | 0 | 26 |
| Comp 0.5mi | 14.37 | 16.13 | 8 | 0 | 73 |
| Comp_sim 0.25mi | 2.74 | 3.58 | 2 | 0 | 28 |

Bookings in our sample are twice more likely to be done on the auction-based (Priceline) than in the posted-price (Hotwire) channel. As shown by Figure 2(a), some of the opaque bookings have negative discounts (34 observations). Although we do not expect opaque rates to be higher than transparent rates, we cannot discard this event and therefore decided to keep the observations in the data.[26] Hotel quality is measured by the seven hotel star-rating levels: 2, 2.5, 3, 3.5, 4, 4.5 and 5 stars. The average hotel quality rating is 3.7 stars and half of the transactions occurred in 4-star

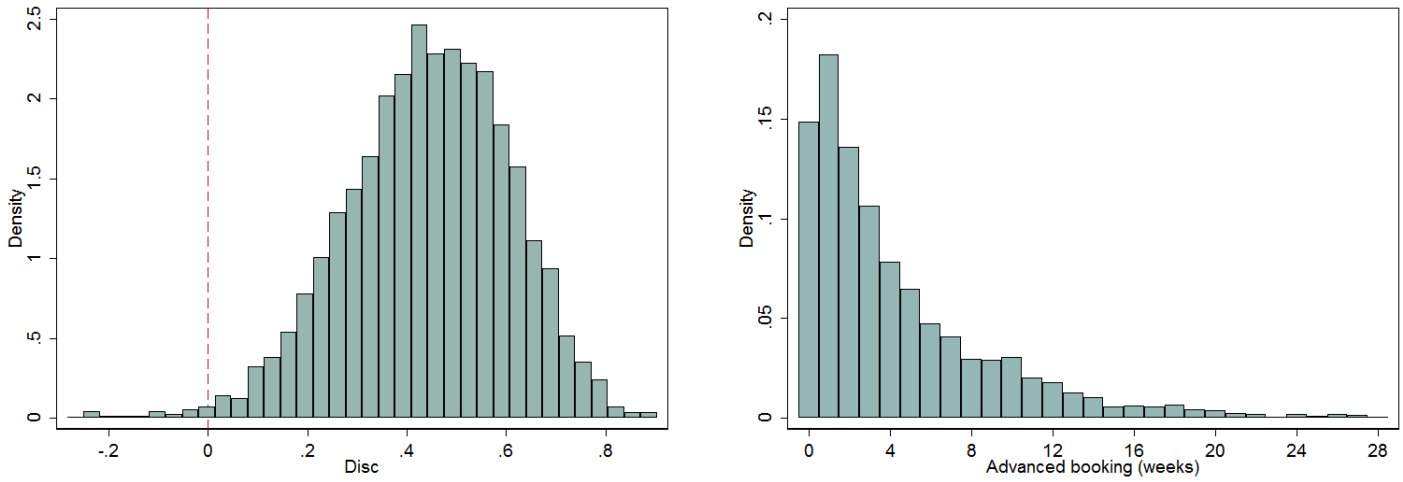
[24] See Table 4 in Appendix B for the definition of each variable.

[25] Hotel stars and city fixed effects explain 13.6% of the variance in discounts and about 45% of the transparent and opaque rates variance.

[26] Negative discounts are expected to be more likely in bookings through Priceline since the possibility of consumer "mistakes" are larger. But only 12 of the 34 bookings with negative discounts were done through Priceline's channel.

and higher hotels. Although the mean quality is similar across platforms, bookings on Priceline are more likely to be of higher quality than those on Hotwire. Moreover, opaque bookings in general are done in hotels of higher quality than that of a random hotel in the market (Figure 3(a) in Appendix A).

Figure 2: Discounts and early bookings distributions




(a) Opaque Discounts: $(1 - p^o/p^\tau)$

(b) Advanced Bookings

We observe significant heterogeneity in the booking dates. The average booking is made about 30 days in advance. The median booking is done 3 weeks in advance and 3/4 of the bookings are done within 6 weeks of the check-in date.[27] The distribution of advance bookings, shown in Figure 2(b), reveals that the opaque channels are open to consumers well in advance to the check-in date. We test whether advance bookings affect the magnitude of the discounts when estimating equation (1) in the next section. Motivated by Figure 5b in Appendix A we allow for non-linear effects of Advance-Booking on Discounts. The dummy variable *Weekend* (stay over on Friday or Saturday night) includes the night before a holiday in the US or Canada to capture demand from non-business travelers. About half of the opaque transactions occur on a weekend. Figure 3(b) in Appendix A complements this statistic with the distribution of bookings over the different days of the week.

The Search Popularity Index ($SPI$) takes values between 0 and 100 and varies across cities and time. Figure 7(b) in Appendix 4 shows the SPI variability in four

[27] The Office of Travel and Tourism Industries in the US reports that the median (average) international airfare reservation in 2010 were done 30 (60) days in advance. Note that these are international bookings and therefore expected to be planned well in advance relative to domestic bookings (http://tinet.ita.doc.gov/).

different cities in our data. The SPI level for a given city fluctuates significantly showing, in some cases, seasonality.[28] We use the dummy varables *Chain* and *Non-refund* as controls. The former identifies hotels that belong to a large and recognizable chain. Presumably, these hotels have more advance reservation system in place and can take advantage of the opaque channel by dynamically updating inventory and adjusting rates with a higher frequency than stand-alone hotels. Most of the opaque bookings (68%) are done in hotels that belong to chains. However, this is not very different from the proportions of chain-hotels in the population (69%).[29] By definition, opaque bookings are non-refundable and therefore should be compared to a non-refundable transparent rate to avoid overestimation of the discounts. The transparent rates that matched the remaining attributes of our opaque booking are more likely to be refundable rates and only in 31% of the cases the transparent rate is non-refundable.

The variables in the bottom of Table 1 vary across markets and are associated with product differentiation, competition and the opacity level of opaque bookings. We calculated the walking distance between all hotels in a given city using Google maps application and used it to determine, for each hotel, the number of hotels within a distance threshold (*Comp*). We repeated this procedure for hotels of equal (*Comp_same*) and similar (*Comp_sim*) star rating (plus/minus 0.5 stars). On average, a hotel has 14 competitors within half a mile of walking distance. Four of those competitors have the same star-rating. The opacity level of a given booking is captured by the opaque area size (*Area size* in squared miles) and the variance of the walking distance between all hotels located within the opaque area (*Dispersion*). Last, 11 percent of our observations fall in airport areas.

Last, Table 2 provides more information regarding the dependent variables in equations 1–3. Discounts are about 9 percentage points larger when opaque bookings are done through auction-based instead of posted-price system. The variance in discounts is similar across opaque platforms and, as Figure 3(a) in Appendix A show, this seems true for higher moments of the distribution of discounts. Note however that the difference in mean discount might not be entirely explained by the selling

[28] The U.S. Travel Association and American Express in their "Destination Travel Insights" provide a ranking with the top 20 cities based on the number of leisure and business travelers in United States. The positions in 2013Q1 for Chicago, Las Vegas and Miami are, respectively, 4, 15, and 8 in the "business destination" ranking and 3, 5, and 8 in "leisure destination" ranking (http://www.ustravel.org/research/destination-insights).

[29] These two numbers are not strictly comparable. It is more common that a hotel belongs to a chain if it has a low star-rating. And opaque bookings are generally done in high quality (thus, less likely to belong to chains) hotels.

Table 2: Summary Statistics–Dependent Variables

| | | | Room rates | | | | Discounts | |
|---|---|---|---|---|---|---|---|---|
| | | | Transparent | | Opaque | | | |
| | | N | Mean | SD | Mean | SD | Mean | SD |
| Platform | Priceline | 2,834 | 189 | 89.66 | 96 | 46.37 | 0.47 | 0.16 |
| | Hotwire | 1,282 | 177 | 85.45 | 105 | 50.74 | 0.38 | 0.17 |
| Stars | 2 | 118 | 95 | 38.71 | 69 | 30.62 | 0.26 | 0.17 |
| | 2.5 | 206 | 106 | 37.07 | 63 | 23.44 | 0.39 | 0.17 |
| | 3 | 602 | 150 | 76.79 | 79 | 45.43 | 0.46 | 0.15 |
| | 3.5 | 998 | 166 | 66.34 | 83 | 35.11 | 0.47 | 0.17 |
| | 4 | 1,725 | 216 | 93.81 | 112 | 49.11 | 0.46 | 0.16 |
| | 4.5 | 130 | 261 | 101.41 | 151 | 66.1 | 0.4 | 0.17 |
| | 5 | 337 | 201 | 67.36 | 122 | 33.01 | 0.37 | 0.14 |
| Chain | 0 | 1,302 | 194 | 91.12 | 107 | 50.13 | 0.42 | 0.17 |
| | 1 | 2,814 | 182 | 87.05 | 95 | 46.42 | 0.45 | 0.16 |

system used. As we mentioned before, Hotwire bookings are done in lower quality hotels than those on Priceline. Similarly, as the table shows, discounts seem to be larger in the higher star-rating segment though the opaque platform used might be contaminating such results. In addition, the distributions of transparent and opaque rates for each star level differ beyond the first moment.[30] Figure 6(a) in the Appendix shows that transparent rates are more spread out than opaque rates. But this is less so for lower quality hotels (2 and 2.5 stars on Figure 6(b)). While the differences in the first moments can explain lower willingness to pay due to the transparent/opaque nature of the product, differences in higher moments can shed light on the nature of competition in each quality submarket. We get back to this point in the next section.

# 4 Results

## 4.1 Discount equation

We start reporting results from the estimation of equation (1). Table 3 presents four specifications where the dependent variable ($Disc_j * 100$) measures the savings by booking in the opaque relative to the transparent channel. All specifications include city-level fixed effects and use robust standard errors clustered at the hotel level. Our preferred specification is given in column (4) where we instrument the competition

[30]The lower rates in 5-star hotels relative to 4.5-star hotels is driven by the fact that all but one of those hotels are located in Las Vegas, a cheaper destination than the average city in our sample.

variable *Comp 0.5mi* using the office rental cost per square feet ($\$psf$).[31] Given that the estimated coefficients do not vary substantially across specification we mainly discuss the results reported in column (4).

The main explanatory variables affecting opaque discounts are platform used and hotel's quality. The coefficient for *Platform* shows that the 9 percentage points difference in unconditional discounts between Priceline and Hotwire (Table 2) is robust to the inclusion of all our covariates. After accounting for non-refundability, the expected discounts are 49.2% for opaque bookings made on Priceline and 39.5% for opaque bookings made on Hotwire. Given the supply-side similarities of these two platforms, we attribute the sign and magnitude of the *Platform* coefficient to the frictions involved in the NYOP bidding system. In particular, the transaction costs imposed to buyers due to the multiple bidding restrictions. Note however that this represents an upper bound of the effect of auction-based, instead of posted-price selling systems. The differences in discounts could also be attributed to the sometimes lower uncertainty faced by Hotwire's users. A hotel room is a bundle of attributes each valued differently by consumers. Some have strong preferences over specific hotel location within the opaque area, others on the amenities available, and other consumers might just care about the hotel identity (e.g., to obtain loyalty rewards). Conditional on buying opaque, and Hotwire releasing more information about the opaque hotel than Priceline (e.g., swimming pool availability), the expected utility for these consumers would be higher if buying through Hotwire. The reason is that buyers can update their—uniform—prior probabilities attached to each hotel in a given opaque area.[32] We argue that the information component of the discount spread across platforms is unlikely to be large. First, as can be observed from Figures 1 and 2 in Appendix A, only in few instances the information regarding an opaque hotel's amenities are different from those expected from a hotel of a given star-rating. Second, as we show below, the opacity level has little effect on the size of the observed discounts.

Better deals are obtained when booking higher quality hotels. Discounts increase on average by 6.62 percentage points for each full unit increase in *Stars*.[33] The boxplot in Figure 3 shows the monotonic relation between predicted discounts, hotel star-

[31] The number of observations in column (4) is reduced because rent data for Las Vegas and Niagara Falls are not available.

[32] Note that the logic is conditional on buying opaque. Buyers with strong preferences over a hotel's amenity would most likely use the transparent channel.

[33] This result is similar across platforms and robust to other non-linear specifications (e.g., star level dummies).

Table 3: Discount Regressions
Dependent variable: $Disc_j * 100$

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Platform (Hotwire=1) | -8.505*** | -8.572*** | -8.604*** | -9.359*** |
| | (0.718) | (0.707) | (0.695) | (0.726) |
| Non–refund | -2.959*** | -2.279* | -2.424** | -2.032* |
| | (0.824) | (0.952) | (0.931) | (0.919) |
| Chain | 0.878 | 0.964 | 0.700 | 0.805 |
| | (1.366) | (1.383) | (1.365) | (1.209) |
| Stars | 4.033*** | 4.772*** | 4.834*** | 6.613*** |
| | (0.879) | (1.051) | (1.100) | (1.256) |
| Airport = 1 | 1.668 | 1.582 | 0.596 | 0.642 |
| | (1.509) | (1.512) | (1.496) | (1.524) |
| Weekend | | 3.212 | 3.383 | 4.640 |
| | | (3.525) | (3.531) | (4.320) |
| Wknd*Stars | | -1.387 | -1.444 | -1.980 |
| | | (0.913) | (0.915) | (1.162) |
| Adv Booking (wks) | | -0.368* | -0.333* | -0.230 |
| | | (0.152) | (0.152) | (0.158) |
| AdvBook$^2$ | | 0.023*** | 0.020** | 0.017* |
| | | (0.006) | (0.007) | (0.007) |
| NonRef*AdvBook | | -0.119 | -0.096 | -0.144 |
| | | (0.129) | (0.127) | (0.127) |
| SPI | | 0.078 | 0.079 | 0.020 |
| | | (0.043) | (0.042) | (0.049) |
| Area size (sqmi) | | | 0.078* | 0.060 |
| | | | (0.036) | (0.042) |
| Dispersion=Var(distance) | | | -1.099** | -0.785* |
| | | | (0.419) | (0.391) |
| Comp 0.5mi | | | -0.077* | -0.177 |
| | | | (0.037) | (0.097) |
| Constant | 41.213*** | 37.987*** | 38.603*** | 34.507*** |
| | (3.677) | (4.399) | (4.665) | (5.081) |
| City fixed effects | Yes | Yes | Yes | Yes |
| R-squared | 0.199 | 0.207 | 0.214 | 0.220 |
| N | 4116 | 4116 | 4109 | 3471 |
| *Underidentification:* | | | | |
| Kleibergen-Paap rk LM statistic | | | | 29.082 |
| Chi-sq (p-val) | | | | 0.000 |
| *Weak identification:* | | | | |
| Kleibergen-Paap rk Wald F statistic | | | | 63.475 |

Robust standard errors clustered at the hotel level in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
2-step GMM with *Comp 5mi* instrumented in Column (4).

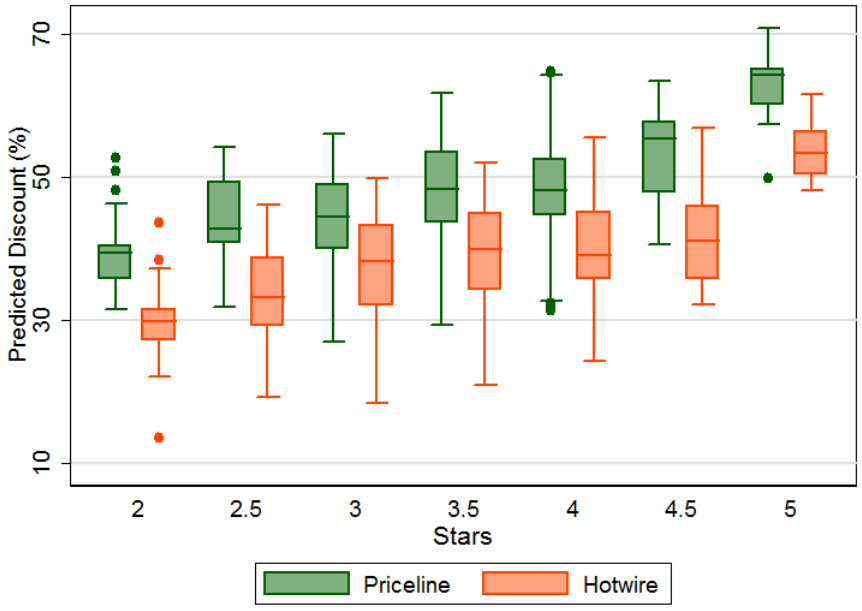
rating and selling platform. Hotel rates in each channel are bounded by consumers' willingness to pay and the marginal cost of providing a room with the exact level being determined by competition in the marketplace. Naturally, all else equal, these bounds increase with a hotel's quality explaining why transparent and opaque rates increase with star ratings. We find that quality affects transparent rates more than opaque rates and this is consistent with the idea that (i) consumers with a lower willingness to pay for quality disproportionately use opaque channels and, (ii) price competition in the opaque market is more intense than in the transparent market. The latter condition is intuitive since the reverse auction implemented by the opaque platform removes product differentiation and therefore, delivers a low wholesale cost to the platform.[34] Note however that condition (ii) alone is not sufficient to explain the larger discounts for higher quality hotels observed in the data. Aside from the wholesale cost of the opaque room, the intermediary must face a more elastic demand in the opaque market for hotels with higher star-rating. Moreover, as the constant gap across platforms in the figure suggests, willingness to pay for quality and transaction costs do not seem to be correlated among the consumer types sorting in the different opaque channels.[35] On a side note, it is interesting that Priceline provides their customers the city-area and star-level where best deals can be obtained. Consistent with our findings, the "best deals" recommended for the bookings in our database are 66% of the time for 4-star hotels and only 0.02% of the time for a 2-star hotel.

Notably, the remaining coefficients in Table 3 are either insignificant or have a very small economic impact on opaque discounts. First, discounts by hotels associated to large chains are not different from those at other hotels. Although the definition used for *Chain* is somewhat subjective, the results do not change when alternative definitions are used (e.g., the ten largest chains in the industry only). Second, it appears that the traditional price discrimination practices in the travel industry, if any, are not different in the transparent and opaque markets. Discounts on check-in dates associated with demand driven by leisure travelers are not different than discounts on other dates. The estimated coefficient for the *Weekend* is positive yet not statistically significantly different from zero. The insignificant coefficient for the interaction term between the variables *Weekend* and *Stars* grants that we are not masking a situation where high and low star-rated hotels have inverse pricing policies regarding weekend and weekdays.

[34] In a truly Bertrand Paradox environment the wholesale cost equals the hotel marginal cost.

[35] We estimated equation 1 including the interaction term between *Platform* and *Stars* and found the coefficient insignificant (not reported and available upon request).

Figure 3: Predicted Opaque and Semi-Opaque Discounts




More importantly, discounts are weakly affected by advanced bookings. A booking done 10 weeks in advance improves the opaque discount by only 1.4 percentage points.[36] This result, combined with the null effect of demand shocks ($SPI$) on opaque discounts and the fact that the average opaque bookings occurs 4 weeks in advance of the check-in date (Figure 2 and Table 2), supports the static price discrimination theories described in Section 2 and rejects the notion that hotels use opaque channels as a last minute resource to dispose of their unsold inventory. Surprisingly, the size of the opaque area on discounts is not statistically significant in column (4). As expected, the coefficient is positive meaning that consumers require larger discounts for bookings that have higher opacity level. But this effect is small: a one standard deviation increase in *Area size* leads to only one tenth of a percentage point increase in opaque discounts. We find that the dispersion in the hotels locations within a given opaque area is associated with lower discounts. All else equal, consumers seem to prefer a more dispersed pattern for hotel locations. We note that the insignificant estimated coefficients for many covariates in the discount regression might be masking considerable action in the transparent and opaque markets if they affect the transparent and opaque rates symmetrically. We therefore postpone the interpretation of the remaining coefficients in Table 3 until after we present the pricing equation results.

[36]The quadratic term used in the specification captures most of the non-linear effects of *Advanced Booking* on Discounts. Figure 7 in Appendix A shows that the residuals from the regression in column (4) are indistinguishable from each other.

The main takeaway from Table 3 is that the discounts offered by hotels fit the predictions from the simple (static) mechanisms of demand segmentation when consumers have heterogeneous preferences among product attributes. As expected, customers booking in opaque markets obtain substantial discounts in exchange for not observing information about their service providers. The discounts in the posted-price platform are lower than in the auction-based platform and reflect the role of transaction costs in segmenting further opaque users. Regardless of platform, opaque discounts increase monotonically with hotel star-rating. Last, the evidence supports the idea that opaque bookings are a tool for price discrimination that is used regularly by hotels and not as a last-minute resource. We now turn to the analysis of the pricing equations to get a better understanding of the mechanisms underlying the discounts results.

## 4.2 Price equations

We now present the results from estimating the pricing equations (2) and (3). Tables 4 and 5 show, for transparent and opaque rates respectively, similar specifications to the ones used when estimated the discount equation. As before, we focus on results reported on column (4) and analyze the estimated coefficients for opaque and transparent rates jointly.

As expected, the major impact on hotel transparent rates—together with destination city—is given by a hotel's quality. Transparent room rates increase by about 44% ($exp(0.369) - 1$) with each additional star in a hotel's quality rating. To put this magnitude in perspective, the largest city fixed effect (Washington, DC) shifts transparent rates by 39% relative to the baseline—and cheapest— city (Atlanta). Consistent with the monotonic relationship between stars and discounts discussed above, opaque rates increase by only 26.5% with hotel's quality.[37] Similarly, the *Platform* coefficient on Table 5 shows that opaque hotel rooms are 16% more expensive when booked in Hotwire's posted-price system rather than through bidding on Priceline.

The price premium for a refundable room is about 10% (Table 4). Refundable rooms offer buyers in the transparent market the opportunity to cancel a booking at no cost before the check-in date. In principle, the premium for refundable rooms in competitive markets would reflect this value option to consumers and therefore converge to zero as the booking date approaches the check-in date (i.e., as consumers'

[37] As with the discount equation estimation, we found no evidence of non-linear effects of star-ratings on transparent or opaque prices.

Table 4: Price Regressions. Transparent Bookings
Dependent variable: $\ln p_j^\tau$

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Non–refund | -0.101*** | -0.104*** | -0.102*** | -0.093*** |
| | (0.018) | (0.021) | (0.021) | (0.022) |
| Chain | 0.019 | 0.024 | 0.026 | 0.023 |
| | (0.028) | (0.028) | (0.029) | (0.029) |
| Stars | 0.366*** | 0.364*** | 0.358*** | 0.369*** |
| | (0.021) | (0.023) | (0.024) | (0.025) |
| Airport = 1 | -0.154*** | -0.151*** | -0.136*** | -0.136*** |
| | (0.033) | (0.033) | (0.032) | (0.031) |
| Weekend | | -0.103 | -0.100 | 0.097 |
| | | (0.090) | (0.090) | (0.081) |
| Wknd*Stars | | 0.023 | 0.022 | -0.044 |
| | | (0.025) | (0.025) | (0.023) |
| Adv Booking (wks) | | 0.007* | 0.006 | 0.009* |
| | | (0.004) | (0.004) | (0.004) |
| AdvBook$^2$ | | -0.000 | -0.000 | -0.000 |
| | | (0.000) | (0.000) | (0.000) |
| NonRef*AdvBook | | -0.001 | -0.001 | -0.003 |
| | | (0.002) | (0.002) | (0.002) |
| SPI | | 0.006*** | 0.006*** | 0.007*** |
| | | (0.001) | (0.001) | (0.001) |
| Comp 0.5mi | | | 0.002* | 0.003 |
| | | | (0.001) | (0.002) |
| Constant | 3.584*** | 3.450*** | 3.453*** | 3.413*** |
| | (0.083) | (0.095) | (0.093) | (0.094) |
| City fixed effects | Yes | Yes | Yes | Yes |
| R-squared | 0.512 | 0.523 | 0.527 | 0.549 |
| N | 4116 | 4116 | 4116 | 3478 |
| *Underidentification:* | | | | |
| Kleibergen-Paap rk LM statistic | | | | 36.378 |
| Chi-sq (p-val) | | | | 0.000 |
| *Weak identification:* | | | | |
| Kleibergen-Paap rk Wald F statistic | | | | 79.631 |

Robust standard errors clustered at the hotel level in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. 2-step GMM with *Comp 5mi* instrumented in Column (4).

uncertainty is resolved). Interestingly, we do not observe such convergence: the coefficient for the interaction term between *Non-refund* and *Adv Booking* is not significantly different from zero (as in Table 3). This result could be consistent with the use of refundable and non-refundable pricing options to screen consumers or exploit frictions in the transparent market.[38] For example, imagine that business-travelers (high valuations and high demand uncertainty) are required by their companies to buy refundable hotel rooms. In such case, the premium for refundable rooms is unlikely to disappear over time.[39]

Anticipating a booking by one week is associated with *higher* transparent and opaque rates (0.9 and 1.2 percent respectively). As Figure 5 in Appendix A shows, hotel rates patterns behave similarly as the booking date approaches the check-in date. The symmetric effect of early bookings explains the insignificant coefficient for *Adv Booking* on opaque discounts. We link the positive and symmetric estimated coefficients for *AdvBook* to traditional peak-load pricing strategies that are not related to demand segmentation. Again, we do not find evidence of price discrimination based on weekend/weekday check-in dates. This is somewhat a surprising result since it is commonly assumed to be an important dimension exploited by the travel industry to segment leisure and business travelers. As we argued before, it is quite possible that hotels do not need to price discriminate within each platform if the existence of opaque channels allows the screening of such consumers.

As in most markets, prices increase as a response to demand shocks. This effect is expected to be more pronounced in those industries where firms are capacity constrained. Our proxy for demand shock *SPI* reveals a significant positive effect on hotel rates. Interpreting the magnitude of the coefficient is not easy since the exogenous variable is a time and cross-section relative index. A one standard deviation increase in *SPI* leads to a 17% increase in hotel room rates. However, it is important to note that demand shocks affect opaque and transparent pricing symmetrically, explaining the null effect of SPI on discounts.

Last, we report the effects of market characteristics on hotel rates. The airport dummy and the number of hotels within half a mile walking distance capture how product differentiation and number of hotels affect prices. Note first that opaque and

[38] Escobari and Jindapon (2012) present a model that introduces the possibility of price discrimination through refundable and non-refundable rates. They also show that refundable and non-refundable airfares converge over time. Watanabe and Moon (2013) link refundable premiums and price discrimination in the airline industry exploiting cross-section variation across markets.

[39] Alternative, assume that demand uncertainty and consumer's willingness to pay are positively correlated and, that consumers' uncertainty is constant and only resolved at the very last minute.

Table 5: Price Regressions. Opaque Bookings
Dependent variable: $\ln p_j^o$

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| Platform (Hotwire=1) | 0.132*** | 0.128*** | 0.128*** | 0.149*** |
| | (0.017) | (0.016) | (0.016) | (0.016) |
| Chain | -0.005 | -0.005 | 0.000 | -0.004 |
| | (0.024) | (0.024) | (0.023) | (0.025) |
| Stars | 0.291*** | 0.278*** | 0.259*** | 0.236*** |
| | (0.022) | (0.024) | (0.025) | (0.028) |
| Airport = 1 | -0.196*** | -0.192*** | -0.151*** | -0.157*** |
| | (0.030) | (0.030) | (0.030) | (0.031) |
| Weekend | | -0.127 | -0.123 | 0.060 |
| | | (0.081) | (0.082) | (0.081) |
| Wknd*Stars | | 0.041 | 0.040 | -0.019 |
| | | (0.022) | (0.022) | (0.022) |
| Adv Booking (wks) | | 0.014*** | 0.012*** | 0.012*** |
| | | (0.003) | (0.003) | (0.003) |
| AdvBook$^2$ | | -0.001*** | -0.000** | -0.000* |
| | | (0.000) | (0.000) | (0.000) |
| SPI | | 0.005*** | 0.005*** | 0.007*** |
| | | (0.001) | (0.001) | (0.001) |
| Area size (sqmi) | | | -0.002** | -0.001 |
| | | | (0.001) | (0.001) |
| Dispersion=Var(distance) | | | 0.003 | -0.008 |
| | | | (0.010) | (0.011) |
| Comp 0.5mi | | | 0.004*** | 0.006** |
| | | | (0.001) | (0.002) |
| Constant | 3.003*** | 2.913*** | 2.963*** | 2.982*** |
| | (0.090) | (0.103) | (0.108) | (0.117) |
| City fixed effects | Yes | Yes | Yes | Yes |
| R-squared | 0.565 | 0.575 | 0.590 | 0.630 |
| N | 4116 | 4116 | 4109 | 3471 |
| *Underidentification:* | | | | |
| Kleibergen-Paap rk LM statistic | | | | 28.806 |
| Chi-sq (p-val) | | | | 0.000 |
| *Weak identification:* | | | | |
| Kleibergen-Paap rk Wald F statistic | | | | 63.330 |

Robust standard errors clustered at the hotel level in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. 2-step GMM with *Comp 5mi* instrumented in Column (4).

transparent rates drop by 16.2 and 14.6 percent in hotels near airports. Aside from hotel quality, customers looking for a room at the airport are usually less concerned about hotel attributes. Moreover, horizontal differentiation by hotels in these markets is less pronounced than in other parts of the city. We should therefore expect tighter competition in the transparent market that pushes down rates in the opaque market.

We find little or none effect of the number of hotels (*comp 0.5mi*) on hotel rates. An additional hotel in the market does not affect transparent rates and increases opaque rates by almost 0.5%. Surprisingly, the coefficients in column (4) are larger once we instrument the number of firms with the office rent per square feet. Two explanations are possible. On the one hand, there is the possibility that our instrument is correlated with some unobservable of the local market not captured by the city fixed-effects or the other covariates. On the other hand, this could be evidence of agglomeration economies, as suggested by (Kalnins, 2006, p.210). In any case, we are mainly interested on the relative impact of competition on opaque rates.

The number of hotels can have two effects on opaque rates. On the one hand, more hotels mean more competition and therefore lower rates. This effect is present in the transparent market although, since product differentiation is absent in the opaque market, competition takes a very different shape. On the other hand, more hotels on a given area affect expected utility for opaque channel users. This latter effect depends on the particular consumer preference. For example, someone attending a conference at hotel X might find the opaque market more attractive if there are, other things equal, more hotels within 0.5mi of the conference hotel. The results from column (4) suggest the latter effect dominates and more hotels in a given area increase opaque rates relative to transparent rates. In any case, the coefficient is very small and therefore not significantly different from zero in the estimation of the discount equation. The variables associated with opacity level (area size and dispersion) in Table 5 are not significant and present a puzzle to our theoretical predictions. The predicted effect for Dispersion is ambiguous but we expected it to be relevant in explaining opaque prices. More importantly, we would expect bookings in larger opaque areas to be associated with more uncertainty and therefore a lower opaque price. Taken together, our findings suggest that consumers require a large discount to be willing to buy a lottery instead of a known product. However, conditional on buying opaque, these consumers are not very sensitive to the opacity level. In other words, variables that affect the distribution of probabilities over the likely outcomes do not change consumers' willingness to pay.

# 5 Conclusions

We used a unique dataset with opaque transactions to document stylized facts and provide insights on the economics of opaque selling. Our results are consistent with hotels using selling channels to segment their demand based on consumers' preferences over hotel attributes and transaction costs. Demand shocks and advance booking affect pricing strategies symmetrically across platforms suggesting that opaque intermediaries are not used as a last minute resource selling channel by hotels. Moreover, there is little evidence of price discrimination within a platform. This is consistent with opaque platforms fully segmenting consumer types and leaving no room for within-platform price discrimination.

# References


Abrate, G., Fraquelli, G. and Viglia, G. (2012). Dynamic pricing strategies: Evidence from european hotels, *International Journal of Hospitality Management* **31**(1): 160–168.

Courty, P. and Liu, W. (2013). Opaque selling: Static or inter-temporal price discrimination?, *Working Paper* .

Dai, W., Jin, G. Z., Lee, J. and Luca, M. (2012). Optimal aggregation of consumer ratings: An application to yelp.com, *Working Paper 18567*, National Bureau of Economic Research.
**URL:** *http://www.nber.org/papers/w18567*

Deneckere, R. J. and McAfee, R. P. (1996). Damaged goods, *Journal of Economics and Management Strategy* **5**(2): 149174.

Escobari, D. and Jindapon, P. (2012). Price discrimination through refund contracts in airlines, *Working Paper* .

Fay, S. (2008). Selling an opaque product through an intermediary: The case of disguising ones product, *Journal of Retailing* **84**(1): 59–75.

Fay, S. and Laran, J. (2009). Implications of expected changes in the seller's price in name-your-own-price auctions, *Management Science* **55**(11): 1783–1796.

Goldmanis, M., Hortasu, A., Syverson, C. and Emre, n. (2010). E-commerce and the market structure of retail industries*, *The Economic Journal* **120**(545): 651–682.
**URL:** *http://dx.doi.org/10.1111/j.1468-0297.2009.02310.x*

HSMAI Foundation (2012). Distribution channel analysis. a guide for hotels.
**URL:** *http://www.hsmai.org*

Jerath, K., Netessine, S. and Veeraraghavan, S. K. (2010). Revenue management with strategic customers: Last-minute selling and opaque selling, *Management Science* **56**(3): 430–448.

Kalnins, A. (2006). Markets: The us lodging industry, *The Journal of Economic Perspectives* **20**(4): 203–3A.

Kimes, S. E. (2002). A retrospective commentary on discounting in the hotel industry: A new approach, *The Cornell Hotel and Restaurant Administration Quarterly* **43**(4): 92–93.

Lewis, M. (2008). Price dispersion and competition with differentiated sellers*, *The Journal of Industrial Economics* **56**(3): 654–678.
**URL:** *http://dx.doi.org/10.1111/j.1467-6451.2008.00348.x*

Phillips, R. (2005). *Pricing and revenue optimization*, Stanford Business Books.

Shapiro, C. and Varian, H. R. (2000). Versioning: The smart way to sell information, *Harvard Business Review* **76**(6): 10614.

Shapiro, D. and Shi, X. (2008). Market segmentation: The role of opaque travel agencies, *Journal of Economics & Management Strategy* **17**(4): 803–837.

Talluri, K. T. and Van Ryzin, G. J. (2005). *The theory and practice of revenue management*, Vol. 68, Springerverlag Us.

Tappata, M. (2011). Strategic vertical market structure with opaque products, *Working Paper* .

Watanabe, M. and Moon, S. (2013). Refundability and price: Empirical analysis on the airline industry, *Available at SSRN 1681337* .

# Appendices

## A Additional Figures and Tables

Figure 1: Hotwire's Secret Hot Rate

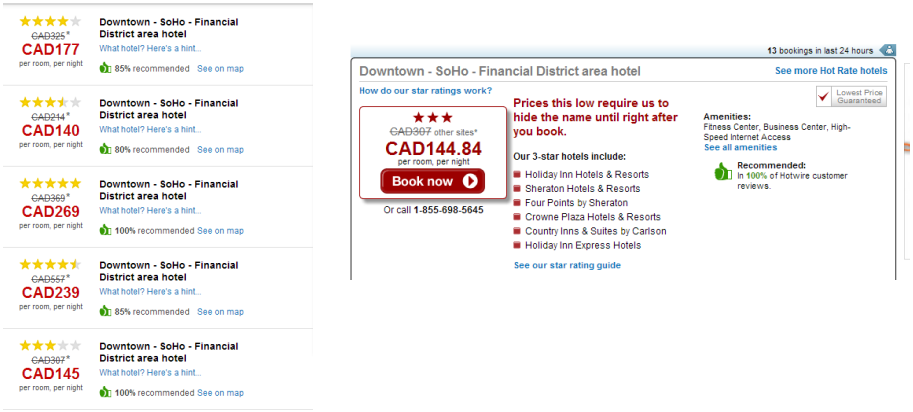



Figure 2: Priceline's Name Your Own Price

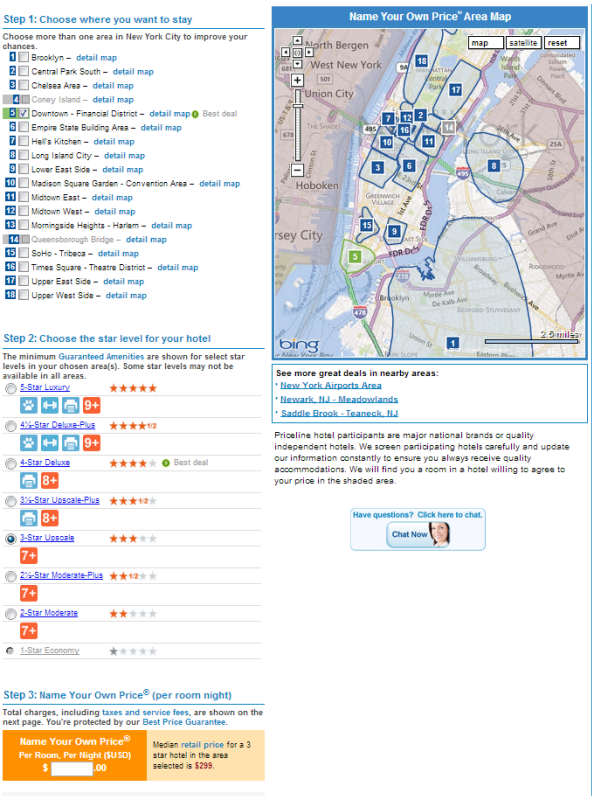
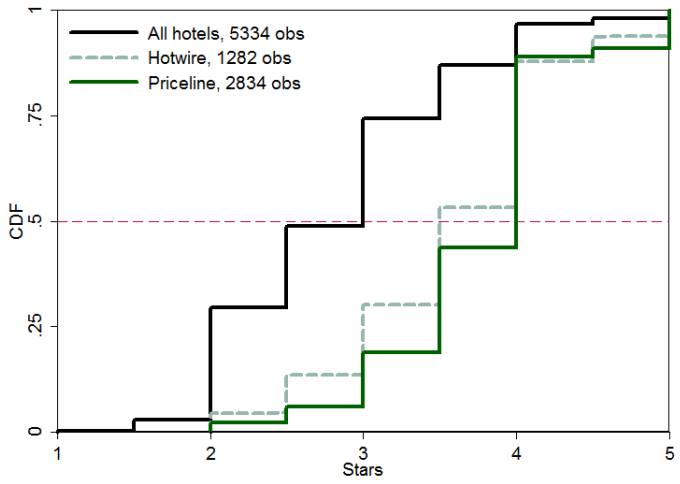
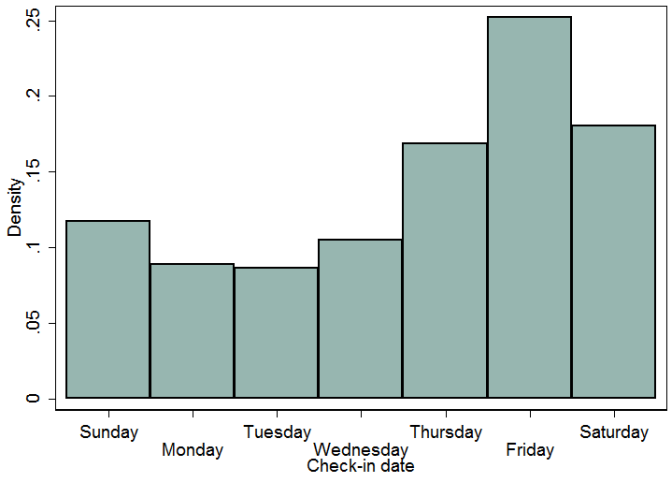


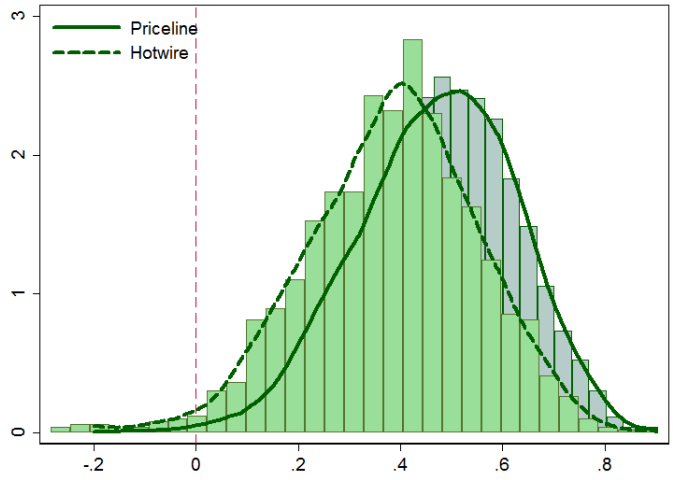
Figure 3: Descriptive Statistics



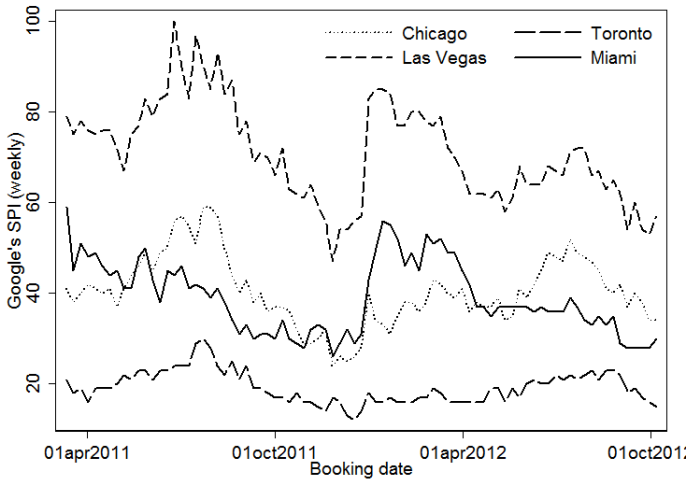

(a) Quality distribution: opaque bookings and hotels




(b) Check-in day of the week




(c) Discounts distribution by platform




(d) Search Popularity Index

Figure 4: Opaque discounts and hotel rates by city

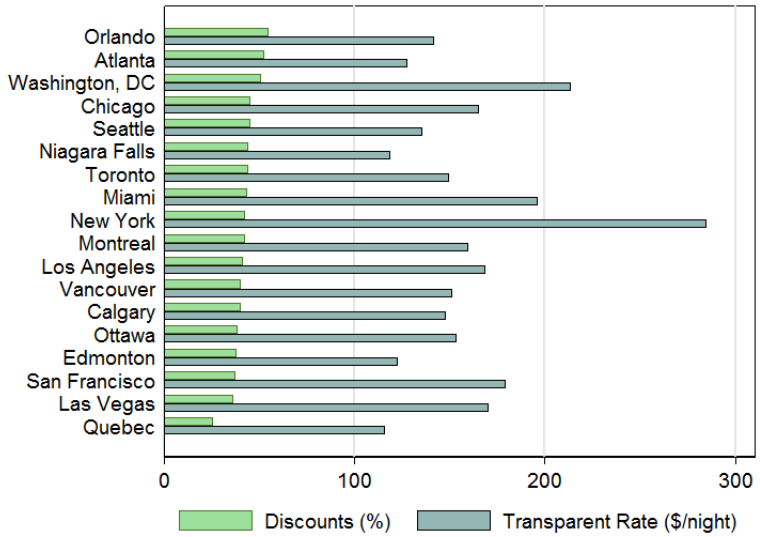
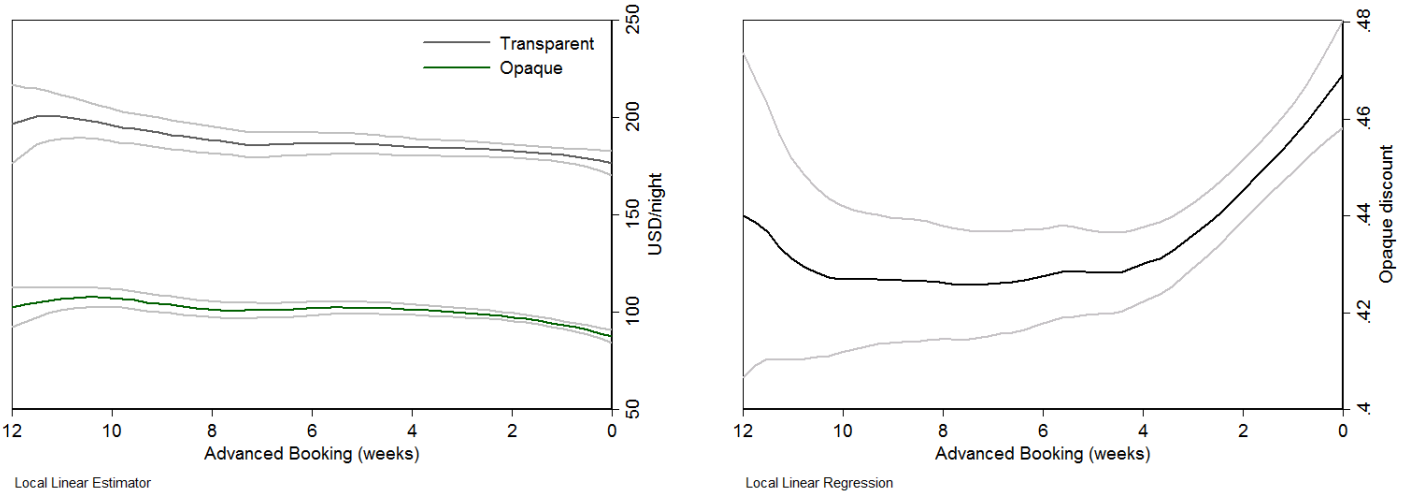


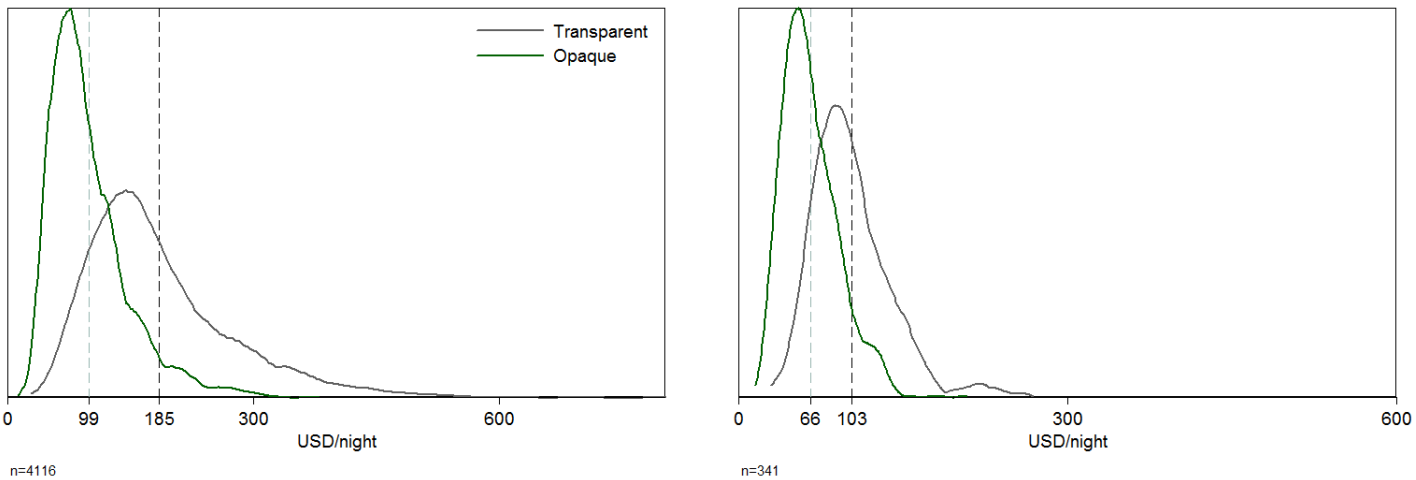

Figure 5: Opaque discounts and hotel rates dynamics




(a) Rates

(b) Discounts

Figure 6: Opaque and transparent room rate densities




(a) All observations

(b) Hotels with stars $< 3$

Table 1: Price Regressions by Platform
Dependent variables: $\ln p_j^\tau$ and $\ln p_j^o$

| | All | | Hotwire | | Priceline | |
|---|---|---|---|---|---|---|
| | $\ln p_j^\tau$ | $\ln p_j^o$ | $\ln p_j^\tau$ | $\ln p_j^o$ | $\ln p_j^\tau$ | $\ln p_j^o$ |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Non–refund | -0.093*** | | -0.081* | | -0.097*** | |
| | (0.022) | | (0.037) | | (0.024) | |
| Platform (Hotwire=1) | | 0.149*** | | | | |
| | | (0.016) | | | | |
| Chain | 0.023 | -0.004 | 0.054 | -0.013 | -0.005 | -0.009 |
| | (0.029) | (0.025) | (0.040) | (0.033) | (0.033) | (0.028) |
| Stars | 0.369*** | 0.236*** | 0.353*** | 0.235*** | 0.365*** | 0.236*** |
| | (0.025) | (0.028) | (0.034) | (0.040) | (0.033) | (0.032) |
| Airport = 1 | -0.136*** | -0.157*** | -0.074 | -0.150*** | -0.164*** | -0.158*** |
| | (0.031) | (0.031) | (0.041) | (0.037) | (0.033) | (0.034) |
| Weekend | 0.097 | 0.060 | 0.093 | -0.005 | 0.079 | 0.109 |
| | (0.081) | (0.081) | (0.121) | (0.120) | (0.113) | (0.101) |
| Wknd*Stars | -0.044 | -0.019 | -0.042 | -0.008 | -0.039 | -0.029 |
| | (0.023) | (0.022) | (0.034) | (0.034) | (0.031) | (0.027) |
| Adv Booking (wks) | 0.009* | 0.012*** | 0.009 | 0.009 | 0.012* | 0.013*** |
| | (0.004) | (0.003) | (0.006) | (0.005) | (0.005) | (0.004) |
| AdvBook$^2$ | -0.000 | -0.000* | -0.000 | -0.000 | -0.000 | -0.000 |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| NonRef*AdvBook | -0.003 | | -0.004 | | -0.004 | |
| | (0.002) | | (0.004) | | (0.003) | |
| SPI | 0.007*** | 0.007*** | 0.005* | 0.005* | 0.008*** | 0.008*** |
| | (0.001) | (0.001) | (0.002) | (0.002) | (0.001) | (0.001) |
| Area size (sqmi) | | -0.001 | | 0.000 | | -0.002* |
| | | (0.001) | | (0.002) | | (0.001) |
| Dispersion=Var(distance) | | -0.008 | | -0.023 | | -0.001 |
| | | (0.011) | | (0.017) | | (0.012) |
| Comp 0.5mi | 0.003 | 0.006** | 0.008*** | 0.008** | 0.000 | 0.004* |
| | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| Constant | 3.413*** | 2.982*** | 3.418*** | 3.197*** | 3.453*** | 2.968*** |
| | (0.094) | (0.117) | (0.143) | (0.164) | (0.121) | (0.134) |
| City fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.549 | 0.630 | 0.537 | 0.582 | 0.552 | 0.655 |
| N | 3478 | 3471 | 1087 | 1087 | 2391 | 2384 |
| *Underidentification:* | | | | | | |
| Kleibergen-Paap rk LM statistic | 36.378 | 28.806 | 29.558 | 23.650 | 24.846 | 19.231 |
| Chi-sq (p-val) | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| *Weak identification:* | | | | | | |
| Kleibergen-Paap rk Wald F statistic | 79.631 | 63.330 | 61.805 | 46.572 | 64.557 | 48.995 |

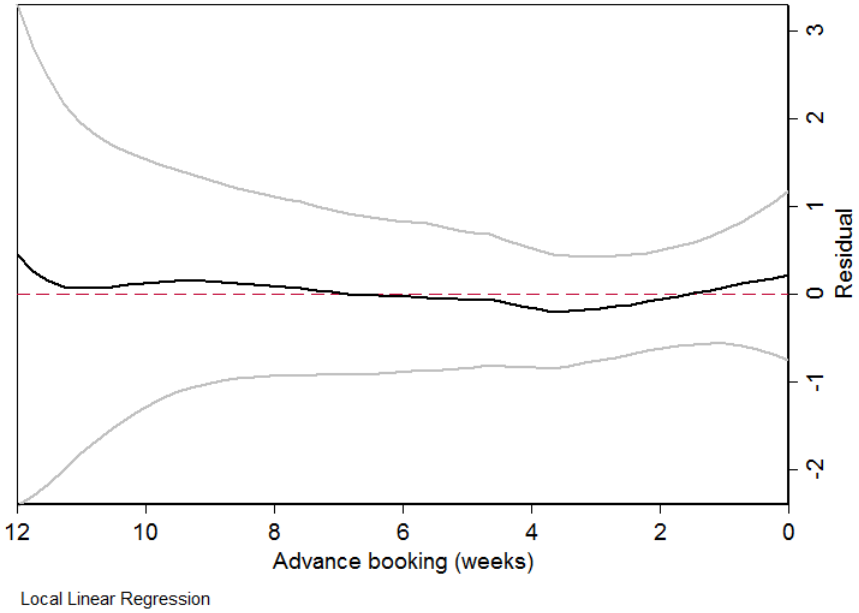
2-step GMM with *Comp 5mi* instrumented in all columns. Robust standard errors clustered at the hotel level in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 2: Price Regressions by Platforms (3SLS)
Dependent variables: $\ln p_j^\tau$ and $\ln p_j^o$

| | All | | Hotwire | | Priceline | |
|---|---|---|---|---|---|---|
| | $\ln p_j^\tau$ | $\ln p_j^o$ | $\ln p_j^\tau$ | $\ln p_j^o$ | $\ln p_j^\tau$ | $\ln p_j^o$ |
| | (1) | (2) | (3) | (4) | (5) | (6 ) |
| Non–refund | -0.066*** | | -0.064** | | -0.068*** | |
| | (0.013) | | (0.023) | | (0.016) | |
| Platform (Hotwire=1) | | 0.163*** | | | | |
| | | (0.009) | | | | |
| Chain | 0.016 | 0.000 | 0.044 | -0.011 | -0.011 | -0.007 |
| | (0.013) | (0.012) | (0.023) | (0.020) | (0.017) | (0.014) |
| Stars | 0.368*** | 0.243*** | 0.352*** | 0.238*** | 0.365*** | 0.244*** |
| | (0.014) | (0.012) | (0.022) | (0.021) | (0.017) | (0.015) |
| Airport = 1 | -0.132*** | -0.152*** | -0.080* | -0.143*** | -0.156*** | -0.155*** |
| | (0.018) | (0.016) | (0.039) | (0.036) | (0.021) | (0.018) |
| Weekend | 0.097 | 0.061 | 0.093 | -0.023 | 0.079 | 0.124 |
| | (0.069) | (0.062) | (0.109) | (0.101) | (0.091) | (0.079) |
| Wknd*Stars | -0.044* | -0.019 | -0.042 | -0.003 | -0.039 | -0.033 |
| | (0.019) | (0.017) | (0.031) | (0.028) | (0.025) | (0.022) |
| Adv Booking (wks) | 0.008** | 0.012*** | 0.008 | 0.009 | 0.010** | 0.013*** |
| | (0.003) | (0.003) | (0.005) | (0.005) | (0.004) | (0.003) |
| AdvBook$^2$ | -0.000 | -0.000** | -0.000 | -0.000 | -0.000 | -0.000* |
| | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) | (0.000) |
| NonRef*AdvBook | -0.003 | | -0.002 | | -0.005 | |
| | (0.002) | | (0.003) | | (0.002) | |
| SPI | 0.007*** | 0.007*** | 0.005** | 0.005*** | 0.008*** | 0.008*** |
| | (0.001) | (0.001) | (0.002) | (0.001) | (0.001) | (0.001) |
| Area size (sqmi) | | -0.001* | | -0.001 | | -0.002** |
| | | (0.000) | | (0.001) | | (0.001) |
| Dispersion=Var(distance) | | 0.003 | | -0.006 | | 0.008 |
| | | (0.005) | | (0.008) | | (0.006) |
| Comp 0.5mi | 0.004*** | 0.007*** | 0.007*** | 0.007*** | 0.002 | 0.006*** |
| | (0.001) | (0.001) | (0.002) | (0.002) | (0.001) | (0.001) |
| Constant | 3.388*** | 2.934*** | 3.421*** | 3.180*** | 3.411*** | 2.915*** |
| | (0.059) | (0.055) | (0.099) | (0.093) | (0.074) | (0.068) |
| City fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| R-squared | 0.548 | 0.626 | 0.537 | 0.582 | 0.554 | 0.652 |
| N | 3471 | | 1087 | | 2384 | |

Generalized Least Squares with *Comp 5mi* instrumented and correlated disturbances across equations. Standard errors in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Figure 7: Advance Bookings and non-linear effects



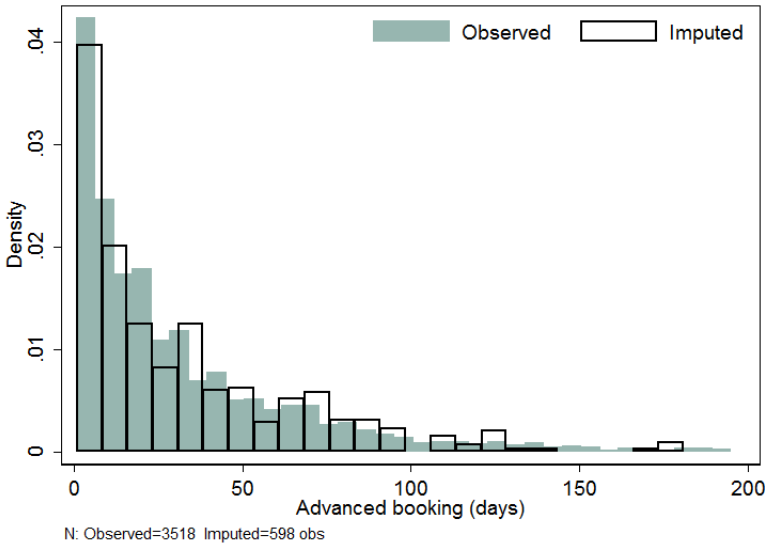
# B Dataset Construction

## B.1 Hotel rates

*Opaque Data:* We collected the opaque bookings from Betterbidding.com. This website has a forum for each city-platform pair where users discuss strategies and report information on successful bookings. Figure 9 shows a snapshot of the Los Angeles–Hotwire forum with 6 different threads. We ignored the "help" related posts and parsed the posts related to successful transactions (i.e. those inside the red boxes). Each one of these post includes a hotel name, opaque area, average rate (net of taxes and fees), check-in date, and length of stay. The users do not report booking dates. However we approximate them with the date the thread was originated.[1]

We dropped observations with booking dates that were made more than 200 days before the check-in date. In many cases (about 14% of the observations), the report dates exceeded the booking date. We performed multiple imputation of booking dates using Stata's ice command and Figure 8 shows the histogram for the observed and the first imputation of advanced booking days. The regressions analysis was done using the first sample of multiple imputation and the complete case sample (i.e. ignoring observations with missing values).

Figure 8: Imputed and observed advance bookings

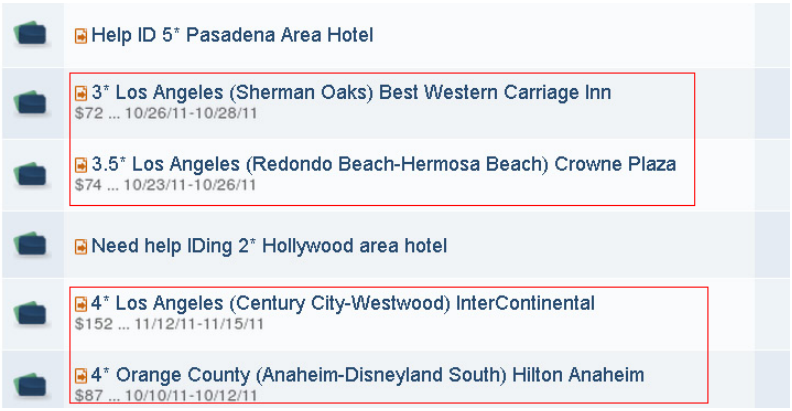



Although hotels price multiresource inventory as the sum of daily rates (Talluri and Van Ryzin, 2005, p. 525), we dropped transactions where the length of stay is over 15 nights because of possible volume discount policies.[2]

[1] Although post dates are not shown in the forum's interface, we can extract it from the html source code.

[2] We analyzed transparent pricing in terms of length of stay (1, 2 and 3 nights) for a sample of 100 hotels in NYC during 6 months and found no evidence of volume discounts. However, it is possible that the additive pricing observed in the transparent market does not apply to opaque bookings.

Figure 9: Betterbidding.com Snapshot: Hotwire–LA Forum




*Matching Opaque and Transparent Data:* The transparent data from Hotels.com includes hotel room rates for one-night stay only. Thus, to match the opaque transactions with the transparent prices we create an average night rate using all the transparent hotel rates for the length of stay of the opaque transaction. For example, consider the Best Western Carriage Inn booking shown in Figure 9. This 2-night booking was made 14 days before the check-in date for an average rate of $72/night.[3] From the transparent data, the 1-night rates posted by the Best Western Carriage Inn on Oct 12 is $92/night for check-in dates Oct 26 and/or Oct 27 and therefore, the opaque discount for this transaction on Hotwire is 22%.

In some cases, the transparent rate for a given hotel was not available for the same booking and check-in dates of the opaque transaction. In this case, we fix the check-in date and look for the closest booking date for which the hotel posted a transparent rate. The booking dates on half of our matched observations differ by less than 2 days. This difference increases to 13 days in the third quartile of the distribution.

*Hotels' quality:* Each platform has their own star rating system and they are remarkably similar (possibly due to legal issues). However, since consumers report the name and stars of their hotels, we use Hotels.com rating. In the few cases (22 out of 924) where a hotel changed its rating during the time-frame of our study we use the star rating in place most of the time.

*Taxes and Booking Fees:* We record opaque and transparent rates net of taxes, fees and commissions. In the case of Hotwire, these are the "base prices" that consumers observe first when searching for deals. Similarly, the opaque rates posted in the Betterbidding.com forum refer to the bids placed on Priceline or the displayed rate by Hotwire.[4] The three platforms collect sales and local taxes. They aggregate taxes with service fees in one line that the buyer observes before confirming a booking. We could not find a statement from each company with the detailed list of taxes and

[3]The post can be seen at http://www.betterbidding.com/index.php?showtopic=144289&

[4]Forum moderators explicitly ask members to post the prices paid before taxes and fees.

Table 3: Opaque transactions matched by cities

| | | Hotels | |
|---|---|---|---|
| City | Obs | Opaque | Transparent |
| Atlanta | 195 | 65 | 496 |
| Calgary | 14 | 10 | 85 |
| Chicago | 458 | 115 | 532 |
| Edmonton | 9 | 3 | 79 |
| Las Vegas | 561 | 44 | 194 |
| Los Angeles | 295 | 84 | 555 |
| Miami | 271 | 85 | 370 |
| Montreal | 89 | 26 | 172 |
| New York | 680 | 123 | 661 |
| Niagara Falls | 31 | 10 | 168 |
| Orlando | 335 | 72 | 407 |
| Ottawa | 26 | 10 | 64 |
| Quebec | 7 | 4 | 100 |
| San Francisco | 255 | 51 | 391 |
| Seattle | 226 | 49 | 253 |
| Toronto | 202 | 48 | 209 |
| Vancouver | 150 | 32 | 155 |
| Washington, DC | 312 | 89 | 444 |
| Total | 4,116 | 920 | 5,335 |

Opaque transactions observed in a given city (Column 1) were booked at one of the "opaque hotel" in Column 2. Column 3 reports the total number of hotels in a given city (source: Hotels.com).

processing fees.[5] We have sampled quotes for each city in each platform to asses the relative differences in commissions across platforms. While the average "taxes and fees" charged by Hotels.com is around 15%, the opaque platforms add about a 24% to the base rate. Priceline's "taxes and service fees" and Hotwire's "tax recovery charges & fees" average 26% and 22% respectively.

## B.2 Hotels and Market characteristics

The transparent data includes rates for 5,335 hotels (924 hotels for which we have opaque transactions) in the 18 cities studied. We collected all characteristics from hotel name, star rating, address, room type. We also calculated walking distances among hotels based on google's driving distance application. Avoiding euclidean distances is important in areas where rivers, highways and airports separate hotels.

## B.3 Office rent data

We use the office rent cost to instrument market structure metrics. The source of our office rent data is Jones Lang LaSalle (JLL), a commercial real estate firm. JLL's quarterly "Office Statistics" reports rent values for markets (usually CBD and suburbs) and submarkets in large metropolitan areas.[6] The report displays the asking rent per square foot for offices of three quality levels (A,B and C).

We recorded the asking rent values reported for the third quarter of 2011. Some submarkets report rent values for only two of the three office qualities. Thus, we created three values by recording the highest, lowest and average in each submarket. We then assigned to each hotel a rent value according to JLL's market and sumbarket boundaries and the hotel quality rating: *Rent Max* if *Stars*>3, *Rent Avg* if *Stars*=3 and *Rent Min* if *Stars*<3). Table B.3 shows the summary statistics for the original and transformed rent data.

## B.4 Google Trends SPI

Google Trends tracks the search volume for the most popular queries used on Google Search. Google constructs a search popularity index (SPI) for each query by *normalizing* and *scaling* the absolute search activity. We collected the SPI values for queries similar to "hotels in Chicago". First, given the region and time frame specified by the user (e.g.: Worldwide, 2009-2011) the share of the total search volume at each point in time is calculated.[7] The scaling process is an extra step required to compare keywords. Consider the keyword V="hotels in Vancouver" searched *worldwide* between January 2009 and January 2011. Weekly frequency is available therefore

[5]From Priceline's website, "The charge for Taxes and Fees varies based on a number of factors including, without limitation, the amount we pay the hotel and the location of the hotel where you will be staying, and may include profit that we retain".

[6]Las Vegas and Niagara Falls are not included.

[7]The frequency available varies with the popularity of the query.

Table 4: Variables definition

| **Variable** | **Definition** |
|---|---|
| Transp Rate ($p_j^\tau$) | Transparent rate posted on Hotels.com |
| Opaque Rate ($p_j^o$) | Opaque rate from Hotwire or Priceline |
| Discount ($D_j$) | Rate discount for opaque room $D_j = 1 - p_j^o/p_j^\tau$ |
| Platform | Dummy variable =1 if opaque booking is done on Hotwire (=0 if Priceline) |
| Adv Booking (wks) | Difference between Check-in and Booking dates (in weeks) |
| Stars | Hotel quality. Values are $\{2, 2.5, ., 5\}$ |
| Weekend | Dummy variable = 1 if check-in date is Friday or Saturday night or a US–Can Holiday |
| Chain | Dummy variable for hotels belonging to chains |
| Non-refund | Dummy variable = 1 if transparent room rate is non-refundable |
| $SPI_{t,c}$ | Google Trends Search Popularity Index for city $c$ and week $t$ (January 2009 – January 2011) |
| Airport | Dummy variable =1 if opaque area is near airport |
| Area size | Opaque areas in Figures 11—13 in squared miles |
| Dispersion | Variance of the walking distance between pairs of hotels located in opaque area |
| Comp Xmi | Number of hotels within X miles (walking distance) |
| Comp sim Xmi | Number of hotels within x miles (walking distance) |
| Comp_same Xmi | Number of hotels of same quality within x miles (walking distance) |
| Comp_sim Xmi | Number of hotels of similar ($\pm$ 0.5 star) quality within x miles (walking distance) |
| Rent X | Asking rent \$psqf by submarket and quality levels $X = \{A, B, C\}$. Source: Jones Lang LaSalle's "Office Statistics" quarterly reports |

Table 5: Asking Rent ($ per square foot, 2011Q3)

| Variable | N | Mean | SD | Med | Min | Max |
|---|---|---|---|---|---|---|
| Rent A | 3,361 | 39.15 | 15.63 | 35.99 | 2.08 | 78.86 |
| Rent B | 3,422 | 31.31 | 10.9 | 31.34 | 1.75 | 53.55 |
| Rent C | 1,341 | 27.84 | 7.06 | 27.09 | 13.47 | 44.04 |
| Rent Max | 3,478 | 39.02 | 15.49 | 35.99 | 2.08 | 78.86 |
| Rent Avg | 3,478 | 33.92 | 12.86 | 31.64 | 1.91 | 66.21 |
| Rent Min | 3,478 | 29.14 | 10.86 | 27.06 | 1.75 | 53.55 |

Source: Jones Lang LaSalle, joneslanglasalle.com.

$t = \{1, 2, .., T = 109\}$. Let $S_j^V$ denote the absolute number of V searches on week j. The normalized search index is

$$I_j^V = \frac{100 S_j^V}{\sum_{t=1}^{109} S_t^V}$$

Naturally, $I_t^V$ is bounded by 0 and 100 and reflects variation over time for a search query. When we use city fixed effects, $I_t^V$ represents changes over time in the popularity of the city.

The second step involves scaling the index such that, when the set of keywords searched is larger than one ($V = \{$"$Vancouver$", "$Chicago$", "$LosAngeles$"), only one keyword–week pair gets a score of 100:

$$SPI_j^{V(1)} = I_j^{V(1)} \frac{100}{max\{I_t^V\}} = \left( \frac{S_j^V(1)}{\sum_{t=1}^{109} S_t^{V(1)}} \right) \left( \frac{100}{max\{I_t^V\}} \right)$$

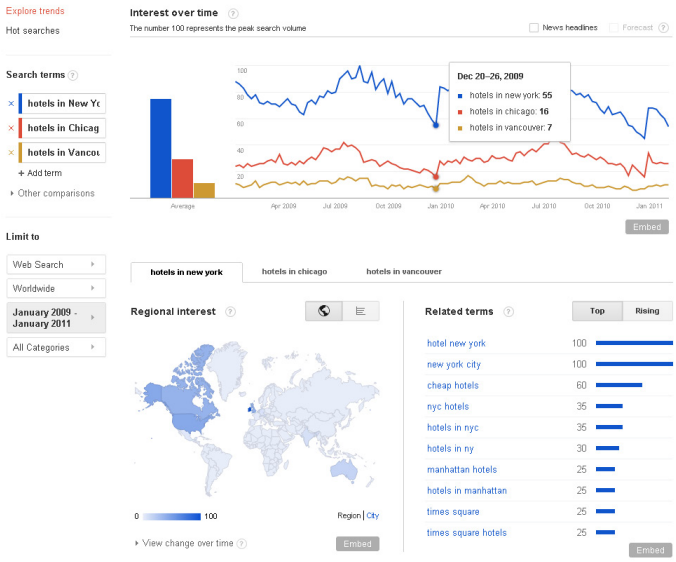
Google trends allows popularity comparisons for different queries at the same time (see Figure 10). All keywords are pooled and the search volumes are normalized by the highest share assuming $t = TxN$ where $N$ is the number of keywords in $V$. That way, the SPI reveals the relative likelihood of a random search among those keywords.[8] In terms of our data, it captures the fixed effect of a city and its city*time effect. We collected Google Trends data for the period between March 6 2011 and Sept 30 2012.

## B.5 Market areas by selling platform

We used Priceline's website to replicate and map the opaque areas in each city. Only two of the opaque hotels in our sample did not fall in the areas shown by Priceline

[8] When downloading data for many queries at a time, the single keyword index can be recovered by dividing a week's score over the max score and multiplying by 100.

Figure 10: Google Trends' snapshot




on May 2013.[9] We used this information to calculate the [10]

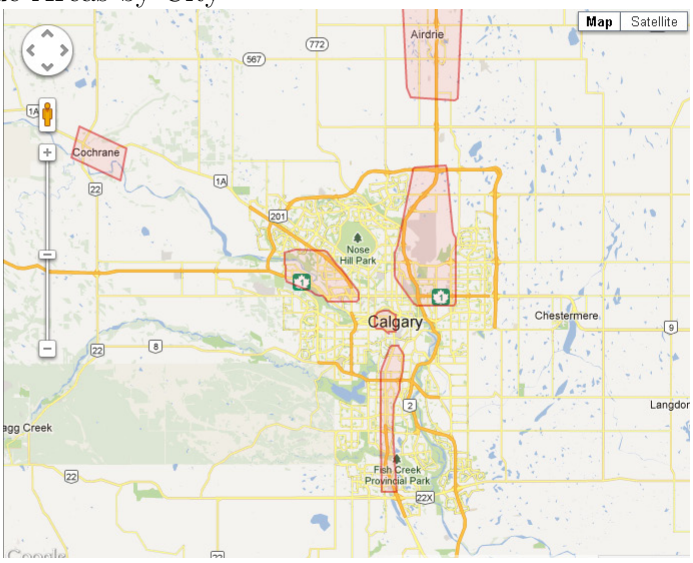
[9]Since the area definitions could have changed from 2012, we allocated these hotels to the closest areas.

[10]We did the same process for Hotwire in Los Angeles and noticed that the areas are almost identical except for the name given by each platform

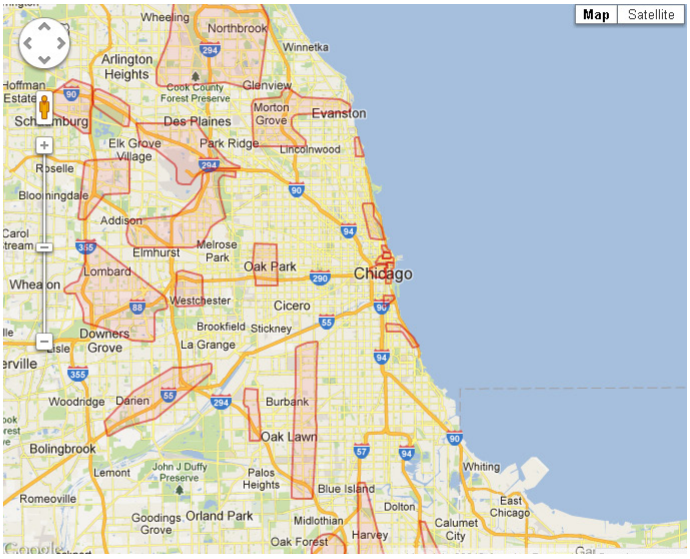
Figure 11: Opaque Areas by City



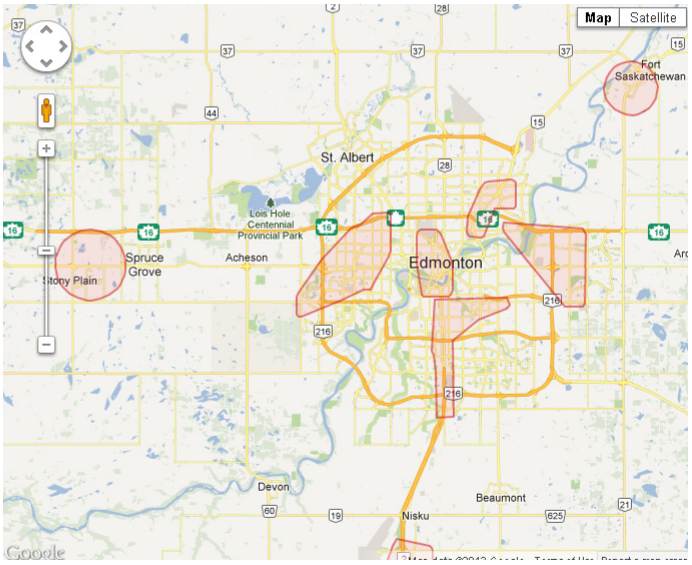

Atlanta



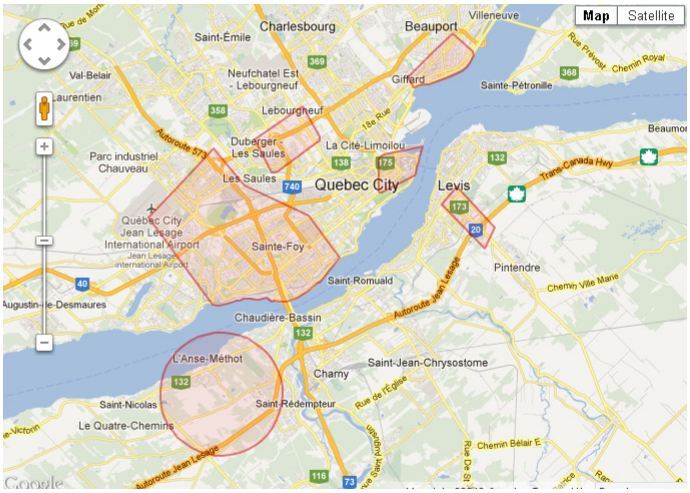

Calgary



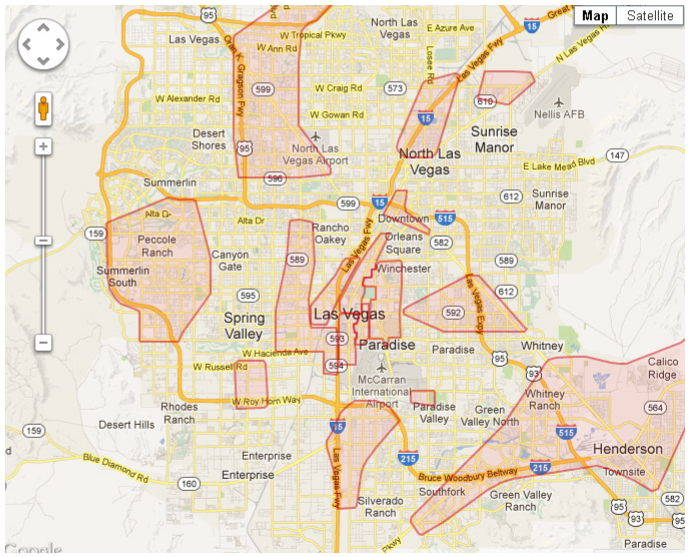

Chicago




Edmonton

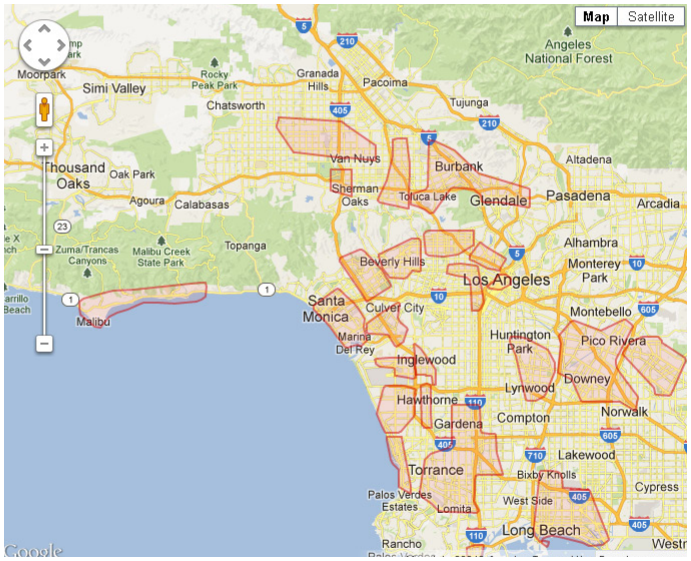



Quebec

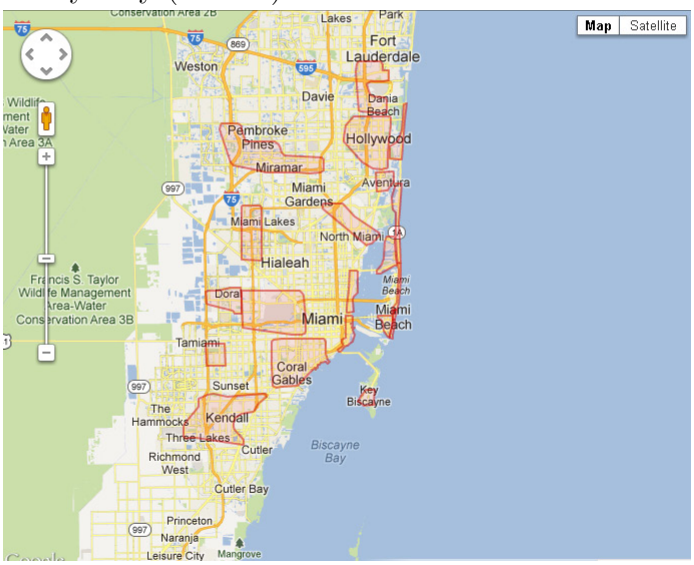



Las Vegas

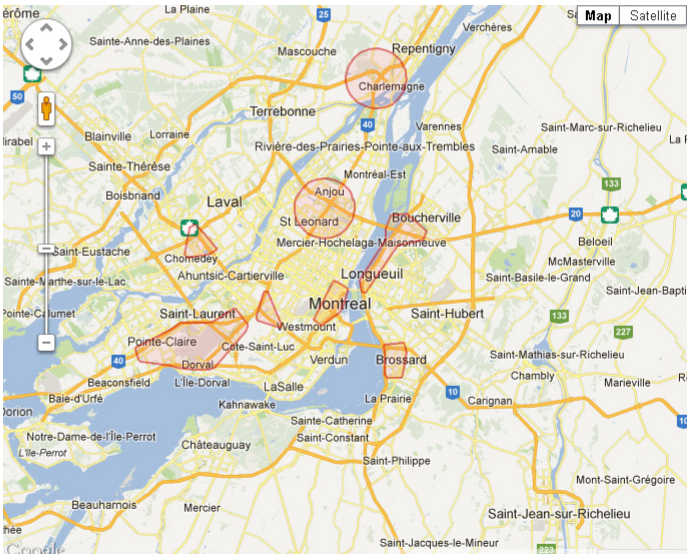
Figure 12: Opaque Areas by City (Cont'd)



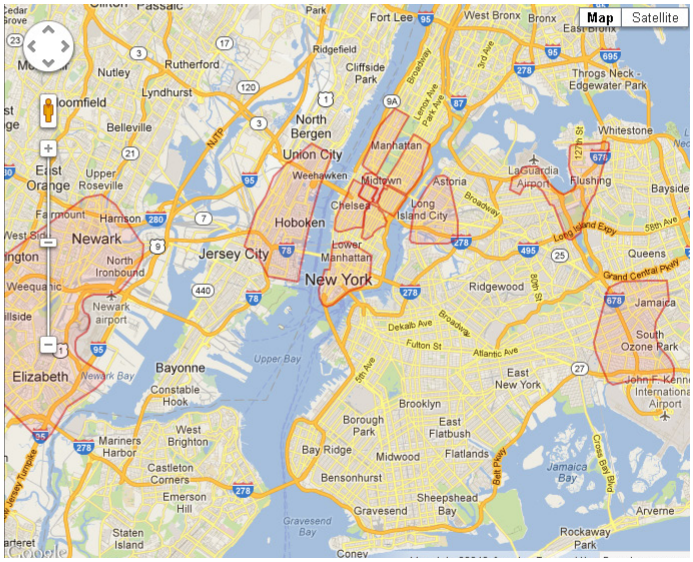

Los Angeles




Miami



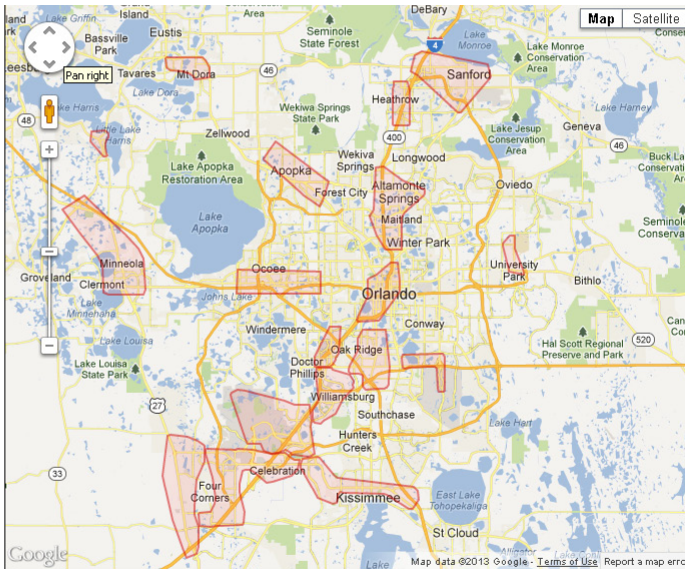

Montreal




New York




Niagara Falls

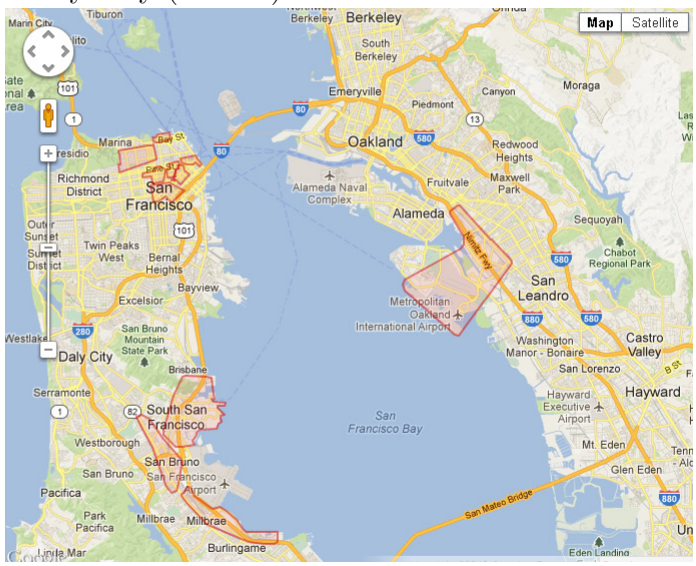



Orlando

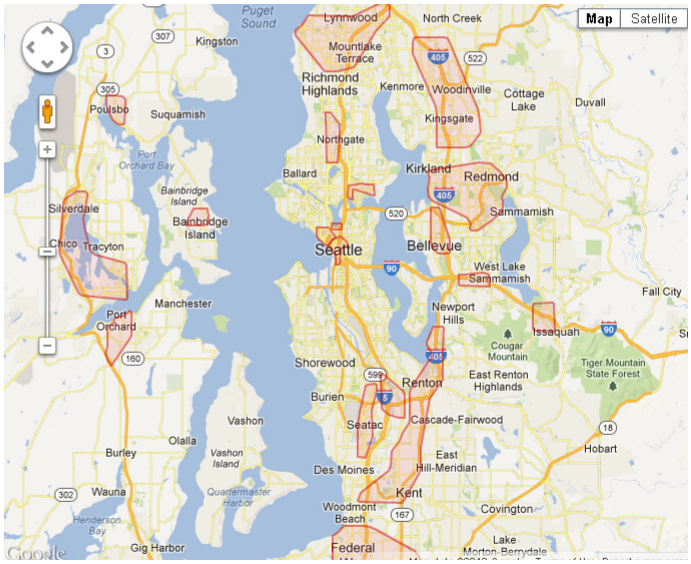
Figure 13: Opaque Areas by City (Cont'd)



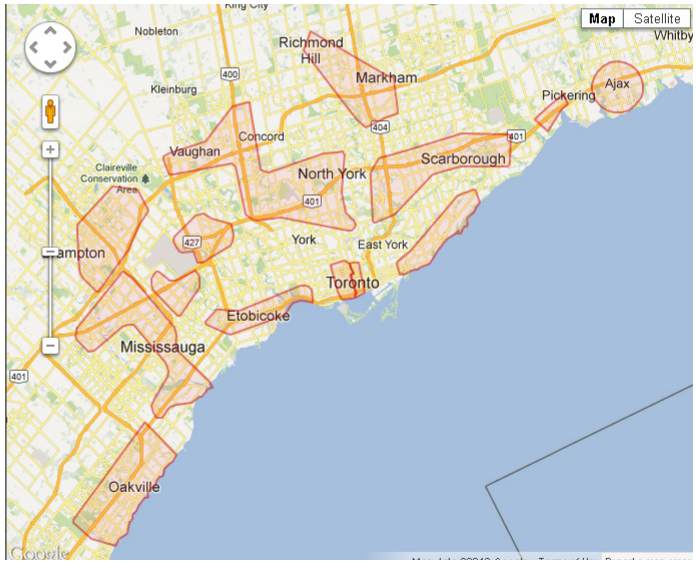

Ottawa



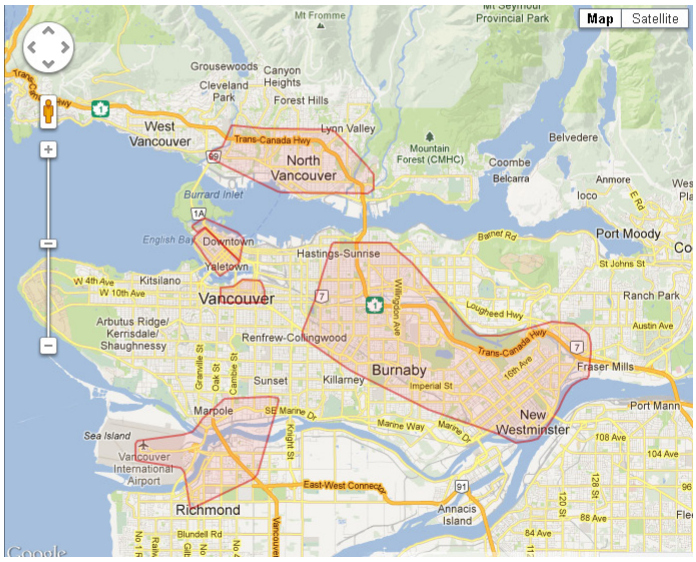

San Francisco



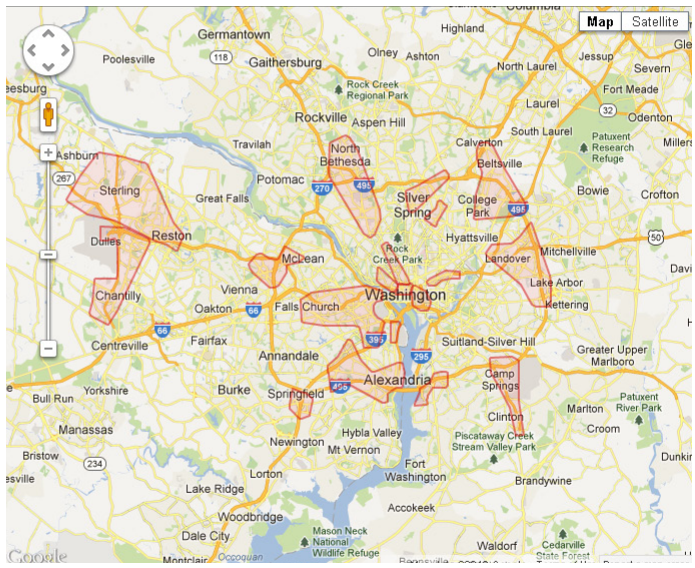

Seattle




Toronto




Vancouver




Washington DC