How Do Online Product Rankings Influence Sellers' Pricing Behavior?*

Luise Eisfeld†

This version: February 23, 2021

Abstract

Products that are displayed more prominently on e-commerce platforms are more likely to be found and purchased by consumers. The algorithms ranking these products, however, may condition a product's position on its price, and can thus intensify, or weaken price competition between sellers. I set up a simple framework in which hotels offering rooms on a booking platform take into account not only the usual, direct effect of their prices on demand, but also the effect of their price on their position in the ranking. Using data scraped from Expedia, I find that for a given hotel, a lower price leads to a more prominent position on the results page. For the demand side, I estimate utility parameters of a sequential search model following Ursu (2018), using additional consumer search data. The parameters on the ranking side as well as the consumer side allow me to simulate how variations to the ranking algorithm affect hotel prices and markups.

Keywords: product rankings, online search, hotel pricing, Online Travel Agents, electronic commerce, online discounts

JEL Classification: D22, D61, D83, L11, L81

---

*I am grateful to Alexandre de Cornière, Daniel Ershov, and Bruno Jullien for their support and guidance. I moreover thank the participants of the TSE Applied Microeconomics Workshop, the TSE IO Student Workshop, the Competition and Innovation Summer School (2019) and EARIE (2019) for their helpful comments. Data obtained via the Wharton Customer Analytics Initiative and sponsored by an anonymous travel website is gratefully recognized. All errors are mine.

†Toulouse School of Economics (TSE). E-mail: luise.eisfeld@tse-fr.eu

Introduction

A fundamental concern of sellers distributing goods or services online is the visibility of their products to consumers. E-commerce websites such as Amazon or Expedia provide access to broad varieties of products offered by numerous third-party sellers. Evaluating the manifold products is costly for consumers. To facilitate product search and discovery, platforms display algorithmically ranked lists of relevant products to users entering a product query. These default rankings have been found to systematically influence the products that consumers become aware of and ultimately purchase (e.g. Ursu 2018, Glick et al. 2014).

While the ranking algorithms used in practice are highly complex, they may condition a product's rank on the price that its seller chooses to set. Trying to guide consumers towards attractive offers, platforms may find it profitable to display lower-priced products in more prominent positions.[1] However, with sellers being aware that lower prices translate into higher visibility, such an algorithm amplifies the competitive pressure between sellers and provides additional incentives to reduce prices. In other contexts, the reverse is happening, with search frictions hampering the comparison of products. Platforms may even steer consumers to more expensive products to extract more revenue per consumer visit (Hagiu and Jullien, 2011). This weakens price competition and enables sellers to charge higher markups (Dinerstein et al., 2018). Rankings therefore affect sellers' price-setting decisions and influence market outcomes in subtle ways.

This paper focuses on these supply-side implications of product rankings, which have been left relatively unexplored by empirical literature that has mostly dealt with demand-side effects of rankings. It studies the empirical setting of hotel rooms offered by online travel agents (OTAs) which appear in an algorithmically ranked ordering. Using scraped data from Expedia, I find that offering a lower price leads to a better positioning of a given hotel in the platform's default list of results. I then use clickstream data to estimate a consumer search model that closely follows Ursu (2018). The estimated ranking parameters and the demand parameters then enable to perform counterfactual simulations. Preliminary results show that hotels set higher prices in a simulated scenario in which the ranking does not take into account supply-side effects. Expedia's

[1] Examples are online travel agents selling hotel rooms. Those agents explicitly state that the competitiveness of hotels' rankings are taken into account in their rankings (see Section 1).

ranking algorithm thus seems to move hotels' pricing decisions into a more competitive equilibrium with lower prices and more promotions offered, compared to a scenario where prices do not affect visibility.

Studying the effect of ranking algorithms on market outcomes within a structural framework is interesting for several reasons. First, it is a well-established fact that search frictions persist to exist even on the Internet, contrary to expectations. A naturally arising question is therefore how these frictions affect market outcomes and especially pricing behavior, which this article attempts to shed light on. Second, online visibility is linked to policy-relevant questions related to the dominance of online platforms, which effectively set the rules of how market participants can interact. For example, the issue of prominence was at the heart of a recent prominent antitrust case in Europe on Google Search (European Commission, 2017). Third, understanding the effect of rankings on prices is of fundamental interest for platforms themselves. Online intermediaries are two-sided markets that crucially depend on attracting both sellers and consumers to be successful. When devising product rankings, platforms need to assess multiple trade-offs on both of these sides. While platforms can evaluate a ranking's effects on consumers' choices by engaging in A/B-testing, the ranking's effect on supply side behavior is much more difficult to evaluate ex ante, as it cannot be revealed by experimentation on subgroups of consumers. The analysis of how rankings affect pricing decision thus requires a structural framework that incorporates demand and supply to shed light on such equilibrium effects. Finally, the hospitality industry is an especially intriguing setting for the study of product rankings, as consumer search is of great importance in this context.[2]

Motivated by empirical evidence stressing the importance of price discounts on OTAs, I propose a theoretical framework in which hotels decide which prices to set and whether to carry out promotional sales. What differentiates this framework from standard supply models is that hotels set prices and promotions in anticipation of

[2] Travellers typically have limited knowledge of hotels and local conditions in a given city, which is the reason why intermediation has been important in the travel sector even prior to the emergence of OTAs. Given the extent of differentiation of hotels and consumer tastes and given that hotels may face thousands of potential competitors, hotels pay a lot of attention to their visibility on these platforms. See the 2017 Position Paper by HOTREC, the European hospitality association, available here: https://www.hotrec.eu/wp-content/customer-area/storage/f5282293ec286d90ba33117497c7c2c6/HOTREC-position-on-the-mid-term-review-of-the-DSM-Strategy-10-October-2017.pdf (accessed May 7th, 2019).

how these decisions affect positions in the ranking. If a hotel marginally increases its price, it thus faces the common "direct" negative impact on demand (conditionally on being seen), as well as an "indirect" marginal impact on demand that is driven by the price's impact on visibility. My ultimate goal is to simulate what demand and hotel prices would be like under a counterfactual ranking of hotels. This requires, first, the estimation of how price and promotion affect a hotel's position, using a linear fixed effects regression with instrumental variables. Second, a model of hotel demand is required that accounts for the fact that consumers are not fully informed, but face search costs and are more likely to click and book hotels that are displayed more visibly. I use Ursu's (2018) model of consumer search, in which each search corresponds to a click on a given hotel. Estimating consumer search costs in an unbiased way, however, is not possible with the actual, relevance-based ranking of hotels that I have at hand. I thus directly employ Ursu's search cost estimates (assuming that they do not change between contexts) and only estimate preference parameters for hotel characteristics. Given these parameter estimates, I can then simulate demand for hotels that are displayed in a given fashion, and conduct counterfactual experiments.

Two datasets are used to carry out the analysis. The primary dataset was scraped from Expedia in early 2019 by carrying out daily queries for a diverse range of travel dates. It entails hotels' ranking positions as well as their pricing and promotion decisions, and thus enables to examine the effect of hotels' pricing decisions on positions. The identification of the marginal impact of prices and promotion on the ranking rests on two components: First, I employ hotel fixed effects that control for any unobserved hotel-specific, time-invariant factors impacting a hotel's average position in the ranking (such as a hotel's unobserved "quality" or the amount of commission it pays to the platform). From the joint variation in positions, prices and promotions for a *given* hotel in the data, I can then estimate the correlation between a hotel's price and promotion with its ranking over query and travel dates. Second, the simultaneity of hotels' price-setting decisions with respect to the ranking are addressed using instrumental variables for prices and promotions.

The second dataset obtained via the Wharton Customer Analytics Initiative (WCAI) details the entire search process of actual consumers who arrive on a travel website and search for a hotel one of the cities that were also scraped in 2009. The data details a strong correlation of position and clicks (driven both by higher relevance of more visible hotels as well as by search costs), confirming prior literature. I moreover find that most clicks and purchases occur under the default ranking, and a remarkable 70% of all users recorded in the data only see search results that have been ranked by default. The stream of consumers' clicks and eventual purchases and the characteristics of hotels observed in the data serve to estimate the demand parameters for the counterfactual scenarios.

I find a significant and economically meaningful effect of a hotel's price on its position that implies that increasing one's price by one dollar leads to a decrease by three positions in Expedia's listing, ceteris paribus. This effect is robust across specifications and instruments that are used, but differs in magnitude depending on which city is considered. This result is interesting in itself: Expedia's ranking algorithm appears to intensify price competition between hotels, so that hotels would be setting higher prices in a situation without such a ranking algorithm. The results moreover indicate a positive correlation of a hotel's rank position and its promotion decision (i.e. promotion associated with being ranked more visibly), although a clear causal link cannot be established due to a lack of sufficiently strong instruments. The consumer side results are in accordance with Ursu (2018). Given the estimates on both the consumer side as well as the ranking side, one can back out hotels' marginal costs from the hotels' optimal pricing condition. Given the estimates, one can then exogenously change variables like the ranking's elasticity with respect to prices or consumers' search costs in the model and simulate how this affects market outcomes. I find that hotels tend to set lower prices if prices are not influential for their respective rankings.

This paper relates to literature dealing with consumer search, platform design, and the industrial organization of hotel booking platforms. A number of recent papers estimate structural models of consumer search in diverse settings, under different assumptions on search behavior (fixed sample search or sequential search), and with different types of data (aggregate or individual-level). The most relevant papers among these use consumer search data from online hotel platforms (Koulayev, 2014; De los Santos and Koulayev, 2017; Chen and Yao, 2018; Ursu, 2018), and have found that search costs in the hotel search context are very significant, with important implications for hotels. These papers also find that consumers gain positive utility from purchasing hotels that are on promotion, which underlines the importance of incorporating promotions into my model. The sequential search model used on the demand side of this model closely follows Ursu (2018). Search costs in this model are essentially defined as the costs of clicking on a given hotel, and are assumed to increase in positions.

The papers estimating models of online search, however, remain ignorant of possible supply-side effects of the rankings or refinement features that they study. Among the rare empirical papers that do shed light on how search frictions on online platforms influence sellers' price-setting is Dinerstein et al. (2018). The authors study the effects of a change in search design on eBay that essentially decreased the search costs of finding the cheapest product, and find a decrease in sellers' prices and markups. While these authors consider a very homogenous product in their supply side model, hotels are both horizontally and vertically differentiated, which may entail interesting, contrasting implications. Further related papers investigate the effect of rankings or search design on entry by sellers (Ershov, 2018), sellers' obfuscation strategies (Ellison and Ellison, 2009), and on the functioning of peer-to-peer platforms where matching is central (Fradkin, 2017).

Finally, this paper is related to studies focusing on competition issues on online hotel booking platforms. A paper that equally focuses on hotels' visibility on these platforms is a study by Hunold et al. (2020), and explores hotels' pricing decisions across booking channels and its impact on hotel rankings. The authors build a model showing that a ranking that maximizes an OTA's short-term profits is not necessarily in accordance with the ranking that maximizes the match value of consumers. An empirical finding in this paper is that setting a lower price on a hotel's *own* website will lead to a worse position in the OTA's rank, as it decreases the hotel's booking likelihood. Using the ranking, a platform can thus discipline hotels in their price setting decisions. My work is complementary to their study. Anecdotal evidence points out that prices *per se*, not only price differentials between booking channels, influence rankings. Having access to a dataset that details actual search behavior further allows me to account for the demand side and study counterfactual scenarios. Further related papers study hotels' price-setting behaviors (Li et al. 2018, Cho et al. 2018), buyer substitution across purchasing channels (Cazaubiel et al. 2018). A paper that focuses on buyer substitution between hotels and Airbnbs and localized competition is Schaefer and Tran (2020).

Section 1 of this paper presents background information about the online hospitality sector and OTAs' rankings of hotels. Section 2 contains the model for the supply side, and Section 3 details the datasets used and provides descriptive statistics. The empirical strategy will be presented and discussed in Section 4, with the results presented in Section 5. Section 6 models counterfactuals. Section 7 concludes.

1 Background

The hotel sales on all major OTAs[3] take place under the so-called "agency model". Within this business model, the supplier (here, the hotel) sets the final price of the product that is sold via the intermediary. The platform merely mediates the transaction between the hotels and consumers by facilitating product search, by giving hotels exposure to potential customers, and by processing the payment. OTAs receive a fixed ad-valorem commission rate for each purchase that is carried out. This rate lies roughly between 10% and 25% according to industry reports. The standard rate is 15%, but can privately re-negotiated and tends to be somewhat lower for large chains (Cazaubiel et al. 2018).

Consumers arriving on an OTA typically enter a city, the number of travellers and arrival and departure dates, and are then confronted with what I call a "listings page". Here, consumers view an ordered list of hotels of typically 55 hotels per page[4] that are ranked by default according to what Expedia calls a "Recommended" ranking. Depending on the number of hotels available at the given location, the ranked hotels can extend to several, possibly hundreds of pages. On the listings page, consumers can sort or filter[5] hotels, and browse to further pages. Clicking on a listed hotel brings a consumer to what is referred to a "hotel page" with more detailed hotel and pricing information.

While OTAs do not disclose the exact functioning of their possibly highly complex default ranking algorithms, it is widely believed to be based on relevance. Moreover, Expedia states that a hotel's price competitiveness (current price, active promotions)

[3]For example, on Booking.com, Expedia, or HRS.

[4]In my data, a few listings pages have less than 55 hotels on the same page. My impression is that it is related to the quality of the Internet connection when accessing the page.

[5]Sorting refers to sorting by price or by guest rating, for example. Filtering means that only hotels of, for example, a certain star rating are displayed.

Figure 1

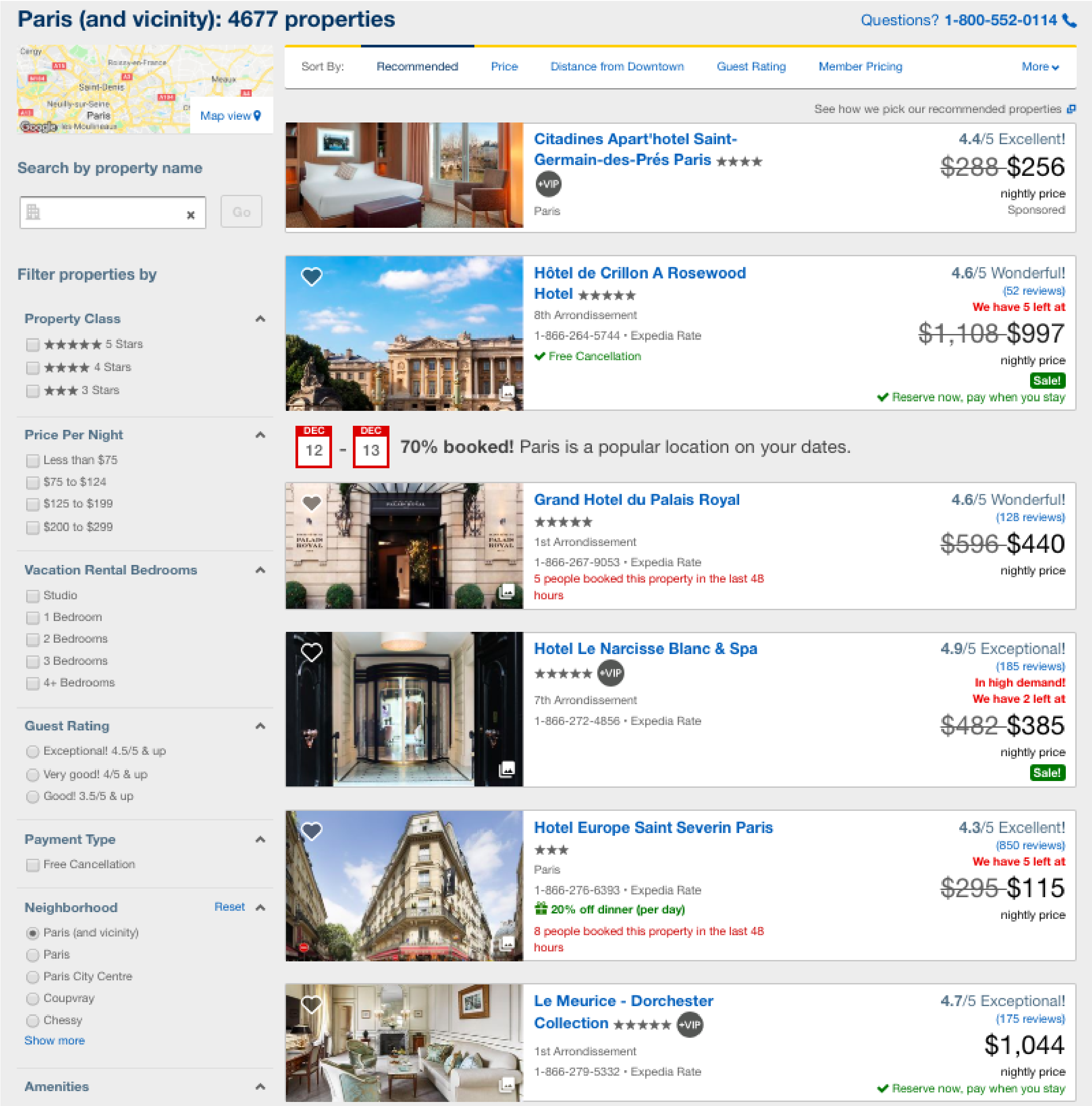


Figure 1: Screen shot of Expedia's listings page, here for Paris, as of February 2019.

are factored into the ranking order (see Figure 8 in the Appendix). Also, Hunold et al. (2020) find evidence that the OTAs Expedia and Booking.com actively demote hotels if they set lower prices on their own or competing OTAs' websites.

There are two main dimensions in which hotels can compete on OTAs in the short run. The first, more obvious one is the pricing dimension. Second, hotels are able to offer promotions on OTAs. This can be a discounted price for a given travel date and room type, or free extra services (such free cancellation or a rebate for meals). What is visible in Figure 1 is that many hotels moreover show both a price in black besides a grey strikethrough price. As of 2019, however, the strikethrough price was no discount given by the hotel as long as there is no "Sale!" tag.[6] Therefore, in this paper, I call a "promotion" any hotel listing that is flagged with a green "Sale!" tag (see Figure 1), and thereby displayed in a somewhat more salient way.[7] The price and a sale tag dummy are then the key explanatory variables in my analysis.

Finally, the listings page also features advertisements run by Expedia TravelAds (see first listing in Figure 1, which says "Sponsored"). Hotels displayed as such sponsored listings pay for every click that occurs and can even specify to target a particular audience. I exclude these sponsored listings from my analysis. Moreover, I focus on the "organic" ranking that is determined by the algorithm only.

2 Model

The centerpiece of the model is a hotel's decision to set prices and promotions. The hotels' first order conditions show that a hotel's price decision hinges on two effects: The direct effect on demand, as well as the indirect effect on position, and thus visibility. On the consumer side, I take as given that consumer search follows the sequential search framework detailed in Ursu (2018). Finally, I am not explicitly modelling a platforms' decision of how to design the ranking algorithm, but discuss the tradeoffs and possible incentives that platforms face in Section 2.1.

2.1 Supply Model

I consider J differentiated hotels in a given market that sell their rooms on an intermediary's website. The intermediary charges an ad-valorem commission fee $\tau \in (0, 1)$ for each purchase it mediates. Hotels have marginal costs c_j. Hotels know that consumers do not only care about prices, but also about discounts that are offered by the hotel. Taking as given the pricing and promotion decisions of their competitors, at a given date q, hotels thus (i) set prices and (ii) decide whether or not to offer a discount for room

[6]Instead, the strikethrough price is, according to Expedia, a comparison price, namely "the third highest price for this room type at this property (with the same length of stay and cancellation policy) that customers have found on our site during a 30 day window around your selected check-in date."

[7]See Expedia Partner Central for more info (last accessed 27/02/2020). The discount in percent that is offered by a given hotel is viewed when hoovering with the cursor above a hotel's price or sale tag.

bookings for date t, such that the resulting decisions form a Bertrand Nash Equilibrium. A discount $d_{jqt} \in \{0, 20\}$ raises demand, but is costly for the hotel[8]. Hotel j's profit, as always, is the product of the markup and the demand $Ms_{jqt}(\cdot)$ (market size $\times$ market share) for query date q and travel date t. What is new is that demand for hotel j in turn is not only a function of the hotel's pre-promotion price p_{jqt} and the promotional discount d_{jqt}, but also of the position r_{jqt} which the hotel is by default displayed in on the OTA's results page. By modelling demand as being dependent on a product's positioning, the model accounts for the stage in which consumers search for products. Products with a worse positioning in the ranking (i.e. a higher r_{jqt}) are less likely to be found by consumers, resulting in lower demand.

Let $(\mathbf{p_{qt}}, \mathbf{d_{qt}}, \mathbf{r_{qt}})$ be the vectors of prices, promotion decisions and rankings for all J hotels in the market. Hotel j's profit maximization problem then writes

$$\max_{p_{jqt}\in\mathbb{R}^+,\, d_{jqt}\in\{0,D\}} \Big(p_{jqt}\cdot(1-\tau) - \big(d_{jqt} + c_{jqt}\big)\Big) Ms_{jqt}\big(\mathbf{p_{qt}}, \mathbf{d_{qt}}, \mathbf{r_{qt}}(p_{jqt}, d_{jqt})\big).$$

Next, define $\tilde{c}_{jqt} \equiv \frac{c_{jqt}}{(1-\tau)}$ and $\tilde{d}_{jqt} \equiv \frac{d_{jqt}}{(1-\tau)}$. Deriving the first order conditions with respect to price and rearranging, one obtains:

$$p_{jqt} = \tilde{c}_{jqt} + \tilde{d}_{jqt} - \frac{s_{jqt}(\cdot)}{\underbrace{\frac{s_{jqt}(\cdot)}{\partial p_{jqt}}}_{\text{direct } (<0)} + \underbrace{\frac{\partial s_{jqt}(\cdot)}{\partial r_{jqt}}\frac{\partial r_{jqt}}{\partial p_{jqt}}}_{\text{indirect } (<0)}} \tag{1}$$

When deciding on prices, hotels thus take into account two types of effects on demand. The "usual" demand ("direct") effect $\frac{\partial s_{jqt}(\cdot)}{\partial p_{jqt}}$ expresses that, other things equal, a lower price will lead to higher demand. Second, prices and promotions affect the average position r_{jqt} which a given hotel is displayed in, which in turn affects how many users become aware of the hotel, click on it, and purchase it. Thus, the "indirect effect" $\frac{\partial s_{jqt}(\cdot)}{\partial r_{jqt}}\frac{\partial r_{jqt}(\cdot)}{\partial p_{jqt}}$ is additionally going to be taken into account by hotels.

[8] I thus assume a hotel's decision of whether to offer a promotion to be binary. The main reason is that discounted hotels are highlighted (no matter how large the discount is), which might already have an effect on users, while the actual percentage reduction is usually only visible when hovering with the pointer over the "sale" button on the results page. Moreover, this fits to the consumer browsing data, in which promotions are also captured by a binary variable. I assume that offering a promotion comes at a cost of 20 euros to a hotel, which is empirically approximately the average discount that hotels offer for Paris on Expedia.

Moreover, as I model giving a discount as a discrete choice, a hotel is willing to offer a discount whenever

$$\Pi_{promo} \geq \Pi_{nopromo}$$

$$\Leftrightarrow \quad \Big(p_{jqt}(1-\tau)-c_{jqt}\Big)\cdot\Big(s_{jqt}\big(\mathbf{p_{qt}},\mathbf{d_{qt}},\mathbf{r_{qt}}(p_{jqt},d_{jqt}=20)\big)-s_{jqt}\big(\mathbf{p_{qt}},\mathbf{d_{qt}},\mathbf{r_{qt}}(p_{jqt},d_{jqt}=0)\big)\Big)$$
$$\geq 20\cdot s_{jqt}\big(\mathbf{p_{qt}},\mathbf{d_{qt}},\mathbf{r_{qt}}(p_{jqt},d_{jqt}=20)\big)$$

Note that again, $\mathbf{d}_{qt}$ enters the market share $s_{jqt}(\cdot)$ both directly and indirectly through affecting $r_{qt}(\cdot)$. Thus, a hotel's willingness to offer a discount depends again not only on how the discount affects the demand directly, but also on how the discount affects the hotel's visibility on the platform.

As a result, if a platform modifies the ranking algorithm - for example by making the algorithm more sensitive to prices, or by ranking hotels completely independent of their prices - hotels will take this into account and price differently. This affects hotels' markups and consumer surplus.

Discussion of Hotel Model

The above model abstracts from the fact that hotels typically distribute rooms via multiple sales channels. According to a monitoring study conducted by the European Commission and participating national competition authorities (2016) which reports survey results of European hotels, chain hotels sold 35% of their rooms via OTAs, whereas for independent hotels this fraction is 42% in 2016. As Hunold et al. (2020) find, OTAs' ranking algorithms in fact seem to take price differentials between the price listed on the OTA and the price listed on the hotels' or other websites into account when ranking hotels, and thus may effectively punish hotels whenever they do not provide the lowest price to the OTA's website. When setting prices on Expedia, hotels are thus likely to take into account their prices and expected demand on other sales channels. I nevertheless believe that the trade-off captured in the above model is likely to be of first order for hotels. Evidence shows that consumers are not very likely to substitute between different distribution channels (Cazaubiel et al., 2018), so that it is adequate to consider the hotel's decision of which price to post on Expedia in isolation, as the model does. Moreover, the ranking algorithms of different OTAs seem to work similarly, therefore providing similar pressure on the price.

Relatedly, the model above also assumes that all hotels are shown and booked via Expedia's default "recommended" ranking. Instead, one could think that hotels differ in the extent to which they react to the ranking: While some hotels might be highly dependent on the Expedia default ranking sales channel, others might derive most bookings from regularly returning guests or via their own website, therefore being rather insensitive to the ranking. The assumption, however, is supported by the fact that consumer search data described below indeed shows that most consumers search for hotels via the default ranking - only around 34% of consumers ever decide to sort or filter any hotels. More than three quarters of all bookings occur under the default ranking.

Moreover, hotels' pricing decisions are in reality a high-dimensional, dynamic problem (see Cho et al. 2018). Hotels are capacity constraint and need to set prices for a range of future dates, across multiple sales channels, and for different room types. I abstract from these dynamic considerations, as they would make the model too complex for my setting.

Promotions should be part of the hotel's pricing problem because promotions on OTAs are very prevalent and likely to influence consumers' booking decisions. On the ranking side, Expedia explicitly states that it takes promotions into account when ranking hotels (see Figure 8 in the Appendix). On the consumer side, models of consumer search have consistently found positive utility parameter for a "promotion" dummy, meaning that consumers seem to derive positive utility from purchasing a hotel that is on promotion. This can be due to the fact that Expedia's red "Sale!" flag makes hotels appear in a more salient way if they are on sale (see Figure 1). Moreover, a higher prices in the hotel market may signal higher quality, so that high-quality hotels may find it optimal to keep baseline prices high but use promotions to sell available rooms.

2.2 Demand Model

To obtain the effect of prices on demand, I employ the sequential search framework that is used in a hotel search context by Ursu (2018). I refer to her paper for a detailed explanation of the model, but will briefly cover a few important features of the model. Consumers in this sequential search model can extract basic hotel information (prices, stars, rating etc.) upon being confronted with a results page listing different hotels. Based on this listing page information, consumers can thus costlessly form an expected utility v_j for each of the hotels, which is parametrized as being linear in a hotel's stars, review score, location score, price, a promotion indicator, and a brand indicator. Consumers can then incur a costly search by clicking on any listing, which will lead to the hotel page where further information can be accessed and where the hotel can be booked. Any such click will reveal a part of a user's utility for a given hotel that is ex ante random, ϵ_j. Hotels ranked further down the page (in less positions) are more costly to click on, which can be interpreted as the costs of scrolling down a page. These search costs are parametrized as $c_j(r_j) = \exp(k + \gamma r_j)$, where r_j here stands for rank of hotel j in a consumer's search, and k is a mean level of search costs. Purchasing decisions are finally made based on the realized utility after search, $u_j = v_j + \epsilon_j$.

The model assumes that consumers' search and purchasing decisions observed in the data follow a model of optimal sequential search, which can be characterized by Weitzman's (1979) optimal sequential search rules. These rules in turn imply bounds on the parameter values for consumers' valuations for price, ranking, location etc. as well as search costs. All in all, the model thus reflects how the ranking impacts demand for a given hotel.

2.3 The Platform's Incentives

The platform's ranking algorithm is a crucial element of the market design that fundamentally influences both sides of the market. On the consumer side, a platform may on the one hand want to assist consumers with finding valuable bargains or high-quality matches (and thus essentially reduce search costs), thereby increasing the website's overall attractiveness in the long run.[9] On the other hand, platforms have the short run objective of maximizing revenues per visit by essentially diverting consumer search (Hagiu and Jullien, 2011). In an OTA setting, the platform may for example find it profitable to display a few pricy (and higher quality) hotels very visibly and thereby induce consumers to buy a more expensive hotel.

[9] This point is also highlighted by Dinerstein et al. (2018) when mentioning that guiding consumer search is one of two key search design objectives, the other one being that a platform may want to foster stronger pricing incentives on sellers.

On the hotel side, the OTA may want hotels to compete fiercely so as to induce them to price competitively and offer promotions, which again makes consumers willing to visit the OTA's website. However, an OTA would not want to encourage competition on prices too much, as hotels might otherwise engage in obfuscation or not even be willing to join the platform in the first place.[10] Moreover, note that the platform's profit in the above framework is the ad valorem fee multiplied by the net price of all bookings incurred, similarly to the model by Hunold et al. (2020). It is not clear whether a platform would want to display a given hotel more visibly when its price is lower than usual, as opposed to it being higher than usual: With a higher price, the platform receives a higher revenue given that the hotel is being booked; however, a higher price would also decrease the likelihood of the hotel being booked, ceteris paribus. One can therefore identify a number of effects of a search platform's ranking algorithm decision on its revenues which may counteract or enforce each other. The empirical estimates below shed some light on how these platforms' (or Expedia's, at least) rankings currently operate.

3 Data

3.1 Ranking Side: Scraped Data

This study uses two kinds of datasets. The first dataset was scraped from Expedia, which is the dominant OTA in the US, for the cities of Budapest, Cancun, Manhattan and Paris. In the month of February and March 2019, I carried out daily hotel queries for trips that begin between one and 250 days later (all in all, for 16 different travel dates or more each day per location), such that the latest travel dates are in November 2019. During each such query, I obtained hotels' default position, pricing, promotions, and additional characteristics from all results pages that appear in a query.[11] I use queries that are empirically being entered by consumers the most often, and are thus the most relevant for hotels setting prices, namely weekend or single night stays for two persons

[10]Causal empiricism indicates that airlines, for example, have appeared to compete more intensely on price during the past years, with fees such as checked luggage or meals increasingly not included in the baseline price that is displayed on search aggregators.

[11]Cookies were cleared after each request. All requests were carried out using the same user agent. A few travel dates that should have been collected on a given query date are missing due to failing Internet connections.

in one room. All in all, the data covers over 6,000 distinct hotels which I identify as active on the platform, most of them (over 2,800) in Paris.[12] As seen on the screen shot in Figure 1, hotels offer various extras to entice consumers (free cancellation, free offers, discounts, a "Sale!" tag, etc.), all of which I extract from the website, allowing me to control for these factors in my analysis.

Table 1 shows the market characteristics by city which indicates some differences especially in the number of active hotels as well as in whether the hotel belongs to a brand.

Table 1: Market characteristics

City	Active Hotels	Avg. stars	Avg. review	Avg. # reviews	Brand (%)
Budapest	1043	3.25	4.11	793.95	10.02
Cancun	944	2.95	3.92	1052.64	7.62
Manhattan	1209	3.53	4.17	2065.28	32.62
Paris	2827	3.27	3.92	1094.18	26.01

The analysis focuses on two explanatory variables that seem highly influential for the ranking and purchasing likelihood: First, the price; and second, a dummy variable indicating whether the hotel is displayed with a green "Sale!" tag. I find substantial variation in hotels' strategic decisions over time. As seen in Table 2, relatively many hotels give a discount or display a "Sale!" tag at any point during the query period - however, for all cities, there is also a large fraction of hotels that never offer a sale during the period of observation. Lastly, to illustrate the seasonality in pricing, Figures 10 to 13 in the Appendix show the variation in average prices across time by star rating of the hotel in the four different cities. Especially hotels with many stars appear to show significant variation of average prices over travel dates, possibly reflecting changing demand patterns. Table 3 moreover displays the between and within variation in positions, prices, and sales in the Paris data (which are used for the estimations below). The table shows that the within-hotel variation in these variables is very large. To illustrate this further, I find that the median hotel's difference between the maximum and the minimum price across different travel dates amounts to 156 euros (not shown

[12] A substantial fraction of hotels was present on the latter pages of the search result, but without any pricing information and without the possibility to book the given hotel. As these hotels never show any pricing information during the observation period, I chose to drop them from my analysis.

in table). The median hotel's difference between the maximum and the minimum rank across different travel dates amounts to almost 1,500 positions.

Table 2: Pricing

City	Active Hotels	Avg. price (€)	% of hotels with "Sale" tag	% of hotels ever with "Sale" tag
Budapest	1,043	87.82	29.06	36.24
Cancun	944	149.71	18.41	24.47
Manhattan	1,209	286.30	18.98	24.81
Paris	2,827	158.64	28.64	48.81

Averages are computed over all observations in each of the cities.

Table 3: Within and Between Variance: Paris data (the sample used for neighborhood instruments)

Variable		Mean	Std.Dev.	Min	Max	Observations
Position	overall	1016	603.1	1	2411	N = 593,984
	between		513.2	35.15	2121	n = 2,182
	within		327.4	-745.8	3044	T-bar = 272.2
Price	overall	152.1	97.60	27	999	N = 591,640
	between		94.67	39.33	998	n = 2,179
	within		40.77	-194.2	999.9	T-bar = 271.5
Sale	overall	0.285	0.451	0	1	N = 593,984
	between		0.362	0	1	n = 2,182
	within		0.268	-0.712	1.281	T-bar = 272.2

3.2 Demand Side: Consumer Search Data

The second dataset stems from one of the major online travel agencies in the US and in the world[13], and was obtained via the Wharton Customer Analytics Initiative (WCAI) at the University of Pennsylvania. The data motivates the analysis, helps to verify the importance of the default ranking, and is used to obtain utility parameters on the demand side. The data contains the search behavior of nearly 18,000 US-based users, and cover all searches for hotels at four local markets (Budapest, Cancun, Manhattan and Paris) that took place on this OTA's website between October 1st and October 15th, 2009. The data amount to roughly 1.3 million observations, where each observation

[13] According to the travel research website tnooz (2009), this website was one of the most visited travel agent websites in the weeks covered by the data. Moreover, it was ranked as being the most used platform worldwide over many years.

corresponds to a given hotel that appeared during a given user's search. The basic dimensions are displayed in Table 4. For each hotel that was viewed, I observe the precise query that was entered along with time and date, whether the hotel appeared in a search for which a refinement action was taken (i.e. sorting by price), browsing behavior (i.e. flipping through pages), and whether it was clicked on (to inquire further information on the hotel page) or purchased (including a time stamp). The data contain basic hotel characteristics such as star rating, price, a binary "promotion" variable, brand and parent company it belongs to (if any), location and name of the hotel, and which position it appeared on. All in all, the dataset gives an insight of consumers' search behaviors at a very granular level and is exhaustive compared to datasets that have been used by other empirical literature[14].

Table 4: Basic data summary: WCAI consumer search data

Market	Observations	Users	Searches	Hotels	Clicks	Purchases
Budapest	353,319	4,947	14,805	276	9,337	262
Cancun	270,071	4,211	11,588	110	7,580	84
Manhattan	350,823	4,203	13,786	543	6,173	111
Paris	330,891	4,425	13,247	1,637	5,787	108
TOTAL	1,305,104	17,769	53,426	2,566	28,877	565

Number of observations, users, searches, hotels, clicks and purchases, by market and in total.

Exploring consumers' search strategies confirms the importance of the default ranking during the search process. Consumers search on average relatively little and rarely flip through results pages. As Table 5 shows that, in Paris, for instance, even searchers who end up buying a hotel end up viewing only 80 out of 1,600 potentially available hotels. Refinement actions (comprising filtering and sorting) are used relatively rarely - the average number of refinement methods chosen is somewhat above one half for people who end up purchasing. Only 30% of all users (and 45% of users who end up buying) ever view hotels that are not ranked by default. In accordance with this, Table 6 shows that most clicks and purchases occur under the default ranking, which stresses the importance of hotels' default position for their revenues. Lastly, Figures 2 and 3 suggestively point to the importance of a hotel's rank for consumers' click and purchase decisions. However, note that these figures do not take into account the fact

[14] For example, Koulayev (2014) does not observe purchases, and Ursu (2018) does not observe refinement decisions or the order of clicks that were made.

that better-ranked hotels are likely also of higher quality, and therefore do not allow for a causal interpretation.[15]

Ursu's (2018) dataset and consumer utility includes a hotel's location score (with better location being equivalent to a higher score), which is likely to be important given that location is likely to be a major dimension for consumers when booking hotel rooms. The scraped data does not include any such score; therefore, I construct a location score by computing hotels' distances to the Louvre museum (which I take as the center of the city) and then assign these distances to discrete location scores that mirror the one of Ursu's (2018) dataset.

Table 5: Search activities for buyers and non-buyers (averaged over users)

	Market	# Searches	# Hotels viewed	# Clicks	# Refinements
Buyers	Budapest	5.19	51.49	5.96	0.51
	Cancun	5.67	41.55	6.53	0.49
	Manhattan	7.07	81.46	7.57	0.64
	Paris	6.33	77.32	6.08	0.73
Non-buyers	Budapest	2.86	42.17	1.71	0.29
	Cancun	2.69	35.21	1.72	0.25
	Manhattan	3.19	52.03	1.34	0.43
	Paris	2.91	49.33	1.22	0.38

Average number of searches, hotels viewed, clicks made, and refinement options chosen in each of the markets are displayed. A new "search" is made whenever any request setting is changed, or a refinement method is chosen.

Table 6: WCAI data: Refinement actions

Ranking	% of users who ever view given ranking	Number of clicks	Number of purchases
Default ranking	97.48	20,648	440
Sort by hotel name	1.52	90	3
Sort by city name	0.56	51	1
Sort by distance	12.63	3,057	34
Sort by star rating	2.89	624	5
Sort by price	17.52	4,565	70
Sort by reviews	1.81	407	12

Table shows the percentage of users who ever see search results that are ranked to a given ranking. (Since many users see the default ranking and make at least one refinement in addition to that, it is natural that the percentages for users does not add up to 100.)

[15] In contrast, Ursu (2018) establishes a causal link using randomly ranked search results.

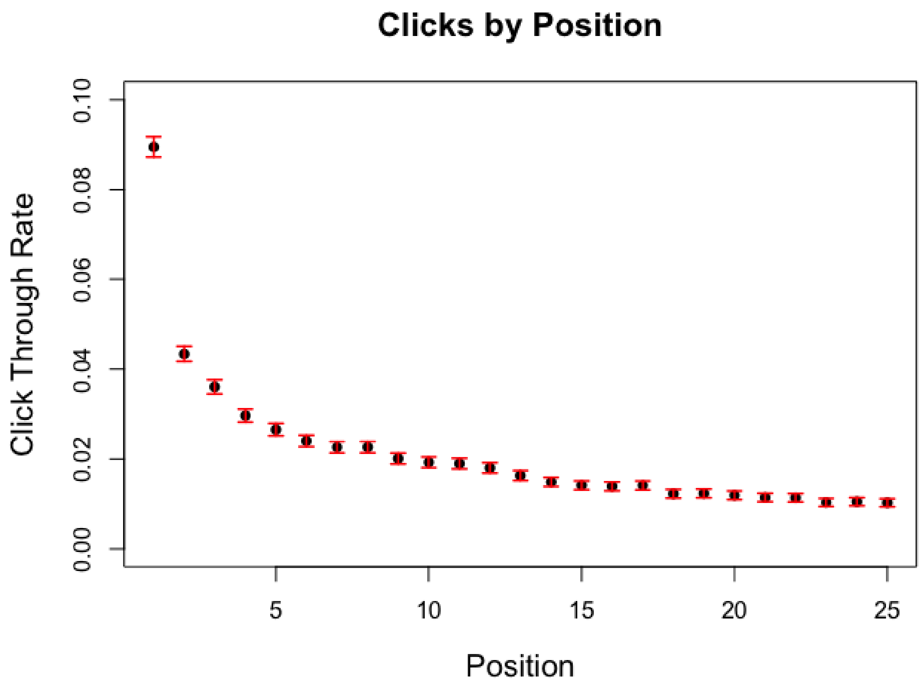


Figure 2: Average click through rate by position (pooling all rankings); confidence intervals at 95% level.

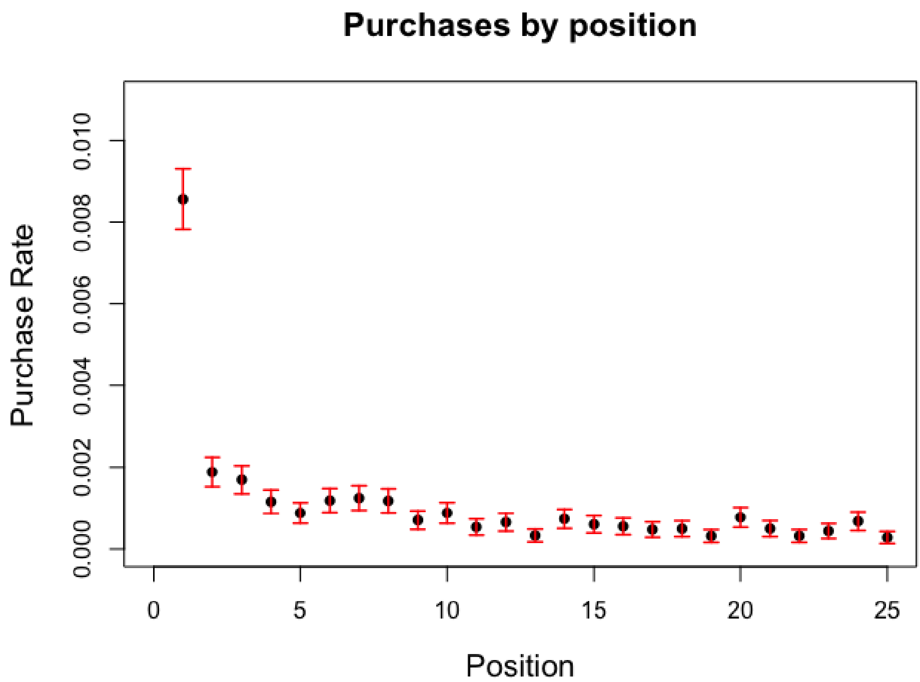


Figure 3: Average purchasing rate by position (pooling all rankings); confidence intervals at 95% level.

3.3 Further Datasets

I obtain further datasets that allow to construct the instrumental variables. First, hotels' addresses are extracted from Expedia as well. These addresses enable me to obtain hotels' geolocation (latitude and longitude) using the Google Maps API.[16] I find that precise geolocation data are available for 82% of the hotels in Paris (and for 94% after removing hotels that I identify as "apartments", as opposed to actual hotels). Second, I obtain publicly available data on Airbnb location information on the zipcode level, as well as prices and availabilities for the next 30 days, from the website http://insideairbnb.com.[17]

4 Empirical Strategy and Results

4.1 Effect of Prices on Ranking: Linear Fixed Effects Regression

To obtain the effect of prices and promotions on a hotel's position, I employ an instrumental variable regression with hotel and travel date fixed effects using the scraped data. The equation of interest writes:

$$r_{jqt} = \alpha p_{jqt} + \beta d_{jqt} + \delta past\_book_{jq} + \gamma_j + \lambda_{qt} + \epsilon_{jqt} \tag{2}$$

where r_{jqt}, p_{jqt} and d_{jqt} indicate hotel j's average position, price and promotion dummy for travel date t and query date q. The coefficients of interest are the price and promotion dummy coefficients, α and β. $past\_book_{jq}$ is a variable counting the number of bookings (for any travel date) that have been made in the last 48 hours prior to scraping a given hotel hotel. It serves to control to some extent for demand shocks that a hotel may experience and that the ranking algorithm may pick up. λ_{qt} accounts for a query date-travel date fixed effects. Hotel fixed effects γ_j allow to control for any unobserved hotel characteristics that are fixed over time. Accounting for hotel fixed effects is crucial: Expedia almost surely takes into account characteristics of a given hotel that are unobserved to the researcher, such as any unobserved "quality" or whatever affects the overall relation between the hotel and the platform, such as the amount of commission

[16] This approach works only limitedly for Cancun, as many addresses are misspecified for that city.

[17] From this dataset, I remove Airbnbs that seem to be inactive (i.e. those that have not had any booking for the past 12 months).

the hotel pays. As long as these do not vary over time, they are accounted for by the fixed effect. The focus of this analysis is therefore on how a *given* hotel's ranking on Expedia reacts to the price it sets, holding the hotel's identity fixed.

4.1.1 Endogeneity and Instrumental Variables

The hotels' positions and prices observed in the data are equilibrium values that are simultaneously determined by both, a hotel's reaction to the ranking algorithm, as well as the algorithm's elasticity with respect to hotels' prices. As a result, the variables of interest in equation (2) are endogenous, and any ordinary least squares estimation would yield biased estimates.[18] A solution is to use an instrumental variable techniques. Valid instruments are variables that exogenously shift price or promotion decisions of a given hotel across observations, and are orthogonal to anything that may influence a hotel's position over time. Instruments commonly used in the empirical IO literature are based on firms' marginal costs. However, hotel prices vary strongly even across days of a week, which cannot reflect marginal costs but rather changing demand across travel dates. Hotels' pricing policies are more likely driven by *opportunity* costs, and are thus higher for high-demand, and lower for low-demand days. This paper provides estimates using different kinds of instruments, each of which have their advantages and disadvantages.

Neighborhood Instruments

As a first instrument, I use the prices and promotion dummies of hotels that are located within 500 meters distance of the focal hotel. The intuition for this instrument stems from empirical observations and theoretical insights concerning hotels' price-setting decisions in practice. Anecdotally, when conducting an interview with a representative of the German hotel association (Hotelverband Deutschland) in February 2020, I was told that hotels often employs revenue management software that suggests hotel prices and takes into account competitors' prices as well as major events taking place in the neighborhood. In the academic literature, Li et al. (2018) find that hotels seem to closely match the prices of those hotels which they believe to be close competitors (i.e. hotels that are located in the same neighborhood and/or have the same star rating as the

[18] Under reasonable assumptions, one would expect the bias to be negative.

focal hotel). Cho et al. (2018) equally find that empirically, competing hotels' prices strongly co-move. Moreover, hotels located close to each other and possibly having the same star rating are likely to experience common demand shocks, for instance because of an event happening in that neighborhood, or due to changing patterns in business and leisure travellers. Indeed, Schaefer and Tran (2020) find a higher substitutability between hospitality businesses located close to each other in the same district, and emphasize the importance of localized competition between hotels. Building up on these empirical facts, the logic of this neighborhood-based instrument is therefore that prices of neighboring hotels constitute a measure of local, short-term demand. Prices of neighboring hotels should thus be correlated with the focal hotel's price, but presumably unlikely to influence the ranking of the focal hotel. Indeed, over 2,500 active hotels are listed on Expedia for Paris that all appear in the ranked list. Therefore, I assume it is be unlikely that a few neighboring hotels' prices alone would influence the focal hotels' position.

All in all, identification rests on the assumption that local demand conditions influence hotels' price setting decisions, and that hotels have better knowledge about these conditions than Expedia does so that local demand conditions do not affect the hotel's rank. If Expedia was able to anticipate future bookings of consumers as good as hotels, then Expedia would possibly display hotels with a high demand shock more visibly while these hotel also raise their prices.

Brand-based Instruments

An observation that has been made in other market settings is that prices tend to be set relatively uniformly across outlets of a certain brand. DellaVigna and Gentzkow (2019) show for US groceries and drug store chains that prices within a given brand show substantially less variation than prices between brands, even though consumer demographics may vary widely between regions. The scraped dataset I use contains 54 different brands of hotels for Paris alone, each of which has an average of 10.9 distinct outlets. I find that the variation of prices and sales within a given brand is much smaller than variation of prices between brands. I therefore use prices and promotion dummies of hotels of the same brand as instrument for the focal hotel's pricing and promotion decision, as these are correlated with the focal hotel's decisions, but should not causally influence the focal hotel's position in the ranking. The identification assumption is that pricing and promotion are being made on a city-wide level and less in response to the ranking. A drawback of this instrument is that its application is of course limited to those hotels which can be identified as brand hotels in the data, which are for Paris only roughly 26% of all hotels.

Airbnb and Events Instruments

The next instrument is based on an intuition that is similar to the one of the neighborhood instruments. First, prices of Airbnbs that are located in proximity (here, in the same zipcode) as the focal hotel are likely to be correlated with the focal hotel's price, but not with the focal hotel's position. As Airbnbs do not offer promotions, I use dummies for certain well-known events (public holidays or the end of the Tour de France for Paris, for instance) happening in the city as an additional instrument. Again, the exogeneity requirement is similar to the one of the neighborhood instrument.

The Airbnb data does not contain any variation of prices across *query* dates. Therefore, I create a "panel" of hotels by collapsing multiple observation of a given travel date (stemming from multiple queries made for that travel date) to a *single observation per hotel per travel date*. The analysis is thus performed on the average values of price, position etc. for a given hotel and travel date, with averages computed over query dates.

A drawback of the instrument is that the variation in Airbnb prices across travel dates is relatively weak. As Figure 4 shows, while there is some seasonal variation across months as well as variation depending on whether the booking is for a weekend or for a weekday, price variation for Airbnbs is still substantially smaller than for hotels.

Dynamic Panel Instruments

A technique that has been employed in the literature especially for production function estimation is the use of dynamic panel instruments. To use these instruments, I am again constructing a "panel" of hotels as in 4.1.1, so that an observation in the data is one hotel - travel date.

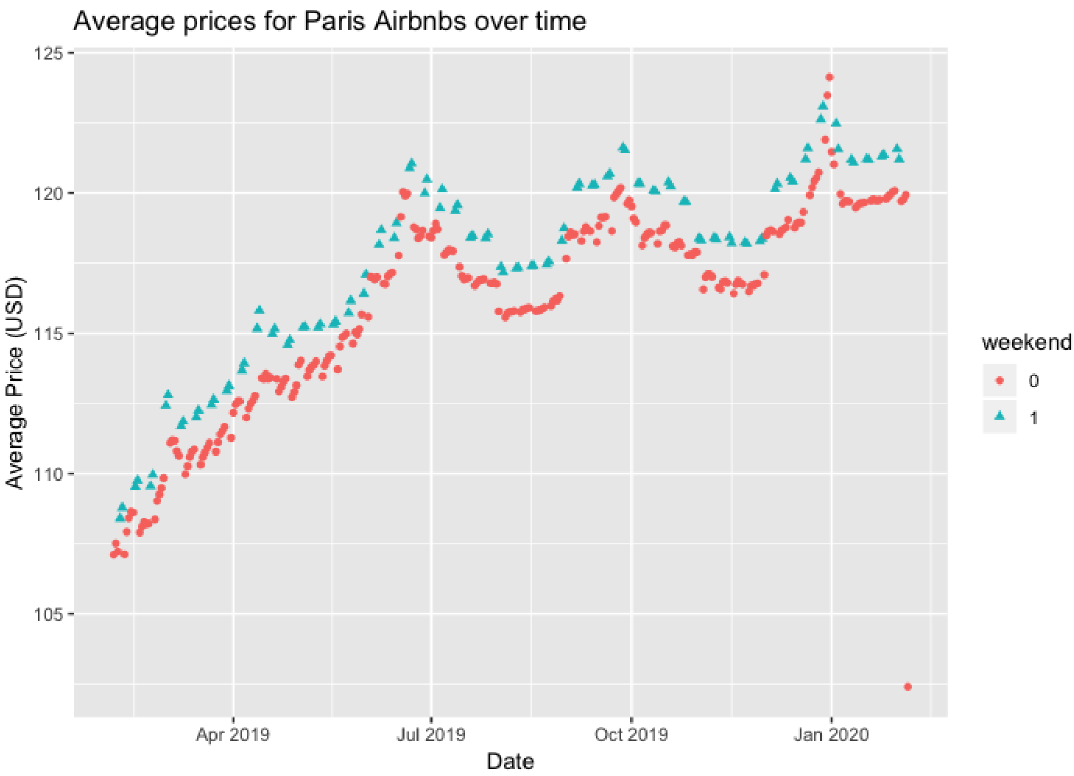


Figure 4: Prices of Airbnb in the whole area of Paris over travel dates.

Then, given J large and T small, under the assumption that $\mathbb{E}[\epsilon_{jt}] = \mathbb{E}[\epsilon_{jt}\epsilon_{js}] = 0$ for $t \neq s$ (i.e. absent serial correlation in the error terms), one can use lagged values of the endogenous variables as instruments in the equation in first differences.

In the given case, taking differences, this results in the following equation that one can estimate:

$$\Delta r_{jt} = \alpha \Delta avg\_p_{jt} + \beta \Delta avg\_d_{jt} + \delta \Delta avg\_past\_book_{jt} + \Delta \lambda_t + \epsilon_{jt} \tag{3}$$

where Δ indicates differences, i.e. $\Delta r_{jt} = r_{jt} - r_{j(t-1)}$. Hotel fixed effects are differenced out with this operation. A standard possibility is now to use second or even further lags of the endogenous variable (i.e. $p_{j,(t-2)}$ and $d_{j,(t-2)}$) as instruments for the variables in this differenced equation. This method is also known as the difference GMM (Arellano and Bond, 1991).

4.2 Demand Side: Utility Parameters

Ursu (2018) uses data from randomly ranked hotels (where a hotel's quality is independent from its position in the ranking) in order to estimate the relevant parameters. Since the listings of hotels observed in the WCAI search data are not randomly ranked, I use the estimated search cost parameter from Ursu (2018) - denoted $\hat{k}$ and $\hat{\gamma}$ in her paper - to avoid any bias due to the relevance-based, default ranking. Thus, I will not estimate the search cost parameters k and γ, but only the utility parameters of hotel observables such as stars or location score.

In principle, one could simply use Ursu's estimated parameters themselves. However, the cities which hotels are located in are not revealed in Ursu's dataset; and consumers who book in Paris might have different tastes for stars, location etc. than consumers booking at other cities. I therefore employ her model to estimate the utility parameters (i.e. valuations of hotel characteristics and outside option) myself, using the consumer search data for Paris.[19] Just like Ursu (2018), I also do not instrument for prices. I thus assume that conditional on a given entered query, the observed price variation is unlikely to be correlated with the error term in the utility, and mostly captured by travel date and query characteristics. The use of the estimates on the demand side enables me to simulate consumers' purchasing choices under any counterfactual ranking of hotels.

5 Results

Both the ranking as well as the consumer side estimations are based on data from Paris.

5.1 Ranking

Tables 7 to 10 display the results. In each table, the results using instruments for prices and sales are juxtaposed with the OLS results based on the same data. All specifications include hotel and query-travel-date fixed effects (travel-date fixed effects only for Tables 9 and 10, which are based on a "panel" of hotels).

[19] The underlying assumption is here that consumers' search costs do not depend which city a consumer is considering. However, I do allow for variation in consumers' valuations for hotel characteristics by estimating utility parameters using consumer search data from Paris directly.

Across specifications, I find a relatively robust positive coefficient for price that, using instruments, lies between 3.5 and 4 for the query-level results in Tables 7 and 8, and between 2.2 and 3.3 for the panel based results in Tables 9 and 10. The coefficient is significantly different from 0 in all specifications. Its value implies that a 1 dollar increase in price would lead to a shift in the hotel's position by two to four places towards the lower end of the page (i.e. an increase in the position number). Given the extent to which hotels' prices vary in practice, this effect is quite substantial, and would imply that hotels may be quite substantially pressured to set lower prices by the default ranking. The ranking's elasticity with respect to hotels' prices therefore seems to be economically meaningful.

For the coefficient of the sales dummy, the results give a very blurry picture. While it is negative and varies between -63 and -46 in all OLS specifications, it jumps up, or down, to a substantial extent once one employs instrumental variables. The reason for this is most likely that the instruments for sale are very weak. Tables 15 to 19 in the Appendix display the first stages, where I regress prices, or the sales dummies, on the instruments and explanatory variables, using the same set of fixed effects. Throughout the first stage results, the coefficient of determination (adjusted R^2) for regressions on the sales dummy are very low, as are the resulting coefficients in the regression. The F-statistics, however, are quite high. In general, it is difficult to draw conclusions from the data as to when hotels decide to offer a promotion, so finding suitable instruments for the promotion variable is very challenging.

All in all, my results suggest that Expedia's hotel ranking intensifies price competition between hotels by pushing hotels to less visible positions if they offer a higher price. The implication is that the elasticity of demand that hotels face is higher compared to a situation in which OTAs rank hotels randomly, for instance. Consequently, hotel markups are lower in the situation with such a ranking algorithm. Another potential implication could be that hotels provide costly add-ons or may try to employ, to some extent, the obfuscation strategies that are explained in Ellison and Ellison (2009).

5.1.1 Robustness

Possible endogeneity of neighborhood instruments As noted above, the exogeneity of the neighborhood instrument relies on the assumption that the prices of neighboring

Table 7: Results: Neighborhood-based instruments

	Dependent variable:					
	position					
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
price	3.494***	3.478***	3.583***	3.619***	3.590***	3.588***
	(0.099)	(0.11)	(0.11)	(0.12)	(0.12)	(0.13)
sale	−47.244***	−460.930***	−49.284***	265.031*	−48.968***	194.581**
	(4.11)	(136.38)	(4.53)	(113.67)	(4.69)	(62.11)
pastbookings_count	−0.743***	−0.451*	−0.881***	−1.098**	−0.952***	−1.095**
	(0.21)	(0.207)	(0.26)	(0.34)	(0.28)	(0.34)
Hotel FE	x	x	x	x	x	x
Query-Traveldate FE	x	x	x	x	x	x
Observations	912,255	912,255	679,806	679,806	632,235	632,235
R^2	0.196	0.136	0.216	0.150	0.217	0.173
Adjusted R^2	0.194	0.165	0.213	0.147	0.214	0.170

Note: *p<0.1; **p<0.05; ***p<0.01

Neighboring hotels are defined as being located within 500 meters of the focal hotel. Column (2) uses as instruments the number of neighboring hotels available on a given query and travel date, and the average price and sales dummy of neighboring hotels at a given travel and query date. Column (4) uses the same instrument, but averages over neighboring hotels with the same star rating as the focal hotel only. Column (6) does again the same, but averages over neighboring hotels that (i) have the same star rating and (ii) have the same value for a dummy indicating whether the hotel is part of a brand or not. Columns (1), (3) and (5 display the OLS results using the corresponding observations that are used in the IV results (number of observations varies due to missing values). Standard errors are heteroskedasticity robust.

Table 8: Results: Brand-based instruments

	Dependent variable:		
	position		
	OLS	IV for price and sale	IV for price and sale
price	3.536***	3.550***	3.592***
	(0.149)	(0.192)	(0.191)
sale	−63.027***	−37.777	−28.081
	(9.663)	(31.847)	(31.564)
pastbookings_count	−0.105	−0.124	−0.130
	(0.243)	(0.244)	(0.245)
Hotel FE	x	x	x
Query-Traveldate FE	x	x	x
Observations	257,735	257,735	257,735
R^2	0.188	0.187	0.187
Adjusted R^2	0.184	0.183	0.183
F Statistic	19,748.450*** (df = 3; 256518)	4,949.646***	5,063.422***
Note:			*p<0.1; **p<0.05; ***p<0.01

Column (1) shows simple OLS results (using only those hotels for which brand-based instruments can be constructed, i.e. hotels that belong to a brand with at least one other outlet in Paris.) Column (2) displays the results when instrumenting for both price and sale and uses as instruments: the average price and sale dummy of the other brand outlets, as well as the number of outlets. Column (3) uses as instrument in addition the average discount in percent, the average of a dummy indicating whether hotel cancellation is free, and the average of a dummy indicating if any free extras are offered. 574 Parisian hotels are included in the data for these results. Standard errors are heteroskedasticity robust.

Table 9: Results: Airbnb and Events × zipcode instruments (based on "panel" of hotels)

	Dependent variable:	
	mean_position	
	(1)	(2)
avg_price	3.221***	2.207***
	(0.114)	(0.143)
avg_sale	−50.475***	44.304
	(5.491)	(52.975)
avg_pastbookings	−1.046*	−1.130*
	(0.443)	(0.498)
Hotel FE	x	x
Date FE	x	x
Observations	446,298	446,298
R^2	0.217	0.209
Adjusted R^2	0.214	0.205
F Statistic	41,140.250*** (df = 3; 444057)	1,947.599***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Events included as IVs (interacted with zipcodes) are national holidays as well as major events that might affect different parts of the city to different extents: fete du travail; Tour de France; Marathon de Paris; Fete de la Musique; Armistice; Toussaint; Fete de la Victoire; Ascension; Pentecote; Easter; National Day; Assomption; Fashion Week. Standard errors are heteroskedasticity robust.

Table 10: Results: Arellano-Bond (based on "panel" of hotels)

	Dependent variable:	
	avg_position	
	panel linear	*panel GMM*
	(1)	(2)
avg_price	3.141***	3.297***
	(0.009)	(0.080)
avg_sale	−48.608***	−169.188***
	(1.564)	(17.453)
avg_pastbookings	−0.489***	0.282***
	(0.142)	(0.008)
Observations	565,525	565,525
# hotels	2,630	2,630
R^2	0.200	
Adjusted R^2	0.196	
F Statistic	47,028.160*** (df = 3; 562605)	
Note:		*p<0.1; **p<0.05; ***p<0.01

Column (1) shows simple OLS results using the hotel panel that is constructed by averaging prices, sale dummies etc. for each hotel-travel date over query dates. Column (2) shows Arellano-Bond estimates using the second lag of price and sale as instrument. P-values of Sargan test and autocorrelation test for Arellano-Bond results are almost 0.

hotels do not causally influence the position of the focal hotel. This assumption would not hold if Expedia is well informed about local demand shocks. If Expedia was able to anticipate a positive shift in demand for rooms in a given neighborhood (for example based on users' search and click behavior), then it might have an incentive to display hotels in that neighborhood very visibly despite their increased prices. This would lead to a downward bias in the estimated price coefficient, as hotels would be ranked at visible positions (low position numbers) despite their high prices.[20] In this exercise, I therefore redefine the definition of a hotel's neighbors by determining that neighbors are hotels located within a ring-shaped area surrounding the focal hotel between the 500 meter and 1 kilometer radius. Prices of these hotels are possibly less likely to causally influence the focal hotel's position, but might still well be correlated with its prices. The results are displayed in Table 11. In specifications (4) and (6), I find that the price coefficient tends to be slightly lower than in the corresponding specifications in Table 7 as well as compared to the OLS results in columns (3) and (5). This is somewhat surprising since, if there is any bias in the Table above, one would expect it to be negative. The instrumented sale dummy coefficients now all have a positive sign, but again widely differ in magnitude.

Preliminary evidence moreover shows that the results are quantitatively somewhat different across destinations, although they all point in the same directions. This is plausible, since the competitive environments across the different cities seem very different (many small hotels (Paris) vs. few large hotels (Cancun); strong vs. weak seasonality; high number of business travellers vs. holiday makers, etc.). Expedia might to some extent adjust its ranking algorithm to the different types of travellers it expects for a given city; and hotels may focus on different segments of demand, and appear thus to engage in different pricing strategies.

5.2 Demand

Table 12 displays the demand estimates which are derived using Ursu's (2018) sequential search model and the WCAI data on consumer searches in Paris. The estimation is based on data of only those searches that contain at least one click (which comprise in

[20] If hotels were more likely to be on sale when demand is low, and Expedia was informed about local demand shifts, the bias in the coefficient of the sale dummy would be biased upwards. However, from the data, the determinants of a hotel's decision to go on sale are not completely clear.

Table 11: Results: Neighborhood-based instruments (“neighbor” = hotels $\in$ [500m, 1,000m) from focal hotel)

	Dependent variable:					
	position					
	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)
price	3.476***	3.377***	3.567***	3.138***	3.559***	3.393***
	(0.098)	(0.166)	(0.106)	(0.188)	(0.110)	(0.124)
sale	−47.536***	−934.468***	−49.161***	−926.684**	−47.829***	−243.046*
	(4.096)	(180.328)	(4.378)	(302.272)	(4.473)	(107.248)
pastbookings_count	−0.696***	−0.137	−0.992***	−0.400	−1.106**	−0.994**
	(0.202)	(0.250)	(0.296)	(0.338)	(0.336)	(0.309)
Observations	924,809	924,809	716,282	716,282	679,565	679,565
R^2	0.195	0.076	0.214	0.077	0.215	0.190
Adjusted R^2	0.193	0.074	0.212	0.075	0.213	0.188

Note: *p<0.1; **p<0.05; ***p<0.01

Neighboring hotels are being defined as lying in a “donut” shaped space between 500 meters and 1,000 meters around the focal hotel. As before, column (2) uses as instruments the number of neighboring hotels available on a given query and travel date, and the average price and sales dummy of neighboring hotels at a given travel and query date. Column (4) uses the same instrument, but averages over neighboring hotels with the same star rating as the focal hotel only. Column (6) does again the same, but averages over neighboring hotels that (i) have the same star rating and (ii) have the same value for a dummy indicating whether the hotel is part of a brand or not. Columns (1), (3) and (5 display the OLS results using the corresponding observations that are used in the IV results (number of observations varies due to missing values). Standard errors are heteroskedasticity robust.

Table 12: Search model estimates using WCAI Paris data. Results based on different starting values and different definitions of "location score".

	Using simple location score:		*Using more complex location score:*	
	(1a)	(1b)	(2a)	(2b)
price ($100)	-0.1757***	-0.17581***	-0.17882***	-0.17882***
	(0.020842)	(0.021658)	(0.0267)	(0.0267)
stars	0.072024***	0.072148***	0.075358***	0.075359***
	(0.022586)	(0.022447)	(0.026354)	(0.026331)
review score	0.033021**	0.033240*	0.034854*	0.034855*
	(0.019051)	(0.020929)	(0.025957)	(0.025974)
chain dummy	-0.047307*	-0.047446*	-0.041674*	-0.041674*
	(0.030171)	(0.030202)	(0.030879)	(0.030880)
location score alt1	0.035425***	0.035443***		
	(0.010758)	(0.010942)		
location score alt2			0.044026***	0.044026***
			(0.012534)	(0.012530)
promotion	0.052148	0.052524	0.049293	0.049288
	(0.046288)	(0.047480)	(0.054909)	(0.054923)
outside option	0.62017***	0.6215***	0.6681***	0.6681***
	(0.000000)	(0.000000)	(0.086314)	(0.086435)
Log-likelihood	-3,739.3	-3,739.3	-3737.8	-3737.8
# individuals	1,051	1,051	1,051	1,051
# observations	31,874	31,874	31,874	31,874

Note: *p<0.1; **p<0.05; ***p<0.01

The definition of the location score for the two columns on the left is based on the distance to the Louvre in Paris. The definition of the location score for the two columns on the right is based on the minimum distance to any of these locations: Sacré Coeur, Louvre, Les Invalides, Hotel de Ville. Results in the first and third column (columns a) are based on using the zero vector as a starting value. Results in the second and fourth column (columns b) are based on using $[0.3, -0.2, 0, 0, 0.2, -0.2, 0]$ as a starting vector.

fact only 16% of all searches): The searches with no clicks are likely to be carried out by customers that may not be "seriously" searching for a hotel or by bots scraping the page.[21] Moreover, as explained above, I use the search cost parameter estimates from Ursu (2018) instead of estimating them again (I use the results derived in column (1) of Table 8 in her paper). Thus, I set $\hat{\gamma} = 0.0044$ as a position parameter, and $\hat{k} = -1.0305$ as the search cost constant. Columns (1a) and (1b) use a location score defined by a hotel's distance to the Louvre in Paris, whereas columns (2a) and (2b) use a location score based on the minimum distances to a variety of touristy landmarks in Paris. Columns (1a) and (2a) use different starting values for the simulated maximum likelihood estimation than columns (1b) and (2b), showing that the starting value does not affect results very much. The price coefficient has the expected sign, being significantly negative and is comparatively large in magnitude. Stars, the location score (with a high index meaning better location), and the outside option are all positive significant. The coefficient of the review score has a positive sign, but is only marginally significant. The promotion parameter has a positive sign, but is insignificant, which reflects Ursu's (2018) results where this coefficient is insignificant in three out of the four destinations. All in all, the results I obtained seem intuitive and are in line with the results of both Ursu (2018) as well as Chen and Yao (2018).

6 Counterfactuals

6.1 Method

As consumers tend to click and purchase only hotels being displayed on the first results page, I carry out counterfactuals and compute the costs only for the first results page, for each query-travel date observed in the data.[22] I moreover only simulate changes in hotels' pricing decisions, and abstract from any changes in hotels' willingness to offer a discount. I first show how the estimates from above (the demand parameters, and the ranking's elasticitiy with respect to prices and positions) along with observed prices imply marginal costs for each hotel observation. Recall that $\hat{\alpha}$ denotes an estimate of $\frac{\partial r_j}{\partial p_j}$,

[21] Information on the exact data cleaning process can be found in the Appendix, Section E.

[22] These are 31,638 observations of 857 unique hotels displayed across 745 distinct queries.

i.e. the effect of a hotel's price on a hotel's position from the tables above. Re-arranging the first order condition (1) yields:

$$\tilde{c}_{jqt} \quad = \quad p_{jqt} \quad - \quad \tilde{d}_{jqt} \quad + \quad \underbrace{\frac{s_{jqt}(\cdot)}{\frac{\partial s_{jqt}(\cdot)}{\partial p_{jqt}} + \frac{\partial s_{jqt}(\cdot)}{\partial r_{jqt}} \cdot \hat{\alpha}}}_{<0} \tag{4}$$

Prices p_{jqt} and promotion decisions $d_{jqt} \in \{0, 20\}$ are observed in the data. To approximately match Expedia's commission rate, I set $\tau = 0.2$. Recall that and $\tilde{d}_{jqt} \equiv \frac{d_{jqt}}{(1-\tau)}$, so that

$$\tilde{d}_{jqt} = \begin{cases} 25 & \text{if hotel is on promotion } (d_{jqt} = 20), \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

Market shares $s_{jqt}(\cdot)$ are computed by simulating 100,000 consumers that are being confronted with a given sequence of hotel listings, and choosing which hotels rooms to book, if any.

The parameters $\frac{\partial s_{jqt}(\cdot)}{\partial p_{jqt}}$ and $\frac{\partial s_{jqt}(\cdot)}{\partial r_{jqt}}$ can only be derived by simulation, as no closed-form solution exists. Starting with the data of observed hotel listings for each given query-travel dates, I compute the following:

- For $\frac{\partial s_{jqt}(\cdot)}{\partial p_{jqt}}$, I compute what an increase in its price of 5% implies for its market share, by simulating 1000 consumers who view the given hotel in query q for date t.
- For $\frac{\partial s_{jqt}(\cdot)}{\partial r_{jqt}}$, I equivalently increase each hotel's position number by 1, and by simulating 1000 consumers I compute how this affects the hotel's market share.

For each hotel, I first average the simulated derivate across all observations of the given hotel to obtain $\frac{\partial \bar{s_j}(\cdot)}{\partial p_j}$ and $\frac{\partial \bar{s_j}(\cdot)}{\partial r_j}$. Assuming that these parameters should be the same across hotels, I then take the average again, obtaining a value for $\frac{\partial \bar{s}(\cdot)}{\partial p}$ and for $\frac{\partial \bar{s}(\cdot)}{\partial r}$, respectively. This results in the following equation to back out the costs:

$$\tilde{c}_{jqt} \quad = \quad p_{jqt} \quad - \quad \tilde{d}_{jqt} \quad + \quad \frac{s_{jqt}(\cdot)}{\frac{\partial \bar{s}(\cdot)}{\partial p} + \frac{\partial \bar{s}(\cdot)}{\partial r} \cdot \hat{\alpha}} \tag{6}$$

Given all parameters and hotel's backed-out costs, one can perturb $\hat{\alpha}$, for instance, and simulate the prices and market shares which hotels are going to face under this scenario. Markups can easily be derived by transforming equation 6:

$$\frac{p_{jqt} - \tilde{d}_{jqt} - \tilde{c}_{jqt}}{\tilde{p}_{jqt}} = \frac{s_{jqt}(\cdot)}{p_{jqt} \cdot (\frac{\partial \bar{s}(\cdot)}{\partial p} + \frac{\partial \bar{s}(\cdot)}{\partial r} \cdot \hat{\alpha})} \tag{7}$$

6.2 Results for costs and interpretation

I obtain $\frac{\partial \bar{s}(\cdot)}{\partial p} = -0.014$ and $\frac{\partial \bar{s}(\cdot)}{\partial r} = -0.032$ from the simulations above. Moreover, I use $\hat{\alpha} = 3.2$ to back out the costs.

Averaged across all hotel observations, I find average costs of 206.4 €, whereas average prices for the sample of hotels are 217.83 € (see Table 13). The substantial variance in backed-out costs is reflective of the substantial variation in prices across hotel observations.

Markups averaged across hotel observations are extremely low. In fact, the model implies markups equal to 0 for 63% of hotel observations. In other words, the model predicts that those hotels set prices equal to marginal cost, and an average markup across all hotel observations of close to 0. In contrast, other research like Cazaubiel et al. (2018) find average markups of 35-43%. The reason for my finding is the peculiarity of the search model generating very sparse demand. Indeed, the 63% of hotel observations for which markups are supposedly 0 also generate 0 demand in a given query, so that by construction marginal costs are predicted to be equal to prices (see equation 6). Only 5% of hotel observations derive a market share of more than 10%, and only 2% a market share of more than 20%. The relatively high search costs contribute to this finding, and possibly also be the lack in consumer heterogeneity in preferences of the sequential search model. In reality, it is unlikely that certain hotels should obtain absolutely 0 demand on Expedia, as hotels also derive bookings through other channels which cannot be taken into account in my demand estimation. Moreover, in my model, a hotel's position in a given query affects the hotel's market share, which in turn affects the hotel's marginal costs. A hotel that is displayed on the bottom of a page in a given query may thus derive no demand, and therefore my model predicts that it should have set prices equal to marginal cost. Finally, markups are low precisely because the model predicts that the algorithm as it is intensifies price competition quite substantially through the additional term $\frac{\partial \bar{s}(\cdot)}{\partial r} \cdot \hat{\alpha}$ that enters the denominator in the first order condition.

These aspects highlight that in this setting, marginal costs cannot be interpreted as physical marginal costs. In reality hotels most likely face very high fixed costs and very low marginal costs of a certain occupation of a room. A hotel's physical marginal costs should not depend on the position in which it appears in a given query. Moreover, hotels' price-setting decisions are in reality dynamic problems since hotels are capacity constrained. When setting prices, hotels optimally take into account their current occupancy and the expected demand of rooms. All in all, I believe that marginal costs should here rather be interpreted as the option value of having a room booked at a given date.

Table 14 shows the distribution of costs, markups and prices across hotels of a given star rating. It is perhaps noteworthy that markups are *higher* for hotels that with 2, 3 or 4 stars, compared to hotels with 5 stars. This is again a result of some very high-priced, luxury hotels having a zero or very small market share, and thus having a low value for $s_{jqt}(\cdot)$.

Table 13: Basic descriptives: prices, costs and markups

	Average	Median	Std Dev
Prices	217.83	150	213.18
Costs	206.4	140	215.0
Markups	8.4 $\times 10^4$	0	0.003

Table 14: Average prices, costs and markups, by star rating

	1 star	2 stars	3 stars	4 stars	5 stars
Prices	120.44	100.88	120.83	151.6	423.83
Costs	114.19	89.36	109.62	142.48	417.24
Markups	3 $\times 10^{-7}$	3.9 $\times 10^{-4}$	4.2 $\times 10^{-4}$	6.7 $\times 10^{-4}$	9.3 $\times 10^{-5}$
# of hotels	4	54	312	403	83

6.3 Simulating Counterfactuals

The model assumptions and estimates derived above allow to assess how the change in the ranking algorithm or other parameters impact equilibrium prices and market shares.

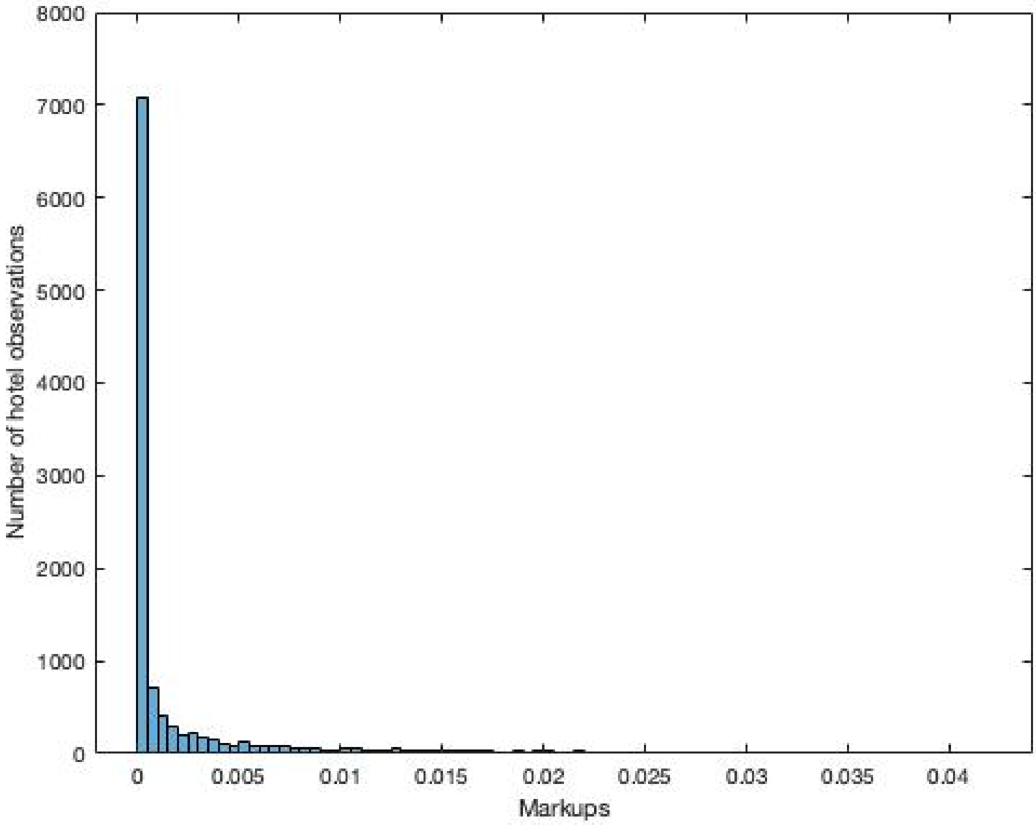


Figure 5: Distribution of markups across hotel observations, after removing the 63% of observations with markups = 0.

I perform counterfactual simulations using 100,000 simulated consumers for those hotel observations with positive market shares (the other hotel observations would not adjust prices in response anyways, since their prices are being fixed equal to marginal costs).

I first simulate prices and market shares under the scenario in which the ranking's elasticity with respect to prices, $\frac{\partial r_j(\cdot)}{\partial p_j} = \alpha$, is set to 32, i.e. increased by a factor 10. I find that in response, as expected, all hotels react by decreasing their price. However, the price increase is very small and amounts to only 23 cents on average. This is a result of the market shares being very often small, so that the effect of a stronger ranking yields only in a limited reaction by hotels. Moreover, recall that hotels *already* are under a lot of pricing pressure given *any* effect from the ranking, which already lowers their markups in contrast to a situation without such a ranking.

In a next step, I therefore compute equilibrium prices and market shares in a scenario in which the ranking does not matter, i.e. $\alpha = 0$. I find much larger effects on prices.

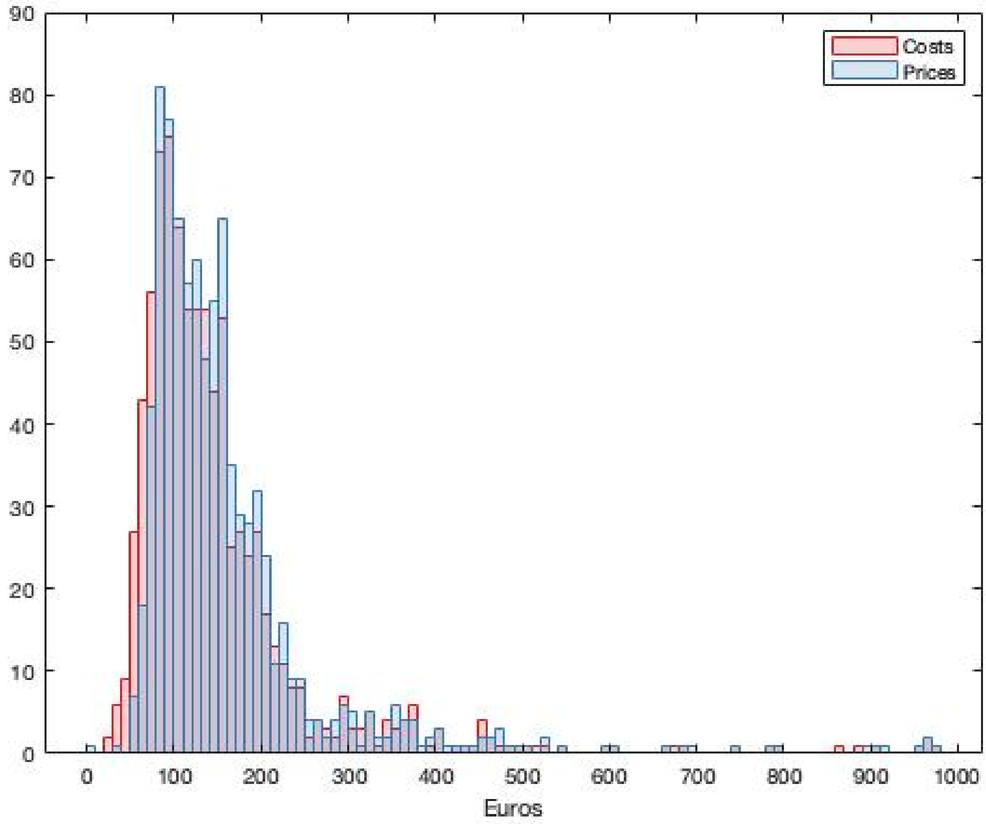


Figure 6: Distribution of costs and average prices across hotel observations.

Whereas the average price increase amounts to only 1,89 euros, 7% of hotels increase their price by more than 10 euros. Overall, while there is significant heterogeneity in hotels' price adjustments. The price adjustments for the majority of hotels are still relatively low, however, due to the small market shares (see Figure 7).

Future work could go further and not only change the elasticity parameter, but also the actual displayed ranking of hotels (and thus show hotels in alternative orderings to the simulated consumers). In contrast, above I vary only the parameter that governs the ranking effect on hotels' prices, without perturbing the ranking itself. This would yield substantially different market shares for hotels. Another interesting exercise would be to compare the current simulations and observed market shares with a scenario in which rankings are random. In addition, it would be interesting to compare the observed prices and market shares to a scenario in which consumers have perfect information (0 search costs) and do not search at all.

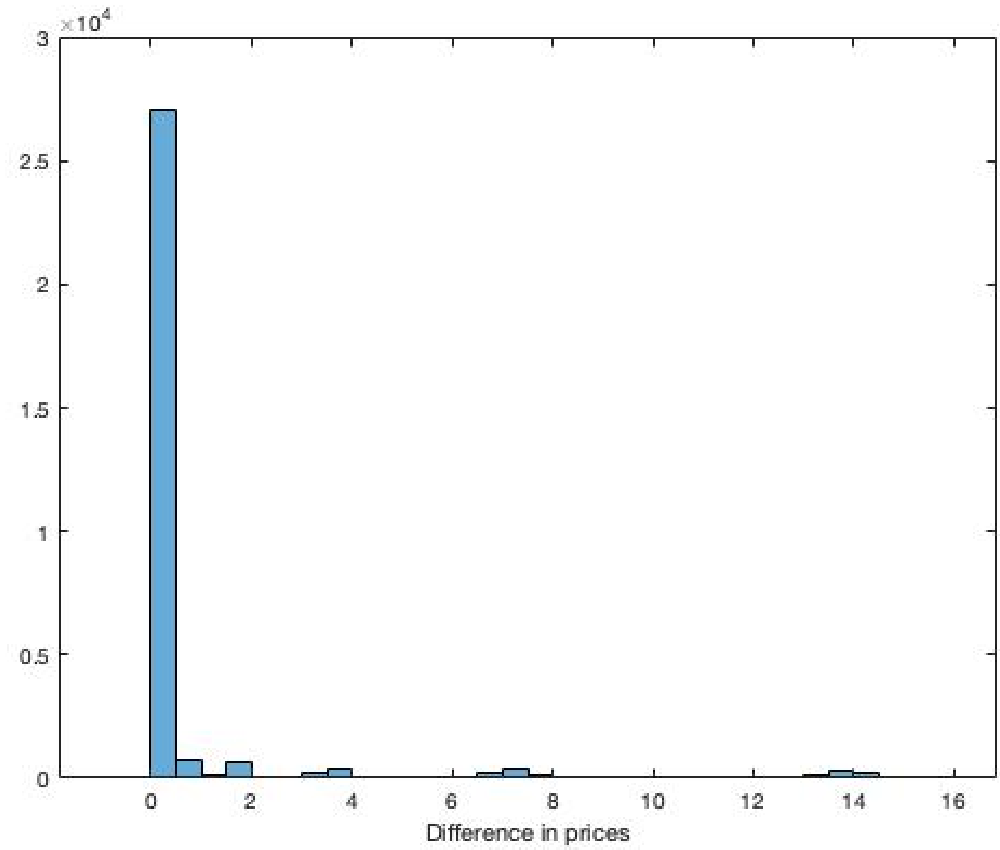


Figure 7: Change in prices (in euros) when effect of ranking is shut off.

7 Conclusion

This paper sheds light on the fact that product rankings on e-commerce platforms do not only affect buyers' clicks and purchases, but also the prices that sellers on these platforms optimally want to set. It therefore highlights that platform design can have intricate consequences for how sellers and buyers interact on an online marketplace, ultimately affecting the overall functioning of the market. Importantly, these equilibrium effects cannot be traced out using A/B experiments that randomize the ranking algorithm across buyers. Instead, they require a structural model of the buyer, as well as the seller side, in order to predict how changes of the ranking algorithm affect sellers' equilibrium prices.

This paper contributes to the literature on product rankings and recommendations, which are ubiquitous on most e-commerce platforms given the cornucopia of goods and services sold online. How these rankings affect the demand-side has been explored by a growing empirical literature, which has established a strong causal link from rankings to the likelihood that a given product is seen and consumed. This, however, implies that the *sellers* of goods and services critically depend on visibility online. The large number of products on online marketplaces and consumers' limited attention leads to competition between sellers for the most visible places online. Rankings that display a product more prominently when it is cheaper therefore influence sellers' price-setting decisions, as sellers then compete even more intensely for visibility.

To shed light on the equilibrium effects of online rankings, I use a simple hotel pricing model on the supply side, and a sequential search model from Ursu (2018) on the demand side. The empirical analysis shows that Expedia's ranking is likely to intensify competition between hotels, and it does so to an economically meaningful extent: Using data from Paris, I find that a hotel is being displayed by 3 positions lower if it increases its price by one dollar. From the observed prices and the estimated search model parameters, the model allows to back out hotels' marginal costs, which will be crucial for comparing markups and prices under counterfactual settings. The fact that the platform intensifies price competition could be good news for consumers, as they would benefit from lower prices on average, and rather bad news for hotels.[23] On the other hand, intensified price competition could also lead to hotels increasingly engaging in obfuscation (for example by offering pricy add-ons or more comfortable rooms that are only visible upon clicking on a hotel), or to a decrease in quality provision by sellers.

The analysis suffers from several limitations. The linear relationship between rank and hotels' prices is a strong abstraction of a platform's ranking algorithm, which is possibly highly nonlinear. The model of hotels' price-setting decisions is equally highly simplified. In practice, hotels engage in dynamic revenue management as their inventory is limited and expected demand may change as a given travel date comes closer. Not only prices, but also how hotels set promotions is not accurately captured by the model as it is. In practice, hotels are highly heterogeneous in the extent to which they are willing to offer a sale and at what times, and it is difficult to rationalize these decisions within a model. Hotels may moreover attract consumers via different channels (for example via their own website, or via the non-default rankings on the online travel

[23]Complaints by hotels about intermediators and rankings supports this idea; see for instance here (in German): https://www.hotellerie.de/go/ranking-bad.

agent). I also fail to find a convincing instrument that shifts a hotel's decision to offer a promotion.

Future research could attempt at solving any of these issues. A further avenue is to explore the impact of horizontal and vertical competition on a given hotel page on hotels' price setting decisions: A hotel displayed next to its closest competitor is likely to face much stronger competition than a hotel that is displayed along with competitors that horizontally and vertically different (e.g. different star rating, different geographical location). Lastly, it would be interesting to analyze whether mandating platforms to increase the transparency of online rankings, which has been demanded by sellers on online marketplaces as well as consumer advocates, benefits or harms the functioning of a market.

References

Arellano, M., and Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. The review of economic studies, 58(2), pp. 277-297.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. Econometrica: Journal of the Econometric Society, 841-890.

Cazaubiel, A., Cure, M., Johansen, B. O., and Vergé, T. (2018). Substitution Between Online Distribution Channels: Evidence from the Oslo Hotel Market. University of Bergen Working Paper No. 08/18. Available at: https://www.uib.no/sites/w3.uib.no/files/attachments/wp_08-18.pdf.

Chen, Y., and Yao, S. (2016). Sequential search with refinement: Model and application with click-stream data. Management Science, 63(12), pp.4345-4365.

Cho, S., Lee, G., Rust, J., and Yu, M. (2018). Optimal Dynamic Hotel Pricing. Working Paper. Available here.

DellaVigna, S. and Gentzkow, M. (2019). Uniform Pricing in U.S. Retail Chains. Quarterly Journal of Economics 134(4), pp. 2011-2084.

De los Santos, B. and Koulayev, S., (2017). Optimizing Click-Through in Online Rankings with Endogenous Search Refinement. Marketing Science, 36(4), pp.542-564.

Dinerstein, M., Einav, L., Levin, J., and Sundaresan, N. (2018). Consumer price search and platform design in internet commerce. American Economic Review, 108(7), pp.1820-1859.

Ellison, G., and Ellison, S. F. (2009). Search, obfuscation, and price elasticities on the internet. Econometrica, 77(2), pp.427-452.

Ershov, D. (2018). The effects of consumer search costs on entry and quality in the mobile app market. Working paper.

European Commission and participating EU competition authorities (2016). Report on the monitoring exercise carried out in the online hotel booking sector by EU competition authorities in 2016. Available at https://ec.europa.eu/competition/ecn/hotel_monitoring_report_en.pdf (accessed 18/02/2021).

European Commission (2017). European Commission - Press release: Antitrust: Commission fines Google EUR 2.42 billion for abusing dominance as search engine by giving illegal advantage to own comparison shopping service. Available at http://europa.eu/rapid/press-release_IP-17-1784_en.htm (accessed 26/02/2019).

Fradkin, A. (2017). Search, Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb (March 21, 2017). Available at: https://ssrn.com/abstract=2939084orhttp://dx.doi.org/10.2139/ssrn.2939084

Ghose, A., Goldfarb, A. and Han, S.P. (2012). How is the mobile Internet different? Search costs and local activities. Information Systems Research, 24(3), pp.613-631.

Glick, M., Richards, G., Sapozhnikov, M., and Seabright, P. (2014). How does ranking affect user choice in online search? Review of Industrial Organization, 45(2), pp. 99-119.
Hausman, J. A. (1996). Valuation of new goods under perfect and imperfect competition. In: The economics of new goods (pp. 207-248). University of Chicago Press.

Hagiu, A., and Jullien. B. (2011). Why do intermediaries divert search? The RAND Journal of Economics, 42(2), pp. 337?362.

Hunold, M, Kesler, R., Laitenberger, U., and Schlütter, F. (2018). Evaluation of best price clauses in online hotel bookings." International Journal of Industrial Organization 61, pp. 542-571.

Hunold, M., Kesler, R., and Laitenberger, U. (2020). Rankings of Online Travel Agents, Channel Pricing, and Consumer Protection. Marketing Science, 39(1), pp. 92-116.

Kim, J.B., Albuquerque, P. and Bronnenberg, B.J. (2010). Online demand under limited consumer search. Marketing Science, 29(6), pp.1001-1023.

Koulayev, S. (2014). Search for differentiated products: identification and estimation. The RAND Journal of Economics, 45(3), 553-575.

Li, J., Netessine, S., Koulayev, S. (2018). Price to compete... with many: How to identify price competition in high-dimensional space. Management Science 64(9): pp.4118-4136.

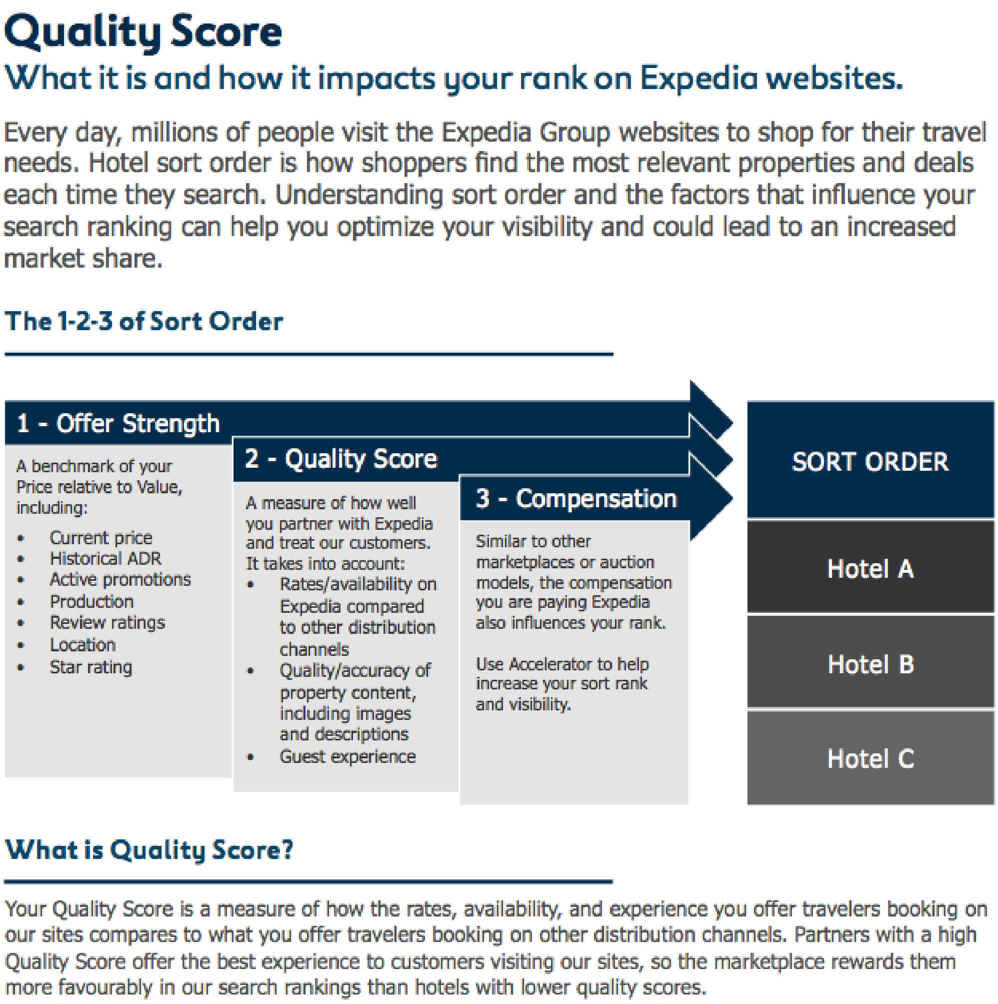
Schaefer, M. and Tran, K. D. (2020). Airbnb, Hotels, and Localized Competition. DIW Discussion Paper 1889. Available here.

Ursu, R. M. (2018). The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions. Marketing Science, 37(4), pp.530-552.

A Expedia's Hotel Ranking

A.1 Information on the Ranking

A.1.1 Quality Score



Quality Score

What it is and how it impacts your rank on Expedia websites.

Every day, millions of people visit the Expedia Group websites to shop for their travel needs. Hotel sort order is how shoppers find the most relevant properties and deals each time they search. Understanding sort order and the factors that influence your search ranking can help you optimize your visibility and could lead to an increased market share.

The 1-2-3 of Sort Order

1 - Offer Strength

A benchmark of your Price relative to Value, including:

- Current price
- Historical ADR
- Active promotions
- Production
- Review ratings
- Location
- Star rating

2 - Quality Score

A measure of how well you partner with Expedia and treat our customers. It takes into account:

- Rates/availability on Expedia compared to other distribution channels
- Quality/accuracy of property content, including images and descriptions
- Guest experience

3 - Compensation

Similar to other marketplaces or auction models, the compensation you are paying Expedia also influences your rank.

Use Accelerator to help increase your sort rank and visibility.

SORT ORDER

Hotel A

Hotel B

Hotel C

What is Quality Score?

Your Quality Score is a measure of how the rates, availability, and experience you offer travelers booking on our sites compares to what you offer travelers booking on other distribution channels. Partners with a high Quality Score offer the best experience to customers visiting our sites, so the marketplace rewards them more favourably in our search rankings than hotels with lower quality scores.

Figure 8: Screen shot of an Expedia's description about how hotels can improve their "offer strength", which in turn affects hotels' "quality scores", which is how hotels were ranked on Expedia as of today. According to Expedia, "Offer Strength" (which includes prices and promotions) is the most important dimension according to which hotels are ranked, while "Compensation" is the least important one (Skift, 2016). Note that compensation is a factor that I control for by taking hotel fixed effects. Source: https://discover.expediapartnercentral.com/wp-content/uploads/2016/12/Expedia_Marketplace-White-Paper-April-2016.pdf (accessed March 13th, 2019).

A.1.2 Further Information on Figure 1

As can also be seen in Figure 1, notifications such as "We have 2 rooms left!" or "8 people booked in the last 48 hours" might be displayed along with the hotel that is on promotion, possibly creating some kind of urgency with consumers. The crossed out price reflects either the standard rate (in case of a sale), or (if displayed without there being a sale tag), according to Expedia, "the third highest price for [the displayed] room type at this hotel with the same length of stay and cancellation policy that customers have found within a 30-day window around the selected check-in date". Some hotels offer free cancellation, membership prices or further benefits like free breakfast. Moreover, some hotels feature the number of people who booked the hotel during the last 48 hours for any travel date, as detailed in listings three and five on Figure 1. All of these information are gathered and can serve as controls.

A.2 Sponsored and Organic search results

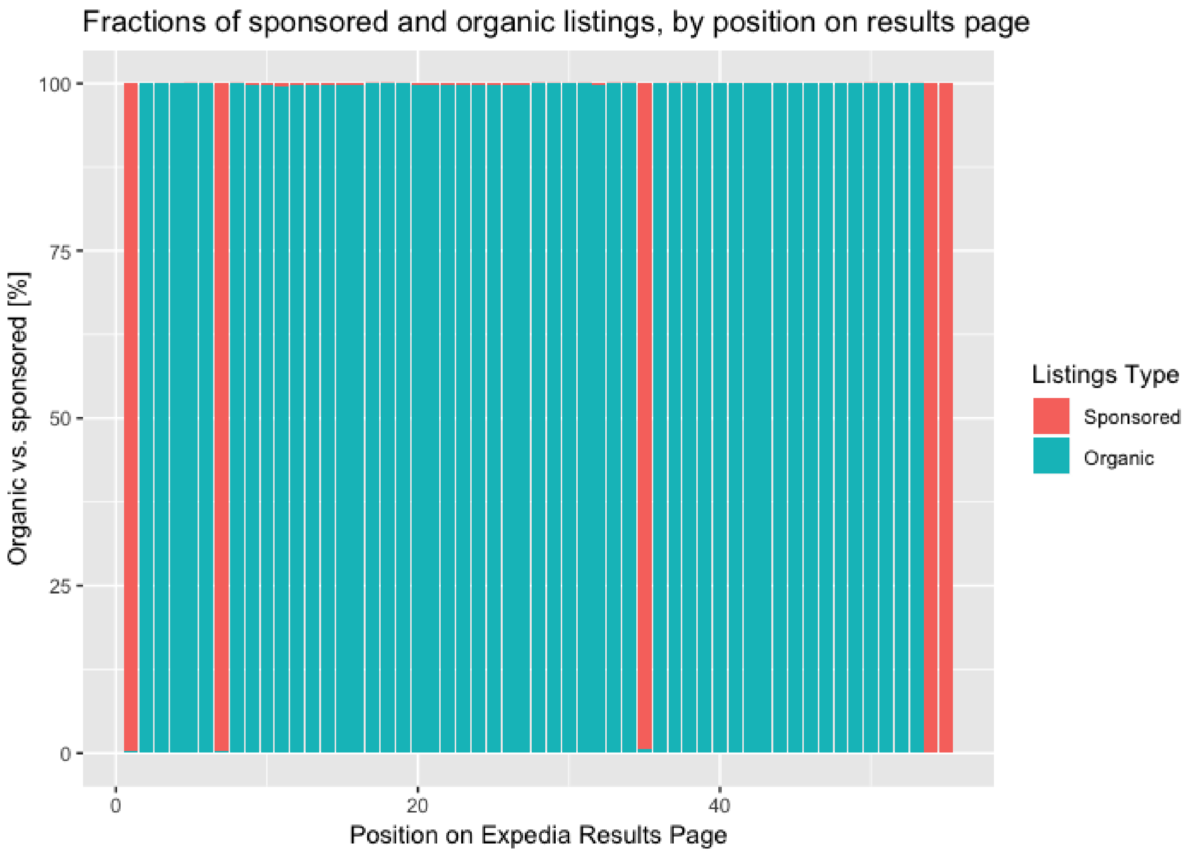


Figure 9: Across all 55 possible positions on a given results page, which percentage of the listings in that position is "sponsored" (meaning that a hotel paid for that slot), and which percentage is a "organic" result? - It becomes evident that positions 1, 7, 35, 54 and 55 are almost always occupied by sponsored hotels paying to be displayed in that slot. As described above, I exclude these from my analysis.

B Sketch of Model, Estimation, and Identification of Optimal Sequential Search (Ursu, 2018)

Consumers' search behavior under optimal sequential search can be described by Weitzman's (1979) Selection, Stopping, and Choice rules. In her paper, Ursu (2018) shows how to estimate consumers' search and purchase decisions jointly in a way that the parameters are consistent with those rules, with the method partly based on work by Kim et al. (2010).

As the next paragraph details, the framework is an adequate description of how hotel search on OTAs in practice: When searching for hotels on an OTA, a user first enters a query and is confronted with an ordered list of hotels, which I call "results page" or "listing page". While basic hotel features (star rating, price, traveller review ratings etc.) are visible from this page, a consumer can click on a given hotel's listing to be referred to what I call a "hotel page". On this hotel page, the consumer finds out further details about the hotel as she can, for instance, view pictures, retrieve information about amenities, or read traveller reviews.

The framework does not endogenize a consumer's decision to refine search results (e.g. by applying filters or by sorting them), or the decision to browse through results pages, or to alter the query. Instead, the model simply poses that, after entering a query, a user is confronted with a listing page containing all the hotels she will view during her complete search sequence. For now, I am thus focussing on estimating consumers' clicks on hotel listings, and purchases only. The model is based on the assumption that every click (which in this framework is a "search") and purchase (or "choice") observed in the dataset has been generated by an optimal sequential search process.

The next sections outline the main components and mechanisms of the model and estimation. I plan to eventually estimate it using the data, but have not done so yet.

Setup

A user i arrives on the OTA and makes a query, specifying location, travel date, and the number of travellers and rooms. She is then confronted with a listing page containing J_i hotels, where J_i is the total number of hotels which i ever views during her complete search sequence on the OTA. Each hotel j that is displayed on the results page has a position value r_{ij}, which is simply the position on the listing page which the hotel was displayed during user i's search. Basic hotel features can be costlessly inferred from this results page. However, as in practice, a consumer needs to click on a given hotel's listing to discover her final valuation for the hotel. Such clicks are costly, with clicks on less prominent and therefore "worse" positions being more costly.

The utility in this model that a given user i derives from hotel j consists of three components:

- The **expected utility prior to search** that can be inferred without incurring costs, v_{ij}, for $j \in \{0, 1, ..., J_i\}$.
- The **expected utility from clicking** on hotel j, $\epsilon_{ij} \sim N(0, \sigma_j^2)$, with $\sigma_j > 0$ for $j \in \{0, 1, ..., J_i\}$.
- The **search (click) costs** $c_{ij}(r_{ij})$, for $j \in \{0, 1, ..., J_i\}$.

The total utility that user i derives from purchasing hotel j is therefore $u_{ij} = v_{ij} + \epsilon_{ij}$. Moreover, I parametrize the search costs as follows to ensure that they are positive:

$$c_{ij}(r_{ij}) = exp(k + \gamma r_{ij})$$

I expect $c'_{ij}(r_{ij}) > 0$. As the data do not contain any information that a user obtains from the hotel page, ϵ_{ij} is unobserved, which is why I assume it follows a normal distribution.

B.0.1 Optimal Search

Sequential search means that after a given click that has been made, the user decides to either click on more options, or to stop searching. If she stops, she will decide which of the searched products (including the outside option) to buy.

A critical role for the optimal sequential search strategy is taken on by the **reservation value** z_{ij} of consumer i for hotel j, defined as:

$$c_{ij} = \int_{z_{ij}}^{\infty} (u_{ij} - z_{ij}) f(u_{ij}) du_{ij} \tag{8}$$

where $f(.)$ is the probability density function of u_{ij}. This implies that the reservation value is the hypothetical utility that would make user i indifferent between searching and not searching product j, given the search costs c_{ij}. According to Weitzman (1979), the optimal search strategy can be characterized by three simple rules.

1. **Selection Rule:** The options should be searched in descending order of the reservation utility.

2. **Stopping Rule:** The consumer should stop searching when the highest utility obtained so far is larger than any reservation value of the unsearched options.

3. **Choice Rule:** The consumer should choose the option that yields the highest utility (including the outside option).

Thus, the selection rule defines the order of the searches, while the stopping rule defines the length of the search. The rules imply a number of inequalities concerning the relationship between reservation values and utilities for the products: Using Ursu's (2018) notation, assume that a user i searches a total of s hotels. Let $R_i(n)$ denote the identity of the hotel with the n-th highest reservation utility, and thus the n'th hotel that was searched. Thus $R_i = [R_i(1), .., R_i(n), ..., R_i(s)]$ is the set of searched hotels and the order in which they were searched. Moreover, let $R_i(0)$ and $j = 0$ denote the outside option.

From Weitzman's (1979) selection rule, we know that, given that user i makes her n'th search, she will optimally pick the hotel that has the highest reservation utility out of all those hotels that have not been searched yet:

$$z_{iR_i(n)} \geq \max_{k=n+1}^{J_i} z_{iR_i(k)} \qquad \forall n \in \{1, ..., J_i - 1\}$$

From the stopping rule, one obtains two separate inequalities. First, user i will make an n'th search when the reservation utility of the product searched in the n'th search exceeds the utility that was revealed from all other searched products (including the outside utility):

$$z_{iR_i(n)} \geq \max_{k=0}^{n-1} u_{ik} \qquad \forall k \in \{0, ..., n-1\}$$

Second, given that user i searches s products, it must be that all hotels that are not searched have a reservation utility that is lower than the maximum of the utility of all searched alternatives, including the outside option:

$$z_{iR_i(m)} \leq \max_{k=0}^{s} u_{iR_i(k)} \qquad \forall m \in \{s+1, ..., J_i\}$$

Last, the choice rule implies that the product that is ultimately chosen must yield a larger utility than any of the other searched options, including the outside option:

$$u_{ij} \geq \max_{k=0}^{s} u_{iR_i(k)} \qquad \forall j \in R_i \cup \{0\}$$

These four inequalities define the probability that a given user i searches in order R_i and purchases product j, and put restrictions on the values for the utility parameters. Given the observed searches and choices for all users, one can derive the joint likelihood. Subsequently, one can estimate the utility parameters and the effect of position on search costs using simulated maximum likelihood estimation[24].

To precisely pin down the mean search costs k, another expression is needed. As Kim et al. (2010) show, from the definition of the reservation utility, one can obtain the following expression:

$$\frac{c_{ij}}{\sigma_j} = \left(1 - \Phi\left(\frac{z_{ij} - v_{ij}}{\sigma_j}\right)\right)\left(\frac{v_{ij} - z_{ij}}{\sigma_j} + \frac{\phi\left(\frac{z_{ij}-v_{ij}}{\sigma_j}\right)}{1 - \Phi\left(\frac{z_{ij}-v_{ij}}{\sigma_j}\right)}\right)$$

where $\phi(.)$ and $\Phi(.)$ are the probability density function and the cumulative distribution function of the standard normal distribution, respectively. Kim et al. (2010) further explain that for a given c_{ij} and σ_j (which will be normalized to 1), one can obtain a unique value of $\frac{z_{ij}-v_{ij}}{\sigma_j}$. By creating a look-up table, one can thus obtain the precise value for $\frac{z_{ij}-v_{ij}}{\sigma_j}$ outside the estimation loop, which allows to compute the exact reservation utility via the expression $z_{ij} = v_{ij} + \sigma_j \frac{z_{ij}-v_{ij}}{\sigma_j}$.

Based on the restrictions described above that result from Weitzman's (1979) rules, one can derive the probability that a consumer searches in a given order and chooses a product j by integrating over the space of values of ϵ that result in the observed pattern of clicks and choices. From there, one can form the log-likelihood, which can finally be estimated I refer to Ursu (2018) for a more detailed discussion on estimation and identification, as I am using the exact same method.

[24]The ϵ_{ij} of the searched options are not going to distributed normally any more.

C Graphs of Prices in Across Cities

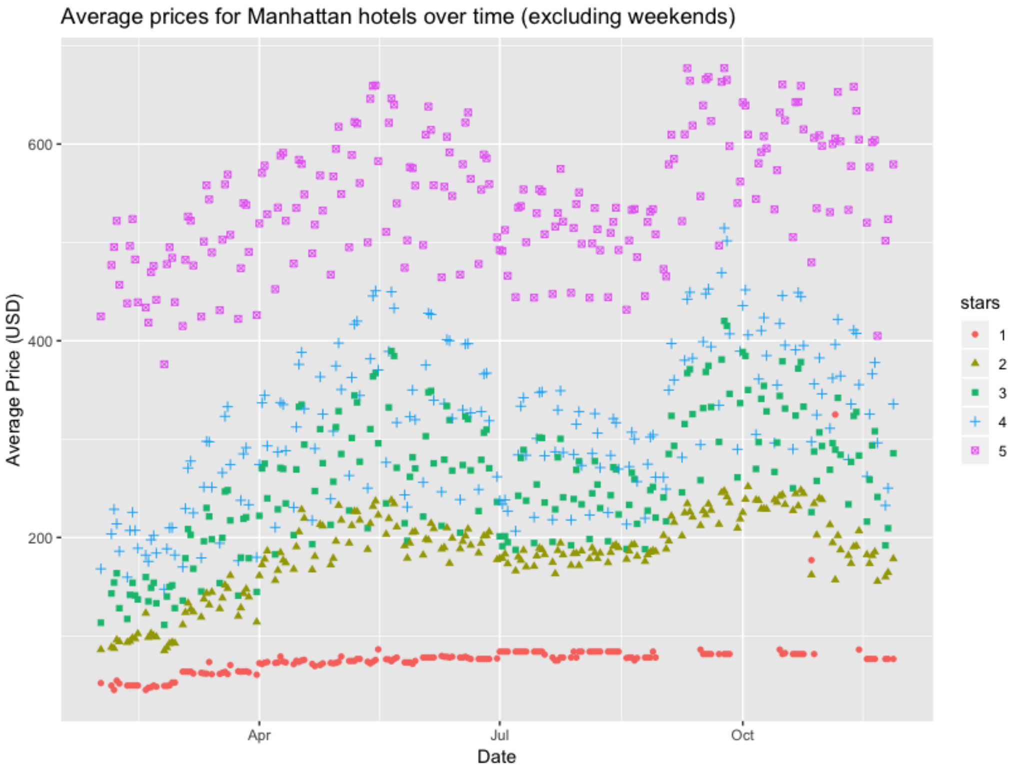


Figure 10: Prices of Manhattan hotels over travel dates.

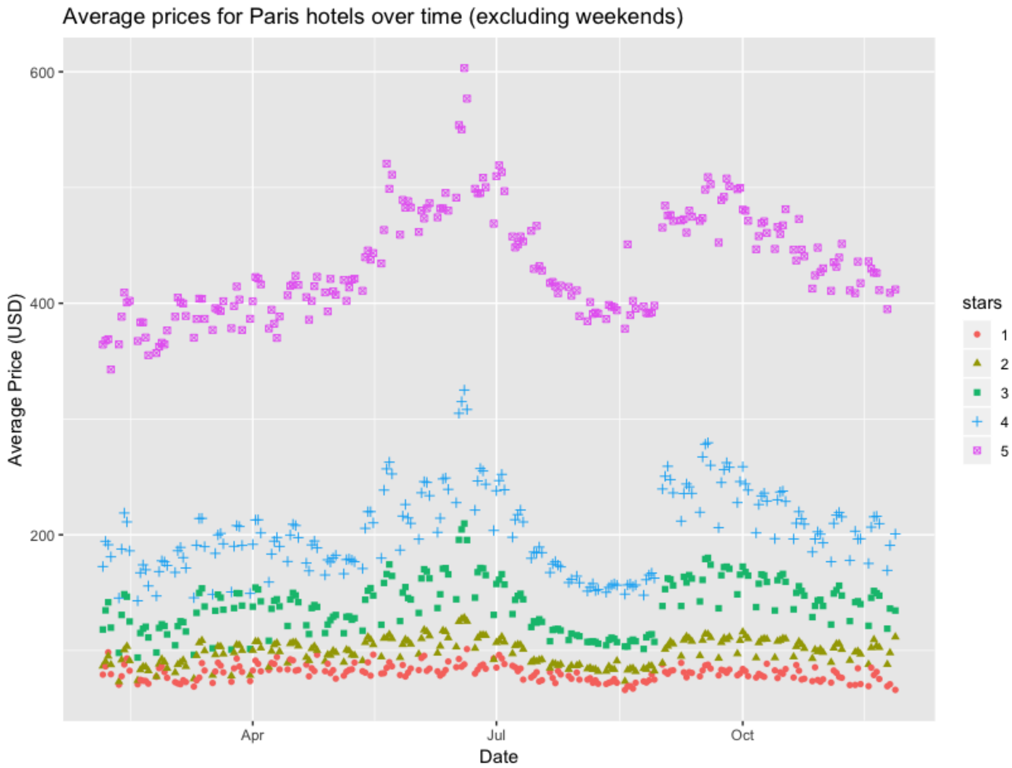


Figure 11: Prices of Paris hotels over travel dates.

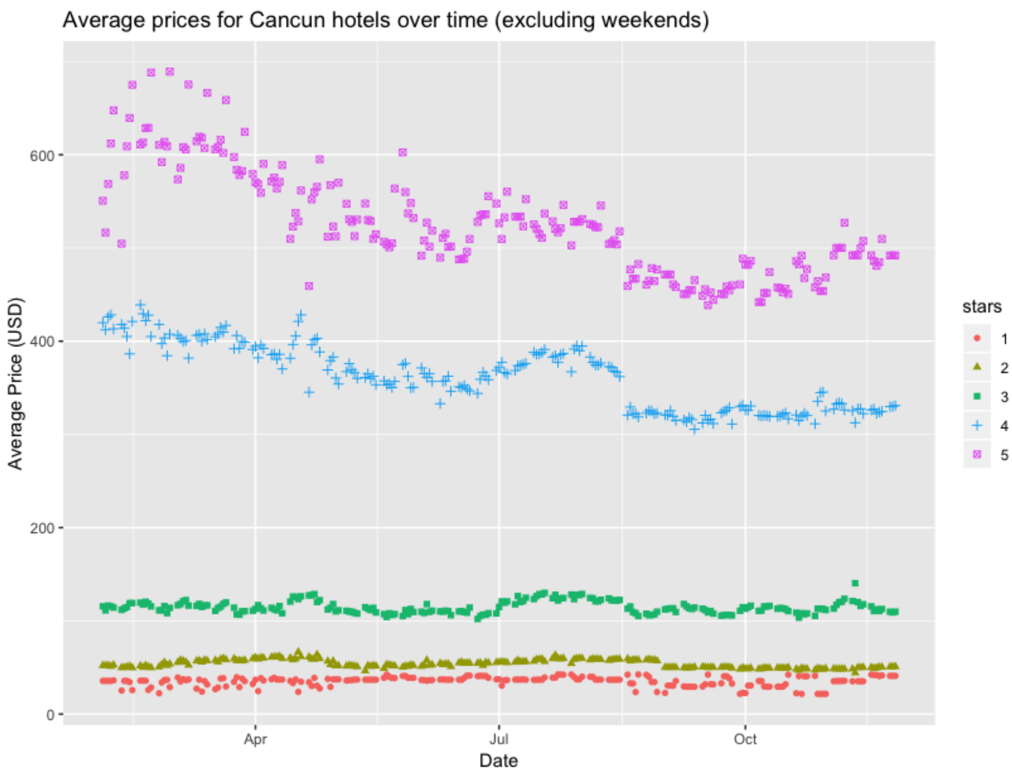


Figure 12: Prices of Cancun hotels over travel dates.

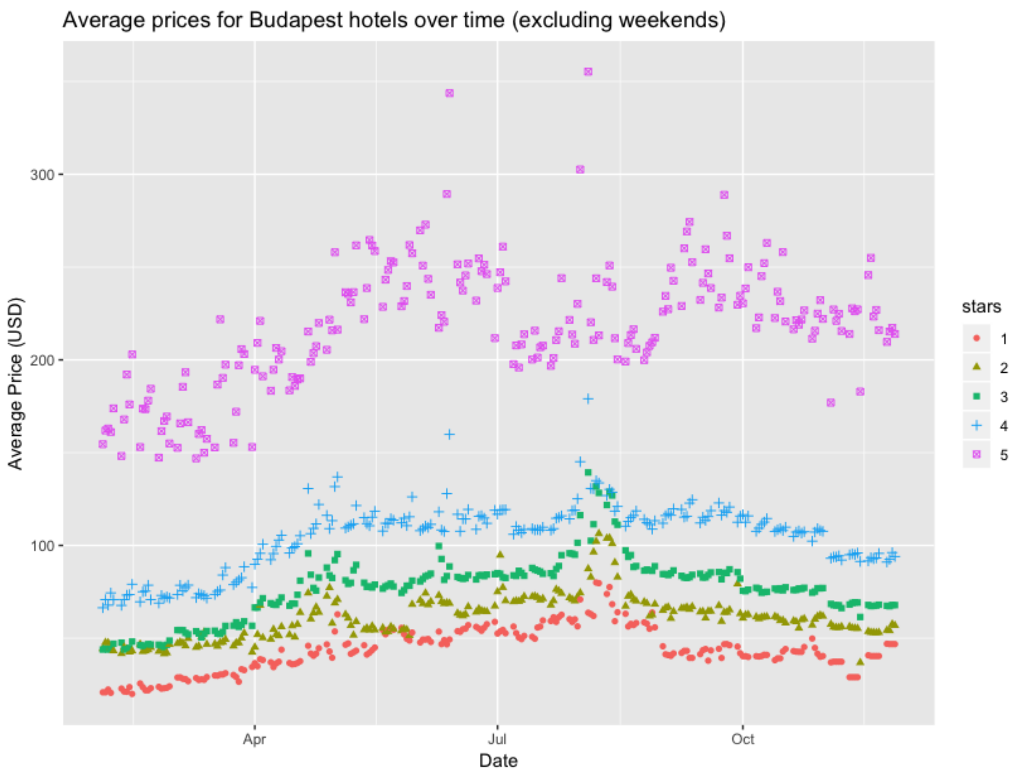


Figure 13: Prices of Budapest hotels over travel dates.

D More Results

Table 15: First stages for prices: neighborhood-based instruments

	Dependent variable:		
	price		
	(1)	(2)	
avg_neighb_price	0.506***		
	(0.002)		
avg_neighb_sale	2.808***		
	(0.461)		
neighb_count	−0.125***		
	(0.006)		
avg_neighb_samestar_price		0.441***	
		(0.002)	
avg_neighb_samestar_sale		2.390***	
		(0.381)	
neighb_samestar_count		−0.397***	
		(0.014)	
avg_neighb_samestar_samebrand_price			
avg_neighb_samestar_samebrand_sale			
neighb_samestar_samebrand_count			−
pastbookings_count	−0.090***	−0.115***	−
	(0.013)	(0.016)	
Observations	924,809	716,282	
R^2	0.057	0.066	
Adjusted R^2	0.054	0.063	
F Statistic	13,808.700*** (df = 4; 922124)	12,534.790*** (df = 4; 714087)	9,959.654*
Note:			*p<0.1; **

Table 16: First stages for sales dummies: neighborhood-based instruments

	Dependent variable:		
	sale		
	(1)	(2)	(3)
avg_neighb_price	−0.00005***		
	(0.00002)		
avg_neighb_sale	0.082***		
	(0.003)		
neighb_count	0.0003***		
	(0.00004)		
avg_neighb_samestar_price		−0.0001***	
		(0.00001)	
avg_neighb_samestar_sale		0.040***	
		(0.003)	
neighb_samestar_count		0.001***	
		(0.0001)	
avg_neighb_samestar_samebrand_price			−0.0001
			(0.0000
avg_neighb_samestar_samebrand_sale			0.053**
			(0.002
neighb_samestar_samebrand_count			0.001**
			(0.0001
pastbookings_count	0.001***	0.001***	0.001**
	(0.0001)	(0.0001)	(0.0001
Observations	928,625	719,940	683,21
R^2	0.001	0.001	0.001
Adjusted R^2	−0.002	−0.002	−0.00
F Statistic	205.420*** (df = 4; 925938)	124.734*** (df = 4; 717743)	167.332*** (df =

Note: *p<0.1; **p<0.05;

Table 17: First stages for prices: Brand-based instruments

	Dependent variable:	
	price	
	(1)	(2)
avg_price_brand	0.433***	0.430***
	(0.003)	(0.003)
avg_sale_brand	−2.596***	−0.369
	(0.693)	(1.015)
brand_count	0.001	0.007
	(0.008)	(0.008)
avg_discount_brand		−0.174***
		(0.057)
avg_freecancel_brand		4.094***
		(0.488)
avg_freeextras_brand		−0.795
		(1.686)
pastbookings_count	−0.020	−0.021
	(0.020)	(0.020)
Observations	257,735	257,735
R^2	0.084	0.085
Adjusted R^2	0.080	0.080
F Statistic	5,907.848*** (df = 4; 256517)	3,388.327*** (df = 7; 256514)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 18: First stages for sales: Brand-based instruments

	Dependent variable:	
	sale	
	(1)	(2)
avg_price_brand	−0.00005***	−0.0001***
	(0.00002)	(0.00002)
avg_sale_brand	0.424***	0.474***
	(0.004)	(0.006)
brand_count	0.0001**	0.0001***
	(0.00005)	(0.00005)
avg_discount_brand		−0.004***
		(0.0003)
avg_freecancel_brand		0.016***
		(0.003)
avg_freeextras_brand		0.026***
		(0.010)
pastbookings_count	0.001***	0.001***
	(0.0001)	(0.0001)
Observations	257,735	257,735
R^2	0.043	0.044
Adjusted R^2	0.039	0.039
F Statistic	2,904.239*** (df = 4; 256517)	1,686.718*** (df = 7; 256514)
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 19: First stages: Airbnb and events × zipcode instruments

	Dependent variable:	
	mean_price	mean_sale
	(1)	(2)
airbnb_p_zip	1.945***	−0.002***
	(0.018)	(0.0001)
airbnb_count_zip	−1.061***	−0.0002
	(0.334)	(0.002)
mean_pastbooking	−0.042	0.0004***
	(0.028)	(0.0002)
Events × zipcode dummies	Yes	Yes
Observations	446,298	448,026
R^2	0.036	0.003
Adjusted R^2	0.030	−0.003
F Statistic	28.393*** (df = 588; 443472)	2.175*** (df = 588; 445197)
Note:		*p<0.1; **p<0.05; ***p<0.01

E Cleaning of WCAI Dataset

I am using the WCAI dataset which details consumer search histories only in order to estimate the demand model's utility parameters for the Parisian data. Overall, my cleaning procedure is similar to the one by Ursu (2018), which she describes in the Appendix. The raw dataset contains 1,546,296 observations of 19,658 distinct user IDs making, all in all, 565 purchases in either Budapest, Cancun, Manhattan, or Paris. After removing a number of data errors (such as duplicated rows), outliers (such as observations of prices below 6$), unusual searches that are likely to be errors (such as searches with an unreasonable long length of stay or for dates in the past or in the far future), I end up with 1,361,377 observations of 17,880 distinct user IDs. Next, I try to remove all those observations that are likely to not have been made by actual humans, but by robots scraping the website. In those searches, all hotels were clicked on or all websites were looked at within a small time frame, and of course no purchase was made. This reduces the number of observations to 1,305,104 observations of 17,769 users who make 51,896 distinct searches. Taking only searches for hotels in Paris, one obtains 330,891 observations of 4,425 users making 12,810 distinct searches. Of these, I again use only a subset, as I only include observations with the following characteristics:

- Searches within the default ranking of the platform, without any minimum star rating specified, and without an explicit hotel name specified (no sorting or filtering);
- Searches which contain at least one click.

In the end, the final dataset that I use for the estimation of the demand parameters contains 31,874 observations, stemming from 1,051 searches. To be able to estimate Ursu's (2018) model using the search data, I moreover create "effective" position numbers: hotels displayed on page 2 in position 3, for instance, are given the position number equal to: "(number of page 1)+position number on page 2". This is in slight contrast to the dataset that Ursu (2018) uses, which only contains observations on click and purchase behavior occurring on the first results page.