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

Luise Eisfeld†

December 11, 2023

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 on a listing's page, however, may condition a product's position on its price. Using web-scraped data on hotels displayed on Expedia and an instrumental variable identification strategy, I find that the ranking algorithm tends to display hotels at *less* favorable positions at times at which they are priced *higher*. I provide a framework that employs these estimates jointly with demand parameters obtained from a sequential search model. I simulate a counterfactual scenario, which reveals that Expedia's ranking algorithm tends to intensify price competition between sellers compared to a random ranking. This increases consumer welfare, but reduces seller profits by decreasing prices by 1.89€ on average. My finding has consequences for two-sided platforms' optimal design of ranking algorithms: in order to foster adoption, platforms should carefully trade off benefits arising to the two sides, and consider equilibrium effects.

Keywords: Product Rankings, Online Search, Hotel Pricing, Online Travel Agents, Electronic Commerce

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

*I am grateful to Alexandre de Cornière, Daniel Ershov, and Bruno Jullien for their support and guidance for this paper and while writing my thesis. I thank Ulrich Laitenberger and Kevin Ducbao Tran for their thoughtful comments. Furthermore, I thank the participants of the LED Young Economist Seminar at UC Louvain, the Competition and Innovation Summer School 2019, EARIE 2019, and of workshops at the Toulouse School of Economics for helpful discussions. Data obtained via the Wharton Customer Analytics Initiative and sponsored by an anonymous travel website is gratefully recognized. All errors are mine.

†University of Lausanne, Faculty of Business and Economics (HEC Lausanne), Extranef Building 230, 1015 Lausanne, Switzerland. Email: luise.eisfeld@unil.ch.

1 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 once users enter 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, Richards, Sapozhnikov, and Seabright (2014)).

Although the algorithms used in practice are very complex, they may make a product's positioning a function of its price. Trying to guide consumers towards attractive offers, platforms may find it profitable to prioritize products that are lower priced, all else equal, by displaying them more prominently.[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 platform may find it optimal to do the reverse, and may steer consumers to higher-margin products so as to extract more revenue per consumer visit (Hagiu & Jullien, 2011). This weakens price competition and enables sellers to charge higher markups (Dinerstein, Einav, Levin, & Sundaresan, 2018). Rankings can therefore affect sellers' price-setting decisions and influence market outcomes in meaningful 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. The empirical setting studied is hotel booking services offered by online travel agents (OTAs). Those intermediaries display hotels that are being algorithmically ranked by recommender systems as consumers enter a query. Most user clicks and purchases occur under these default rankings. Using web-scraped data from Expedia, I find that offering a lower price leads to a better positioning of a given hotel on the platform's listings page. I then use consumer search and transactions data to estimate a model of sequential consumer search that closely follows Ursu (2018). The estimated ranking parameters and the demand parameters then enable to perform counterfactual simulations. Preliminary results show that hotels will set somewhat higher prices under a ranking that does not take into account supply-side effects, although losses in consumer surpluses remain limited. Expedia's ranking algorithm thus seems to move hotels' pricing decisions into a more competitive equilibrium with lower prices, compared to a scenario where prices do not affect visibility.

Studying the effect of ranking algorithms on market outcomes within a structural framework is in-

[1] This is the case in the context studied in this paper: Online travel agents explicitly state that the competitiveness of hotels' offerings (in particular, prices) are taken into account in their rankings (see Section 2 as well as Appendix A.1). A further example is provided by L. Chen, Mislove, and Wilson (2016), who find that prices are the most important feature predicting both Amazon's ranking of sellers of a given product, as well as which seller gets to be displayed prominently in the "Buy Box".

teresting and important 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 sellers' pricing behavior and the functioning of markets, which this article attempts to shed light on. Second, this question is relevant from a public policy perspective. Online platforms essentially act as gatekeepers that control what information users view and how market participants interact. The issue of prominence in particular was at the heart of a high-profile antitrust case in Europe regarding Google Search[2], but our knowledge on how visibility affects market outcomes is still limited. My results also inform about the strength of competition in the market for online booking services, which was subject to antitrust investigations in several European countries.[3] 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, these platforms need to assess the trade-offs on both sides. While platforms can evaluate how a particular ranking affects consumers' choices by varying recommended products across subgroups of consumers (A/B-testing), such experiments are not going to shed light on the ranking's effect on *supply side* behavior. Instead, analysis of how rankings affect pricing decision 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.[4]

On the supply side, I thus propose a framework in which the economic agents are hotels deciding what prices to set. What differentiates this framework from standard supply models is that hotels set prices in anticipation of how these decisions affect their respective 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 demand and hotel prices under a counterfactual ranking of hotels. This requires, first, an estimate of how price and promotion affect a hotel's position. I obtain this estimate using a linear fixed effects regression with instrumental variables. Second, this requires estimates of how position and price impact demand. On the demand side, one therefore needs a model that accounts for the fact that consumers are not fully informed, but face search costs, and are

[2]See https://tinyurl.com/ycxpwncx for the European Commission's press release on this case (accessed 03/05/2022).

[3]In a sequence of decisions, the German Bundeskartellamt, for instance, prohibited the use of 'best price' clauses by OTAs, emphasizing the restriction of competition between OTAs that such practices would entail: see https://tinyurl.com/4e4j446j and https://tinyurl.com/55pkau9u (accessed 02/04/2023). Also see European Commission (2016) for a report on a monitoring excercise on the hotel booking sector by the European Commission and several European antitrust authorities.

[4]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, which emphasizes the importance of visibility in this context: 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 07/05/2019).

more likely to click and book hotels that are displayed more visibly. I use Ursu (2018)'s model of consumer search, in which each search corresponds to a click on a given hotel, and search ("click") costs are increasing in a hotel's position. 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 calibrate the search cost parameter to the estimates Ursu (2018) obtains under a random ranking of hotels (assuming that search costs have not changed over time), and only estimate preference parameters for hotel characteristics. Given these 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 first dataset was web-scraped from Expedia over two months 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, hotel fixed effects are used in order to 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 with its ranking over query and travel dates. Second, the simultaneity of hotels' price-setting decisions with respect to the ranking are addressed with the help of instrumental variables for prices.

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 in the same cities that I web-scraped. The data reveal 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. I moreover find 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. 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: my estimates reveal that a one-dollar increase in a given hotel's price implies a decrease (towards less visible positions) by roughly three positions on Expedia's listings page, ceteris paribus. This effect is qualitatively robust across specifications and across different sets of instrumental variables. 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 sales promotion decision (i.e., a sales promotion associated with being ranked more visibly), although a clear causal link cannot be established, likely due to a lack of sufficiently strong instruments. The consumer side results are qualitatively in accordance with Ursu (2018). Given the estimates on both the consumer side as well as the ranking side, one can solve for hotels' marginal costs from the hotels' optimal pricing condition. I then consider exogenous perturbations to variables such as the ranking's elasticity with respect to prices, 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.

Literature. This paper relates to previous research dealing with platform design, consumer search, and the industrial organization of the market for online booking intermediaries. Over the past few years, there have been a few notable contributions that study supply-side effects of platform and search design in online contexts. Like the present paper, those contributions estimate demand models that can account for search frictions that consumers are facing.

Dinerstein et al. (2018) analyze the effects of a search design change on eBay that essentially decreased users' search costs of finding the cheapest product. The authors consider sellers of a very homogenous product, and estimate that the re-design led to a decrease in sellers' prices and markups. In contrast, in my setting, hotels are both horizontally and vertically differentiated, and I focus my attention on the ranking algorithm, as opposed to the search design more generally. Further, Lee and Musolff (2021) find that Amazon tends to make demand more elastic and thus intensifies price competition. This, however, comes at the expense of lower entry by third-party sellers in the long run. Teng (2022) studies a change in the search algorithm of Apple's App Store, which led to reduced prominence of Apple's own apps. Instead of prices, the author analyzes the updating frequency of mobile applications, and finds that the average app quality would be higher if Apple restrained from self-preferencing its own apps on its platform.[5] Ershov (2022) finds that a change in the search design of the Android app store triggered new entry (and thus more variety), but led to congestion externalities that dominated consumers' gains from variety. In contrast to the aforementioned authors, I consider a market in which prices seem to be sellers' main strategic variables, whereas entry tends to be stable.[6] Bar-Isaac and Shelegia (2022) take on the platform's perspective, and theoretically study the trade-offs a platform faces when deciding between allocating visibility to sellers by using an auction, or alternatively by using an algorithm, in different contexts. Further related papers investigate sellers' obfuscation strategies online (Ellison & Ellison, 2009), and the design of peer-to-peer platforms where matching is central (Fradkin, 2017).

Next, this research is related to a number of papers estimating structural models of consumer search in diverse settings, under different assumptions on search behavior (fixed sample search or sequen-

[5] In addition, both Lee and Musolff (2021) and Teng (2022) also study the prevalence and the welfare implications of self-preferencing on Amazon and on the Apple app store, respectively.

[6] Whereas hotels can in practice de-list from online platforms, this tends to happen extremely rarely. Hotel bookings have been commonly mediated by third parties even before the advent of the internet.

tial 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 (Y. Chen & Yao, 2016, De los Santos & Koulayev, 2017, Koulayev, 2014, Ursu, 2018), and have found that search costs in the hotel search context are very significant, with important implications for hotels. 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 modeled as a function of positions.

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, Kesler, and Laitenberger (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. By employing the ranking strategically, 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. (Cazaubiel, Cure, Johansen, & Vergé, 2020) find that buyer substitution across purchasing channels is relatively slow, therefore lending support to the assumption that a single dominant platform can be viewed as a single market. Further related papers study hotels' price-setting behaviors (Cho, Lee, Rust, & Yu, 2018, Li, Netessine, & Koulayev, 2017), the effect of online platform price parity clauses (Hunold, Kesler, Laitenberger, & Schlütter, 2018, Mantovani, Piga, & Reggiani, 2021), or on buyer substitution between hotels and Airbnbs and localized competition is Schaefer and Tran (2020).

Roadmap. Section 2 of this paper presents background information about the online hospitality sector and OTAs' rankings of hotels. Section 3 contains the model for the supply side, and Section 4 details the datasets used and provides descriptive statistics. The empirical strategy will be presented and discussed in Section 5, with the results presented in Section 6. Section 7 models counterfactuals. Section 8 concludes.

2 Background

The hotel sales on all major OTAs[7] take place under the so-called "agency model". Under this business model, the supplier (i.e., the hotel) sets the final price of the product that is sold through the intermediary. The platform merely mediates the transaction between the hotels and consumers by facilitating

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

product search, by giving hotels exposure to potential customers, and by processing the payment.[8] OTAs receive an *ad-valorem* commission rate for each purchase that is carried out on their platform. On Expedia, the standard rate is 15%, but can privately re-negotiated and tends to be somewhat lower for large chains (Cazaubiel et al., 2020). Industry reports quote rates between 10% and 25%.[9]

Consumers arriving on an OTA make a query by entering a city, the number of travellers, and arrival and departure dates, and are then confronted with what I call a "listings page". On Expedia, consumers then view an ordered list of typically 55 hotels per page 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[10] hotels, and browse to further pages. Clicking on a listed hotel takes a consumer to what is referred to as 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, the ranking is widely believed to be based on relevance. Moreover, Expedia states that a hotel's price competitiveness (current price, active promotions) are factored into the ranking order (see Figure 10 in the Appendix).

Hotels' main strategic variable seems to be the nightly rate for a room, which is displayed very prominently on OTAs' listings page. 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!" flag.[11] Therefore, in this paper, I call a "sale" whenever a listing is flagged with a green "Sale!" flag (see Figure 1), and thereby displayed in a somewhat more salient way.[12] The price is the key explanatory variables in my analysis, and the key strategic decision my model incorporates.[13] More facts and insights on hotels' promotional decisions can be retrieved in Appendix D.3.

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 will later exclude these sponsored listings from my analysis, and thus focus on the "organic" ranking that is determined by the algorithm only (see

[8] At least some OTAs also provide revenue management and pricing or promotion tools to hotels.

[9] See, e.g., https://prenohq.com/blog/ota-commission-rates-expedia-booking-com-more/ (accessed 02/05/2021).

[10] 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.

[11] 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".

[12] 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 flag.

[13] The sale flag indicator is considered in some specifications of the ranking analysis.

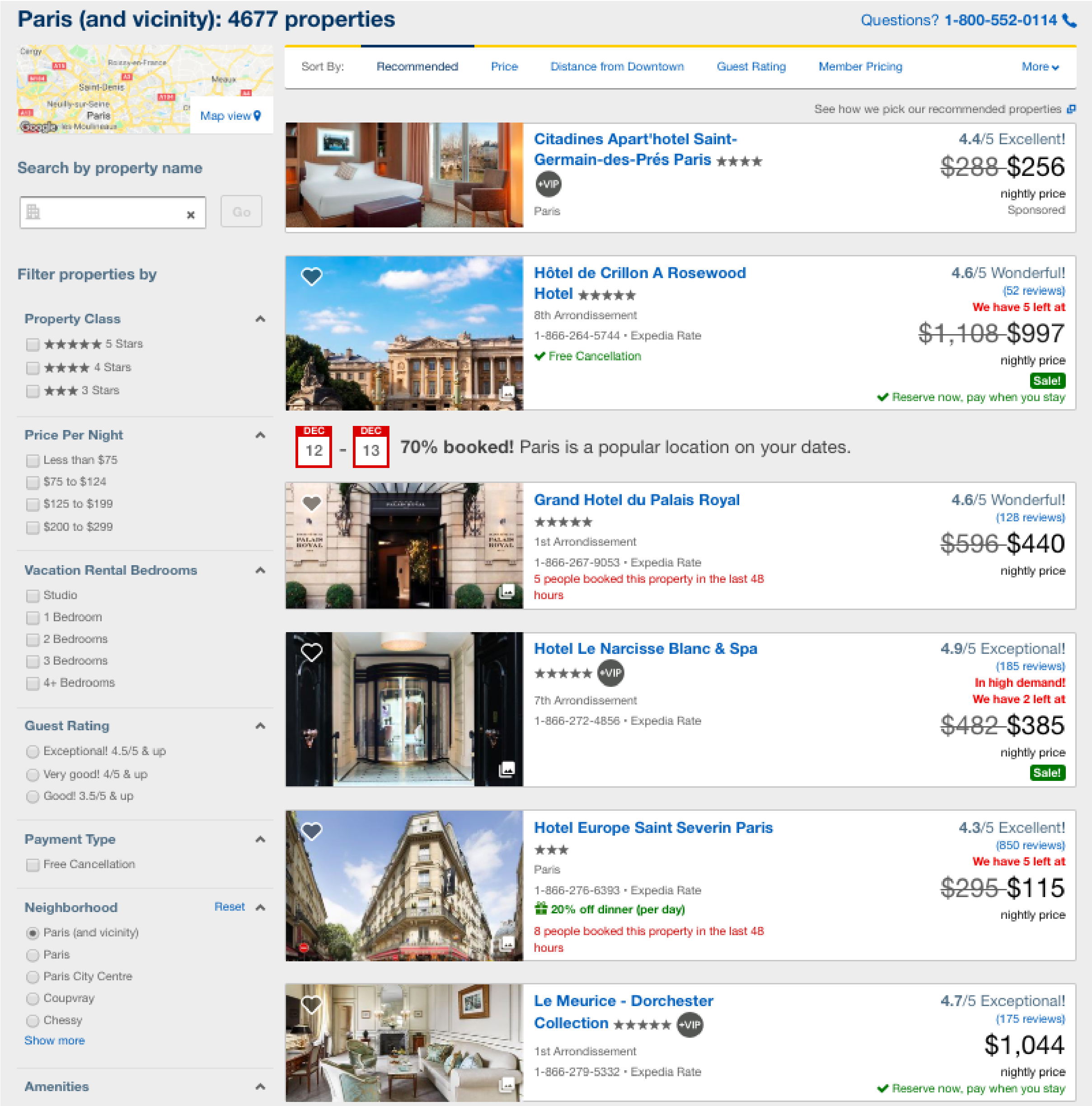

Figure 1: Exemplary screen shot of Expedia's listings page for hotels in Paris, as of February 2019.

Appendix C.2).

3 Model

The centerpiece of the model is a hotel's pricing decision. The hotels' first order conditions show that optimal prices depend on the magnitudes of two effects: the common direct effect on demand (i.e., as hotels raise prices, demand decreases, *ceteris paribus*), and, secondly, the indirect effect on a hotel's visibility (propagated via the ranking algorithm) on demand. On the consumer side, I take as given that consumer search follows the sequential search framework detailed in Ursu (2018). I do not explicitly model a platforms' decision of how to design the ranking algorithm, but discuss the tradeoffs and possible incentives that platforms face in Section 3.1.

3.1 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. On a given query date q and for travel date t (with $t \geq q$), hotels have marginal costs c_{qjt}.[14] Taking as given the pricing decisions of their competitors, at query date q, hotels set prices for room bookings for date t, such that the resulting decisions form a Bertrand Nash Equilibrium among all hotels. Hotel j's profit is (as always) the product of the markup and the demand. 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}, but also of the position r_{jqt} which the hotel is by default displayed in on the OTA's results page. By employing a demand model in which demand depends on a product's position, the full 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.[15]

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

$$\max_{p_{jqt} \in \mathbb{R}+} \Big(p_{jqt} \cdot (1 - \tau) - c_{jqt}\Big) M s_{jqt}\big(\mathbf{p_{qt}}, \mathbf{r_{qt}}(p_{jqt})\big).$$

Next, define $\tilde{c}_{jqt} \equiv \frac{c_{jqt}}{(1-\tau)}$. Deriving the first order conditions using with respect to price (using the chain and the product rule) and re-arranging, one obtains:

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

[14] These marginal costs are likely composed of the opportunity costs of not selling a room for a given night. See discussion in Section 7.2.

[15] In an alternative version of the model considered in Appendix B, I extend the model such that hotels not only choose prices, but also discounts.

When deciding on prices, hotels thus take into account two types of effects on demand. The "usual" (*direct*) demand 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 $\frac{\partial s_{jqt}(\cdot)}{\partial r_{jqt}} r'_{jqt}(p_{jqt})$ is additionally going to be taken into account by hotels.

As a result, if a platform modifies the ranking algorithm – for example, by employing an algorithm that reacts more sensitively to prices, or by ranking hotels completely independent of their prices – hotels will take this into account and set prices differently. This affects hotels' profits and consumer surplus.

Discussion of Supply Model

The above model abstracts from the fact that hotels typically distribute rooms via multiple sales channels. According to a monitoring study conducted by European Commission (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., 2020), 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 constrained 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.

Possibly, promotions should also be a part of the hotel's decision 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 10 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. On the other hand, it is unclear how exactly hotels use promotions in this setting: sales promotions do not correlate very much with demand (see Figure 5), and the different types of promotions employed on Expedia tend to be opaque (see Appendix D.3). To be sure, I do consider a model version that endogenizes hotels' promotion decisions in Appendix B.

3.2 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). Consumers' search behavior under optimal sequential search can be described by Weitzman (1979)'s Selection, Stopping, and Choice rules. 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, Albuquerque, and Bronnenberg (2010).

In this sequential search model, consumers 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 "lower" positions) are more costly to click on, which can be interpreted as the costs of scrolling down a page. 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 (1979)'s 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.

The framework appears to be 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 on the listings 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, just like in Ursu (2018), the model poses that, after entering a query, a user is confronted with a listing page containing all the hotels she can search. With help of the model, one thus estimates parameters that govern consumers' click and purchase decisions only, under the assumption that each click ("search") and each purchase ("choice") observed in the dataset has been generated by an optimal sequential search process.

3.2.1 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:

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

where k is a mean level of search costs, and γ is the additional effect of position on search costs. 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 follow Ursu (2018) in assuming that it follows a normal distribution.

3.2.2 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:

$$c_{ij} = \int_{z_{ij}}^{\infty} (u_{ij} - z_{ij}) f(u_{ij}) du_{ij} \tag{2}$$

where $f(\cdot)$ 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 (2018)'s 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 (1979)'s 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[16].

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 (1979)'s rules, one can derive the probability that a consumer searches in a given order and chooses a product j by integrating over

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

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.

3.3 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.[17] On the other hand, platforms have the short run objective of maximizing revenues per visit by essentially diverting consumer search (Hagiu & 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.

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.[18] 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.

4 Data

I use two main datasets. The first dataset is consumer search and transactions data stemming from consumer queries made in 2009 on one of the major online travel agencies in the US and in the world[19], and

[17] 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.

[18] Casual 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.

[19] According to the travel research website tnooz, this website was one of the most visited travel agent websites as of October 2009. Moreover, it was ranked as being the most used platform worldwide over many years. See `https://www.tnooz.com/article/us-travel-site-crunch-data-week-end-october-17-2009/` (accessed 03/03/2019).

was obtained via the Wharton Customer Analytics Initiative (WCAI). As all observations are generated by actual consumer queries, this dataset does not yield sufficient variation in the rankings and prices of a given hotel over query and travel dates that would allow to estimate the relationship of hotels' prices and rankings. I therefore additionally web-scrape data from Expedia in 2019 that yields variation in hotels' rankings and prices across query and travel dates. Both datasets are available for four cities – Manhattan, Budapest, Cancun, and Paris –, but I focus the analysis on Paris only.[20]

4.1 Demand Side: Consumer Search and Transactions Data

The consumer search data motivates the analysis, helps to verify the importance of the default ranking, and is used to obtain utility parameters on the demand side. This dataset records the search behavior of nearly 18,000 US-based users on a major OTA, 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 corresponds to a given hotel that appeared during a given user's query. 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 query 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. Both clicks and purchases include a time stamp. A "query" is defined to be a combination of request (i.e., typically location and travel dates entered), refinement action, and page.[21] A hotel is considered to be "viewed" if it was displayed anywhere on the current listings page.[22] The basic dimensions are displayed in Table 1. 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[23].

Exploring consumers' search strategies confirms the fundamental importance of the default ranking during the search process. On average, consumers seem to have small consideration sets, and rarely flip through results pages. Table 2 shows that in Paris, even eventual buyers of a hotel room end up viewing less than 80 out of 1,600 potentially available hotels, and click on six of them. Refinement actions (comprising filtering and sorting) are rarely used – on average, roughly three quarters of consumers who end up purchasing take one refinement action. Only 30% of all users (and 45% of users who end

[20] A benefit of focusing on Paris is that geolocation data is especially well-covered for the subset of web-scraped hotels in Paris. In contrast, for Cancun for instance, many addresses are misspecified and the geolocation cannot be recovered.

[21] Thus, browsing to the next listing page is a new "query", as is any re-ranking or filtering action.

[22] As is common in these datasets, I am not able to take into account whether a consumer actually scrolled to a given hotel and looked at it.

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

Table 1: Basic data summary: WCAI consumer search data

Market	Observations	Users	Queries	Hotels	Clicks	Purchases
Budapest	353,051	4,946	14,794	276	9,337	262
Cancun	270,071	4,211	11,588	110	7,580	84
Manhattan	350,183	4,197	13,772	543	6,048	111
Paris	328,926	4,424	13,204	1,637	5,787	108
TOTAL	1,302,231	17,762	53,358	2,566	28,752	565

Number of observations, users, queries, hotels, clicks, and purchases, by market and in total. Note: I only use observations from Paris in the analysis.

Table 2: Search activities for buyers and non-buyers (averaged over users)

	Market	# Queries	# 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	42.12	1.71	0.29
	Cancun	2.69	35.21	1.72	0.25
	Manhattan	3.19	51.96	1.31	0.43
	Paris	2.90	49.12	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 "query" is made whenever any request setting is changed, or a refinement method is chosen. Note that I only use observations from Paris in my subsequent analysis.

up buying) ever view hotels that are not ranked by default (not shown in table).

In accordance with this, Table 3 shows that most views, clicks and purchases occur under the default ranking. Again, this stresses the importance of hotels' position in OTAs' default ranking 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 that better-ranked hotels are likely also of higher quality, and therefore do not allow for a causal interpretation.[24]

Appendix D.1 provides further descriptive facts on how users search for hotels online, focusing on Paris.

4.2 Ranking Side: Web-Scraped Data

The second dataset is web-scraped from Expedia, which is the dominant OTA in the US. In the months of February and March 2019, I carried out daily hotel queries for hotel stays in the cities covered in the search and transactions data (i.e., Budapest, Cancun, Manhattan and Paris). I conduct queries for trips beginning between one and 250 days after the query date (all in all, for 16 different travel dates

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

Table 3: WCAI data: sorting behavior and importance of default ranking

Ranking	% of users who ever view given ranking	Number of clicks	Number of purchases
Default ranking	97.48	20,523	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	17.53	4,565	70
Sort by reviews	1.81	407	12

Left column shows the percentage of users who ever see search results that are ranked to a given ranking, using data from all four cities. (Since many users see the default ranking and may take refinement actions in addition to that, it is natural that the percentages for users does not add up to 100.)

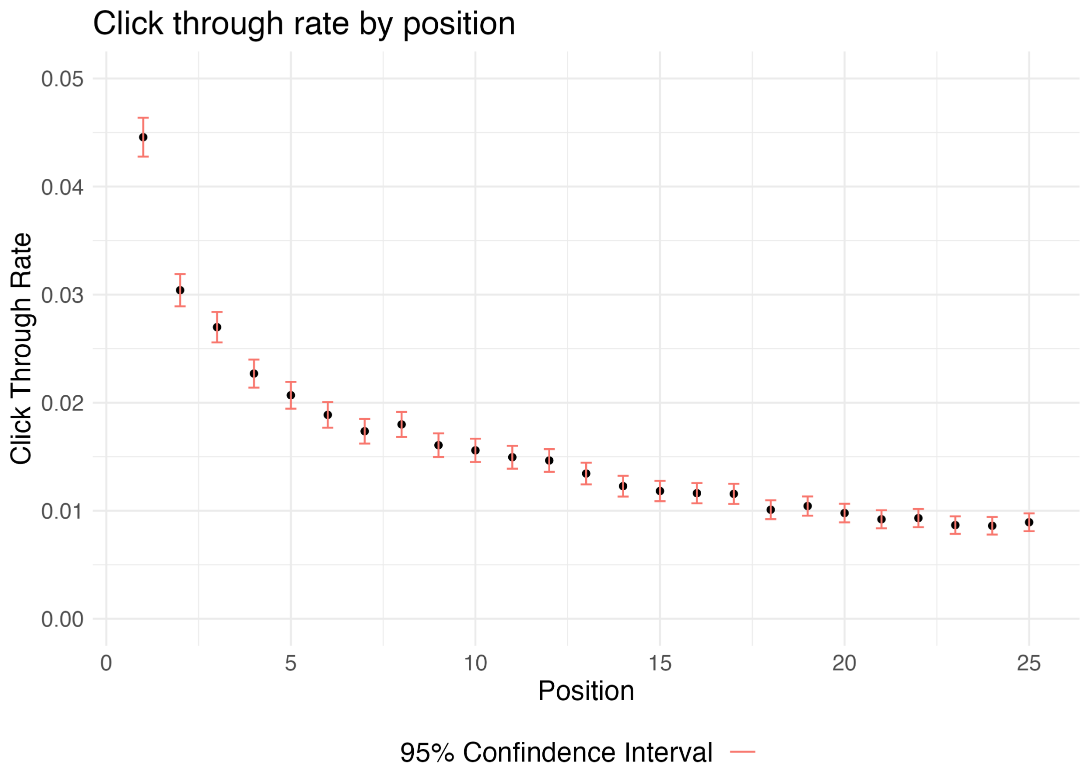

Figure 2: Average click through rate by position on a given results page. I exclude searches in which fewer than 20 hotels were displayed on a given page (which would occur if strict filtering methods are used, or if the query specifies the property name). This plot uses observations of clicks and positions occurring under any ranking: when focusing when focusing on data generated by the default ranking only, or alternatively any other than the default ranking, the results look qualitatively extremely similar (but are shifted downwards for the default ranking, and upwards for other rankings).

or more each day per location), such that the latest travel dates are in November 2019.[25] During each such query, I obtain hotels' default position, pricing, promotions, and additional characteristics from

[25] I do so to reflect the observation that users' booking windows are also very heterogeneous, as shown in the WCAI dataset, see Appendix D.1.

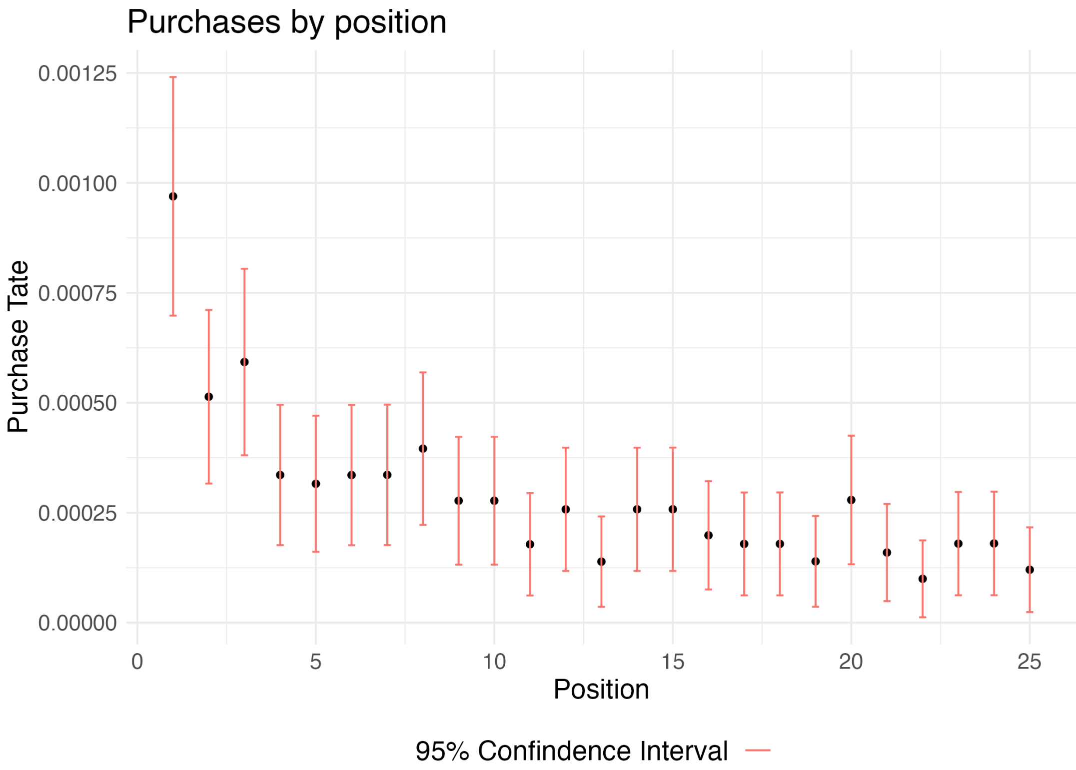

Figure 3: Average purchasing rate by position on a given results page. I exclude searches in which fewer than 20 hotels were displayed on a given page (which would occur if strict filtering methods are used, or if the query specifies the property name). This plot uses observations of clicks and positions occurring under any ranking: the findings look qualitatively extremely similar when focusing on default ranking. However, the downward trend becomes very noisy when considering only queries under alternative rankings, due to a low number of purchases ever occurring under alternative rankings.

all results pages that appear in a query.[26] I employ search parameters that are empirically most often being entered by consumers (see Appendix D.1), and are thus the most relevant for hotels setting prices, namely weekend or single night stays for two persons in one room. All in all, the data cover over 6,000 distinct hotels which I identify as active on the platform, over 2,800 of them in Paris. As seen on the screen shot in Figure 1, hotels offer various extras to entice consumers (free cancellation, free offers, discounts, a "Sale!" flag, etc.). I collect all these information from the website, too, allowing me to potentially control for these factors in my analysis. In the resulting dataset, an observation is a hotel j that the "scraper" views at query date q for a potential single-night stay starting at travel date t.

The analysis focuses on hotels' decisions on which prices to set, which is a variable that seems influential for both a hotel's rank as well as the purchasing likelihood.[27] I find substantial variation in hotels' pricing decisions over time.

[26]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. See Appendix C.2 for further details.

[27]Endogenizing a hotel's decision to offer a sale is difficult due to the opacity of the types of sales that can be offered, and a lack of understanding how hotels make sales decisions or how pricing and sales decisions interact; see the discussions in Appendices B and D.3.

As seen in Table 4, the distribution of prices across observations seems to have a long right tail. Moreover, relatively many hotels display a "Sale!" flag at any point during the query period. Appendix D.3 provides further descriptive results on hotels' promotion decisions. Figure 4 shows the variation in average prices across time by star rating of the hotel in Paris, and strikingly illustrates the strong seasonality in pricing. Especially those hotels that have four or five stars appear to show significant variation of average prices over travel dates, possibly reflecting changing demand patterns. Interestingly, the month of August – typically the travel month in France during which certain amenities as well as sights are closed – shows very little day-to-day variation in prices.

Table 5 displays the between and within variation in positions, prices, and sales in the Paris data (which are used for the estimations below). The within-hotel variation in these variables is very large: the within-variance of prices amounts to 65 euros, and the within-variance in position to 341. To illustrate this high variation within hotels further, I find that the median hotel's difference between the maximum and the minimum price across different travel dates amounts to 164 euros (not shown in table). The median hotel's difference between the maximum and the minimum rank across different travel dates amounts to almost 1,700 positions.

Table 4: Pricing behavior across cities

City	Active Hotels	Median price (€)	Avg. price (€)	% of hotels with "Sale" flag	% of hotels ever with "Sale" flag
Budapest	1,045	71	277.03	29.05%	36.17%
Cancun	972	85	180.02	18.11%	23.77%
Manhattan	1,222	259	300.87	18.80%	24.55%
Paris	2,842	135	168.09	28.55%	48.63%

Averages are computed over all query-travel date observations in each of the cities.

Figure 5 shows the fraction of hotels with a "Sale" flag at any travel date, separately for hotels of different star ratings.[28] Note that for observations in the month of February and March, the booking window is very short, as scraping took place in those months. In those months, the fraction of hotels with sales offers is very low across all star ratings, amounting to 10-20%. The fraction of hotels on sale subsequently grows and stabilizes from May onwards, and varies considerably between hotels of different star ratings: interestingly, both 1- and 5-star hotels rarely offer sales. In contrast, 4, 2, and 3-star hotels offer sales more often, amounting to up to 35% of observations for 3-star hotels. As pointed out previously, promotions on Expedia are relatively opaque, with the "Sale" flag possibly having different meanings that can only be apparent when hovering above it with a mouse, and thus potentially requiring greater effort by users. These promotions are studied in greater detail in Appendix .

Ursu (2018)'s dataset and her specification for consumer indirect utility includes a hotel's location score (with better location being equivalent to a higher score), which is likely to be important given that

[28] Note that 36% of hotels in Paris do not have any star rating and are thus excluded from the plot.

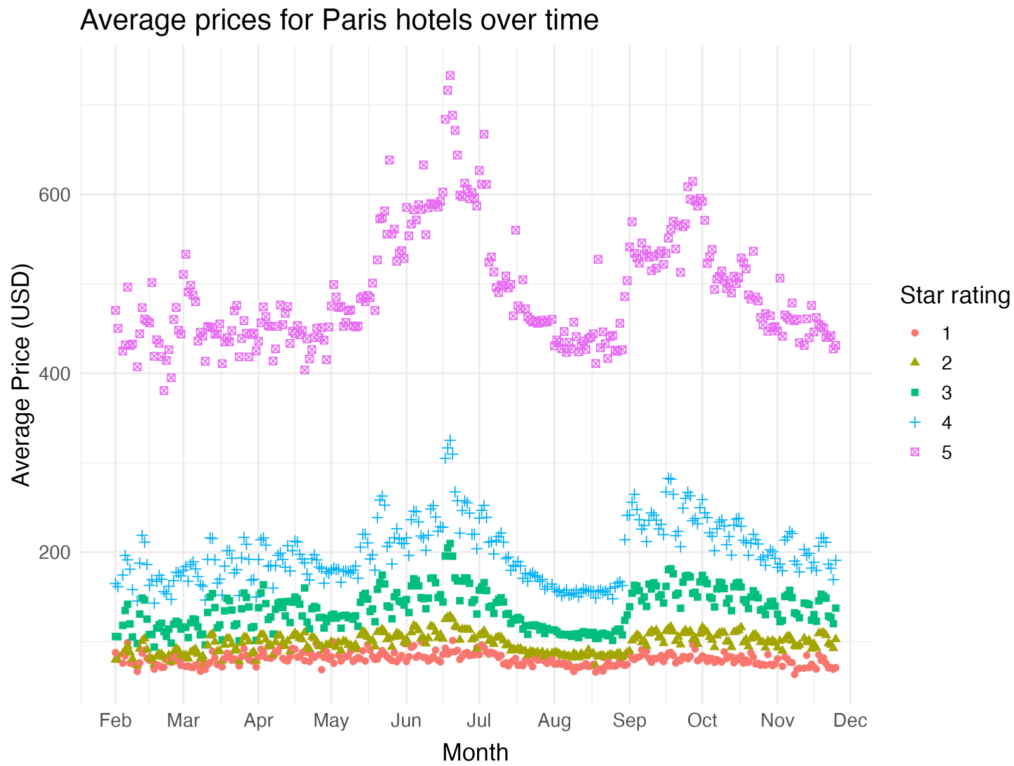

Figure 4: Prices of Paris hotels over travel dates.

Table 5: Within and Between Variance, Paris data

Variable		Mean	Std.Dev.	Min	Max	Observations
Position	overall	1020.6	608.6	1	2466.0	N = 1,202,289
	between		604.2	44.5	2396.8	n = 2,842
	within		340.7	-897.0	3006.4	T-bar = 423.0
Price	overall	168.1	271.9	27	13768.0	N = 1,202,289
	between		355.1	39.33	12460.1	n = 2,842
	within		65.3	-3020.1	12079.8	T-bar = 423.0
Sale	overall	0.286	0.451	0	1	N = 1,202,289
	between		0.346	0	1	n = 2,842
	within		0.271	-0.713	1.283	T-bar = 423.0

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 (2018) dataset.

Appendix C.3 compares the WCAI to the web-scraped dataset.

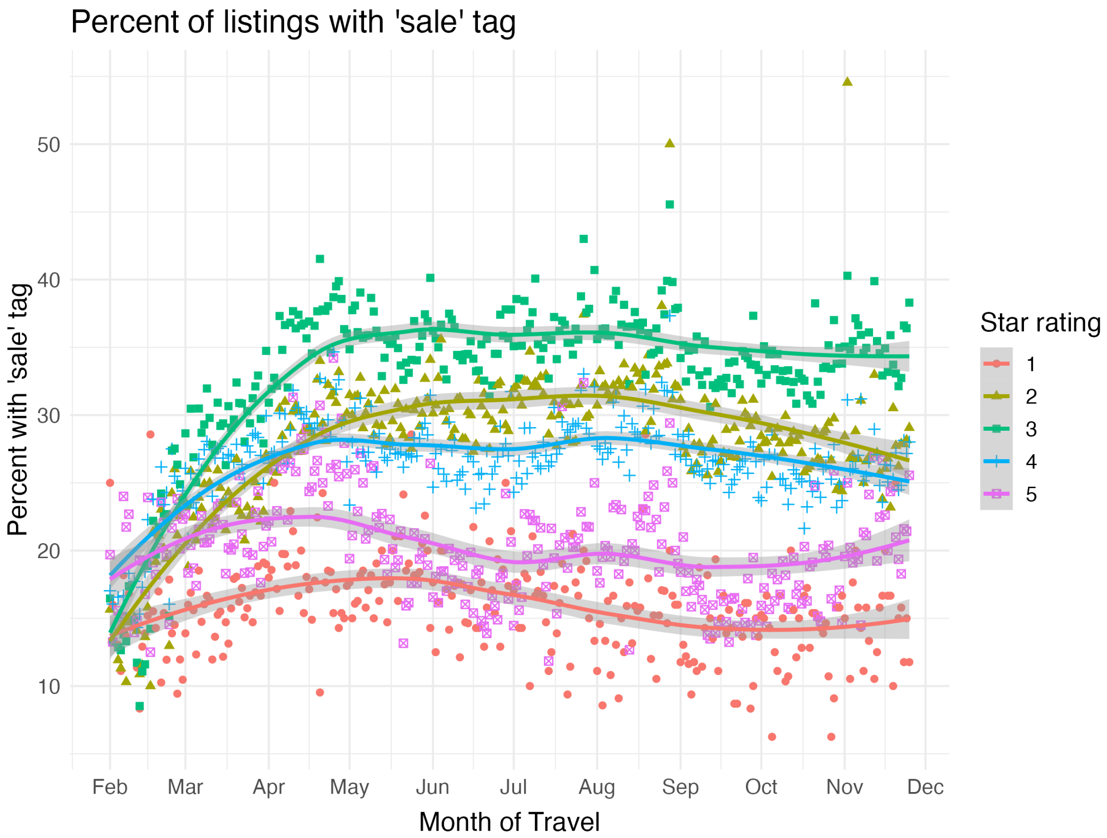

Figure 5: Average percent of listings with "Sale" flag at a given travel date for hotels in Paris. Locally estimated scatterplot smoothing is used for plotting the smoothed lines. Note that for observations in February and March, the booking window is very short. The plot looks very similar when looking at the average percentage of hotels offering any percentage discount greater than 0.

4.3 Further Datasets

I obtain further datasets that allow to construct the instrumental variables. First, hotels' addresses are web-scraped from Expedia as well. I then use the Google Maps API to obtain hotels' geolocation (latitude and longitude). 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 58,067 Airbnbs in Paris from the website insideairbnb.com. For each Airbnb, the data contain information on the Airbnb's zip code it is located in, as well as price and availability from beginning Feburary 2019 for the following 365 days.[29]

[29] I drop Airbnbs that have not had any booking for the past 12 months, as those seem to be inactive, and end up with 37,885 Airbnbs located in 50 zip codes. I then construct average prices at the zip code travel date level.

5 Empirical Strategy

5.1 Effect of Prices on Ranking: Linear Model with Fixed Effects

The first step of the analysis is to estimate the effect of a given hotel's price on that hotel's position in the ranking, denoted $r'_{jqt}(p_{jqt})$ in the theoretical model of Section 3. To capture this relation in a high-level sense, I employ a linear model with hotel and query-travel date fixed effects. I use observations of hotels, their pricing decisions, and their respective rankings observed in the web-scraped data to identify the parameters of interest. The equation of interest writes:

$$r_{jqt} = \alpha p_{jqt} + \beta d_{jqt} + \delta past\_book_{jq} + \gamma_j + \lambda_{qt} + \epsilon_{jqt} \tag{3}$$

where r_{jqt}, p_{jqt} and d_{jqt} indicate hotel j's average position, price and an indicator variable for specifying whether a hotel is on sale[30] for travel date t and query date q. The coefficient of interest is the price coefficient, α, which can be viewed as an estimate for $r'_{jqt}(p_{jqt})$.[31] $past\_book_{jq}$ is a variable counting the number of bookings (for any travel date) that have been made within a 48-hour window prior to scraping a given hotel hotel and that is displayed on some listings, as seen in Figure 1. It serves to control (to some extent) for demand shocks that a hotel may experience and that the ranking algorithm may pick up. The promotion dummy d_{jqt} allows to control for whether or not a hotel's listing has a "Sale!"-flag, which may affect the hotel's ranking. Hotel fixed effects γ_j allow to control for any unobserved hotel characteristics that are fixed over time. Employing these 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 factors that affect the business relation between the hotel and the platform, such as the amount of commission the hotel pays. As long as commissions or quailty do not vary over time – and this is unlikely to be the case for the 2-month period during which the web-scraping took place –, the fixed effects can account for them. The focus of this analysis is therefore on how a *given* hotel's average ranking on Expedia – as observed by the web-scraper making the query (and which proxies for the average customer) – reacts to the price it sets, holding the hotel's identity fixed. λ_{qt} is a query date-travel date fixed effect accounting for seasonal demand patterns, for instance.

5.1.1 Endogeneity and Instrumental Variables

The hotels' positions, prices, and sales decisions observed in the data are equilibrium values that are simultaneously determined by both, a hotel's reaction to the behavior of the ranking algorithm, as well as the algorithm's elasticity with respect to hotels' prices. As a result, the variables of interest in Equa-

[30] As indicated by having a "Sale!" flag, as shown in Figure 1. The sale flag typically denotes a percentage reduction in the price. See Appendix D.3 for a discussion.

[31] If we were to endogenize hotels' promotion decision as well, we would also be interested in β.

tion (3) are endogenous, and any ordinary least squares estimation would yield biased estimates.[32] I use instrumental variable techniques to resolve this endogeneity concern. Valid instruments are variables that exogenously shift price or promotion decisions of a given hotel across booking or query dates, and that are orthogonal to anything that may influence that hotel's position. Instruments for prices 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.

To understand hotels' pricing decisions in greater detail, I conducted an interview with a representative of the German hotel association (*Hotelverband Deutschland*) in February 2020. I was told that hotels often employ revenue management software. This software suggests prices and takes into account competitors' prices as well as major events taking place in the neighborhood, and provides a rationale for employing the below instruments. Overall, I provide estimates using different sets of instruments, each of which have their advantages and disadvantages.

Neighborhood Instruments. As a first instrument, I use the prices and sales indicator variables of hotels that are located within 500 meters distance of a focal hotel. The rationale why this should be a relevant instrument stems from both the anecdotal evidence from my abovementioned interview with the German hotel association, as well as academic literature. Li et al. (2017) 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 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.

Using this set of instruments, the identification rests on the assumption that local demand conditions influence hotels' price setting decisions, but that hotels have better knowledge about these conditions than Expedia does, so that local demand conditions do not affect the hotel's rank.[33] The instrument is not valid if Expedia was able to anticipate future bookings of consumers as well as hotels, and thus

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

[33] One might also think that, during times of high demand, hotels set prices to match each others' rates and have their rooms occupied at high prices, and might be less concerned about their ranking at those times. Also, note that over 2,500 active hotels are listed on Expedia for Paris that all appear in the ranked list, and hotels are capacity constrained and can only be allocated into one of the ranking's slots, so that there must be some randomness in how hotels are being ranked.

display hotels with a high demand shock – and thus at a time at which their prices are high – more visibly.

Airbnb 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 zip code) as the focal hotel are likely to be correlated with the focal hotel's price, but not with the focal hotel's position. To identify both the price as well as the sales parameter, I additionally employ the interaction of an indicator variable indicating certain weeks in August in which hotel prices are not very dispersed (see Figure 4) with the zip code.[34] The argument for the exogeneity of this instrument is similar to the one of the neighborhood instrument.

The Airbnb data do 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. Denoting $\bar{p}_{jt}$ the average of p_{qjt} over all query dates q (and analogous for the other variables), I thus estimate the parameters of the following model:

$$\bar{r}_{jt} = \alpha^{avg\_q}\bar{p}_{jt} + \beta^{avg\_q}\bar{d}_{jqt} + \delta^{avg\_q}\overline{past\_book}_{jq} + \gamma_j^{avg\_q} + \lambda_t^{avg\_q} + \eta_{jt} \tag{4}$$

A drawback of this instrument is that the variation in Airbnb prices across travel dates less strong. As Figure 6 shows, while there is some seasonal variation across months, price variation for Airbnbs is still substantially smaller than for hotels. Moreover, I find that Airbnbs tend to be priced higher on Fridays or Saturdays, whereas hotels are priced highest on Tuesdays and Wednesdays, reflecting demand coming from different types of customers.

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. My web-scraped dataset contains observations of 39 different hotel brands with at least two outlets for Paris alone. 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 sales indicator variables of hotels of the same brand as instrument for the focal hotel's pricing and sales decision, as these are correlated with the focal hotel's decisions, but should not causally influence the focal hotel's position in the ranking.

[34] Anecdotally, August is a holiday month in France, with many attractions and even amenities in Paris being closed. Based on Figure 4, it does seem like competition between hotels works in some sense differently during those months. In the past, I have also employed indicator variables 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, interacted with the zip code. As these instruments tend to not be very relevant, I have excluded them in my current version of the results.

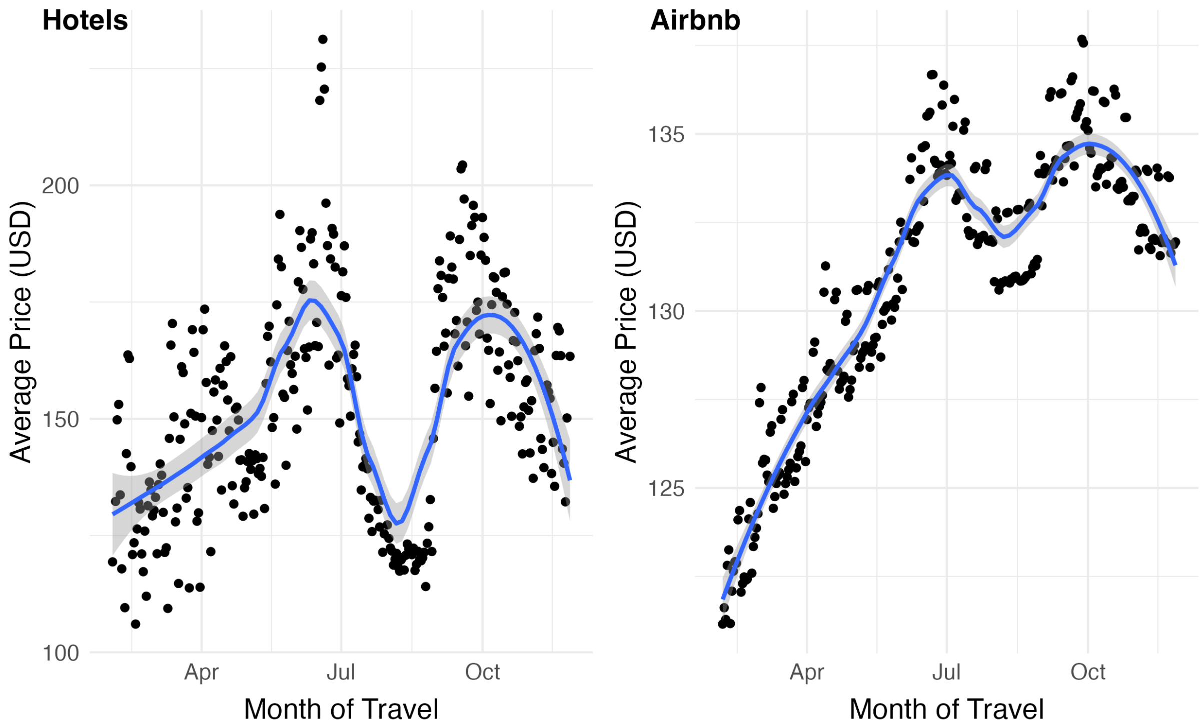

Figure 6: Average prices of hotels and Airbnbs over travel dates. Locally estimated scatterplot smoothing is used for plotting the smoothed lines. Beware that the y-axes have different scales. I drop 5-star hotels to reduce noise.

The identification assumption is that pricing and sales decisions are being made on an at least city-wide level, and therefore less in response to the ranking. A drawback of this instrument is that its application is of course limited to those hotels which indeed are brand hotels, therefore accounting for a subset of only 25% of observations.

Dynamic Panel Instruments. A technique that has been employed in the literature especially for production function estimation (see Arellano and Bond (1991)) is the use of dynamic panel instruments. To use these instruments, I am again aggregating the hotel-query date-travel date panel, so that each observation corresponds to one hotel-travel date observation averaged across queries. 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 the lagged difference of hotel prices as instrument for the current prices.

5.2 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 pages observed in the WCAI search data are not randomly ranked, I employ the estimated search cost parameters 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 will only employ her model for estimating the utility parameters of hotel observables such as stars or location score.[35] Just like Ursu (2018), I also do not instrument for prices, and instead 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. I refer to Ursu (2018) for details of the log-likelihood estimation. Overall, the demand side estimates enable me to simulate consumers' purchasing choices under any counterfactual ranking of hotels.

6 Results

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

6.1 Ranking

Tables 6 to 9 display the results of the effect of a hotel's price on the rank position which it is displayed in. Each table provides a specification that accounts for a sales indicator in addition to a hotel's price, or does not. Moreover, in each table, results using instruments are juxtaposed with the OLS results stemming from the same data. All specifications include hotel and query-travel-date fixed effects (or travel-date fixed effects, respectively, for Tables 7 and 9, which are based on the somewhat aggregated version of the data).

Across specifications, I find a robust positive coefficient for price that, using instruments, ranges between 3.4 and 3.8 for the query-level results in Tables 6 and 8, and between 1.5 and 2.3 for the results based on the more aggregate travel-date panel in Tables 7 and 9. The coefficient is significantly different from 0 in all specifications, and, with the exception of columns (3) and (4) in Table 7, increases in magnitude once instruments are employed. Its value implies that a 1 dollar increase in price would lead to a shift in the hotel's position by roughly one to four positions 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 not minor, 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 more blurry picture. While it is negative and varies between -74 and -104 in the OLS specifications, it jumps up (or down, respectively, when using the brand-based instruments in Table 8), to a substantial extent once instrumented for, peaking at -480 in Table 7. The reason for this is most likely that the instruments for sale are relatively weak, and that it is more challenging to identify parameters with two, instead of one, endogenous variable.

[35]The underlying assumption is 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 the hotel preference parameters using consumer search data from Paris directly.

Interestingly, the number of bookings in a 48-hour window tends to increase visibility of a given hotel, indicating that the ranking algorithm may indeed take popularity into account.

All Tables report the F-test statistics for a test of relevance of the excluded instruments on price and sales respectively. All instruments pass the weak instruments rule of thumb, although the Arellano-Bond type instruments tend to be the weakest. The corresponding first stages are reported in Tables **??** to 22 in Appendix E.

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).

Table 6: Results using neighborhood-based instruments

	Dependent variable:			
	position			
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
price	2.007***	3.463***	1.973***	3.394***
	(0.216)	(0.146)	(0.214)	(0.150)
1{sale}			−74.041***	−290.785***
			(4.348)	(105.237)
# bookings past 48h	−0.945***	−0.820***	−0.894***	−0.617***
	(0.203)	(0.194)	(0.201)	(0.204)
Query×travel-date FE	✓	✓	✓	✓
Hotel FE	✓	✓	✓	✓
1st stage F-stat of excl. instruments on price		104.2		77.25
1st stage F-stat of excl. instruments on sale				85.58
Observations	914,814	914,814	914,814	914,814
Adjusted R^2	0.730	0.711	0.731	0.701

Standard errors clustered at query×travel-date level. *p<0.1; **p<0.05; ***p<0.01

Column (2) uses as instruments for price the number of neighboring hotels available on a given query and travel date, and the average price of neighboring hotels at a given travel and query date. Column (4) uses as instruments for price and the sale indicator the number of neighboring hotels and the average price and sales indicator. Neighboring hotels are defined as being located within 500 meters of the focal hotel.

Robustness: Neighbors based on "donut"-shaped area around focal hotel. As noted above, the exogeneity of the neighborhood instrument relies on the assumption that the prices of neighboring 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

Table 7: Results using Airbnb and 1{August}×zip code instruments (based on more aggregated travel-date panel of hotels)

	Dependent variable:			
	average position across queries			
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
price	1.908^{***}	2.066^{***}	1.880^{***}	1.477^{***}
(average across queries)	(0.338)	(0.171)	(0.336)	(0.198)
1{sale}			-76.533^{***}	-492.178^{***}
(average across queries)			(8.274)	(146.941)
# bookings past 48h	-1.177^{**}	-1.174^{**}	-1.143^{**}	-0.966^{**}
(average across queries)	(0.510)	(0.506)	(0.494)	(0.451)
Travel-date FE	✓	✓	✓	✓
Hotel FE	✓	✓	✓	✓
1st stage F-stat of excl. instruments on price		286.99		117.88
1st stage F-stat of excl. instruments on sale				72.91
Observations	447,378	447,378	447,378	447,378
Adjusted R^2	0.787	0.787	0.788	0.760
Standard errors clustered at travel date level.			*p<0.1; **p<0.05; ***p<0.01	

The dependent variable is hotel j's average position in Expedia's listings page for travel date t across web-scraped queries. Instruments used for mean_price in column (2) are the price and availability of Airbnbs located in the same zip code as a focal hotel. Instruments used for mean_price and mean_sale in column (4) are both the price and availability of Airbnbs located in the same zip code as a focal hotel, as well as an indicator variable for the weeks in August with little price variation interacted with the zip code.

Table 8: Results using brand-based instruments

	Dependent variable:			
	position			
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
price	1.199^{*}	3.757^{***}	1.179^{*}	3.743^{***}
	(0.652)	(0.284)	(0.645)	(0.284)
1{sale}			-103.698^{***}	-37.077
			(16.042)	(33.557)
# bookings past 48h	-0.235	-0.197	-0.155	-0.169
	(0.271)	(0.254)	(0.262)	(0.250)
Query×travel-date FE	✓	✓	✓	✓
Hotel FE	✓	✓	✓	✓
1st stage F-stat of excl. instruments on price		54.28		77.93
1st stage F-stat of excl. instruments on sale				1126.23
Observations	258,011	258,011	258,011	258,011
Adjusted R^2	0.738	0.651	0.739	0.652
Standard errors clustered at query×travel-date level.			*p<0.1; **p<0.05; ***p<0.01	

Column (2) uses as instruments for price the average price charged by other hotels in Paris of the same brand. Column (4) uses as instruments for price and the sales indicator the average price and average sales indicator employed by other hotels in Paris of the same brand.

Table 9: Results using Arellano-Bond instruments (based on more aggregated travel-date panel of hotels)

	Dependent variable:			
	average position across queries			
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
price	1.813^{***}	2.349^{***}	1.788^{***}	2.322^{***}
(average across queries)	(0.361)	(0.436)	(0.358)	(0.435)
1{sale}			-76.228^{***}	-95.585^{***}
(average across queries)			(8.645)	(28.868)
# bookings past 48h	−0.341	−0.326	−0.320	−0.300
(average across queries)	(0.270)	(0.260)	(0.262)	(0.251)
Travel-date FE	✓	✓	✓	✓
Hotel FE	✓	✓	✓	✓
1st stage F-stat of excl. instruments on price		20.79		36.28
1st stage F-stat of excl. instruments on sale				21.62
Observations	451,366	451,366	451,366	451,366
Adjusted R^2	0.795	0.793	0.796	0.794
Standard errors clustered at travel date level.			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

The dependent variable is hotel j's average position in Expedia's listings page for travel date t across web-scraped queries. Column (2) uses the lagged difference of prices as an instrument for a hotel's average price on travel date t. Column (4) uses the lagged differences of prices and sales indicator variable as instruments for a hotel's average price and sales indicator on travel date t.

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.[36] 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 10. The instruments are somewhat weaker than in Table 6, as indicated by the lower first stage partial F-test of the excluded instruments, but still considered strong enough. The instrumented price coefficient in column (2) is very similar as in Table 6, whereas the coefficient on the instrumented sales indicator in column (4) increases substantially in magnitude, possibly due to weaker instruments.

Robustness: Aggregated panel. I estimate the parameters using the neighborhood-based as well as the brand-based instruments also on the more aggregate hotel-travel date panel (i.e., the panel used for the Airbnb and Arellano-Bond type instruments above). The results, displayed in Tables 23 and 24 in Appendix F, are extremely similar to my findings above.

[36] 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 10: Results using neighborhood-based instruments (“neighbor” = hotels ∈ [500m, 1,000m) from focal hotel)

	Dependent variable:			
	position			
	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)
price	2.008***	3.521***	1.974***	3.316***
	(0.214)	(0.166)	(0.213)	(0.170)
1{sale}			−73.811***	−874.567***
			(4.314)	(134.050)
# bookings past 48h	−0.890***	−0.759***	−0.845***	−0.209
	(0.180)	(0.171)	(0.178)	(0.204)
Query×travel-date FE	✓	✓	✓	✓
Hotel FE	✓	✓	✓	✓
1st stage F-stat of excl. instruments on price		54.05		40.52
1st stage F-stat of excl. instruments on sale				94.99
Observations	927,322	927,322	927,322	927,322
Adjusted R^2	0.732	0.713	0.734	0.573

Standard errors clustered at query×travel-date level. *p<0.1; **p<0.05; ***p<0.01

Column (2) uses as instruments for price the number of neighboring hotels available on a given query and travel date, and the average price of neighboring hotels at a given travel and query date. Column (4) uses as instruments for price and the sale indicator the average price and sales indicator. Neighboring hotels are defined as being located in a “donut” shaped circle of 500 to 1,000 meters of the focal hotel.

Robustness: Linear probability model. The above linear regressions assume that a hotel’s position can be linearly related to its price. I explore alternative specifications by employing as dependent variable an indicator variable specifying whether a given hotel is displayed at the top 10 (or top 5, respectively) positions. The results are shown in Appendix F Tables 25 and 26 using neighborhood-based instruments. Clearly, the effect of price retains its sign across specifications, i.e., a higher price is associated with a decline in the probability of being displayed very visibly. For the parameter on the sales dummy, it looses significance once instruments are employed.

Robustness: Other travel destinations. Preliminary evidence shows that the results are quantitatively somewhat different across destinations, but qualitatively the same. It is plausible that they might be different, since the competitive environments across the different cities seem very different (many small hotels (Paris) vs. few large hotels (Cancun); strong vs. weak seasonality in demand; 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.

Table 11: Parameter estimates of the search model, using WCAI Paris data. Results based on different starting values and different definitions of "location score".

	Using Louvre-based location score:		*Using more complex location score:*	
	(1a)	(1b)	(2a)	(2b)
price ($100)	-0.176***	-0.176***	-0.179***	-0.179***
	(0.021)	(0.022)	(0.027)	(0.027)
stars	0.072***	0.072***	0.075***	0.075***
	(0.023)	(0.022)	(0.026)	(0.026)
review score	0.033**	0.033*	0.035*	0.035*
	(0.019)	(0.021)	(0.026)	(0.026)
chain dummy	-0.047*	-0.047*	-0.042*	-0.042*
	(0.030)	(0.030)	(0.031)	(0.031)
location score alt1	0.035***	0.035***		
	(0.011)	(0.012)		
location score alt2			0.044***	0.044***
			(0.013)	(0.013)
sale	0.052	0.052	0.049	0.049
	(0.046)	(0.047)	(0.055)	(0.055)
outside option	0.620***	0.622***	0.668***	0.668***
	(0.087)	(0.087)	(0.086)	(0.086)
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.

Table 12: Fixed parameters, search model

search cost constant (k)	-1.0305
search cost position parameter (γ)	0.0044

Note: Parameters are taken from Ursu (2018) and are being held constant across specifications of the demand model.

6.2 Demand

Table 11 displays the demand estimates which are derived using Ursu (2018)'s 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 fact, only 16% of all searches): 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.[37] Moreover, as explained above, I use the search cost parameter estimates from Ursu (2018) instead of estimating them again (in particular, I employ the results displayed 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 (see Table 12). 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.[38] 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 sales parameter has a positive sign, but is insignificant, which reflects Ursu (2018)'s results where this coefficient is insignificant in three out of the four destinations. All in all, the results I obtain seem intuitive and are in line with the results of both Ursu (2018) as well as Y. Chen and Yao (2016).

7 Counterfactuals

The goal of the counterfactuals is to study hotels' price-setting decisions and consumers' welfare gains or losses under a counterfactual ranking algorithm. I only simulate changes in hotels' pricing decisions, and abstract from any changes in hotels' willingness to offer a sale. As above, this is motivated by the lack of understanding about what determines hotels' decisions to offer a discount.

Recall that I do not observe demand in my web-scraped dataset. To compute market shares, I therefore employ the structural demand parameters estimated on the 2009 consumer search and purchase dataset. I then simulate consumers that arrive on Expedia and are confronted with a list of hotels corresponding to the listings observed in the web-scraped dataset from 2019. The underlying assumption is that consumers' preference parameters have not changed over time.

[37] Information on the exact data cleaning process can be found in the Appendix, Section C.1.

[38] Accounting for hotel stars and checkin date, the zip code of the Louvre – 75001 – persistently features the highest priced hotels, possibly indicating a premium consumers are willing to pay for the perceived centrality of that location.

7.1 Method

To conduct the counterfactual simulations, I make a simplification in that I carry out counterfactuals only for observations of hotels displayed on the first results page. This is motivated by the fact that consumers hardly ever click or purchase hotels on later results pages, leading to almost zero demand for hotels displayed on those pages.

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 $r'_{jqt}(p_{jqt})$, i.e., the effect of a hotel's price on a hotel's position obtained from the estimations above. Re-arranging the first order condition (1) yields:

$$\tilde{c}_{jqt} \quad = \quad p_{jqt} \quad + \quad \underbrace{\frac{s_{jqt}(\cdot)}{\frac{\partial s_{jqt}(\cdot)}{\partial p_{jqt}} + \frac{\partial s_{jqt}(\cdot)}{\partial r_{jqt}} \cdot \hat{\alpha}}}_{<0} \tag{5}$$

Prices p_{jqt} are observed in the data. To approximately match Expedia's commission rate, I set $\tau = 0.2$. As the web-scraped dataset does not contain demand data, 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 query-travel date observations to obtain $\overline{\frac{\partial s_j(\cdot)}{\partial p_j}}$ and $\overline{\frac{\partial s_j(\cdot)}{\partial r_j}}$. Assuming that these estimates are the same across hotels, I then take the average again, obtaining a value for $\overline{\frac{\partial s(\cdot)}{\partial p}}$ and for $\overline{\frac{\partial s(\cdot)}{\partial r}}$, respectively. I can then back out costs using the following equation:

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

Given all parameters and hotels' cost estimates, one can perturb $\hat{\alpha}$ and simulate prices and market shares under different scenarios. Markups can easily be derived by transforming Equation 6:

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

7.2 Results for costs and interpretation

The simulations yield $\overline{\frac{\partial s(\cdot)}{\partial p}} = -0.014$ and $\overline{\frac{\partial s(\cdot)}{\partial r}} = -0.032$. 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. (2020) 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 obtain absolutely no demand on Expedia. 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 itself 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, as well as the substantial price fluctuation over time that we saw in Figure 4, highlight once again 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 likely to be 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	8.4×10^{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	3×10^{-7}	3.9×10^{-4}	4.2×10^{-4}	6.7×10^{-4}	9.3×10^{-5}
# of hotels	4	54	312	403	83

7.3 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. 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 decrease 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. 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 9).

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

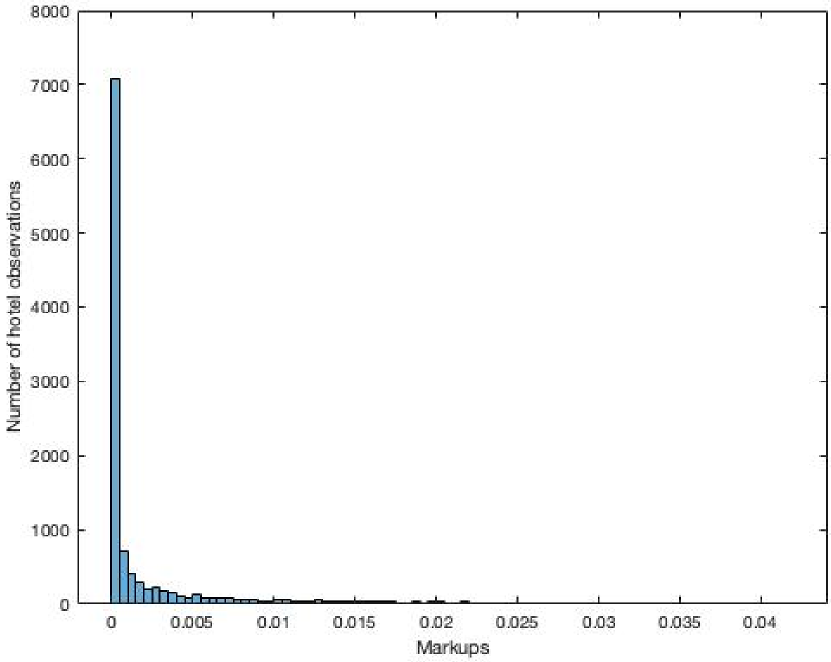

Figure 7: Distribution of markups across hotel observations, after removing the 63% of observations with markups = 0.

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.

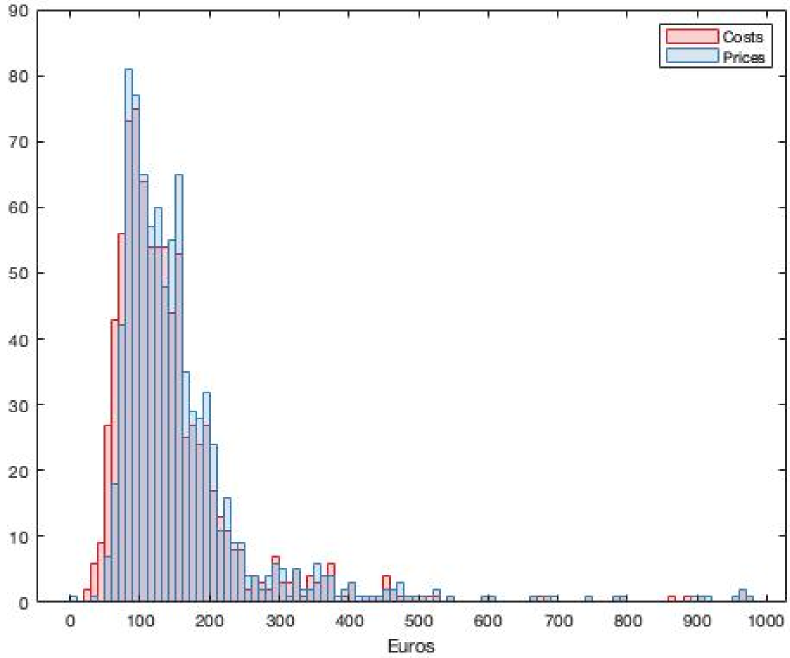

Figure 8: Distribution of costs and average prices across hotel observations.

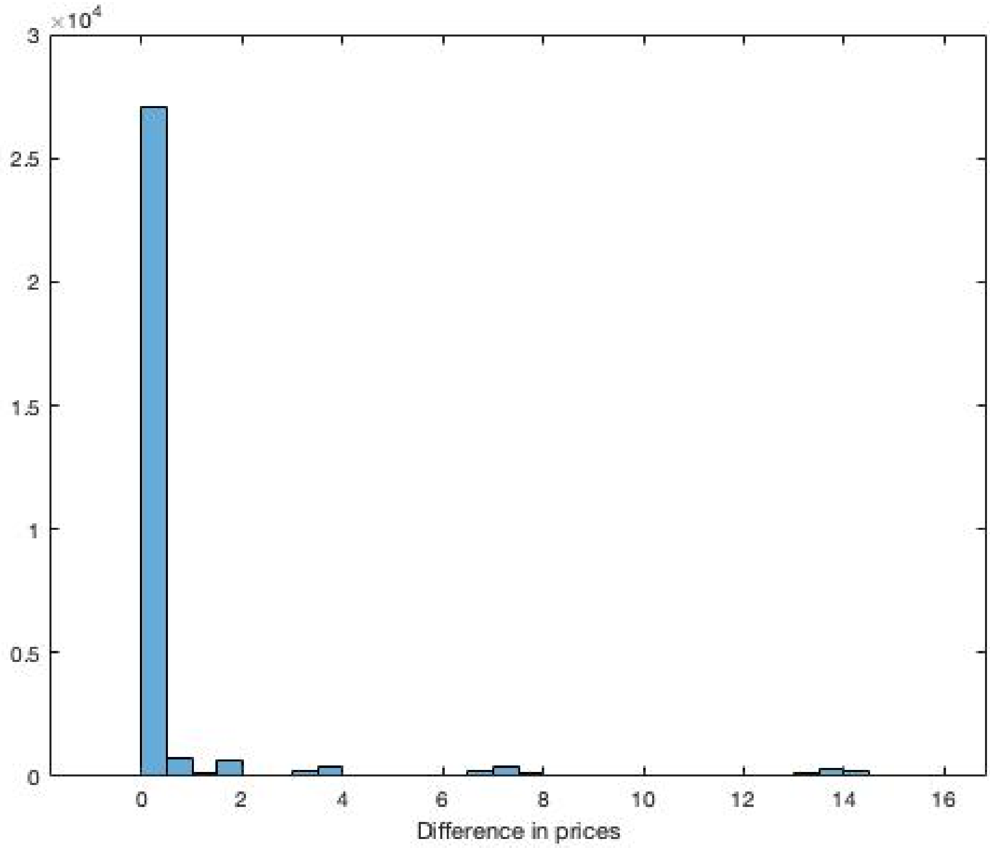

Figure 9: Change in prices (in euros) when effect of ranking is shut off.

8 Conclusion

The seemingly infinite virtual shelf space available online has given rise to a cornucopia of products competing for consumers' attention. As consumers face search costs, online platforms play an essential role in designing algorithms that make products discoverable and that guide consumer search. As visibility is important to generate demand, online sellers may adjust their strategic decision-making so as to be displayed more favorably by the algorithms.

In the context of Expedia, I find that the algorithms that these intermediaries employ tend to display hotels more visibly at times at which they are priced lower. This tends to intensify price competition among hotels, thus moving the market into a somewhat more competitive equilibrium. The structural model of the demand as well as the supply side allows to simulate how hotels would set prices under a counterfactual ranking algorithm. I find that, if prices did not affect hotels' visibility, hotels would on average increases prices by 1.89€. 7% of hotels increasing their prices by more than 10€.

These results and the tools I use are, first, important for managers of online platforms: the equilibrium effects my analysis sheds light on can *not* be traced out by merely using A/B experiments that randomize the hotel rankings across website visitors. Instead, investigating these effects requires a structural model of the buyer, as well as the seller side, allowing to predict how changes of the ranking algorithm affect sellers' equilibrium prices. My results suggest that sellers tend to be squeezed by Expedia's ranking algorithms, whereas consumers are enticed by providing higher competition. Recall that platforms, however, crucially need to attract both seller, as well as buyer-side in order to subsist. My finding therefore has managerial implications for platforms seeking to trade off between providing value to the two sides, while at the same time also being profitable in the short as well as long run.

Second, my results are interesting for antitrust authorities, which have in the past been concerned about online booking platforms' market dominance and behavior with respect to hotels. My finding that booking platforms intensify price competition provide suggestive evidence that online platforms provide additional value to the consumer side – at the expense of hotels, however. This is aligned with the complaints that hotel associations in many countries have voiced regarding the dominance of online platforms.

The analysis suffers from several limitations. The linear relationship between rank and hotels' prices is an abstraction of a platform's ranking algorithm, which is possibly nonlinear and very complex. The model of hotels' price-setting decisions is equally 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. Promotional sales practices seem to be widely used, but how hotels choose promotions in addition to prices is difficult to explain with a model. My model also abstracts from the fact that hotels might attract consumers via different channels (for example via their own website, or via the non-default rankings on the online travel agent).

Future research could attempt at solving any of these issues. Moreover, this paper could benefit from a more explicit model of the platform's decision process in how to optimally trade off between providing benefits to the seller as well as the buyer side, and what its incentives are when designing the algorithm. Lastly, policymakers as well as consumer advocates have emphasized the idea of mandating platforms to increase the transparency of their rankings. I leave the investigation of the effects of such a policy change for future research.

References

Arellano, M., & Bond, S. (1991). Some tests of specification for panel data:monte carlo evidence and an application to employment equations. *Review of Economic Studies*, *58*(2), 277–297.

Bar-Isaac, H., & Shelegia, S. (2022). Monetizing Steering. *Working Paper*, 1–25. Retrieved from https://www.dropbox.com/s/d3l0zv4kvhfmoox/Monetizing.pdf?dl=0

Cazaubiel, A., Cure, M., Johansen, B. O., & Vergé, T. (2020). Substitution between online distribution channels: Evidence from the Oslo hotel market. *International Journal of Industrial Organization*, *69*, 102577. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167718719301055

Chen, L., Mislove, A., & Wilson, C. (2016). An Empirical Analysis of Algorithmic Pricing on Amazon Marketplace. *WWW '16: Proceedings of the 25th International Conference on World Wide Web*, 1339–1349.

Chen, Y., & Yao, S. (2016). Sequential Search with Refinement: Model and Application with Click-Stream Data. *Management Science*, *63*(12), 4345–4365. Retrieved from https://doi.org/10.1287/mnsc.2016.2557 doi: 10.1287/mnsc.2016.2557

Cho, S., Lee, G., Rust, J., & Yu, M. (2018). Optimal Dynamic Hotel Pricing. *Working Paper*. Retrieved from https://economics.sas.upenn.edu/index.php/system/files/2018-04/hp_final_update.pdf

De los Santos, B., & Koulayev, S. (2017). Optimizing click-through in online rankings with endogenous search refinement. *Marketing Science*, *36*(4), 542–564.

DellaVigna, S., & Gentzkow, M. (2019). Uniform Pricing in U.S. Retail Chains. *The Quarterly Journal of Economics*, *134*(4), 2011–2084.

Dinerstein, M., Einav, L., Levin, J., & Sundaresan, N. (2018). Consumer Price Search and Platform Design in Internet Commerce. *American Economic Review*, *108*(7), 1820–1859.

Ellison, G., & Ellison, S. F. (2009). Search, Obfuscation, and Price Elasticities on the Internet. *Econometrica*, *77*(2), 427–452.

Ershov, D. (2022). Variety-Based Congestion in Online Markets: Evidence from Mobile Apps. *Forthcoming in American Economic Journal: Microeconomics*. Retrieved from https://drive.google.com/file/d/1behYlPeLP0XVO6591U0oWm3KbwC4_6bU/view

European Commission. (2016). *Report on the monitoring exercise carried out in the online hotel booking sector by EU competition authorities in 2016* (Tech. Rep.). Retrieved from https://ec.europa.eu/competition/ecn/hotel_monitoring_report_en.pdf

Fradkin, A. (2017). *Search, Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb.* Retrieved from http://andreyfradkin.com/assets/

SearchMatchingEfficiency.pdf

Glick, M., Richards, G., Sapozhnikov, M., & Seabright, P. (2014). How Does Ranking Affect User Choice in Online Search? *Review of Industrial Organization*, *45*(2), 99–119.

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

Hannak, A., Sapiezynski, P., Khaki, A. M., Lazer, D., Mislove, A., & Wilson, C. (2017). Measuring Personalization of Web Search. *Proceedings of the 22nd International Conference on World Wide Web*. Retrieved from http://arxiv.org/abs/1706.05011 doi: 10.1145/2488388.2488435

Hunold, M., Kesler, R., & Laitenberger, U. (2020). Rankings of Online Travel Agents, Channel Pricing, and Consumer Protection. *Marketing Science*.

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

Kim, J. B., Albuquerque, P., & Bronnenberg, B. J. (2010). Online Demand Under Limited Consumer Search. *Marketing Science*, *29*(6), 1001–1023.

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

Lee, K. H., & Musolff, L. (2021). Entry Into Two-Sided Markets Shaped By Platform-Guided Search. *Working Paper*. Retrieved from https://lmusolff.github.io/papers/Entry_and_Platform_Guided_Search.pdf

Li, J., Netessine, S., & Koulayev, S. (2017). Price to Compete … with Many: How to Identify Price Competition in High-Dimensional Space. *Management Science*, *64*(9), 4118–4136. Retrieved from https://doi.org/10.1287/mnsc.2017.2820 doi: 10.1287/mnsc.2017.2820

Mantovani, A., Piga, C. A., & Reggiani, C. (2021). Online platform price parity clauses: Evidence from the EU Booking.com case. *European Economic Review*, *131*, 103625. Retrieved from https://www.sciencedirect.com/science/article/pii/S0014292120302555

Schaefer, M., & Tran, K. D. (2020). Airbnb, Hotels, and Localized Competition. *Working Paper*. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3676907

Teng, X. (2022). Self-preferencing, Quality Provision, and Welfare in Mobile Application Markets. *Working Paper*.

Ursu, R. M. (2018). The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions. *Marketing Science*, *37*(4), 530–552. Retrieved from https://doi.org/10.1287/mksc.2017.1072

Weitzman, M. L. (1979). Optimal Search for the Best Alternative. *Econometrica*, *47*(3), 641–654. Retrieved from http://www.jstor.org/stable/1910412

A Expedia's hotel listings page

A.1 Information on the ranking

Expedia gives some guidance to hotels on what affects hotels' visibility online, but the precise workings of the ranking algorithms are proprietary and possibly very complex. As seen in the screen shot in Figure 10, according to Expedia, hotels can improve their "Offer Strength", which in turn affects "Quality Scores" determining how a hotel will be ranked on Expedia as of 2019. Further, the tourism industry website Skift notes that, 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 (see `https://tinyurl.com/54k3k397`, accessed 03/05/2023). Finally, Hannak et al. (2017) find that, as of 2014, Expedia's ranking differs quite substantially from one user to the next, and find that a lot of this is due to A/B testing. Note that compensation is a factor that I control for by taking hotel fixed effects, as long as those do not vary over the 2-month period in which I scraped the data.

A.2 Further note regarding Figure 1

As can 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 flag), 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.

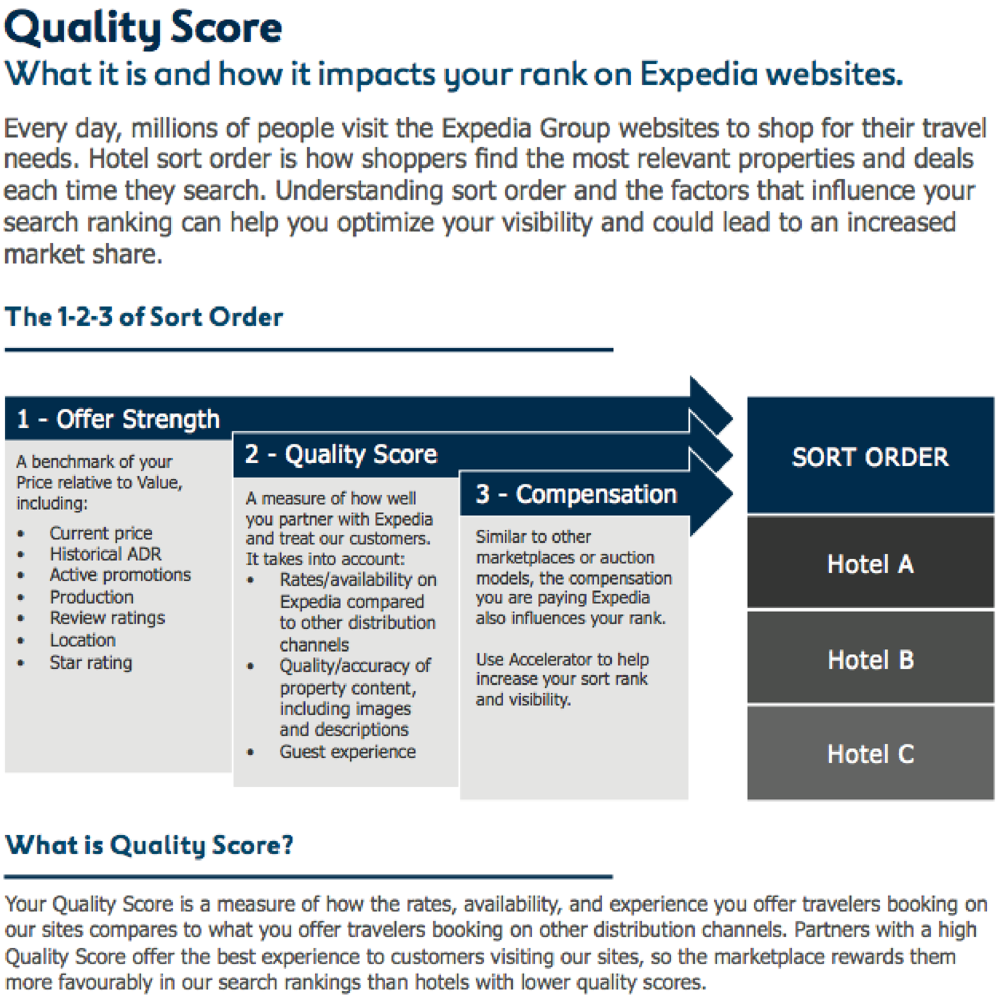

Figure 10: Determinants of a hotel's "Quality Score", which is a crucial input into a hotel's ranking, as of Expedia. Source: `https://discover.expediapartnercentral.com/wp-content/uploads/2016/12/Expedia_Marketplace-White-Paper-April-2016.pdf` (accessed 13/03/2019).

B Version of the model in which hotels decide on sales promotions

Previous literature has found that consumers do not only take into account prices, but also *in addition* value discounts that are offered by the hotel (Y. Chen & Yao, 2016, De los Santos & Koulayev, 2017, Koulayev, 2014, Ursu, 2018). At the same time, however, promotions on Expedia tend to be very opaque (see Appendix D.3). Without a behavioral model of consumer choice, or a dynamic pricing problem in which promotions serve to make last minute changes to prices, it is difficult to understand why a hotel would want to offer a promotion *in addition* to a price. Moreover, if promotions are valuable to consumers, hotels would want to *always* show promotions, and raise prices somewhat so as to account for that.

Despite these conceptual difficulties, in this Section, I provide a version of the model in which hotels not only choose prices, but also have the ability to choose a binary variable that will give a discount on the price shown. 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 bookings for date t, such that the resulting decisions form a Bertrand Nash Equilibrium.

I model the decision to provide a discount as a binary choice, i.e., $d_{jqt} \in \{0, 20\}$: the 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.

A discount has the potential to raise demand, but is costly for the hotel. As in the model in the main text, hotel j's profit is the product of the markup and the demand $Ms_{jqt}(\cdot)$ (market size $\times$ market share). 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

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

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 using with respect to price (using the chain and the product rule) and re-arranging, one obtains:

$$p_{jqt} = \tilde{c}_{jqt} + \tilde{d}_{jqt} - \frac{s_{jqt}(\cdot)}{\underbrace{\frac{\partial s_{jqt}(\cdot)}{\partial p_{jqt}}}_{\text{direct } (<0)} + \underbrace{\frac{\partial s_{jqt}(\cdot)}{\partial r_{jqt}} r'_{jqt}(p_{jqt})}_{\text{indirect } (<0)}} \tag{8}$$

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" $\frac{\partial s_{jqt}(\cdot)}{\partial r_{jqt}} r'_{jqt}(p_{jqt})$ is additionally going to be taken into account by hotels.

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 $r_{jqt}(\cdot)$. 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 set different prices and promotions. This affects hotels' markups and consumer surplus.

C Data Construction

C.1 Preparation of WCAI Dataset

Data used in descriptive analysis. The analysis employs the WCAI dataset on consumer searches and purchases for Paris in order to estimate the demand model. 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.

- I first remove a number of data errors (such as duplicated or empty rows), outliers (in particular, observations of prices below $6; queries for nine or more people; queries for 5 or more kids; a few observations of hotels with prices above 10,000US$), and unusual searches that are likely to be errors (these are: searches with an unreasonable long length of stay or for dates in the past or in the far future; searches in which no prices are recorded at all; or observations of a hotel listing that are the identical except for the number of reviews; or searches in which hotels show up that are not actually located in one of the regions). I also remove a small number of observations of hotels that almost never display a price when they show up, or observations of hotels that display an unresonably unusual price (possibly displayed in a different currency). I end up with 1,361,377 observations of 17,880 distinct user IDs (and, as before, 565 purchases).
- Next, I try to identify all those searches 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, with never any purchase made. This reduces the number of observations to 1,302,231 observations of 17,762 users (and, as before, 565 purchases).

I save the resulting dataset; it serves as a basis for all descriptive analyses carried out in this paper. When taking only searches for hotels in Paris, one obtains 328,926 observations of 4,424 users making 108 purchases.

Data used for demand model estimation. For the purpose of estimating a model, I again use only a subset of this dataset. I employ data from Paris. Further, 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.

As I cannot observe whether a given consumer has done previous searches on or off the platform, I assume that each search corresponds to a new customer arriving on the page. In the end, the final dataset that I use for the estimation of the demand parameters contains 1,051 distinct searches. To be able to estimate Ursu (2018)'s 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".[39]

C.2 Preparation of Web-Scraped Dataset

In the web-scraped dataset, each observation corresponds to a hotel that showed up in a certain query. There is a certain number of hotel observations in which no prices are displayed: this typically occurs for hotels that are fully booked, which are displayed on later pages of the results page. I remove all observations with no prices displayed.

Sponsored and organic listings. Expedia typically displays 55 hotels per results page.[40] Aside from the ranking algorithm that ranks all hotels, hotels on Expedia can pay to appear at a given position in its ranking: in this case, Expedia discloses that the given hotel is "sponsored", as can be seen in Figure 11. A sponsored hotel will in addition to that be displayed in the organic ranking as well. The organic listing will presumably continue to be a function of the hotel's price, whereas the sponsored listing will be determined by the hotel's willingness to pay for the slot. My analysis focuses on Expedia's organic ranking; therefore, I collect information about whether a given hotel appears in a sponsored listing, and subsequently exclude sponsored listings from my analysis. As seen in Figure 12, across all all results pages, sponsored listings are essentially limited to positions 1, 7, 35, 54 and 55, which are almost always occupied by sponsored hotel displays. In contrast, all other positions are almost always occupied by organically ranked hotels.

Table 15 shows the basic dimensions of the entire web-scraped dataset, after removing sponsored listings and listings without pricing information.

Table 15: Basic dimensions of web-scraped dataset, after removing (1) sponsored listings, and (2) listings with no price information. The number of observations for each city differs because the number of hotels per city differs a lot (see last column).

City	Observ	First query (date & time)	Last query (date & time)	# query dates scraped	# travel dates scraped	# hotels scraped
Budapest	319,796	2019-02-01 16:54:00	2019-04-01 15:40:00	784	310	1,045
Cancun	271,254	2019-02-02 14:50:00	2019-03-31 22:44:00	737	306	972
Manhattan	306,660	2019-01-30 18:39:00	2019-04-01 10:42:00	746	306	1,222
Paris	1,202,289	2019-02-01 18:04:00	2019-04-01 12:43:00	745	302	2,842

[39] In contrast to this, Ursu (2018) employs a dataset which only contains observations on click and purchase behavior occurring on the first results page.

[40] 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.

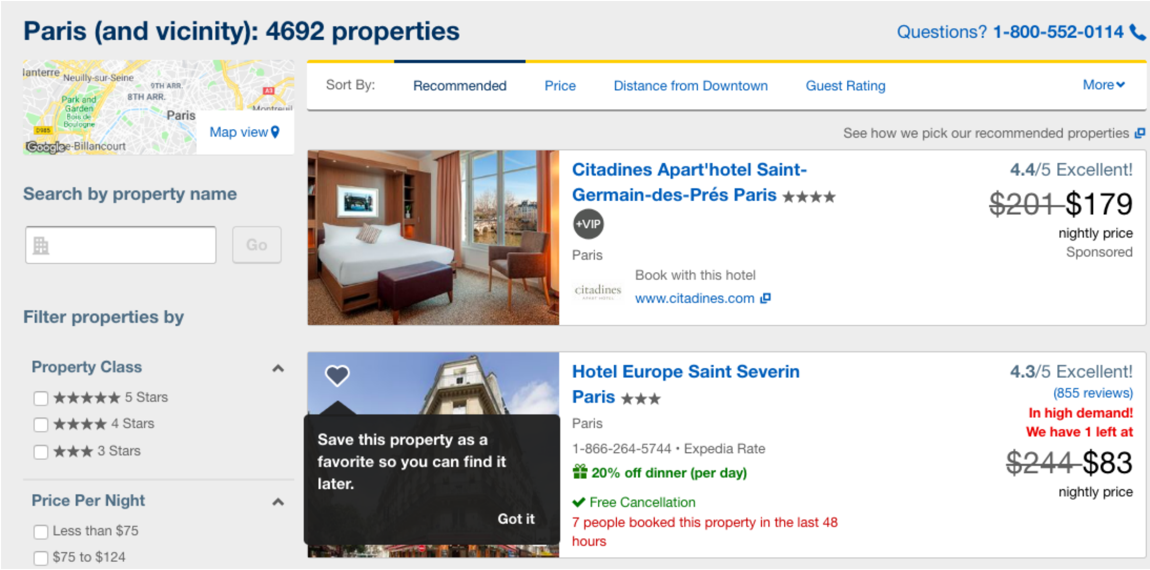

Figure 11: Screen shot taken from Expedia's listings page. It can be seen that the hotel in position 1 is sponsored (written below its price).

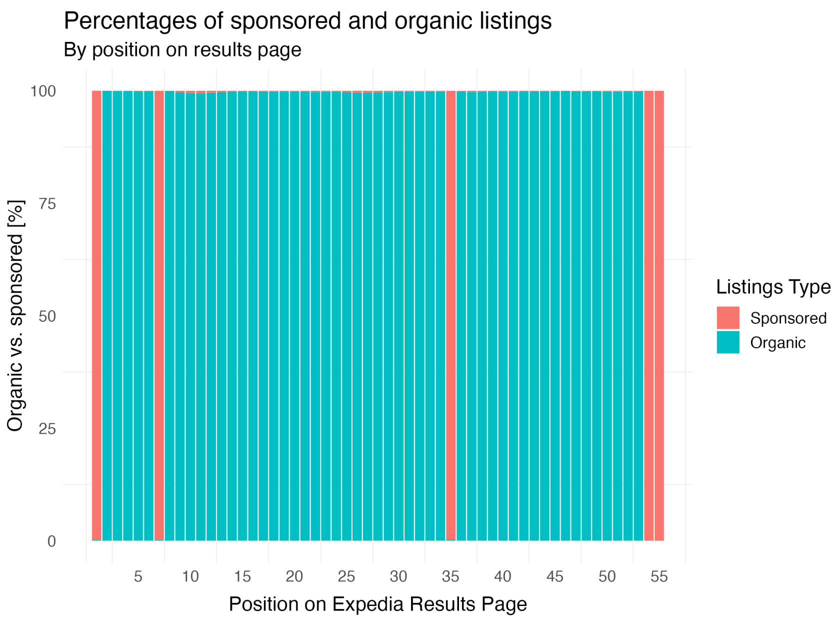

Figure 12: Percentage of "organic", as opposed to "sponsored", listings across observations of all 55 possible positions on Expedia's results pages.

C.3 Comparing hotels observed in the WCAI to hotels observed in the web-scraped dataset

Table 16 shows the market characteristics by city in the web-scraped, as well as in the WCAI dataset. There are notable differences in market structure across cities, but especially also when comparing the web-scraped to the WCAI dataset. The web-scraped and the WCAI dataset may differ for two reasons: (1) the markets have changed from 2009 to 2019; (2) the WCAI dataset is generated by actual consumer queries, and it is thus unlikely that the observations are giving an accurate picture of the full market. Indeed, the numbers of active hotels indicate that the web-scraped data captures a much higher number of hotels: in the case of Budapest, the number of hotels scraped in 2019 is over three times as high, for instance. The WCAI dataset moreover contains a much higher share of branded hotels.

Table 16: Market characteristics for the web-scraped dataset, as well as WCAI dataset.

	City	# Hotels	Avg. stars	Avg. review	Avg. # reviews	Brand (%)
Web-scraped (2019)	Budapest	1,045	3.25	4.11	333.20	10.02
	Cancun	972	2.95	3.92	588.22	7.49
	Manhattan	1,222	3.54	4.18	2,198.83	32.73
	Paris	2,842	3.27	3.93	377.78	25.93
WCAI (2009)	Budapest	276	3.25	4.15	17.44	25.21
	Cancun	110	3.31	4.17	252.91	31.63
	Manhattan	543	2.85	3.96	117.35	42.51
	Paris	1,637	2.56	3.77	15.98	40.91

There are two explanations formally why market characteristics differ across the datasets: (1) The market has changed over time. (2) The WCAI dataset is generated from actual consumer queries, and thus likely does not capture the entire market.

D Further Descriptive Analysis

D.1 Further aspects of users' search behavior, using data for searches in Paris only

The following descriptives provide further evidence of how users search for hotels online, and motivate my choice of search parameters employed when web-scraping. Figure 13 shows that the majority of users looking for accommodation in Paris are asking for stays of one to four nights. Note that as of 2009, this OTA did not necessarily require users to specify travel dates in order to see results in the listings page.

Figure 14 shows a large amount of heterogeneity in the booking window. The uptick at the booking window of 70 to 100 days may be driven by the popularity of travel dates around Christmas and New Year's in Paris (recall that all queries are made in the first 2 weeks of October).

Figure 15 displays the number of adult travelers for whom rooms are searched for, and Figure 16 the number of rooms. Note that as of 2009, 2 adults was the default in this OTA. In 84% of all searches made, the user did not specify the name of the property (in the remaining 16% of queries, the users tend to enter brand names like Best Western or Mercure, or landmarks like Louvre or Eiffel). As seen in Figure 17, in the vast majority of searches, there is no minimum star rating specified (i.e., no filtering of specific hotels takes place).

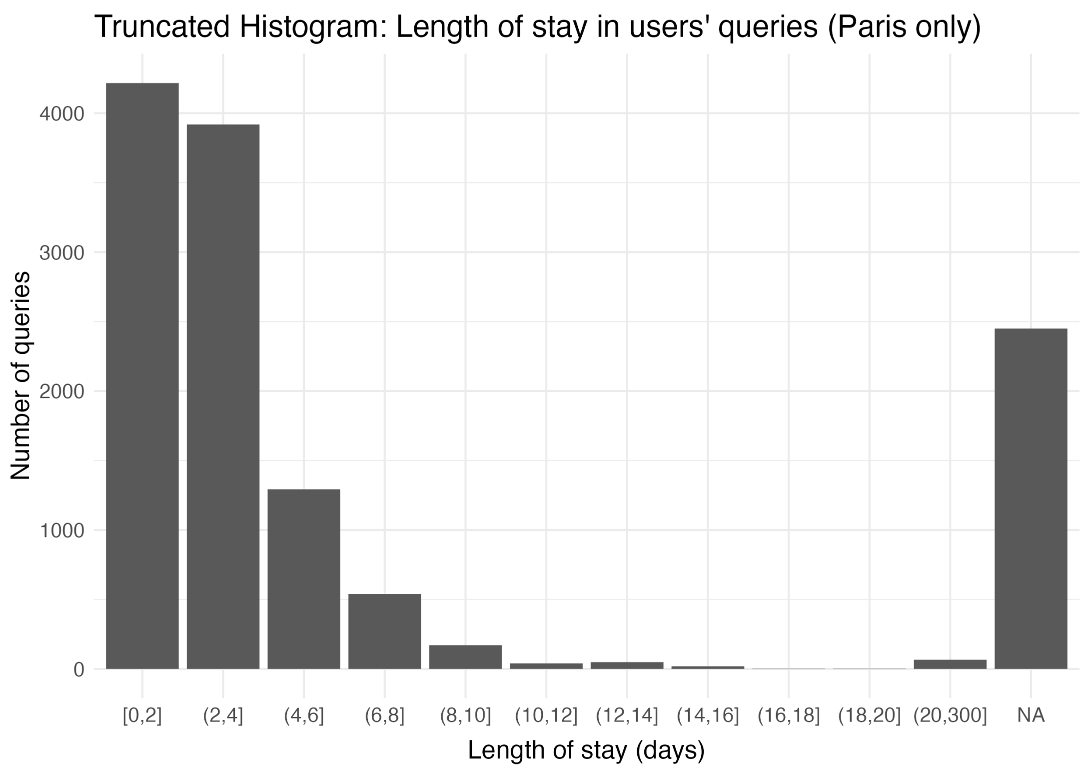

Figure 13: Length of stay asked for across user queries.

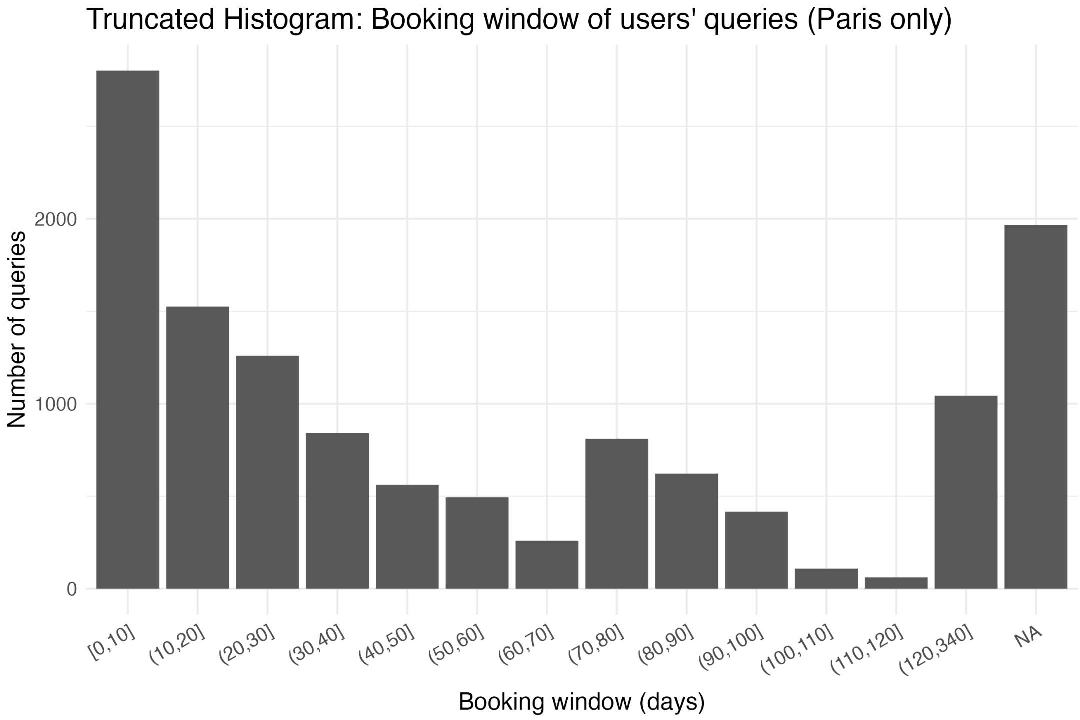

Figure 14: Booking windows across user queries.

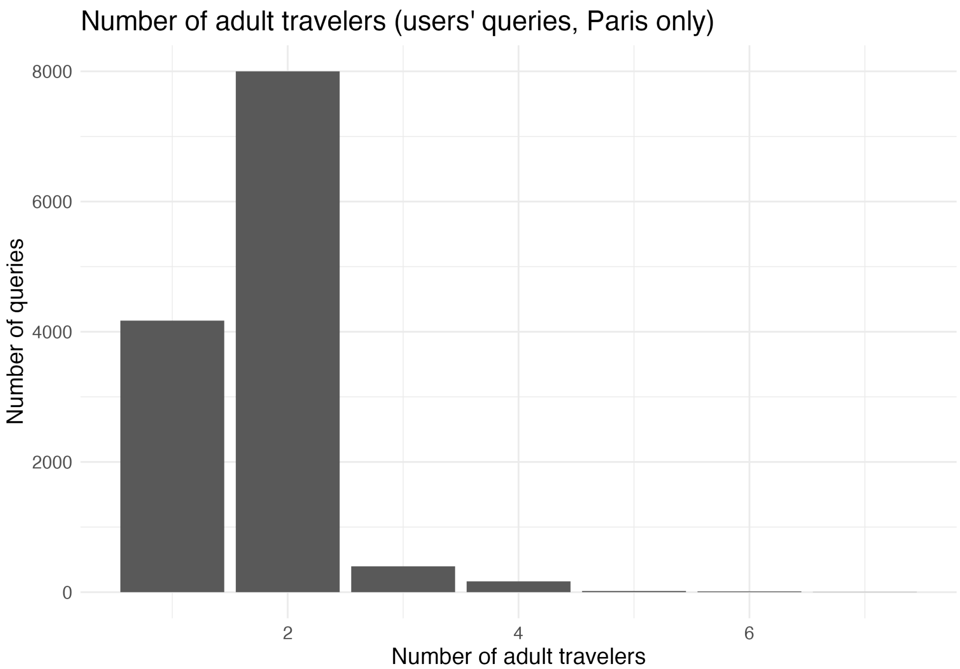

Figure 15: Number of adult travelers. Note that as of 2009, searching for 2 adults was the default option provided by this OTA.

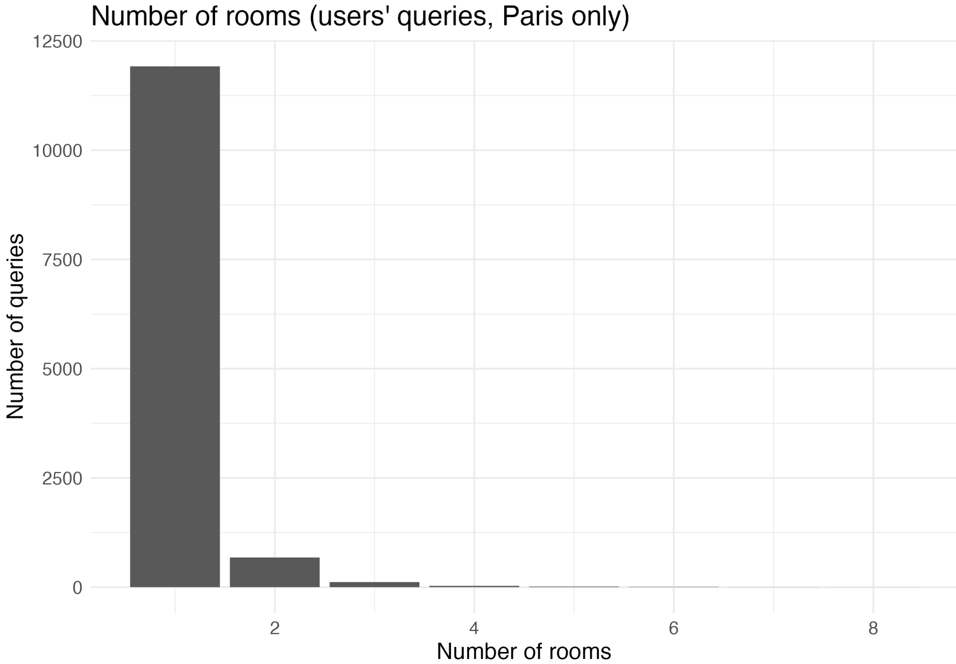

Figure 16: Number of rooms searched for.

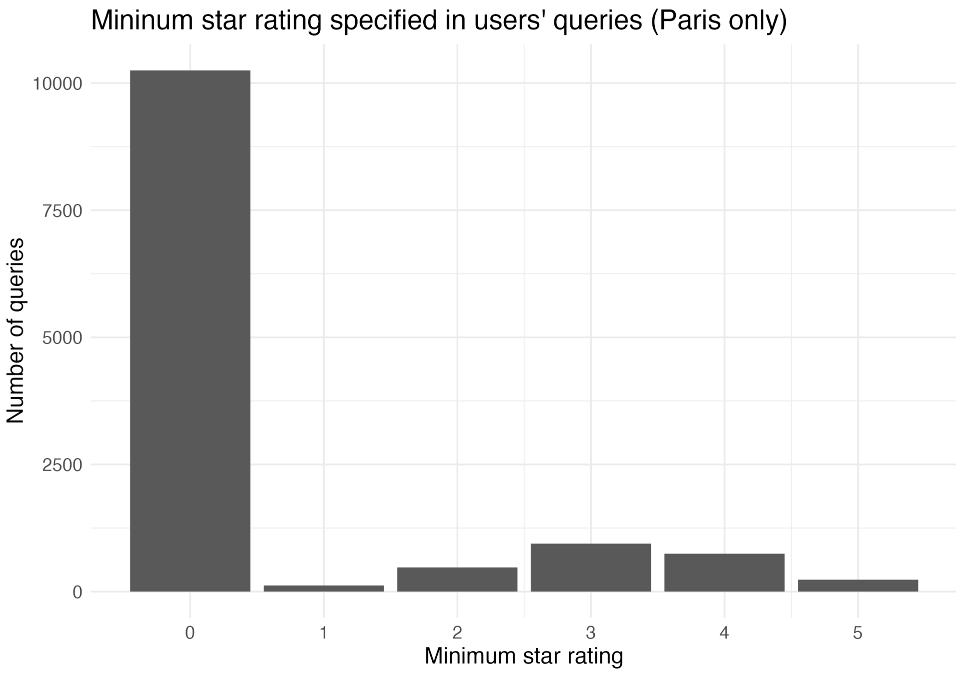

Figure 17: Minimum star rating specified by user.

D.2 Hotels' pricing decisions

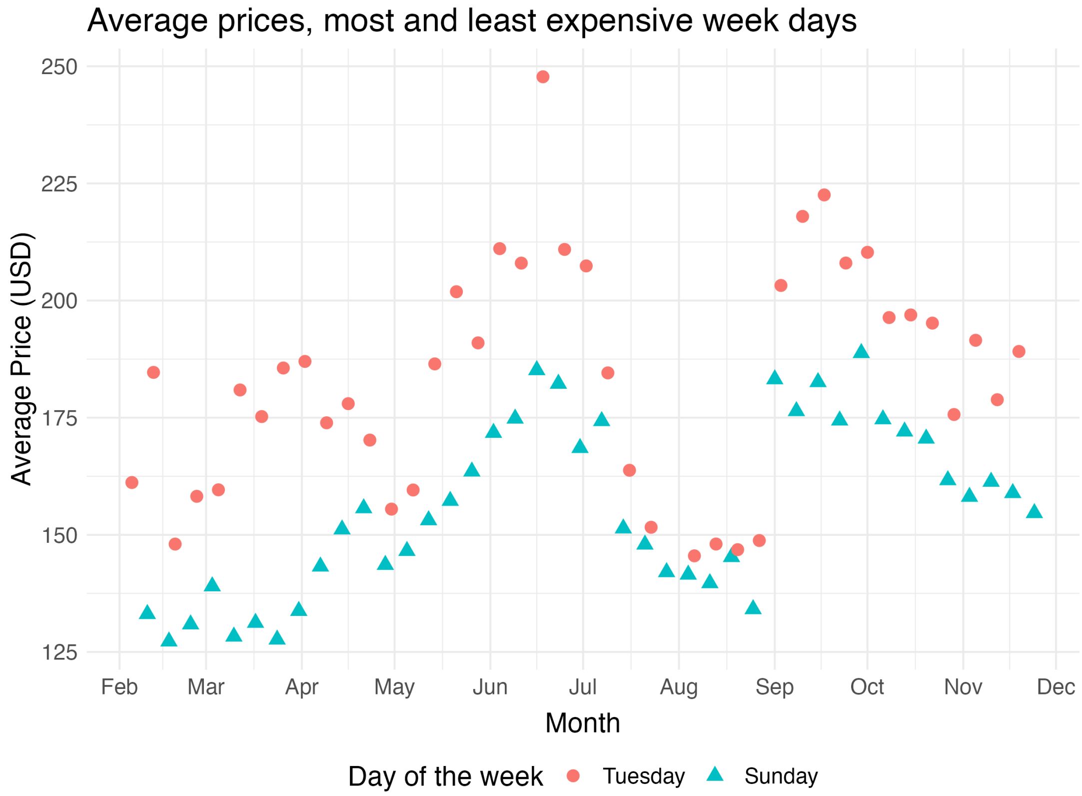

Figure 18: Average prices for most and least expensive days of the week.

D.3 Hotels' promotion decisions

Hotels on Expedia are able to offer various kinds of special promotions, some of which are opaque even after the investigating these. First, hotels can be displayed with a green "Sale!" flag (as seen in Figure 1). When hovering above the flag, a toll tip may *or may not* open up that describes the promotion in place. The promotion in place may for instance be a a discount given on the price, or a special "Expedia rate".[41] Aside from the "Sale!" flag, other flags or highlighted pieces of text can be seen, such as "Tonight only", special membership rates, or free extras. Table 17 displays the fraction of hotels in a given city that ever offer a given type of discount. Overall, 84% of hotels in Paris are observed offering *any* kind of promotion across queries.

Figure 19 considers the average fraction of hotels offering free cancellation. In general, 2-star hotels are the most likely to offer free cancellation, followed by 1-star hotels, although there is considerable variation for these types of hotels. 3-, 4-, and 5-star hotels are the least likely to offer free cancellation. Interestingly, for short booking windows, free cancellation seems to be offered *more* frequently for 4-

[41] Expedia wrote, as of 2019: "Expedia Rate properties may qualify for special promotions and coupon redemption. Expedia Rate requires that your credit card be charged for the full payment upon reservation. Special offers may apply to specific room/unit types and have additional terms and conditions."

Table 17: Percentage of hotels that *ever* offer a given promotion in any given query.

City	# Hotels	% of hotels ever offering given promotion							
		"Sale" tag	Dicount (in %) >0	"Tonight only"	Membership rate	Free cancellation	Free extra	Expedia rate	*Any promo*
Budap	1,045	36.2	35.9	0.8	37.6	49.4	2.1	60.4	65.8
Canc	972	23.8	22.7	0.4	23.0	38.9	8.4	32.6	47.4
Manh	1,222	24.5	23.6	0.3	24.1	48.7	2.2	51.8	58.6
Paris	2,842	48.6	46.6	0.5	26.2	63.4	11.7	78.4	84.1

and 5-star hotels and subsequently declines, but *less* often for 1-, 2-, and 3-star hotels; i.e., the gradient at the beginning of the observation period differs between the different groups of hotels.

Figure 20 shows that the percentage of hotels offering "Expedia rate" is smooth over time, and increasing in the number of stars. (The exception are hotels with 1-star ratings, for which this measure is very noisy due to a low number of observations, and thus not plotted.)

Figure 21 plots the average discounts offered by hotels, conditionally on offering any discount greater than zero. There is essentially no variation in the discount rate for hotels of 2 or more stars, neither over time nor between the groups. In contrast, 1-star hotels seem to behave very heterogeneously, although this finding could again be partly driven by a low number of observations.

All in all, these plots may give us some insight into how hotels engage in dynamic revenue management over booking windows. However, as mentioned in Appendix B, in addition to the opacity, there are several conceptual difficulties regarding hotels' decisions of whether to offer any promotion. I therefore do not endogenize hotels' promotion decisions in my main model.

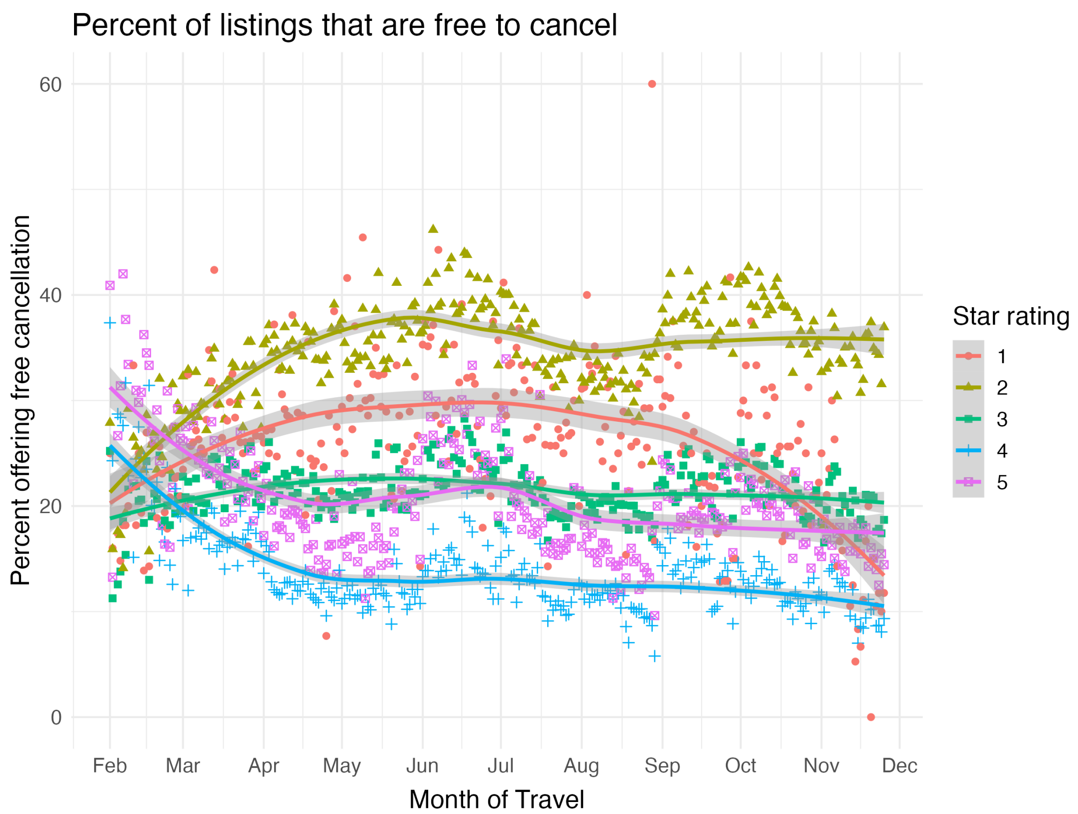

Figure 19: Average percent of listings offering free cancellation at a given travel date for hotels in Paris. Locally estimated scatterplot smoothing is used for the lines. Note that for observations in February and March, the booking window is very short.

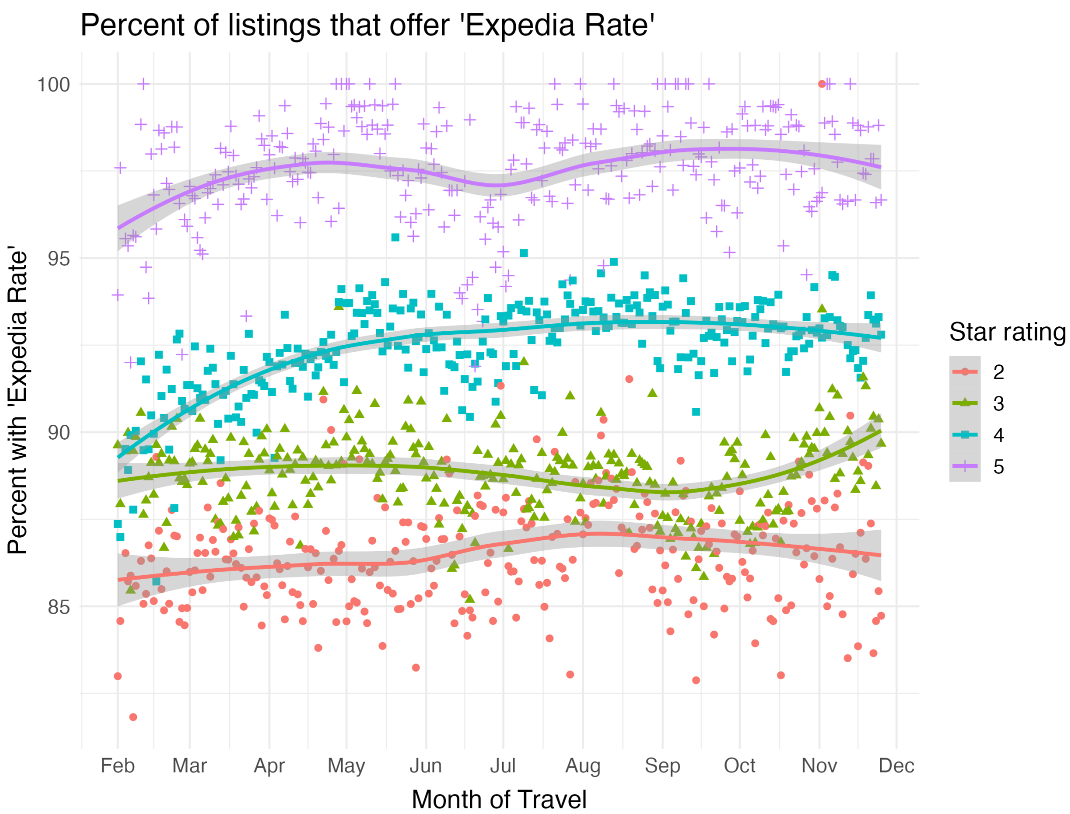

Figure 20: Average percent of listings with "Expedia Rate" at a given travel date for hotels in Paris. Locally estimated scatterplot smoothing is used for the lines. Note that for observations in February and March, the booking window is very short.

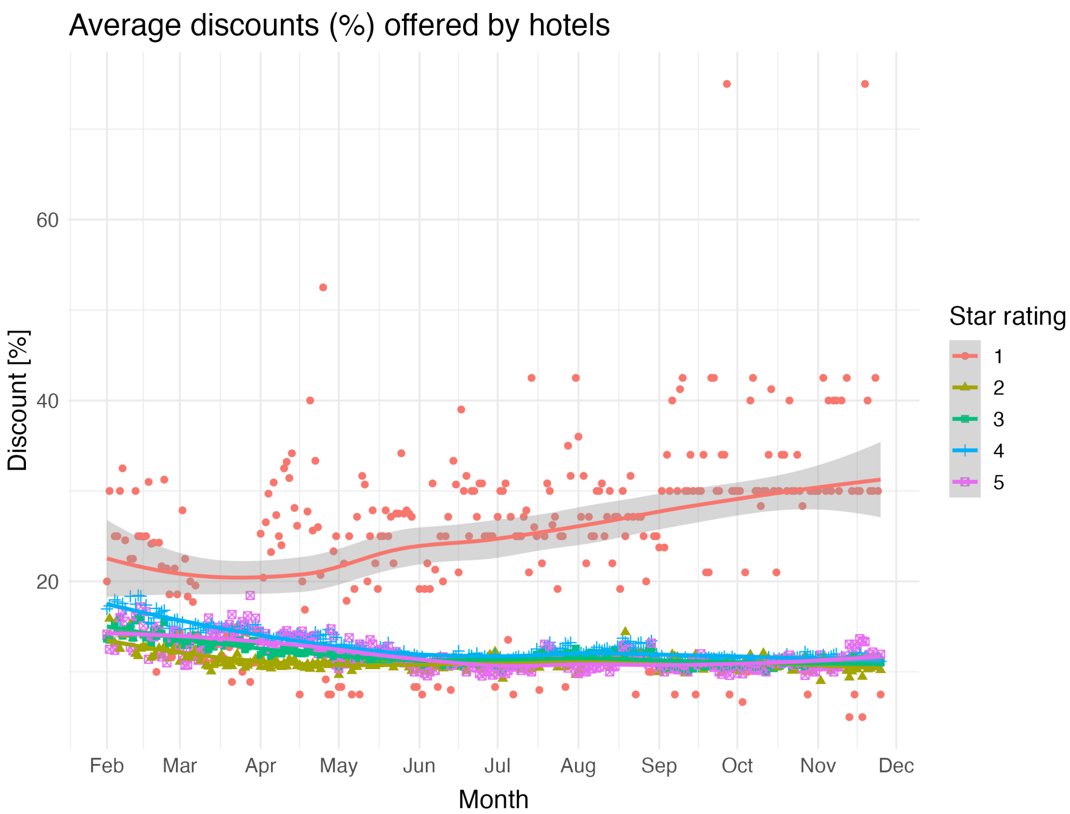

Figure 21: Average discount offered by hotels (in percent) for a given travel date for hotels in Paris. Locally estimated scatterplot smoothing is used for the lines. Note that for observations in February and March, the booking window is very short.

E First Stages

Table 18: First stages, neighborhood-based instruments

	Dependent variable:		
	price		sale
	(1)	(2)	(3)
avg_neighb_price	0.300***	0.301***	−0.0001***
	(0.065)	(0.066)	(0.00001)
avg_neighb_sale		1.945	0.052***
		(1.888)	(0.004)
neighb_count	−0.348***	−0.347***	0.0003***
	(0.054)	(0.054)	(0.0001)
pastbookings_count	−0.082***	−0.082***	0.001***
	(0.023)	(0.023)	(0.0001)
Query×travel-date FE	✓	✓	✓
Hotel FE	✓	✓	✓
F-stat excl. instruments	104.2	77.25	85.58
Observations	914,814	914,814	914,814
Adjusted R^2	0.821	0.821	0.659

Standard errors clustered at query×travel-date level. *p<0.1; **p<0.05; ***p<0.01

Neighboring hotels are defined to be within a 500m radius of focal hotel

Table 19: First stages, neighborhood-based instruments ("donut" definition)

	Dependent variable:		
	price		sale
	(1)	(2)	(3)
avg_neighb_price	0.287***	0.287***	−0.00004**
	(0.078)	(0.078)	(0.00002)
avg_neighb_sale		0.746	0.082***
		(1.387)	(0.005)
neighb_count	−0.165***	−0.165***	0.0003***
	(0.029)	(0.029)	(0.0001)
pastbookings_count	−0.081***	−0.081***	0.001***
	(0.022)	(0.022)	(0.0001)
Query×travel-date FE	✓	✓	✓
Hotel FE	✓	✓	✓
F-stat excl. instruments	54.05	40.52	94.99
Observations	927,322	927,322	927,322
Adjusted R^2	0.819	0.819	0.661

Standard errors clustered at query×travel-date level. *p<0.1; **p<0.05; ***p<0.01

Neighboring hotels are defined to be located within a donut-shaped ring ∈[500m, 1,000m) radius of focal hotel.

Table 20: First stages, Airbnb and 1{august}×zip code instrument

	Dependent variable:		
	mean_price		mean_sale
	(1)	(2)	(3)
airbnb_p_zip	2.050^{***}	2.013^{***}	-0.002^{***}
	(0.086)	(0.082)	(0.0001)
airbnb_avail_zip	-0.015^{**}	-0.013^{*}	0.00002
	(0.007)	(0.007)	(0.00002)
mean_pastbooking	-0.022	-0.023	0.0004^{**}
	(0.041)	(0.041)	(0.0002)
zip code×1{August}		✓	✓
Travel date FE	✓	✓	✓
Hotel FE	✓	✓	✓
F-stat excl. instruments	286.99	117.88	72.91
Observations	447,378	447,378	447,378
Adjusted R^2	0.827	0.827	0.752

Standard errors clustered at travel date level. *p<0.1; **p<0.05; ***p<0.01

Table 21: First stages, brand-based instruments

	Dependent variable:		
	price	mean_price	sale
	(1)	(2)	(3)
avg_price_brand	0.259^{***}	0.258^{***}	-0.00003^{***}
	(0.035)	(0.035)	(0.00001)
avg_sale_brand		-4.808^{***}	0.425^{***}
		(1.413)	(0.009)
pastbookings_count	-0.012	-0.011	0.001^{***}
	(0.041)	(0.041)	(0.0001)
Query×travel-date FE	✓	✓	✓
Hotel FE	✓	✓	✓
F-stat excl. instruments	54.28	77.93	1126.23
Observations	258,011	258,011	258,011
Adjusted R^2	0.661	0.661	0.663

Standard errors clustered at query×travel-date level. *p<0.1; **p<0.05; ***p<0.01

Table 22: First stages, Arellano-Bond type instruments

	Dependent variable:		
	mean_price		mean_sale
	(1)	(2)	(3)
mean_price_diff2	0.134***	0.134***	−0.00001
	(0.029)	(0.030)	(0.00001)
mean_sale_diff2		−0.706	0.096***
		(0.669)	(0.016)
mean_pastbooking	−0.025	−0.025	0.0003*
	(0.034)	(0.034)	(0.0002)
Checkin date FE	✓	✓	✓
Hotel FE	✓	✓	✓
F-stat excl. instruments	20.79	36.28	21.62
Observations	451,366	451,366	451,366
Adjusted R^2	0.967	0.967	0.773

Standard errors clustered at checkin date level. *p<0.1; **p<0.05; ***p<0.01

F Robustness Checks

Table 23: Results using neighborhood instruments, employing more aggregate hotel-travel date panel

	Dependent variable:			
	average position across queries			
	(1)	(2)	(3)	(4)
price	2.069***	3.124***	2.041***	3.191***
(average across queries)	(0.340)	(0.108)	(0.338)	(0.121)
1{sale}			−70.597***	454.136**
(average across queries)			(8.036)	(180.621)
# bookings past 48h	−0.736**	−0.710**	−0.712**	−0.864**
(average across queries)	(0.318)	(0.299)	(0.310)	(0.386)
Travel-date FE	✓	✓	✓	✓
Hotel FE	✓	✓	✓	✓
1st stage F-stat of excl. instruments on price		122.66		134.98
1st stage F-stat of excl. instruments on sale				45.65
Observations	508,956	508,956	508,956	508,956
Adjusted R^2	0.783	0.774	0.784	0.734

Standard errors clustered at travel date level. *p<0.1; **p<0.05; ***p<0.01

The dependent variable is hotel j's average position in Expedia's listings page for travel date t across web-scraped queries. Column (2) uses as instruments for price the number of neighboring hotels available on a given query and travel date, and the average price of neighboring hotels at a given travel and query date. Column (4) uses as instruments for price and the sale indicator the number of neighboring hotels and the average price and sales indicator. Neighboring hotels are defined as being located within 500 meters of the focal hotel.

Table 24: Results using brand-based instruments, employing more aggregate hotel-travel date panel

	Dependent variable:			
	average position across queries			
	(1)	(2)	(3)	(4)
price	1.783***	3.768***	1.758***	3.750***
(average across queries)	(0.312)	(0.176)	(0.310)	(0.177)
1{sale}			−75.236***	89.926**
(average across queries)			(7.705)	(35.480)
# bookings past 48h	−0.585**	−0.553**	−0.566**	−0.577**
(average across queries)	(0.280)	(0.249)	(0.273)	(0.260)
Travel-date FE	✓	✓	✓	✓
Hotel FE	✓	✓	✓	✓
1st stage F-stat of excl. instruments on price		99.98		50.28
1st stage F-stat of excl. instruments on sale				453.18
Observations	562,732	562,732	562,732	562,732
Adjusted R^2	0.798	0.765	0.799	0.764

Standard errors clustered at travel date level. *p<0.1; **p<0.05; ***p<0.01

The dependent variable is hotel j's average position in Expedia's listings page for travel date t across web-scraped queries. Column (2) uses as instruments for price the average price charged by other hotels in Paris of the same brand. Column (4) uses as instruments for price and the sales indicator the average price and average sales indicator employed by other hotels in Paris of the same brand.

Table 25: Linear probability model, using neighborhood-based instruments

	Dependent variable:			
	P(1{position ≤ 10})			
	(1)	(2)	(3)	(4)
price	−0.0001***	−0.00004***	−0.0001***	−0.00004***
	(0.00001)	(0.00001)	(0.00001)	(0.00001)
1{sale}			0.004***	−0.011
			(0.0003)	(0.009)
# bookings past 48h	0.001***	0.001***	0.001***	0.001***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Query-travel date FE	✓	✓	✓	✓
Hotel FE	✓	✓	✓	✓
1st stage F-stat of excl. instruments on price		104.2		77.25
1st stage F-stat of excl. instruments on sale				85.58
Observations	914,814	914,814	914,814	914,814
Adjusted R^2	0.173	0.172	0.173	0.170

Standard errors clustered at query×travel-date level. *p<0.1; **p<0.05; ***p<0.01

Table 26: Linear probability model, using neighborhood-based instruments

	Dependent variable:			
	P(1{position $\leq$ 5)			
	(1)	(2)	(3)	(4)
price	-0.00003^{***}	-0.00002^{***}	-0.00003^{***}	-0.00002^{***}
	(0.00000)	(0.00001)	(0.00000)	(0.00001)
1{sale}			0.002^{***}	-0.003
			(0.0002)	(0.007)
# bookings past 48h	0.0004^{***}	0.0004^{***}	0.0004^{***}	0.0004^{***}
	(0.00004)	(0.00004)	(0.00004)	(0.00004)
Query-travel date FE	✓	✓	✓	✓
Hotel FE	✓	✓	✓	✓
1st stage F-stat of excl. instruments on price		104.2		77.25
1st stage F-stat of excl. instruments on sale				85.58
Observations	914,814	914,814	914,814	914,814
Adjusted R^2	0.142	0.142	0.142	0.141

Standard errors clustered at query$\times$travel-date level. *p<0.1; **p<0.05; ***p<0.01