Consumer Data and Consumer Welfare: Evidence from the Hotel Booking Market.

Andrey Minaev*

This version: April 21, 2021
Please click here for the latest version

Abstract

I study how the information a search intermediary has about consumer preferences impacts the market. Consumers participate in costly search among different sellers' products, relying on the rankings order provided by the intermediary based on their preferences. Better product targeting affects consumer search and purchases, which, in turn, changes the seller pricing incentives. I considered these aspects by modeling both sides of the market under various ranking algorithms used by the intermediary. On the demand side, I develop a model consumer costly search and purchase joint decision. On the supply side, I model the firms' pricing game. To estimate the demand and supply models, I utilized a rich dataset provided by Expedia, which includes consumer search and purchase data and information on the hotels and prices they charge. I find that if the intermediary uses data on consumers' preferences to provide them personalized rankings of products, consumers, on average, experience a 3.6% ($4.9) utility decrease due to increased transaction prices, a 0.8% ($1.1) utility gain due to a reduction in search spending, and 0.5% ($0.7) utility gain due to finding a better-fitted hotel.

JEL classification: D12, D83, L13, L83.

Keywords: Big Data, Consumer Search, Online advertising, E-commerce, Intermediaries, Platforms.

*Department of Economics, University of North Carolina at Chapel Hill. E-mail: andrey@unc.edu.
Acknowledgments: The earlier version of this paper was presented at the 2020 Annual Oligo Workshop and Consumer Search Seminar 2020. I would like to thank my advisors Fei Li and Brian McManus for providing immeasurable guidance and support. I am grateful to Gary Biglaiser, Luca Maini, Jonathan Williams and Andrew Yates as well as participants of the industrial organization seminar at the University of North Carolina at Chapel Hill for providing helpful feedback. I deeply appreciate Alina Malkova for helpful comments and support. Any errors are mine.

1 Introduction

Platforms like Google, Amazon, Facebook, Expedia, etc., collect enormous amounts of data about consumers' preferences and behavior. Although they claim to use these data to provide better services to customers, a discussion has recently been raging about whether we should allow these tech giants to collect and use our personal data? While the central part of this discussion is about privacy and human rights, it also raises economic questions. How does it change competition between firms that advertise on platforms? How does it change market prices? How does consumer welfare change? Does it really help provide the best service to consumers or simply increase tech giants' potential to control markets?

Consumers often are initially uninformed about the quality of the products available on the market. They may conduct a costly search to learn about product qualities, and in many cases, these searches are facilitated by information intermediaries. For example, online platforms as Amazon and Expedia provide consumers ranked lists of products. Nowadays, in the Internet era, consumers conduct much lower search costs and have access to a much wider set of products to choose from. Therefore, consumers as never before are dependent on platforms steering their search for products that provide a ranking of products. Using personal consumer data on preferences helps platforms to provide more accurate rankings to consumers. This paper highlights how market outcomes change if platforms collect and use personal consumer data on preferences.

Due to the presence of the search frictions, consumers explore not all products before making purchase decisions. As a result, the platform's ranking algorithm's change leads to a change in consumer demand function since consumers are more likely to explore products on higher positions in the ranking ceteris paribus. Better ranking helps consumers easier and faster find better-suited products, reducing search expenditures and procuring a better product match. However, if consumers change search behavior in equilibrium, sellers also change their behavior. With a better ranking, consumers find well-suited products higher in the list and have lower incentives to search further, which shrinks their consideration sets and changes the demand elasticity, which, in turn, relaxes competition between sellers and changes their pricing strategies. Thus the effect of better ranking on consumer welfare is ambiguous without additional analysis.

In this paper, I address how the market prices, consumer and economic welfare, and the quality of the purchased products change if the platform can provide consumers better product rankings based on personal consumer preferences. I compare market outcomes in two different cases: in the first case, the platform provides the personalized rankings of products to consumers based on their personal preferences; in the second one, the platform provides the common ranking to all consumers based on the aggregated data of all consumers preferences.

To address these questions, I utilize the dataset provided by Expedia.[1]. It includes consumers' search and purchase data and information on the hotels observed by consumers after filling a search query. I provide the equilibrium model to investigate market outcomes' change under the platform's different ranking mechanisms. To analyze consumer demand, I construct the structural model of optimal consumer choice with the search frictions based on the classical Weitzman (1979) model, where consumers conduct sequential search and on each step, after exploring the hotel, make a decision whether to explore another one and if yes, then which hotel to explore next. Conditional to this demand, I model hotels' pricing game and use it to estimate hotels' costs. Last, using estimation results of demand and supply sides, I run simulations to evaluate the market outcomes under the platform's different ranking mechanisms.

This paper is the first attempt to estimate the equilibrium model in such a setting. Previous empirical works do not model firms' strategic pricing response on the change of platform's ranking mechanism and estimate only the welfare effects due to the change in consumers' search and purchase behavior. Part of the reason for that is computational difficulty in simulating the change in firms' pricing decisions due to the complicated nature of the demand correspondence accounting for search frictions. I overcome this difficulty by applying findings of Choi et al. (2018) and Moraga-González et al. (2018), which allows me to translate the pricing game among the sellers into a familiar discrete-choice problem. The equilibrium model allows me to estimate the change in market prices and get more accurate results. In contrast to previous research, I show that personalized ranking is harmful to consumers despite the decrease in search expenditures.

I find that under the personalized ranking, consumers experience on average .8% ($1.1) utility gain due to a reduction in search intensity compared to the common ranking case since consumers find better-suited products in higher positions. Besides, due to better ranking, consumers on average are able to find better-suited hotels, which increases their

[1]The dataset was originally provided for the Kaggle competition Expedia provided the allowance to use the dataset for academic purposes after the competition was finished.

utility on average by 0.5% ($0.7). On the other hand, consumer utility reduces on average by 3.6% ($4.9) due to increased prices in the case of personalized ranking comparative to the common ranking case. The resulting effect is summarized as an average loss of 2.3% ($3.1). Simultaneously, less price-sensitive consumers might experience more than 11% ($15) utility gain, and more price-sensitive consumers lose more than 15% ($20) of utility.

This study results might argue in the discussion of policy implementation regarding collecting and using personal consumer data. In contrast to previous research, my results show that personal data usage is harmful on average for consumers. Although they might help provide better service to consumers, the market power shifts toward the supply side disproportionately, increasing market prices by higher amounts than consumers' gain. Simultaneously, consumer personal data usage raises economic welfare by reducing search expenditures and helping consumers find better-suited products. Hence, to forbid platforms from collecting and using personal consumer data might not be optimal because it would reduce economic welfare. Direct money transfers to consumers for the data they share with companies might be a better solution.

1.1 Contribution to the Literature

Consumers often have to search among different products before deciding which one to purchase. The search behavior might be influenced by the way the products are presented to consumers. If one of the products is more prominent than others, consumers might find it optimal to start the search from this product. For example, Meredith and Salant (2013) and Ho and Imai (2006) find that candidate's vote share increases if the they are listed first in the ballot.

This paper adds to the literature studying the effect of rankings on consumer search and purchase decisions. Several recent papers estimate consumers' demand parameters and search costs using the demand model based on the classical Weitzman (1979) sequential search model. Consumer search was firstly empirically analyzed by Kim et al. (2010). Additionally, Honka and Chintagunta (2017), Chen and Yao (2017) and Ursu (2018) extended their analysis to model search and purchase joint decisions. Later, Kim et al. (2017) discusses the method of computational burden decrease by providing semi-closed-form expressions for the probability of choice in Weitzman (1979) search model, applying a probit model of sequential search.

I contribute to this branch of the literature in two directions. First, I provide the approach to translate consumer joint search and purchase decision to a standard discrete choice model, using findings of Choi et al. (2018) and Moraga-González et al. (2018), which dramatically lowers the computational complexity of estimation by providing closed-form choice probabilities. Second, my paper is the first attempt to model the market's supply side in such settings to the best of my knowledge. I explicitly model the pricing game among sellers and analyze the price change under different rankings. My results show that consumer-specific rankings are harmful to consumer surplus, in contrast to all aforementioned papers.

The online sponsored-search studies are another branch of literature that discusses how the ranking of alternatives affects consumer search and purchase behavior (e.g. Ghose and Yang (2009), Athey and Ellison (2011), Agarwal et al. (2011), Ghose et al. (2014), Jeziorski and Segal (2015)). These studies have found that advertisements in lower positions of the paid rankings consistently get lower click-through rates. This literature branch is concentrated on the analysis of consumer click and purchase behavior and does not consider seller pricing. This literature might also benefit from my study's findings showing that better product targeting might be harmful to consumer utility because it shifts market power toward the supply side and leads to an increased price.

Furthermore, my results add empirical evidence to recently growing literature discussing the effect of information on competition on markets with horizontally differentiated products. Elliott and Galeotti (2019) show that an information designer can suppress competition by segmenting the market. Jones and Tonetti (2019) in contrast show it is socially optimal when consumers, rather than firms, own and trade their data. Other studies (Roy (2000), Iyer et al. (2005) and Gaelotti and Moraga-Gonzalez (2008)) show that information allows firms to target consumers and segment the market, which soften price competition. However, De Corniere (2016) shows that, targeting leads to more intense competition when consumers actively search for products. The literature mentioned above is solely theoretical, and this paper contributes to it providing empirical evidence of information disclosure effect on firms competition.

The rest of the paper is organized as follows: In section 5 I provide the motivating example. Section 6 introduces the empirical demand and supply model used in this study.

The details of the dataset are discussed in section 3. Section 7 provides the results of estimation. In section 8 I provide the main results – market simulations under different data allowance policies. Section 9 is a concluding remark.

2 The Online Travel Agent Industry Background

Here I provide the main details of the online travel agency industry that are relevant to this article. In 2013 (the year relevant to the dataset used in this study), the American online travel agency (OTA) booking market had a revenue of $157 billion, accounting for 80% of the total online booking market. Expedia was the largest OTA on the market and combined with Booking.com, Orbitz and Travelocity accounted for 95% of all OTA bookings.[2]

OTAs provide consumers an ordered list of third-party sellers of hotel rooms. In order of competition with rivals, each OTA tries to ensure a better consumer experience to their customers and puts better-suited products higher in the lists shown to consumers. OTAs rank different hotel rooms according to consumers' preferences based on the room's characteristics such as price, hotel star rating, location, etc. Such a business model makes it impossible to sellers to affect their positions in rankings directly.

This paragraph provides details on the process consumer follows booking a hotel room on Expedia. At first, the consumer fills the query on the Expedia site specifying trip details such as travel dates, the room type, the location of the hotel, the desired room price, etc. Conditional on consumer's query, Expedia provides an ordered list of hotel rooms that match consumer's preferences. Consumer observes this ordered list and might click on any room to explore additional information by navigating to a sub-page of the chosen room. After that consumer might either book this room or come back to the previous page to explore another room or leave the Expedia site without booking.

3 Data

The dataset used in this study was provided by Expedia for the Kaggle contest in 2013. The dataset is organized as the set of search results presented to consumers in response

[2] http://economist.com/news/business/21604598-market-booking-travel-online-rapidly-consolidating-sun-sea-and-surfing

to their queries. Each consumer observes the set of hotel rooms matching his preferences according to the search query specifications. In addition to hotels' quality characteristics, prices, and positions in the ranking of hotels in a set shown to each consumer, the dataset contains consumer purchase and search behavior: there is explicitly observed which hotels consumers clicked on to get extra information and which, if any, they booked.

The advantage of the data, allowing to study consumers' search behavior, is that the dataset includes not only purchases of consumers but also all clicks they make. The disadvantage is that the dataset does not contain info on the additional information consumers observe after clicking the hotel page. Unfortunately, the dataset does not provide unique IDs for consumers, hence, I can not link different queries made by the same consumer. On the Expedia site, consumers can filter the resulting list of hotels or apply the custom ranking according to price, quality, location, etc. However, the dataset contains only search queries ordered according to default Expedia algorithms. One of the main advantages of the dataset is that besides search impressions from the default Expedia algorithm, it contains search impressions where the hotels are randomly sorted, which helps to study the effect of ranking on hotel attractiveness for the consumer.

The data summary statistics at the hotel and the query level are provided in Table 1. The median hotel has three stars and a reviews score of 4 out of 5. On average the hotel room in those hotels costs \$156 per night. Most hotels are chain hotels and only 35% are independent hotels. The desirability of a hotel's location is represented by an Expedia location score ranging between 0 and 7, which primarily captures the distance of the hotel from downtown but also takes into account amenities nearby. The score for an average hotel in the dataset is 3.09. In a query results, a median consumer sees 31 hotel displayed on the page. The median consumer travels with no children and looks for one hotel room for two adults for two days. The dataset contains 231,7181 clicks, where 72,813 clicks are conducted under the Random ranking. Each search query result includes at least one click. There are approximately 7% of search queries results have two or more clicks. This suggests for high consumer search costs. Around 66% of all consumers book a room after the search. The total number of transactions in the data is 135,546 where only 4,891 are conducted under the Random ranking. An average displayed hotel is \$13 (\$22) more expensive than clicked (booked) ones and has a lower review ranking and a lower number of stars.

Table 1: Hotel and Query Summary Statistics

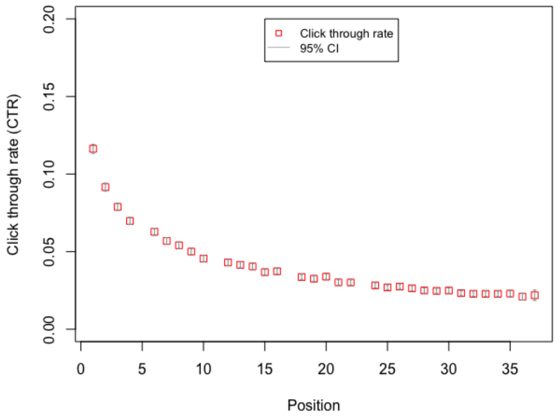
	Observations	Mean	Median	SD	Min	Max
Hotel level						
Price	5,511,851	156.49	129.00	101.28	10	1000
Stars	5,383,647	3.31	3.00	0.88	1	5
Review Score	5,505,786	3.86	4.00	0.91	0	5
Chain	5,511,851	0.65	1.00	0.48	0	1
Location Score	5,511,851	3.09	3.00	1.52	0	7
Promotion	5,511,851	0.24	0.00	0.43	0	1
Query level						
Number of hotels displayed	206,657	27.12	31.00	8.10	5	38
Trip length (days)	206,657	2.42	2.00	1.98	1	40
Booking window (days)	206,657	39.26	18.00	53.89	0	498
Saturday night (percent)	206,657	0.50	1.00	0.50	0	1
Adults	206,657	2.00	2.00	0.90	1	9
Children	206,657	0.39	0.00	0.79	0	9
Rooms	206,657	1.12	1.00	0.44	1	8
Total clicks	206,657	1.12	1.00	0.61	1	25
Two or more clicks (percent)	206,657	0.07	0.00	0.25	0	1
Transaction	206,657	0.66	1.00	0.48	0	1
Random ranking (percent)	206,657	0.31	0.00	0.46	0	1

4 Reduced Form Evidence

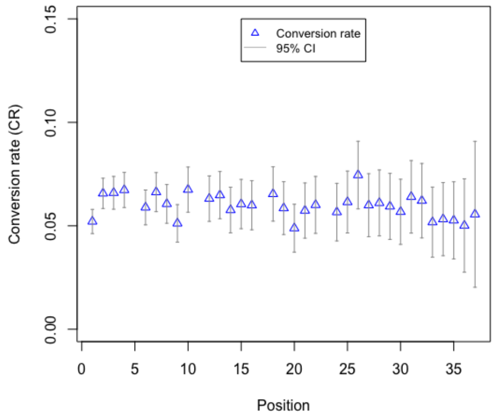
This section presents reduced-form evidence of the hotel's ranking effect on consumer search and purchase behavior. To illustrate the main behavior patterns, I use the part of the dataset where the hotels were shown to consumers in random order without accounting for their fit to consumers' preferences. That allows omitting endogeneity bias in a general Expedia ranking since under a general ranking, Expedia tries to put better hotels on the top of the list. Figure 1a depicts the click-through as a function of the hotel's position in the list, that is, the probability the hotels was clicked and explored (searched) by a consumer conditional on it was shown to him. The data suggest hotels in higher positions are explored more often, which suggests the ranking affects consumers' search behavior. Figure 1b shows there is no significant position of ranking on the conversion rate, e.g., the probability the hotels were booked if explored. As a result, I conclude that the hotel's position on the screen does not change the valuation of the hotel by the consumer. However, the position still affects the unconditional probability of purchase through the probability that the hotel will be included in the consumer's consideration set. So, ranking affects what consumers search, but conditional on search, it does not affect purchases.

Figure 1: Hotels' Click Through and Conversion rates. Randomly sorted queries.

(a) Click Through Rate



(b) Conversion Rate



Note: The click-through rate and the conversion rate (the purchase rate conditional on click) over positions for the case when the lists of hotels, presented to consumers were formed randomly without accounting to the utility provided by hotels.

Click-through and conversion rates under the general Expedia's ranking are provided in Figure 2a and Figure 2b respectively. It shows that under Expedia's puts better-suited to consumer preferences hotels on the top of the list, higher-ranked hotels get more clicks and bookings conditional on a click, increasing the effect of ranking.

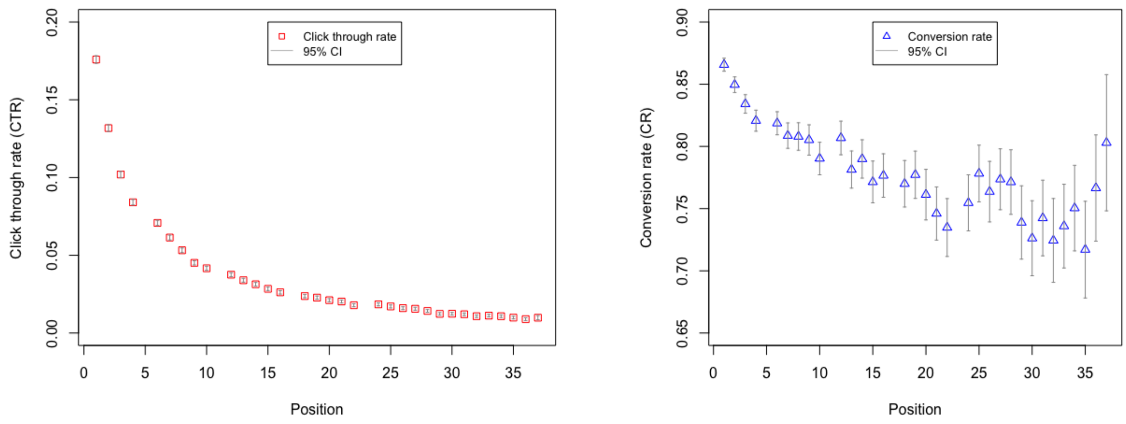
5 Motivating Example

In section 4, I discussed how the change in the hotels' ranking leads to consumer behavior change. If consumers change their search and purchase behavior, hotels also will adjust their pricing strategies accordingly. As an illustration of the logic of the mechanism of how the ranking affects prices, here I discuss a simple theoretical example. The example's main objective is to demonstrate the difference in prices firms charge when the platform can provide the personal ranking to each consumer based on consumer's preferences and when the platform has to provide the common ranking to all consumers based on the aggregate preferences of these consumers.

The economy consist of two firms A and B, selling products a and b respectively, unit mass of consumers and the platform. Each consumer has a unit demand and does

Figure 2: Hotels' Click Through and Conversion rates. Queries sorted by Expedia ranking.

(a) Click Through Rate *(b) Conversion Rate*



Note: The click-through rate and the conversion rate (the purchase rate conditional on click) over positions for the case when the lists of hotels, presented to consumers were formed according to Expedia's algorithm accounting to the utility provided by hotels.

not have any outside option. The platform is the only place where the consumers can purchase the product. Consumers do not observe the entire product matching quality and pay the search cost to explore it. Though, prior to the search, consumers observe the part of the product's matching quality and observe the second part after the search. Consumers can not purchase the product without exploring it and paying the search cost. The platform guides the consumers' search process providing the ranking of products and placing products with higher potential matching qualities on top positions in the ranking. More detail about the platform's role is provided below. Firms compete in prices and set them optimally conditional on consumers' behavior. Firms' objective is to maximize profit. The marginal costs of both products are normalized to zero.

If the consumer i purchases product j, his utility:

$$U_{ij} = u_{ij} - p_j = \delta_{ij} + \epsilon_{ij} - p_j,$$

where δ_{ij} and ϵ_{ij} are parts of utility observed prior and after the search, respectively, and p_j is the price of product j. ϵ_{ij} is assumed to be a random draw from the exponential distribution with parameter 1 and be uncorrelated among consumers and firms.

Consumers are different in their valuations of products. ϵ_{ij} is iid across consumers and products, though consumers value differently δ_{ij}, the product's part of utility observed prior to search. Two-thirds of consumers (labeled Consumer 1) have preferences $\delta_{ia} = \delta$, and $\delta_{ib} = 0$, while the remaining one-third of consumers (labeled Consumer 2) have preferences $\delta_{ia} = 0$, and $\delta_{ib} = \delta$. Consumers' product values are illustrated in Table 2.

Table 2: Consumers' products values

Products	Consumer 1	Consumer 2
a	$\delta + \epsilon_{ia}$	$0 + \epsilon_{ia}$
b	$0 + \epsilon_{ib}$	$\delta + \epsilon_{ib}$

As mentioned above, the platform guides consumer's search process by providing the ranking of products and placing on higher positions products with higher potential matching qualities. Due to ϵ_{ij} are i.i.d among consumers and products, the platform attempts to place on the higher position the product with higher δ_{ij}. This exercise aims to compare market outcomes in two scenarios: first, the platform can provide the personal ranking of products to each given consumer, and second, the platform has to provide the same ranking to all consumers. In the first scenario, the platform will place the product a in a higher position for two-thirds of consumers (Consumer 1) and product b for the remaining one-third of consumers (Consumer 2). In the second scenario, the best the platform can do is place product a higher for all consumers. The rankings under two scenarios are represented in Table 3.

Table 3: Positions of products under common and personal rankings

Position	Common ranking		Personal ranking	
	Consumer 1	Consumer 2	Consumer 1	Consumer 2
1	a	a	a	b
2	b	b	b	a

In accordance with the literature, I let the search cost differ over positions. Consumers pay zero cost to explore ϵ_i of the product placed in the first position, while consumer i have to pay search cost s_i to explore ϵ_i of the product placed in the second position. s_i is assumed to be a random draw from the standard uniform distribution $U[0, 1]$ and be uncorrelated among consumers.

Choi et al. (2018) shows that as a result of optimal search and purchase decisions, rational consumer purchases the product with the highest $w_{ij} - p_j$, where w_{ij} is defined in Equation 1.

$$w_{ij} = \min\{u_{ij}, r_{ij}\}, \tag{1}$$

where r_{ij} is the reservation utility of product j for consumer i, i.e. such utility level that the consumer i is indifferent between obtaining utility r_{ij} immediately and visiting seller j. The mathematical definition of reservation utility r_{ij} is provided as a solution of Equation 2 in r_{ij}.

$$s_{ij} = \int_{r_{ij}}^{\infty} (u - r_{ij}) dF(u) = \int_{r_{ij}-\delta_{ij}}^{\infty} (\epsilon - r_{ij}) dF(\epsilon) \tag{2}$$

Due to the assumption that $\epsilon \sim Exp(1)$, Equation 2 can be solved in closed-form and the reservation utility can be decomposed into a utility observed prior to search component and a search cost component:

$$r_{ij} = \delta_{ij} + log\left(\frac{1}{s_{ij}}\right) \tag{3}$$

As a result, the Equation 1 can be rewritten as

$$w_{ij} = \delta_{ij} + \min\left\{\epsilon_{ij}, log\left(\frac{1}{s_{ij}}\right)\right\}, \tag{4}$$

Due to ϵ_{ij} are i.i.d. over consumers and products and s_{ij} depends only on the position of the product in the ranking but not the identity of the product itself, the distribution of the second additive part in the equation above depends only on the position of the product in the ranking. If the product j is listed on the first position, then $s_{ij} = 0$, and hence $\min\{\epsilon_{ij}, log\left(\frac{1}{s_{ij}}\right)\}$ follows an exponential distribution with parameter 1. If the product j is listed on the second position, then $s_{ij} \sim U[0, 1]$, which makes $log\left(\frac{1}{s_{ij}}\right)$ follow the exponential distribution with parameter 1, and hence $\min\left\{\epsilon_{ij}, log\left(\frac{1}{s_{ij}}\right)\right\}$ follows an exponential distribution with parameter 2. The distribution of w's is summarized in Table 4.

As shown in Choi et al. (2018), each consumer purchases the product with a higher realization of w. As a result, the demands of firm A and firm B can be expressed as

Table 4: The distribution of w's under different rankings.

Position	Common ranking		Personal ranking	
	Consumer 1	Consumer 2	Consumer 1	Consumer 2
1	a:$w_{1a} - \delta \sim Exp(1)$	a: $w_{2a} \sim Exp(1)$	a:$w_{1a} - \delta \sim Exp(1)$	b:$w_{2b} - \delta \sim Exp(1)$
2	b: $w_{1b} \sim Exp(2)$	b:$w_{2b} - \delta \sim Exp(2)$	b: $w_{1b} \sim Exp(2)$	a: $w_{2a} \sim Exp(2)$

shown in Equation 5.

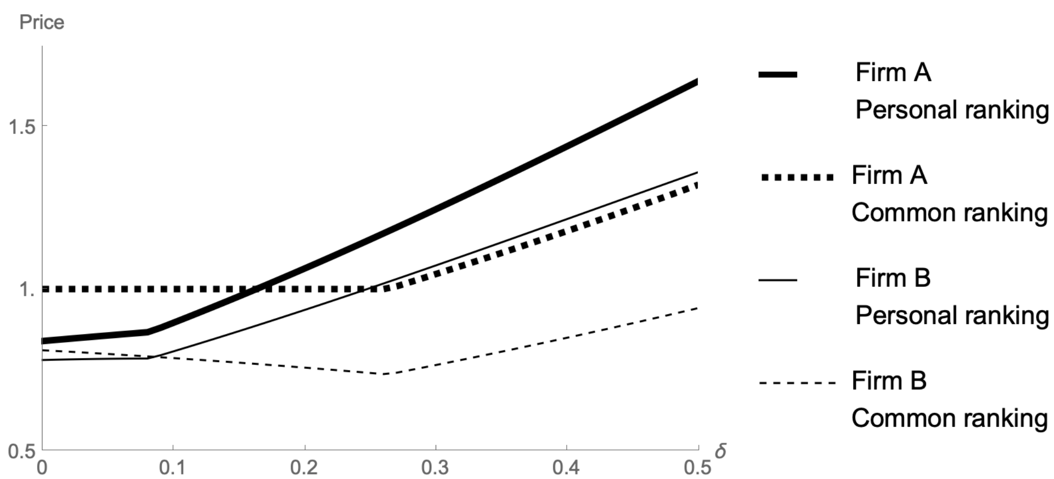
$$
\begin{aligned}
&D_A(p_A, p_B) = \frac{2}{3}Pr(w_{1a} + \delta - p_A > w_{1b} - p_B) + \frac{1}{3}Pr(w_{2a} - p_A > w_{2b} + \delta - p_B) \\
&D_B(p_A, p_B) = 1 - D_A(p_A, p_B)
\end{aligned} \tag{5}
$$

Note that firms have different demand functions under the common and personal rankings due to w_{2a}, w_{2b} have different distributions under the common and personal rankings. Each demand function is the probability that one exponential variable with a given parameter is lower than another exponential variable with another given parameter; hence it can be expressed as a probability distribution function of a random variable that follows the Laplace distribution.

As Quint (2014) showed, due to the distribution of w's is log-concave, there exists a unique equilibrium, which is in pure strategies, in the pricing game among the sellers. Standard FOC conditions determine the price equilibrium. In this setting, the FOC condition is a transcendental equation and can not be solved in the closed form, so I provide numerical solution results on the Figure 3.

As we see, depending on the value δ of the level of products horizontal differentiation, firms might charge higher or lower prices in the case of personalized ranking comparative to the common ranking case. This might be explained by the fact that the transition from the common ranking to the personalized ranking involves two changes in the firm's pricing incentives, summarized by the following two effects. The first effect provides incentives to both firms to increase prices. In the case of the personal ranking, compared to the common ranking case, consumers on average find a well-suited product in the first position, which lowers their incentives to search further. This leads to a decrease in the competition between firms, and as a result, both firms have an incentive to increase prices regardless of their position in the common ranking. The second effect affects firms pricing decisions heterogeneously depending on their ranking position in the common ranking. As Armstrong (2017) shows, when prices are observed prior to the search, they can influence a consumer's search order. Firm A, shown on the first positions under the common ranking, has zero search cost and does not need to keep prices low to attract consumers to explore its product. Under the personal ranking, firm A is shown on the second position for one-third of consumers, which provides incentives to decrease the price. Firm B, shown in the second position under the common ranking, needs to keep its prices low; otherwise, consumers will not explore its product due to search costs. Under the personal ranking firm B is shown to one-third of consumers on the first positions. As a result, it has a lower incentive to keep prices low under the personal ranking. As the level of product horizontal differentiation increases, the advertising effect becomes less important since consumers have stronger preferences toward one of the products. Hence as δ increases, firm A has more incentives to increase the price. For firm B, both effects provide an incentive to increase the price for any level of δ, but for very low δ, firm B in equilibrium decreases price in response to a dramatic decrease in product A price.

Figure 3: Prices as functions of δ.



Note: Firms' prices in the case of personalized ranking and common ranking for the market settings, discussed in section 5.

The example's main point is demonstrated on Figure 3, which highlights that the permutation of product positions in ranking alone is enough to change the market outcomes. Firms charge different prices if the platform is allowed to rank products according to personal consumers' preferences rather than use the common ranking to all consumers. Besides, the difference in price between two ranking mechanisms depends on the level of products' differentiation.

6 Empirical Model

6.1 Modeling of the Platform's Information

This section explains how I model the information about consumers' preferences that the platform uses to rank products under different ranking paradigms: the common ranking, the personal ranking, and the random ranking.

By analogy with the example from the previous section, the platform observes δ's, the part of utility observed by the consumer prior to the search. δ is a convolution of objective product characteristics weighted on consumer's sensitivity to them. More precisely, product j utility that consumer i observes prior to search is

$$\delta_{ij} = \alpha_i p_j + \boldsymbol{\beta_i'} \boldsymbol{x_j},$$

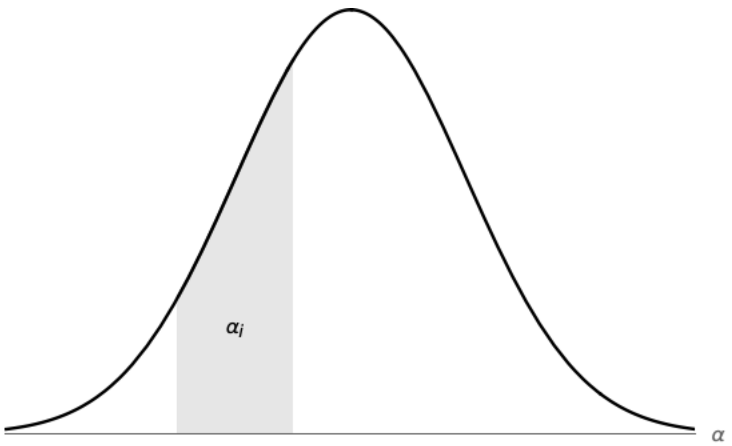
where p_j and $\boldsymbol{x_j}$ are price and the vector of objective product's characteristics observed prior to exploring the product's page. In the case of hotels, $\boldsymbol{x_j}$ might contain such characteristics as hotel star rating, review score, chain identity, location, snd, etc. α_i and $\boldsymbol{\beta_i}$ describe consumer's sensitivity to price and mentioned characteristics.

In general, two different consumers value differently the same objective properties of the product. In the case of hotels, different consumers might, for example, have different favorite hotel chains and have different sensitivity to the price of the hotel room. As a result, different consumers have different αs and $\boldsymbol{\beta}$s, labeled as α_i and $\boldsymbol{\beta_i}$, showing their affiliation to consumer i. The set of α_i's and $\boldsymbol{\beta_i}$'s of all consumers on the market form the distribution with means $\bar{\alpha}$ and $\bar{\boldsymbol{\beta}}$ and variances σ_α and $\boldsymbol{\Sigma_\beta}$.

By saying that the platform knows personal consumer preferences, I assume that the platform knows some information about individual α_is and $\boldsymbol{\beta_i}$s. In the extreme case, the platform knows the actual values of α_i and $\boldsymbol{\beta_i}$ for each given consumer. In a more realistic scenario, illustrated on Figure 4, the platform knows in what part of a distribution bell α_i and $\boldsymbol{\beta_i}$ are positioned. Both scenarios allow estimating δ_{ij} for each given consumer, which is different from the mean among populational δ_j.

If the platform is allowed to use the information about consumer's personal preferences to form rankings, the platform can rank the products to each given consumer i, placing products with higher δ_{ij}'s on higher positions, what, as we saw in the previous section, leads to market prices change. If the platform, on the contrary, is not allowed to use the information on the personal preferences, then it has to use only information on aggregated preferences, $\bar{\alpha}$ and $\bar{\boldsymbol{\beta}}$, which lead to identical ranking to all consumers.

Figure 4: Example of the platform's information



6.2 Demand Side

The response to each consumer's query contains J different hotels (indexed by $j = 0, 1, 2, \ldots J$, where 0 stand for the outside option). The utility consumer i derives from hotel j is given by:

$$u_{ij} = \alpha_i p_j + \boldsymbol{\beta_i'} \boldsymbol{x_j} + \xi_j + \epsilon_{ij}, \tag{6}$$

where the variable p_j stands for the price of hotel j and the vector $(\boldsymbol{x_j}, \xi_j, \epsilon_{ij})$ describes different hotel attributes that consumer values. α_i and $\boldsymbol{\beta_i'}$ denote consumer-specific price coefficient and a vector of tastes parameters. As usual, $\boldsymbol{x_j}$ includes a 1 to allow for a constant term in the utility function. I assume that the consumer observes the hotel attributes contained in $\boldsymbol{x_j}$ without searching. The variable ϵ_{ij} measures the match between consumer i and hotel j and is independently and identically distributed across consumers and hotels. Each ϵ_{ij} is a draw from Gumbel distribution with location and scale parameters 0 and 1 (Type I Extreme Value), as is common in choice models. I assume that ϵ_{ij} captures hotel's characteristics that can be observed only after exploring the hotel page. I assume that the econometrician observes the hotel characteristics $\boldsymbol{x_j}$ but does not observe characteristics ξ_j and matching value ϵ_{ij}. The variables ξ_j are often interpreted as unobserved by econometrician quality, and, since quality is likely to be correlated with the price of a hotel, this will lead to the usual price endogeneity problem, which I treat with the standard control function approach (Train (2009)). The price and the quality characteristics x_0, ξ_0 of the outside option are assumed to be equal to zero.

It is important to note that the consumer's purchase decision and actual consumption happen not at the same time moment. Consumers book a hotel room in advance and visit the hotel after some time. As described in section 3, the median time between booking and staying in the hotel (booking window) is 18 days in the observed dataset. Consumers make decisions on what hotels to book, conditional on prices and availability of hotels presented at the booking date. Unfortunately, the dataset does not contain any sort of consumers' IDs and does not allow tracking consumers' decisions in time, making it impossible to introduce any dynamics in modeling consumers' decisions. If a consumer does not book any hotel after conducting a search, I assume the consumer leaves the market with an outside option and does not return to the platform in the future.

Consumers differ in their value of hotel characteristics. Parameters α_i and $\boldsymbol{\beta_i}$ differ across consumers in order to capture consumer heterogeneity in tastes. These parameters are assumed to follow the multivariate normal distribution, i.e.

$$\begin{bmatrix} \alpha_i \\ \boldsymbol{\beta_i} \end{bmatrix} = N\left(\begin{bmatrix} \alpha \\ \boldsymbol{\beta} \end{bmatrix}, \begin{bmatrix} \sigma_\alpha & 0 \\ 0 & \boldsymbol{\Sigma_\beta} \end{bmatrix} \right), \tag{7}$$

where $\boldsymbol{\Sigma_\beta}$ is a diagonal matrix, i.e., I assume that consumer demand elasticities are independent.

Following the mainstream consumer search literature, I assume consumers do not initially know the exact utility they derive by booking each of the available hotels and incur a search cost to learn them. To be more specific, I assume that before searching a consumer i knows (i) hotel characteristics p_j and $\boldsymbol{x_j}$ for each hotel j, (ii) the distribution $F(\epsilon)$ of match values ϵ_{ij}, including the outside option ϵ_{i0}. Consumer i searches by visiting j hotel's page and learning the value of the matching parameter ϵ_{ij} incurring the search cost associated with this hotel.

Consumers search sequentially with costless recall, i.e., they determine after each visit to a hotel's page whether to book any of the inspected hotels so far, continue searching, or opt-out for the outside option. The outside options' price and characteristics are normalized to zero; hence, the outside option u_{i0} equals ϵ_{i0} and follows Type I Extreme

Value distribution. Let $s_{n_{ij}}$ denote the search cost of consumer i for visiting page of the hotel j, where n_{ij} is the position of the hotel j in the list of hotels shown to the consumer i by the platform. In section 6.2.2 I discuss why the cost of exploring the hotels depends on its position in the rank rather than the hotel's identity. The search cost associated with the outside option is assumed to be zero. As a result, each consumer knows the value of his outside option $u_{i0} = \epsilon_{i0}$ without paying any search cost.

6.2.1 Optimal Consumer Sequential Search

The utility function in Equation 6 can be rewritten as

$$u_{ij} = \delta_{ij} + \epsilon_{ij}, \tag{8}$$

where δ_{ij} is the mean utility consumer i derives from hotel j and ϵ_{ij} is TIEV random shock. As explained above, the consumer knows δ_{ij} but has to search to discover ϵ_{ij}. The match values ϵ_{ij} follow TIEV distribution, which is the same for all consumers and hotels, and is given by $F(\epsilon)$ with pdf $f(\epsilon)$.

Since I allow for consumer-specific taste parameters, the distribution of consumer i's utility u_{ij} from a given hotel j differs across consumers. This leads to the usual aggregation problem I need to deal with. Since the utility shock ϵ_{ij} is an iid draw from TIEV distribution, the utility distribution for hotel j faced by consumer i is

$$F_{ij}(u) = F(u - \delta_{ij}), \tag{9}$$

that is, the distribution of u_{ij} is Gumbel distribution with a location parameter δ_{ij} and scale 1.

Following Weitzman (1979), I define $H_{ij}(r)$, the expected gains to consumer i from exploring the hotel j when the best utility the consumer has found so far is r:

$$H_{ij}(r) = \int_{r}^{\infty} (u - r) dF_{ij}(u) \tag{10}$$

If consumer i's expected gains are higher than the cost $s_{n_{ij}}$ he has to incur to explore the hotel j, it's optimal for him to explore the hotel j. Correspondingly, I define the reservation value r_{ij} as the solution to the equation

$$H_{ij}(r_{ij}) = s_{n_{ij}} \tag{11}$$

Notice that H_{ij} is strictly decreasing so Equation 11 has a unique solution. Therefore H_{ij} is an invertible function.

$$r_{ij} = H_{ij}^{-1}(s_{n_{ij}}) \tag{12}$$

Note that r_{ij} is a scalar and that for each consumer i, there is one such scalar for every hotel j. Moraga-González et al. (2018) shows that the reservation value can be decomposed into a mean utility component and a search cost component:

$$r_{ij} = \delta_{ij} + H_0^{-1}(s_{n_{ij}}), \tag{13}$$

where

$$H_0(r) \stackrel{\text{def}}{=} \int_r^\infty (u - r)dF(u) = \gamma - r + \int_{e^{-r}}^\infty \frac{e^{-t}}{t}dt, \tag{14}$$

where in the last equation, the fact that ϵ_{ij} is TIEV random variable is used. γ here is the Euler constant. The outside option's reservation utility equals positive infinity since the cost of exploring the outside option is normalized to zero.

Weitzman (1979) demonstrates that the optimal search strategy for a consumer i consists of visiting sellers in descending order of reservation values r_{ij} and stopping search as soon as the best option encountered so far (which includes the outside option) gives a higher utility than the reservation value of the next option to be searched. This optimal search strategy can be characterized by the following search rules:

1. **Selection rule.** If a hotel is to be explored, it should be that hotel with the highest reservation utility.

2. **Stopping rule.** Terminate search whenever the maximum utility observed (including the outside option) exceeds the reservation utility of every unsearched option, i.e.

2.1 If the consumer explores a hotel, his reservation utility from that hotel exceeds his utility from all already searched hotels, including outside option.

2.2 The maximum utility among all searched hotels is higher than the utilities of all unsearched ones.

3. Choice rule. Once the search is terminated, the consumer will choose the hotel with the highest utility among those searched, including the outside option.

The rules **2.2** and **3** rely only on the information what hotels consumer explored and which one finally booked, while rules **1** and **2.1** requires the data of the order in which consumers explores alternatives. Expedia's dataset does not include information on the order in which the consumer visits hotels' pages. Jeziorski and Segal (2015) showed that users click ads in a nonsequential order which makes it unreasonable to assume any given order of search (e.g., assume that consumers search in the order of ranking positions). Given that some consumers explore up to 25 hotels, the number of possible search orders for these consumers is $25! \approx 10^{25}$, which makes it computationally impossible to model the search order. To address this challenge, I adapt recent findings from the theoretical search literature by Armstrong (2017) and Choi et al. (2018) and its application by Moraga-González et al. (2018) that make it possible to compute the buying probability of a given alternative without having to go explicitly through the myriad of possible ways in which a consumer may end up considering the alternative in question.

6.2.2 The Effect of Ranking

As discussed in section 3, Expedia's dataset contains impressions where the hotels were sorted randomly, which allows separating the effect of hotels' positioning on the consumer behavior from the effect of hotels' attractiveness. The right panel of Figure 1 shows that the conversion rate does not depend on the position itself, which is an argument that the position the hotel is presented does not affect consumers' utility. The left panel shows that the Click-through rate is decreasing over positions, which is an argument that the hotel's position affects consumer's search behavior.

Given consumer's optimal search strategy, described in section 6.2.1, the effect of the ranking on consumers' choice can be rationalized only in one of the following situations. The ranking affects either consumers' search behavior by affecting reservation utilities r_{ij}

associated with the hotels, or it affects consumers' purchasing behavior through affecting the actual utilities u_{ij} consumers derive from booking the hotels. According to Equation 8 and Equation 13, there are three potential ways how the ranking can affect the reservation or actual utilities – by affecting the utility prior to search (δ_{ij}), the portion of utility realized after the search (ϵ_{ij}), and the search cost ($s_{n_{ij}}$).

Ursu (2018) showed using the dataset discussing in this study, that the rank of a hotel in the list provided to consumer's query has the effect only on the search cost associated with the hotel and does not have any effects on δ_{ij} and ϵ_{ij}. Therefore, the ranking affects the reservation utility and, in turn, the optimal searching and purchasing decisions only through an effect on the displayed hotel's search cost, which is the model used in this paper. Ursu's arguments are mainly based on the observation that the probability the consumer books the hotel, conditional on exploring it, does not depend on the hotel's position, as shown on Figure 1b. She concludes that the hotel's position in the rank does not affect how the consumer values the hotel and only affects the probability the hotel appears in his consideration set.

6.2.3 Probabilities of Purchase

For each consumer i and hotel j define a random variable w_{ij}, *effective utility*, as a minimum of the utility u_{ij} and the reservation utility r_{ij}.

$$w_{ij} \stackrel{\text{def}}{=} \min\{u_{ij}, r_{ij}\} = \delta_{ij} + \min\{\epsilon_{ij}, H_0^{-1}(s_{n_{ij}})\}. \tag{15}$$

Choi et al. (2018) showed that if the consumer conducts a sequential search, he purchases product i with the highest value of w_{ij} among all products. This result's intuition is as follows: If the reservation utility r_{ij} is too low, the product is never even explored by a consumer. If the actual utility u_{ij} is too low, the consumer will not purchase the product even if examined. As a result, consumer decision depends on the minimum of these two.

According to that, a consumer's purchase decision can be described as in the discrete-choice model. However the consumer decision is based on newly introduced *effective utilities* w_{ij}, rather than utilities u_{ij} or reservation utilities r_{ij}. Obviously, w_{ij} is related to utilities u_{ij}. As $s_{n_{ij}}$ approaches to 0, w_{ij} tends to u_{ij} since $H_0^{-1}(s_{n_{ij}})$ converges to ∞). Intuitively, consumers make a fully informed decision if there are no search costs associated with exploring products and gathering the information (i.e., $w_{ij} = u_{ij}\ \forall i$). Hence consumer purchases the best product among all alternatives. If the search cost associated with only the product j becomes relatively high, keeping all other search costs neglectable, making the product j less attractable to explore and hence decreases its chances to be purchased. According to Equation 15, $H_0^{-1}(s_{n_{ij}})$ associated with the product j decreases leading to decrease of w_{ij} and r_{ij}. Hence ϵ_{ij} becomes less important since the consumer is less likely to explore this product at all. If search costs of all products uniformly grow arbitrarily large, then consumers make a purchase decisions based only on values δ_{ij} observed prior to search since consumer either explore the product with the highest δ_{ij} and find it not profitable to incur the search cost to explore the next one, or do not search at all and leave the market with the outside option.

Accordingly, the probability that buyer i books hotel j can be expressed as:

$$P_{ij} = Pr(w_{ij} \geq \max_{k \neq j} w_{ik}) = \int \left(\prod_{k \neq j} F^w_{ik}(x) \right) f^w_{ik}(x) dx \tag{16}$$

The distribution of $w_{ij} = \min\{u_{ij}, r_{ij}\}$ can be obtained by computing the CDF of the minimum of two independent random variables. This means that

$$F^w_{ij}(x) = 1 - (1 - F^r_{ij}(x))(1 - F_{ij}(x)) \tag{17}$$

where F^w_{ij} and F^r_{ij} are the CDF's of w_{ij} and r_{ij}, respectively. Recall that $F_{ij}(x)$ is the CDF of u_{ij}, which has been specified above in Equation 9.

To obtain the reservation values distribution, I use Equation 12.

$$F^r_{ij}(x) = Pr(r_{ij} < x) = Pr(H_{ij}(r_{ij}) > H_{ij}(x)) = Pr(s_{ij} > H_{ij}(x)) = 1 - F^s_{ij}(H_{ij}(x))$$

Substituting this into Equation 17 gives

$$F^w_{ij}(x) = 1 - F^s_{ij}(H_{ij}(x))(1 - F_{ij}(x)) \tag{18}$$

Equation 18 provides a relationship between the search cost distribution and the distribution of the w's. Assuming the right search costs distribution, any needed distribution of w's can be obtained. Moraga-González et al. (2018) shows that if

$$F_{ij}^{s}=\frac{1-exp(-exp(-H_0^{-1}(s)-\mu_{ij}))}{1-exp(-exp(-H_0^{-1}(s)))}, \tag{19}$$

where μ_{ij} is a consumer-hotel specific parameter of the search cost distribution, then CDF of w_{ij} is given by Gumbel distribution:

$$F_{ij}^{w}(x)=exp(-exp(-(x-(\delta_{ij}-\mu_{ij})))) \tag{20}$$

Given Equation 20, P_{ij} in Equation 16 has a closed form:

$$P_{ij}=\frac{exp(\delta_{ij}-\mu_{ij})}{1+\sum_{k\in J\setminus 0} exp(\delta_{ik}-\mu_{ik})}, \tag{21}$$

where 1 in denominator is due to for the outside option $w_{i0}=\min\{u_{i0},r_{i0}\}=u_{i0}=\epsilon_{i0}$, since the search cost for outside option equals zero and hence $r_{i0}=\infty$. As a result the effective utility of the outside option w_{i0} follows TIEV distribution.

Finally, the unconditional choice probability can be obtained from P_{ij} in Equation 21 by integrating out the consumer-specific variables. Denoting by θ_i the vector of all consumer-specific random variables in P_{ij}, the probability that hotel j is booked is the integral

$$P_j=\int P_{ij}dF^{\theta}(\theta_i)=\int \frac{exp(\delta_{ij}-\mu_{ij})}{1+\sum_{k\in J} exp(\delta_{ik}-\mu_{ik})}dF^{\theta}(\theta_i) \tag{22}$$

As discussed in Section 6.2.2, consumer-hotel specific parameter of the search cost distribution μ_{ij} depends not on the identity of the hotel, but its position in the ranking. I model μ_{ij} as $\mu_{ij}=log(1+e^{\gamma\cdot n_{ij}})$, where n_{ij} is the position of the hotel j in ranking shown to the consumer i.

6.3 Supply Side

At the moment t' each hotel j sets the price $p_{jtt'}$ for a given hotel room at a given night t to maximize the expected profit of such sale, conditional on the prices and characteristics of rivals and the opportunity cost $c_{jtt'}$ and the hotel-specific ad-valorem fee f_j charged by the platform. As discussed at subsection 6.2, consumer consumption and purchase decision are spaced in time. At the moment t' consumer books a hotel room to stay in at the moment t. The median booking window in the dataset equals 18 days.

This aspect makes the hotel's pricing decision dynamic. By selling the room today, the hotel loses the opportunity to sell this room tomorrow to another consumer for a potentially different price. While I do not model it explicitly, the hotel's dynamic price decision is captured by the opportunity cost. It is important to note the fundamental difference between the opportunity cost and the marginal cost. Opportunity cost represents the cost of selling the room at the moment the query was submitted, which in addition to the marginal cost for room serving, includes the cost of not having this room available in the future.

The hotel j profit is:

$$\Pi_{jtt'} = \big((1 - f_j)p_{jtt'} - c_{jtt'}\big) D_{jtt'}(p_{jtt'}) \tag{23}$$

The expected demand of hotel j can be expressed as

$$D_{jtt'}(p_{jtt'}) = \int P(buy|\theta)(p_{jtt'}) dF^{\theta}(\theta) \tag{24}$$

where $P(buy|\theta)(p_{jtt'})$ is a probability that consumer with demand parameter θ purchases the product of the firm j. This probability depends on the position of the hotel in the hotel ranking shown to the consumer. Equation 24 can be rewritten as

$$D_{jtt'}(p_{jtt'}) = \int \left(\sum_{positions} P(buy|\theta, position)(p_{jtt'}) \cdot \mathbb{1}(position|\theta)(p_{jtt'}) \right) dF^{\theta}(\theta), \tag{25}$$

where $\mathbb{1}(position|\theta)(p_{jtt'})$ is an indicator function of the hotel j be shown on the position $position$ in $i's$ consumer ranking and can be expressed as:

$$\mathbb{1}(position)(p_{jtt'}) = \begin{cases} 1 & \text{if } \delta_j = \delta^{(position)} \\ 0 & \text{if } \delta_j \neq \delta^{(position)} \end{cases}$$

where $\delta^{(position)}$ is a $position$ order statistic of δs, shown to the consumer i.e. $position$ largest δ among δs of hotels in the query response.

Choi et al. (2018) shows that as a result of optimal search and purchase decisions, rational consumer purchases the product with the highest $w_{ij} - p_j$, where w_{ij} is defined in Equation 1. Hence $P(buy|\theta, position)(p_{jtt'})$ in Equation 31 can be expressed as

$$\begin{aligned}
P(buy|\theta, position)(p_{jtt'}) &= Pr(w_{ij} \geq \max_{k \in J_i} w_{ik} \big| \theta) = \\
&= Pr\left(\delta_{ij} + \min(\epsilon_{ij}, H_0^{-1}(s_{n_j})) \geq \max_{k \in J_i} \left[\delta_{ik} + \min(\epsilon_{ik}, H_0^{-1}(s_{n_k}))\right] \big| \theta\right) = \\
&= \frac{exp(\delta_{ij}(\theta) - \mu_{ij})}{1 + \sum_{k \in J_i} exp(\delta_{ik}(\theta) - \mu_{ik})}
\end{aligned} \tag{26}$$

Since the platform tends to put better-fitted hotels in higher positions, the probability that the hotel j is shown on the given position depends on the utility the consumer i derives from booking this hotel. As a result, if the hotel increases room price, there are two effects on its demand. First, it decreases the hotel's chances to be shown in a high position, and second, for any position, it decreases the probability the hotel is booked, as described in Equation 27.

$$\begin{aligned}
\frac{\partial D_{jt}(p_{jtt'})}{\partial p_{jtt'}} = \int \Bigg(&\sum_{positions} \frac{\partial P(buy|\theta, position)(p_{jtt'})}{\partial p_{jtt'}} \cdot \mathbb{1}(position|\theta)(p_{jtt'}) + \\
&+ \sum_{positions} P(buy|\theta, position)(p_{jtt'}) \cdot \frac{\partial \mathbb{1}(position|\theta)(p_{jtt'})}{\partial p_{jtt'}} \Bigg) dF^{\theta}(\theta)
\end{aligned} \tag{27}$$

Profit maximizing hotel j sets the price according to the following equation:

$$p^*_{jtt'} = \frac{c_{jtt'}}{1 - f_j} - \frac{D_j(p_{jtt'})}{\frac{\partial D_j(p_{jtt'})}{\partial p_{jtt'}}} \quad \forall j \tag{28}$$

It is essential to discuss how the price the hotel charges affects its demand. There are two effects. First, the price affects whether the hotel will be included in the consumer's consideration set. Since the platform wants to put on the higher positions hotels that provide higher utility to consumers, the price increase moves the hotel down the list, increasing the cost of exploring this hotel and reducing the reservation utility. In addition to that, the reservation utility explicitly depends on price through the part of utility observed prior to the search. Moreover, the price also affects purchase probability conditional on the consideration set since it affects the utility level u that the consumer derives booking the hotel's room. To summarize, price changes the hotel's demand by affecting the hotel's probability of appearing in the consumer's consideration set and the probability of being booked conditional on the consideration set.

7 Estimation

7.1 Demand Side

7.1.1 Estimation Strategy

The probability that a random consumer purchase the product of firm j was provided in section 6.2.3 in Equation 22 as $P_j(\theta)$, where $\theta = (\alpha, \sigma_\alpha, \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}, \gamma)$ is a set of population distribution parameters.

Hence, the log-likelihood function can be obtained as:

$$LL(\theta) = \sum_i \sum_j d_{ij} log(P_j(\theta)), \tag{29}$$

where $d_{ij} = 1$ if the consumer i books the hotel j and zero otherwise. There is no closed-form solution for the integral in Equation 22. Hence, I replace $P_j(\theta)$ with the simulated choice probability $\tilde{P}_j(\theta)$. This approach results in the following simulated log-likelihood:

$$SLL = \sum_i \sum_j d_{ij} log(\tilde{P}_j(\theta)). \tag{30}$$

To simulate $\tilde{P}_j(\theta)$ I draw many values of θ, plug them into P_{ij} and average over the resulting logit probabilities. Both the numbers of observations and simulations must go to infinity to guarantee that the maximum simulated likelihood estimate of $\hat{\theta}$ be a consistent estimator of true parameter θ. However, Börsch-Supan and Hajivassiliou (1993) show that for polychotomous choice problems, MSL provides accurate parameter estimates, even with a small number of simulations. I use 10,000 simulations, and that is considered sufficiently more than small for this type of problem.

7.1.2 Identification

In this section I discuss how the model parameters $\theta = (\alpha, \sigma_\alpha, \boldsymbol{\beta}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}, \gamma)$ can be recovered using the variation in consumer behavior observed in the dataset. The model's parameters include the mean and variation of consumers' utility parameters and position effect on the search cost. Consumer tastes for hotel characteristics are identified from consumers' choice conditional on the consideration set. If the consumer clicked on several hotels and booked one of them (or none), this hotel (or outside option) provides a higher utility to the consumer. Different lists of hotels are presented to different consumers which provides variation sufficient for identification consumers' heterogeneous tastes. The consumer's search behavior is also useful for identifying utility parameters because consumers explore only hotels with high enough utility observed before search. Disparities in the search and booking frequencies are used to identify the position effect on the search cost. If the hotel is explored frequently but rarely purchased after exploration, it has low search cost and provides low utility. On the contrary, the hotel, which rarely explored but often purchased after exploration, has high search cost and provides high utility.

7.1.3 Monte Carlo Simulations

This section describes simulation results to show that the estimation strategy described in subsubsection 6.2.3 works well to recover consumers' taste and search cost parameters. For simulation purposes, I generate a dataset of 1,000 consumers, each searching among 30 hotels. Hotel characteristics (Quality and Price) are assumed to be drawn from a multivariate log-normal distribution. Table 5 presents the result of Monte Carlo simulations. The true parameters are given in the first column, and the estimation results are in the second one. Based on the results, we can conclude that provided estimation method is effective in recovering true demand parameters. In the next section, I apply the method to real data provided by Expedia to estimate the utility and search parameters of consumers who participated in the hotel search and booking.

7.1.4 Empirical Results

I apply the estimation strategy, derived in section 6.2.3 to estimate consumer's demand, using the data provided by Expedia. The results of the estimation are provided in Table 6.

The results show that the search cost is significant. It has important implications

Table 5: Monte Carlo Simulation Results

	True values	Estimated values
Utility		
Price	-1	-0.9608^{*}
		(0.0525)
Price heterogeneity	0.3	0.2849^{*}
		(0.0430)
Quality	2	1.8941^{*}
		(0.0883)
Quality heterogeneity	0.6	0.5335^{*}
		(0.0542)
Search cost		
Position effect	0.1	0.0856^{*}
		(0.0065)

Note: Stars indicate estimates significant at the 99% level.

for consumer search behavior. Hotels that appear lower in the ranking of slots have lower chances of being searched. Placing hotels with high expected utility levels in more prominent positions may reduce the cost of each search. Hence, the existence of search cost makes ranking especially beneficial for consumers.

Consumers demonstrate considerable heterogeneity in their hotel attributes' sensitivities, especially in the hotel location and chain affiliation. As a result, using the personalized ranking might have a big impact on consumers' search and utility. Consumers may use alternative refinement methods that prioritize more important attributes. I will further explore the effect of heterogeneity on the market structure in the policy simulation section.

7.2 Supply Side

7.2.1 Estimation Strategy

The point of interest is hotels' opportunity costs $c_{jtt'}$, which vary among hotels and queries, and hotel-specific fees f_j charged by the platform and vary among hotels only. For the simulation purpose, it is not necessary to estimate both the cost and fees, but only the ratio $\frac{c_{jtt'}}{1-f_j}$ because, as described in Equation 28, hotels set prices conditional on this ratio.

Under the existing Expedia algorithm, the hotel's position does not depend on the

Table 6: Estimated Demand Parameters

	Mean	Standard Deviation
Utility		
Constant	4.1167***	6.2731***
	(0.1742)	(1.152)
Price ($100)	−2.4881***	0.6499***
	(0.1203)	(0.1259)
Star rating	1.2369***	0.0166
	(0.0705)	(0.1703)
Review score	0.1118	0.0916
	(0.0895)	(0.171)
Location score	0.0737	0.9509***
	(0.1006)	(0.0726)
Chain dummy	0.7346***	1.6514***
	(0.2656)	(0.4475)
Search cost		
Position	0.0625***	–
	(0.0032)	–

Note: Stars indicate estimates significant at the 99% level.

consumer's characteristics. Hence the Equation 25 can be rewritten as

$$D_{jt}(p_{jtt'}) = \sum_{positions} \left(\left[\int P(buy|\theta, position)(p_{jtt'}) dF^{\theta}(\theta) \right] \cdot \mathbb{1}(position)(p_{jtt'}) \right) \tag{31}$$

Using the consumers demand characteristics estimates from section 7.1.4, Equation 26 can be expressed as a function of hotel's j price p_j. Under the assumption that the hotel knows on what position it will be shown conditionally on price, the demand, described in Equation 31 can be expressed as a function of the hotel's j price.

Finally, the first order condition, provided in Equation 28 can be used to estimate the parameter $\frac{c_{jtt'}}{1-f_j}$. These parameters are used later in section Counterfactual Simulations to run simulations for different data allowance policies.

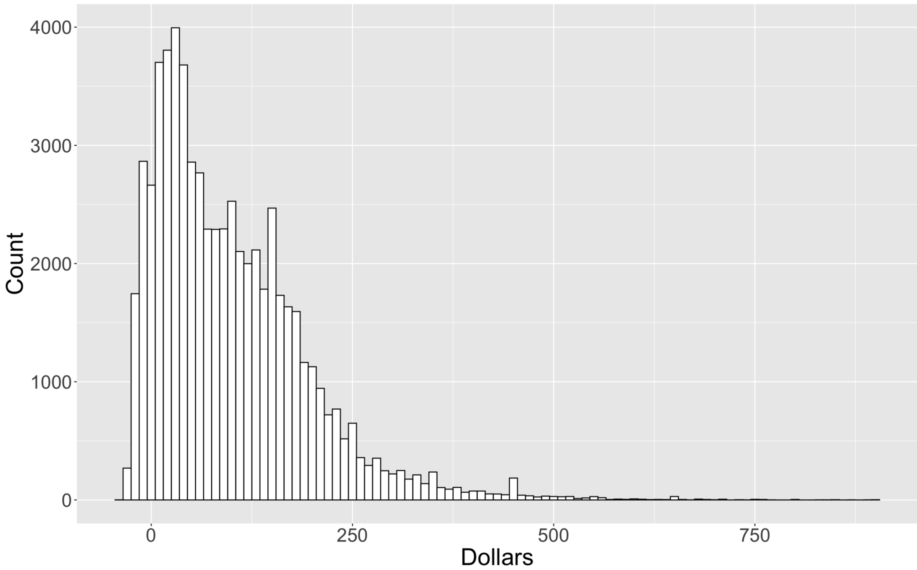
7.2.2 Empirical Results

The estimation strategy described in the previous chapter allows recovering hotels' opportunity costs. The histogram of hotels' opportunity costs is represented on Figure 5. I use this estimation in the next section to get counterfactuals results and estimate the change in hotel pricing under the personal and common rankings.

It is important to note that around 9% of the opportunity cost in the data is negative.

As discussed in Supply Side, the opportunity cost captures the dynamic nature of the hotel's pricing problem and represents the cost of selling the room when the query was submitted. If the hotel expects that in the future, the equilibrium price on the market is going to decrease, for example, because of an increase in competition, the opportunity cost of selling the room right now might be negative. Also, selling the room for a low price, the hotel might expect the consumer to write a positive review, which increases the future hotel's competitiveness and might be considered an investment.

Figure 5: Hotels' Opportunity Costs Histogram $\left(\frac{c_{jtt'}}{1-f_j}\right)$

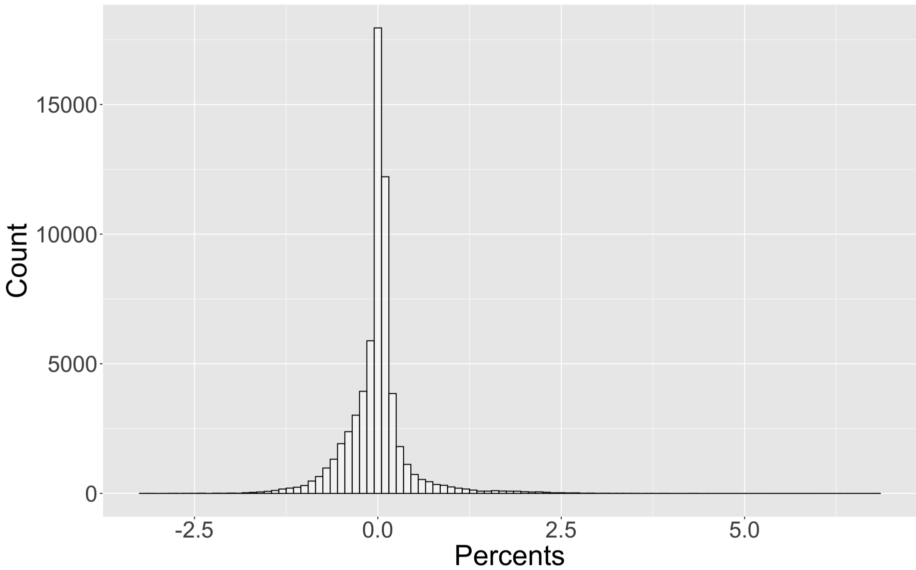


Note: Estimated distribution of hotels' opportunity costs. Opportunity cost is negative for 9% of rooms.

8 Counterfactual Simulations

In this section, I discuss the details of counterfactual simulations. Using the demand and firms' opportunity costs estimations, provided in sections 7.1.4 and 7.2.2 respectively, I simulate firms' pricing decisions under two different data usage policies and compare results. In the first one, I allow the platform to use consumers' personal data to provide the personal ranking to each consumer. In the second one, the platform is allowed to use only aggregated data of all consumers and provide the same ranking to all consumers. In the first case, consumers find better-suited hotels in higher positions, affecting consumers' search behavior and, thus, hotels' demand function. This leads to different optimal prices under different raking mechanisms. Figure 6 shows the histogram of the change in price each firm charges under the personal and common rankings.

Figure 6: Percentage Price Change. Personal vs Common rankings

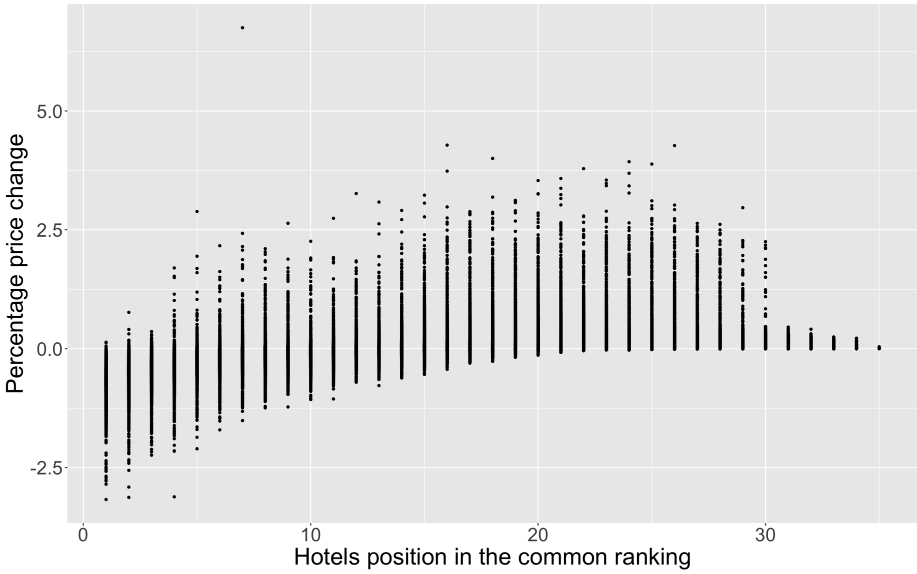


Note: The histogram of the percentage price change with switching from the common ranking to the personalized one.

In the case of the personal ranking, compared to the common ranking case, consumers find better-suited hotels in higher positions, which lowers their incentives to search and decreases the average number of searched hotels. This effect leads to a decrease in the competition between hotels, and as a result, all hotels have an incentive to increase the price regardless of their position in the common ranking. The second effect affects hotels' pricing decisions heterogeneously depending on their ranking position in the common ranking. As (Armstrong, 2017) shows, if prices are observed prior to search they can be used to influence a consumer's search order. The hotels shown on high positions under the common ranking have low search costs and do not need to keep prices low to attract consumers to explore them. Under the personal ranking, these hotels are shown in lower positions for some consumers, which provides incentives to decrease the price. The hotels shown in low positions under the common ranking need to keep their prices low. Otherwise, consumers will not explore them due to their high search costs. Under the personal ranking, these hotels are good-suited for some consumers and shown to them in the high positions. As a result, these hotels have a lower incentive to keep prices low under the personal ranking.

Figure 7 and Figure 8 illustrate the heterogeneity of the sum of two effects over the positions of hotels in the common ranking. Figures show that hotels in higher positions in the common ranking have higher incentives to decrease prices.

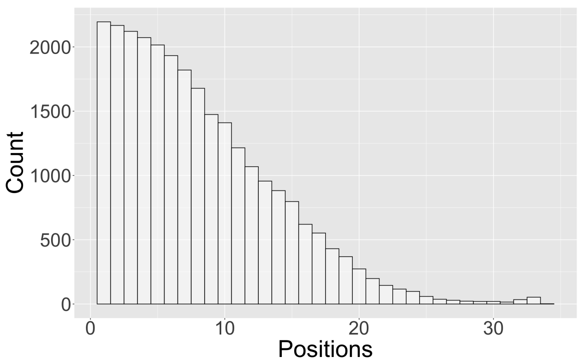
Figure 7: Percentage Price Change by position in the common ranking



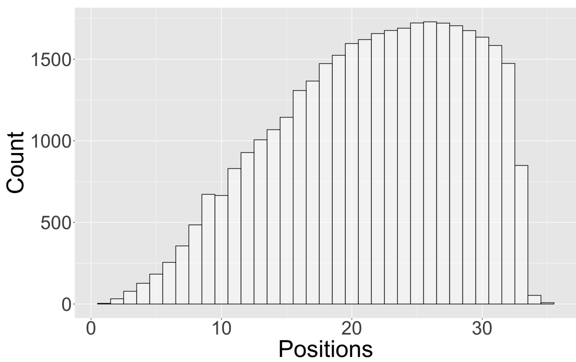
Note: The histogram of the percentage price change by position in the common ranking. Switching from the common ranking to the personalized one.

Figure 8: Positions of the hotels which increase and decrease prices respectively if the platform applies the personal ranking

(a) Positions of hotels which charges lower prices under the personal ranking



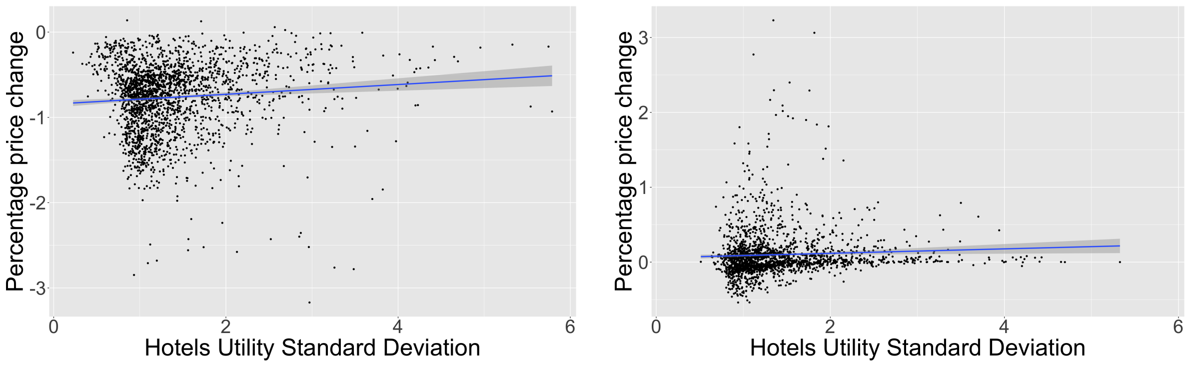
(b) Positions of hotels which charges higher prices under the personal ranking



As discussed previously, all hotels have incentives to charge higher prices under the personal ranking due to consumers find better-fitted hotels in higher positions and explore fewer hotels, which lowers the competition between hotels. This effect increases with the level of hotel horizontal differentiation. Figure 9 shows that if the consumer observes a higher variation of hotel utilities in the query, the first effect has a bigger magnitude and the hotels have higher incentives to increase the price.

Figure 9: Price change. Personal vs Common rankings

(a) Position #1 *(b) Position #15*



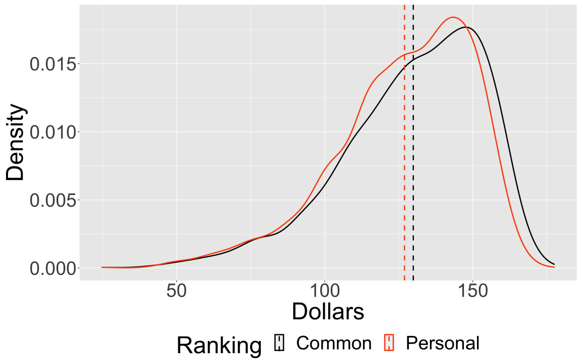
Note: The percentage price change by the measure of the horizontal differentiation of the hotels in query. Switching from the common ranking to the personalized one.

The change of the ranking mechanism has two effects on consumer utility. In addition to the price change discussed above, the consumer finds better-suited hotels in higher positions, which leads to a reduction in search expenditures. The first effect is summarized on Figure 10. On average, due to the price increase, consumers lose \$4, or 3% of their utility if the platform applies the personal ranking, comparative to the common one. More sensitive to price, consumers lose more, and less sensitive ones lose less utility as illustrated on Figure 10b.

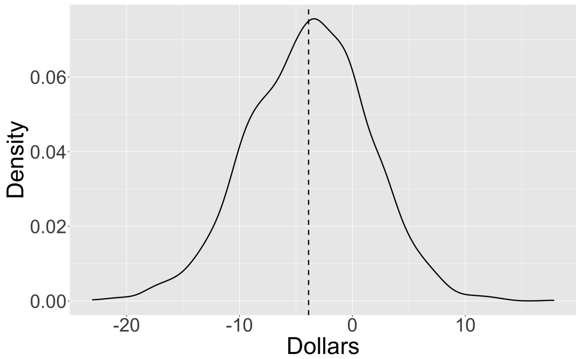
The second effect is represented on Figure 11, which shows that the booked hotels' average position decreases under the personal ranking. On average, consumers save \$1 of search expenditures if the platform applies the personal ranking, compared to the common one.

Figure 10: Consumers' utility. Personal vs Common rankings

(a) Consumers utility distribution. Personal vs Common rankings

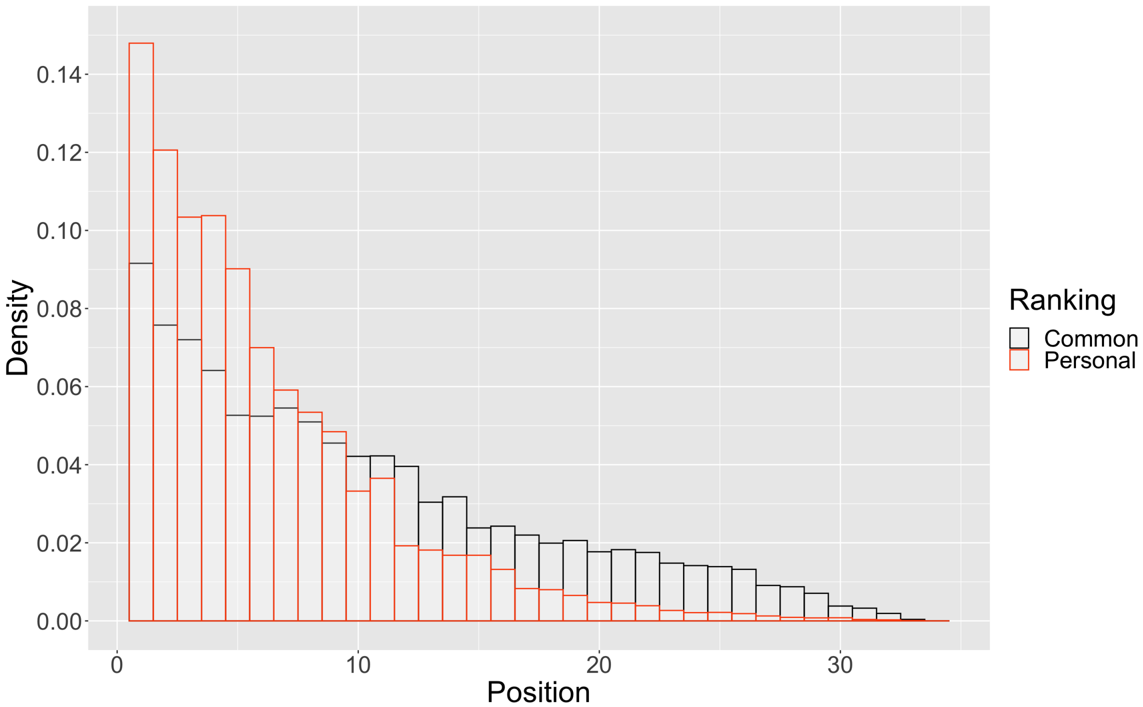


(b) Distribution of utilities difference. Personal vs Common rankings



Note: The left panel provides distributions of consumers' utilities under two rankings. The right panel provides the distribution of the difference in consumers' utility under two rankings.

Figure 11: Positions of booked hotels. Personal vs Common rankings



Note: The figure provides the histogram of positions of the hotels booked by consumers under personalized and common rankings.

9 Concluding Remark

This paper studies the influence of the consumers' personal information, aka big data, on markets. Consumers are often uninformed about the quality of the products available on the market and have to conduct a costly search to learn it.

In many markets, consumers search costly among alternative options before making a purchase. The way to present products to consumers impacts their search and purchase behavior and hence the market outcomes.

This paper contributes to the literature studying the change in firms' competition due to a change in consumer behavior caused by a change in platforms' ranking mechanisms. To discover empirical results, I use a rich dataset, which contains consumers' search and purchase decisions. In contrast to previous research, my results show that personal data usage is harmful on average for consumers. Although data usage might help provide better service to consumers by reducing search expenditures and procuring a better product match, the market power shifts toward the supply side disproportionately, increasing market prices by higher amounts than consumers' gain.

The fact that the platform uses consumer's personal preference data to provide him a better products ranking allows a consumer to spend less effort to find a suitable product and save on average .8% of utility ($1.1) by the reduction of search expenditures and increase utility by .5% ($0.7) by booking a better hotel. However, the reduction of search intensity reduces the competition between firms, providing them incentives to raise prices. As a result, consumers lose 3.6% of utility ($4.9) on average due to the price increase. The resulting effect is negative in contrast to all previous empirical studies, which did not account for transaction price change due to the change of hotels' competition.

Methodologically, this study contributes to the literature by providing a computational method of analyzing firms' pricing game in case of the demand function formed by consumers who search costly among alternatives and form their consideration sets endogenously. To my knowledge, this was computationally impossible before applying in this paper modern theoretical findings.

References

Agarwal, Ashish, Kartik Hosanagar, and Michael D Smith (2011), "Location, location, location: An analysis of profitability of position in online advertising markets." *Journal of marketing research*, 48, 1057–1073.

Armstrong, Mark (2017), "Ordered consumer search." *Journal of the European Economic Association*, 15, 989–1024.

Athey, Susan and Glenn Ellison (2011), "Position auctions with consumer search." *The Quarterly Journal of Economics*, 126, 1213–1270.

Börsch-Supan, Axel and Vassilis A Hajivassiliou (1993), "Smooth unbiased multivariate probability simulators for maximum likelihood estimation of limited dependent variable models." *Journal of econometrics*, 58, 347–368.

Chen, Yuxin and Song Yao (2017), "Sequential search with refinement: Model and application with click-stream data." *Management Science*, 63, 4345–4365.

Choi, Michael, Anovia Yifan Dai, and Kyungmin Kim (2018), "Consumer search and price competition." *Econometrica*, 86, 1257–1281.

De Corniere, Alexandre (2016), "Search advertising." *American Economic Journal: Microeconomics*, 8, 156–88.

Elliott, Matthew and Andrea Galeotti (2019), "Market segmentation through information." *Available at SSRN 3432315.*

Gaelotti, Andrea and Jose Luis Moraga-Gonzalez (2008), "Advertising, segmentation and prices." *International Journal of Industrial Organization*, 26, 1106–1119.

Ghose, Anindya, Panagiotis G Ipeirotis, and Beibei Li (2014), "Examining the impact of ranking on consumer behavior and search engine revenue." *Management Science*, 60, 1632–1654.

Ghose, Anindya and Sha Yang (2009), "An empirical analysis of search engine advertising: Sponsored search in electronic markets." *Management science*, 55, 1605–1622.

Ho, Daniel E and Kosuke Imai (2006), "Randomization inference with natural experiments: An analysis of ballot effects in the 2003 california recall election." *Journal of the American Statistical Association*, 101, 888–900.

Honka, Elisabeth and Pradeep Chintagunta (2017), "Simultaneous or sequential? search strategies in the us auto insurance industry." *Marketing Science*, 36, 21–42.

Iyer, Ganesh, David Soberman, and J Miguel Villas-Boas (2005), "The targeting of advertising." *Marketing Science*, 24, 461–476.

Jeziorski, Przemyslaw and Ilya Segal (2015), "What makes them click: Empirical analysis of consumer demand for search advertising." *American Economic Journal: Microeconomics*, 7, 24–53.

Jones, Charles I and Christopher Tonetti (2019), "Nonrivalry and the economics of data." Technical report, National Bureau of Economic Research.

Kim, Jun B, Paulo Albuquerque, and Bart J Bronnenberg (2010), "Online demand under limited consumer search." *Marketing science*, 29, 1001–1023.

Kim, Jun B, Paulo Albuquerque, and Bart J Bronnenberg (2017), "The probit choice model under sequential search with an application to online retailing." *Management Science*, 63, 3911–3929.

Meredith, Marc and Yuval Salant (2013), "On the causes and consequences of ballot order effects." *Political Behavior*, 35, 175–197.

Moraga-González, José L, Zsolt Sándor, and Matthijs R Wildenbeest (2018), "Consumer search and prices in the automobile market." *Working Paper*.

Quint, Daniel (2014), "Imperfect competition with complements and substitutes." *Journal of Economic Theory*, 152, 266–290.

Roy, Santanu (2000), "Strategic segmentation of a market." *International Journal of Industrial Organization*, 18, 1279–1290.

Train, Kenneth E (2009), *Discrete choice methods with simulation*. Cambridge university press.

Ursu, Raluca M (2018), "The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions." *Marketing Science*, 37, 530–552.

Weitzman, Martin L (1979), "Optimal search for the best alternative." *Econometrica: Journal of the Econometric Society*, 641–654.