Title
Consumer Perceptions of Sponsored Listing and their Impact on Online Marketplaces

Author
Rallabandi, Kalyan Chakravarthy

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Consumer Perceptions of Sponsored Listing and their Impact on Online Marketplaces

A dissertation submitted in partial satisfaction
of the requirements for the degree
Doctor of Philosophy in Management

by

Kalyan Chakravarthy Rallabandi

2022

ABSTRACT OF THE DISSERTATION

Consumer Perceptions of Sponsored Listing and their Impact on Online Marketplaces

by

Kalyan Chakravarthy Rallabandi
Doctor of Philosophy in Management
University of California, Los Angeles, 2022
Professor Brett William Hollenbeck, Chair

In this dissertation I explore consumer perceptions towards sponsored listings, a digital marketing innovation. Chapter 1, shares the background details that provide the motivation for my dissertation.

In the first essay (chapter 2), I note that consumer perceptions of these digital ads are influenced by a signal of quality due to the sponsored status and by a bias against it due to a natural preference for organic ("non-sponsored") listings. In this essay I propose to reconcile these opposing mechanisms and analyze, from a consumer behavior perspective, the evolution of the combined effect of these two mechanisms. In doing so I am able to address how the presence of sponsored listings affect consumers' engagement with the online marketplace. Using reduced form models, I first show evidence to support the presence of these two opposing mechanisms and then show how the net combination of these two mechanisms manifests in the empirical context of online hotel bookings. I then estimate a structural model of sequential search to rationalize the combined effect, while accounting for potential endogeneity bias due to targeted advertising. Using counterfactual analysis I compare how sponsored listings fare against organic listings towards consumer engagement

with the online marketplace. I find that if sponsored listings are replaced with organic listings at the top of the page, then (a) the total clicks are likely to improve by 34% and 30% on laptops and mobiles respectively and, (b) conditional on clicking, the booking rate is likely to improve by 44% and 39% on laptops and mobiles respectively. Additionally, consumer engagement with the online marketplace is at its optimal best with organic listings at the top ranks and sponsored listings at the middle & lower ranks. As search cost increases, I find evidence that sponsored listings play a direct information role and become valuable for all concerned (marketplace, consumer and seller). Finally, I find that consumer welfare is higher under the marketplace assigned ranking with only organic listings (and no sponsored advertising

In the second essay (chapter 3), I investigate the mechanism(s) behind these sponsored advertisements and assess the impact of sponsored listings along the consumer shopping and purchase process. I do this using a structural model of consumer sequential search and a novel data set from an online travel agent. Since online marketplaces return rank ordered lists in response to consumer searches, I assume that based on historical browsing, consumers form priors regarding the quality at each rank. In response to the noisy signal of quality conveyed by sponsored ads, consumers are assumed to update their quality priors in a Bayesian fashion. I find that sponsored listings induced rank (position) change has a major impact on user awareness and that advertisers use sponsored listings for not just lift but also for prominent locations where consumer attention experiences spikes. I also find that advertising disclosure associated with sponsored listings has an impact, so that consumer consideration and choice probabilities are impacted at the top ranks (positions) of a page. This is however not the case at other ranks (positions) further down the page.

The dissertation of Kalyan Chakravarthy Rallabandi is approved.

Stephan Seiler

Dominique Hanssens

Robert Zeithammer

Brett William Hollenbeck, Committee Chair

University of California, Los Angeles

2022

To My Mother & Father, my first teachers;

To My Wife, Karthika, for her patience;

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I am grateful to my family and friends for their constant support. I am indebted to my doctoral advisor Brett Hollenbeck & to rest of my committee members Robert Zeithammer, Dominique Hanssens and Stephan Seiler for their support, guidance and insightful observations. I would also like to thank other UCLA faculty members including Anand Bodapati, Randy Bucklin, Sushil Bikhchandani, Peter Rossi, Sanjay Sood, Suzanne Shu, Stephen Spiller, Hal Hershfield, and Andres Terech, for their kind advice & guidance. Thanks also to Craig Jessen, Kristin Christian, Lydia Heyman, April Barfield, Christian Yirgu, Ameyalli Martinez, Bethany Lucas and Crystal Hwang for their wonderful administrative support. My tenure at UCLA wouldn't have been the same without the company of my colleagues Dan Yavorsky, Prashant Chintapalli, Elicia John, Kate Christensen, Keunwoo Kim, Wayne Taylor, Charlene Chu, Jonathan Lim, Bobby Nyotta, Julia Levine, Sherry He, David Dolifka and Daniel Mirny.

I would like to thank the UCLA Data Science Center support team for providing access to their server and server support. I thank Expedia.com for providing the data and an anonymous data scientist at the data provider firm for making this research collaboration possible. In addition, participants at the 2021 Frontiers in Service Conference, 2022 Winter AMA Conference and seminar participants at University of South Carolina as well as quite a few of the previously mentioned UCLA colleagues and faculty have provided helpful suggestions; thank you all.

VITA

2000–2004	B.Tech. Computer Science & Information Technology, JNTU Hyderabad
2004–2009	Software Engineer, Infosys Technologies Ltd
2009–2011	M.B.A. Marketing & Finance, Indian Institute of Management Lucknow
2011–2013	Assistant Manager, ITC Ltd.
2013–2013	Area Sales Manager, Pepsico Holdings India Ltd.
2013–2014	Graduate Diploma Economics, University of London
2014–2015	M.A. Economics, The University of Texas at Austin
2015–2020	Anderson Fellowship, UCLA Anderson School of Management
2016–2021	Teaching and Research Assistant, UCLA Anderson School of Management

CHAPTER 1

Sponsored Listings and Consumer Behavior

The online two-sided marketplaces, which have become the backbone of modern e-commerce, constantly evolve and create multiple touch points/interfaces with consumers. Digital marketing is ever evolving in response to the persistent innovations in online landscape. Most of the marketing literature in this regard has focused on the marketplace strategy and/or the causal impact of such innovative marketing interventions. And while these questions are relevant from the supply-side revenue perspective of the online marketplace, from an optimization perspective it is also important to account for how consumers are likely react to these interventions and how these new methods assimilate into the digital marketing landscape. These would be interesting not only from an academic viewpoint but also because as the demand side of these platforms, consumer reaction to these interventions can be critical.

The focus of my dissertation is to understand consumer perceptions regarding the digital marketing interventions. I have chosen for my analysis a form of digital advertisement referred to as *sponsored listing*, which caters to the third party seller's need for prominence on a two-sided platform. Thus while sponsored listings generate revenue for the online marketplace,how consumer's perceive them and behave in their presence is unclear.

In the two essays of this dissertation I study consumer perception and behavior towards sponsored listings. The first essay explored the impact of sponsored listings on user engagement with the platform. I compare user perceptions for organic and sponsored listings, This work is likely to be useful to the managers of online marketplaces. In the second essay, I explore the impact of sponsored advertising disclosure on consumers' consideration and choice probabilities. It is likely to be useful from the perspective of the third party vendors (and their advertising strategy) on the online marketplaces.

CHAPTER 2

Consumer Perceptions of Sponsored Listing and their Impact on Online Marketplaces

2.1 Introduction

2.1.1 Overview

Digital advertising revenue is expected to grow to $129.3 billion in 2019. It is estimated that advertising revenues contribute between 6% to 16% of gross revenues across firms in the online retail industry.[1] For example, the world's largest online travel retailer, Expedia, made 10% of its 2018 gross revenue (of $11.2 billion) from online advertising.[2] The stakes have grown considerably as far as the online marketplace advertising is concerned.[3]

Usually there are hundreds of available offerings aggregated by the online marketplaces. In response to consumers' queries, online marketplaces return a rank ordered list of available offerings. The rank ordering is often based on some confidential proprietary algorithm and the ranks assigned by the algorithm are referred to as being organic. Existing literature, academic (Ghose and Yang (2009); Ghose, Goldfarb, and Han (2012); Yang and Ghose

---

[1]https://www.investopedia.com/ask/answers/041015/how-important-advertising-revenue-internet-sector.asp.

[2]https://www.phocuswire.com/Expedia-earnings-full-year-2018

[3]https://www.geekwire.com/2019/report-shows-amazon-taking-digital-advertising-market-share-google-facebook-duopoly/

(2010); Jeziorski and Segal (2015)) and otherwise[4], has documented the importance and the benefits of being at the top of a ranked list. Properties ranked higher experience higher click through rates (CTRs) and conversions, since in the online context consumers start browsing mostly from the top of the rank ordering. The scrolling effort that users have to expend, interpreted as a type of search cost (Ghose, Goldfarb, and Han (2012)), disadvantages the listings with larger ranks.

Online marketplaces monetize the presence of such search costs in the online browsing environment by allowing non-organic or sponsored listings.[5] The firms which sell the products have the option of improving the prominence of their product by choosing to advertise on the online marketplace. The advertised product, referred to as sponsored listing, is shown at a better rank. The rank is better vis-à-vis the organic listing for the same firm. It is fairly intuitive that the value of being prominent (due to being sponsored listing) and search costs are complementary, i.e. higher search cost implies a greater value to being prominent. This is yet to be demonstrated on sponsored listings which are intermingled with organic listings.

To fully understand the impact of the sponsored listings on the online marketplace, it is important to understand consumers' perception of these sponsored listing. As per economic theory, sponsored listings are likely to be perceived positively by consumers. As per informational view (Bagwell, 2007) sponsored listings not only convey direct information about existence of a product but also serve as sources of indirect information by providing signals of the quality of the product. Notwithstanding the evidence provided by economists, the behavioral literature, in stark contrast, has shown, in different contexts, that consumers are less likely to trust sponsored listings.

Thus in this paper I propose to reconcile both these opposing mechanisms, while studying how consumers' perceive sponsored listings? Specifically I document the evolution of these two opposing effects, a positive (signal) effect and negative (push-back for non-organic) effect,

[4]https://blog.advertising.expedia.com/search-marketing-rank-relevance

[5]https://searchsolutions.expedia.com/how-travelads-works/travelads-sponsored-listings-faq

and using a structural model of consumer search study how the net effect, i.e. net of the positive signal effect and the negative non-organic effect, evolves along the rank ordering. I find that (a) consumers have a natural preference for organic listings at the top of the page, (b) the push back against sponsored listings weakens significantly further down the page and the indirect information effect makes sponsored listings valuable at middle and lower ranks, (c) demonstrate the complementary relationship between sponsored listings and search costs by showing that under environments characterized by high search frictions, effectiveness of sponsored listings, make them valuable for the buyer, seller and the online marketplace, and (d) consumer welfare is higher when the marketplace assigned default rank order does not include sponsored listings. I do this using a novel dataset, from a popular online travel intermediary, with detailed data on consumer shopping and purchase in a rank ordered environment and in the presence of sponsored listings.

2.1.2 Sponsored Listings & Mechanisms at Play

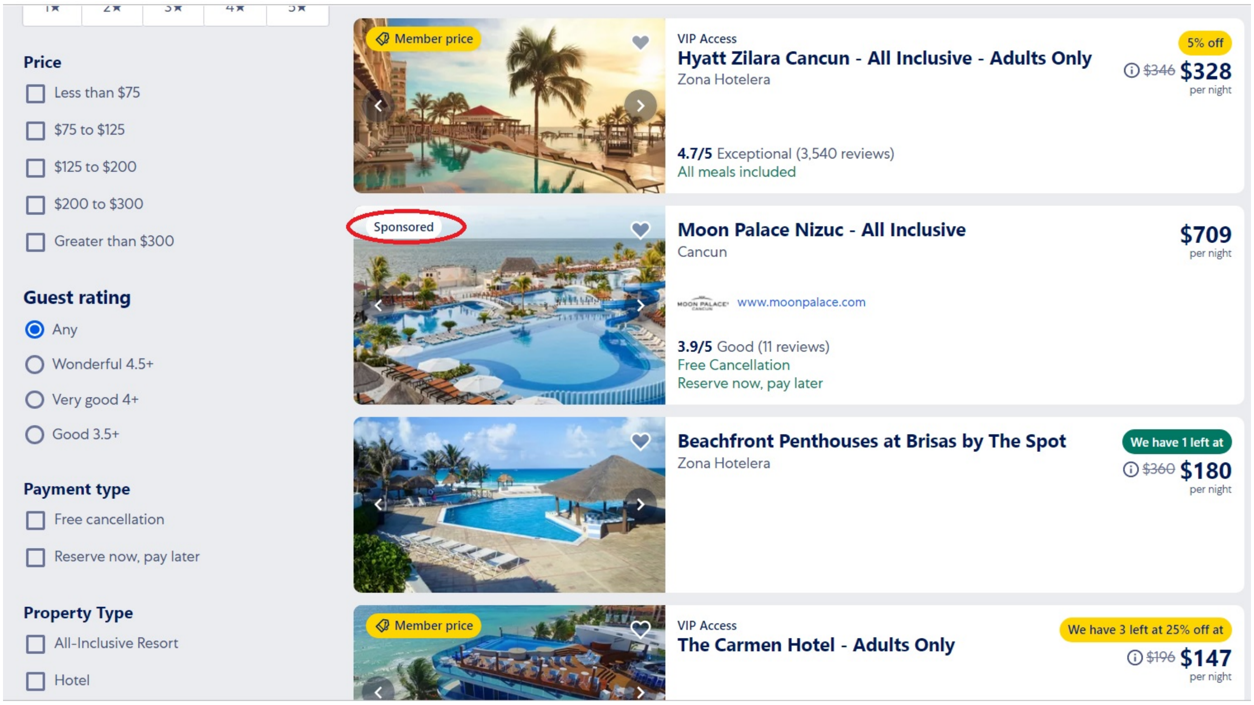
How do sponsored listings work? Sponsored listings are inter-mixed with organic listings as figure 2.1 shows. There are a fixed number of slots per page that are assigned for sponsored listings.[6] The fees for such advertising is usually charged on a CPC (cost per click) basis.[7] When users click on any listing (sponsored or organic), they are directed to the landing page for the product which provides additional product details (usually also includes product reviews). Users usually purchase a product from the landing page. In order to be legally compliant (in U.S.), online marketplaces need to disclose the sponsorship status of a listing (whenever applicable). Thus the sponsorship status of any listing is readily accessible to consumers.

Given the importance of prominence, sponsored listings provide the most value to the

[6] For e.g. Expedia.com displays 50 properties per page and usually shows sponsored listings on ranks 1, 7, 49 and 50

[7] https://advertising.expedia.com/solutions/products/sponsored-listings

Figure 2.1: Sponsored listings on Expedia.com - interspersed with organic listings



advertiser by providing a mechanism to improve visibility. And although sponsored listings provide an additional revenue stream for online marketplaces, any assessment of the impact of such listings on the online marketplaces will be incomplete without a thorough understanding of the consumer search and purchase behavior at these marketplaces, in the presence of sponsored listings. It has been argued that sponsored listings are particularly appealing due to the high degree of situational relevance, while catering to the consumers' search and purchase intent. And that the likelihood of the sponsored listings being perceived as an interruption is lower in the context of consumers' shopping process. In this paper I show that this is not necessarily true and that under certain rank conditions sponsored listings do adversely impact consumer engagement with the online marketplace.

The motivation for this research is based on the existing theories concerning advertising and sponsored content. As per economic theory advertising can convey direct information about the existence and attributes of a product (Anderson and Renault, 2005) and it can also convey indirect information about the product quality (Kihlstrom and Riordan, 1984). In the context of sponsored listings, the direct effect can be thought of as the prominence enhancing effect. As per the theory on indirect information, firms with superior quality (low-cost) seek to advertise to signal their superior quality (efficiency) (Nelson (1970, 1974)). From a supply perspective, it is fairly intuitive to see that this explains the correlation between advertising and quality. Prior literature has empirically shown, from a supply perspective, the existence of this correlation(e.g. Moorthy and Zhao (2000)). Nelson's signaling theory has been extended to the consumer level where it has been shown that even when the advertisement is uninformative and is used purely to drive salience (attention grabbing), ad-spend signals brand quality (Kirmani (1990)). More recently, Sahni and Nair (2020) show that sponsored listings on an online restaurant search portal were perceived by consumers as signals of quality which served to enhance consumer utility. They did this by showing that ad-disclosure effect exists for sponsored listings.

The behavioral literature on consumer trust of sponsored listings has shown that consumers generally have a lower trust of the sponsored listings compared to their organic counterparts (Ma, Liu, and Hossain (2013)). Surveys in the non-academic domain have also documented that consumer trust in sponsored ads is lower than in organic listings.[8] Interestingly, research in this area shows that even while the sponsored and organic listings may be equally relevant, consumers tend to prefer organic listings more than sponsored listings (Jansen (2007)). The literature in this area further documents that consumers are suspicious of sponsored links (Jansen and Resnick (2006)) and rated sponsored links as lower quality (Hotchkiss (2004)).

Setting these two mechanisms side by side, the following can be appreciated: The direct & indirect information that is conveyed by sponsored listings makes them useful to con-

[8]https://www.nielsen.com/us/en/insights/report/2015/global-trust-in-advertising-2015/

sumers and hence sponsored listings should perform better than non-sponsored listings as they enhance consumer utility, i.e. firms which use this option are better off due to the lift in the rank and also due to the perceived quality that is signaled by the sponsorship disclosure. The "non-organic" nature of the listings, however, is bound to inhibit the consumer preference for such sponsored listings i.e. leads to a dis-utility effect. The net of these two mechanisms will manifest itself in the form of consumer engagement with sponsored listings on the online marketplace. The effect due to these mechanisms is likely to vary based on, among other factors, the rank at which the sponsored listing displayed. This is because the position effect, is likely to interact with each of the two effects (i.e. positive utility enhancing effect and the negative dis-utility effect) inherent in sponsored listings. So it would be fair to say that the net impact (net of the two effects) will pan out differently at different ranks. Accordingly, in this paper I seek to provide evidence for these effects at play empirically. I also plan to study how the net impact of consumer perceptions, regarding quality signaling of sponsored listings and the bias against sponsored listings, pans out for sponsored listings as their prominence varies. The aim of this paper is to study the impact of consumer perceptions, regarding a listing's sponsored status, on the engagement with online marketplace. In order to truly optimize the return to the online marketplace due to sponsored listings, it is important to understand the effect of sponsored listings from the consumers' perspective. Understanding the difference in the users' response towards sponsored and organic listings along the rank ordering, will enable the online market platform to better align itself towards the core (retail) business and also optimize the sponsored listing mechanism (redesign ranking and/or reconsider the pricing). The empirical context for this research is the online booking of hotels for vacation stay on Expedia.com.

In what follows, I will first provide reduced form evidence to show that the sponsored status of a listing influences consumers' perceptions of its quality. I then provide reduced form evidence to show that, (a) the positive consumer response to sponsored listings is offset by some sort of negative effect (suspicion towards for non-organic listings) and (b) this net of the positive and negative consumer responses to sponsored-listings is interacting with rank, i.e. at the top of the page, as expected, the performance of organic listings is better in comparison to the sponsored listings. However, for properties ranked lower down the page, sponsored listings seem to be outperforming their organic counterparts. This effect is statistically significant and holds after controlling for other observable characteristics associated with the properties (hotels). The reduced form evidence seems to support the intuition from the two theories, i.e. the performance of sponsored listings is positive and declines with the rank of the property, and the performance of sponsored listings in comparison with organic listings is moderated by a "bias" against the sponsored nature of such listings. I then formulate and estimate a structural model of sequential search to rationalize the empirical pattern due to the counteracting effects - the advertising-quality effect and preference for organic effect. The structural model endogenizes the drop in CTR with rank by making search cost a function of the rank at which a product is listed. Additionally, I handle the tricky problem of endogeneity associated with targeted advertisements by using a unique feature of my data which is correlated with advertising cost.

The manifestation of prominence in the online context is through the rank of an ordered list. The cognitive behaviors that could potentially explain existence of the rank effect and serial position can be rationalized through search cost (notional or actual). Since I am interested in analyzing consumer behavior in the presence of sponsored listings on online marketplaces, exploring the factors that influence or are likely to influence consumer search and purchase behavior is important. To this end, I use a structural model of *consumer search* behavior, which explicitly models for search cost. Having estimated consumer utilities from the revealed preferences of real users (through click and purchase observations), the use of a structural model is conducive for analyzing user behavior under various counterfactuals (Jeziorski and Segal (2015)). The choice of *consumer search* model(Hong and Shum (2006); Kim, Albuquerque, and Bronnenberg (2010)), allows for the limited information search behavior of users due to the presence of search costs (psychological or otherwise). Using this model, I account for the consumer search (clicking) and purchase behavior and how it is impacted by the presence of sponsored listings.

2.1.3 Contribution & Managerial Relevance

This paper contributes to the growing literature on sponsored listings used on online marketplaces (Sahni and Nair (2020); Long, Jerath, and Sarvary (2018)). One of the few papers which explores the quality signaling role of sponsored listings. However, this is perhaps the first paper which has attempted to account for bias against sponsored products, using observational data. Sharma and Abhishek (2017) is the closest in terms of the research question, but crucial differences exist. Unlike Sharma and Abhishek (2017) which uses a field experiment, I use observational data and a structural model of costly sequential search. Also unlike Sharma and Abhishek (2017), I have explicitly attempted to account for and explain the bias against sponsored listings in my model. Choi and Mela (2019) has also attempted to study the effect of sponsored listings on online marketplaces, but unlike them, I have ignored the supply side and modeled the impact of advertising through the demand side model. The demand side search model used in Choi and Mela (2019) is based on the ordered search theory (Arbatskaya (2007)), whereas I use the search model based on the random search Weitzman (1979).[9] Another crucial difference between my paper and the two papers mentioned above is in the empirical context. This paper is likely to be of relevance while optimizing the gains from sponsored listings for online marketplaces. An understanding of the interplay between the two effects and the net impact on sponsored listings is crucial for online marketplaces. Further, this research is likely to be useful in optimizing the rank ordering of properties from the perspectice of a recommender system designed (De los Santos and Koulayev (2017)) to improve consumer search.

[9] In this model consumers order the search based on reservation utilities. The relation between reservation utilities & search cost and the use of rank fixed effects in search cost specification should be able to rationalize consumers' sampling firms in a rank ordered fashion.

2.2 Relevant Literature and Hypotheses

2.2.1 Sponsored listings as quality signal

Nelson (1970, 1974) explored the mechanism behind informational advertising, impact of direct & indirect information on search & experience goods respectively and put forth the idea that one of the three reasons advertising provides indirect information for experience goods is that advertising signals product quality.

There are papers which have explored this hypothesis from supply side. Kirmani and Wright (1989) was among one of the earliest papers in marketing to confirm Nelson's quality signal theory using experiments. Moorthy and Zhao (2000) find that for search goods, advertising expenditure is positively correlated with perceived product quality, even after accounting for objective quality, price, and market share. A few papers have also explored the signaling hypothesis from consumer perspective. Moorthy and Hawkins (2005) find that ad repetition can influence perceived quality. Sahni and Nair (2020) use a field experiment to find evidence for a utility enhancing ad disclosure effect of advertisement for restaurants. This is interpreted as evidence supporting the quality signal due to advertisements that is perceived by consumers. In spite of the empirical evidence favoring the quality signal role of advertising, the empirical literature on the returns to paid search seems inconclusive. For example, while Dai and Luca (2016) find increased Yelp page views due to ads and also increased purchase intent, Blake, Nosko, and Tadelis (2015) find negative average returns to the ads on eBay. Sharma and Abhishek (2017) find a small negative impact of advertising on the platform for electronics and that organic listings displaced by the sponsored listings experience an improved CTR. These results seem to imply that the category in which ads and sponsored listings are used also seem to make a big difference as suggested by Animesh, Ramachandran, and Viswanathan (2007) who use a SEC (Search, Experience and Credence) framework to show different perceptions to ads in different categories. For instance, the empirical context of Sahni and Nair (2020) is a restaurant delivery platform and is a relatively low consideration purchase occasion. And it may be argued that consumer response to advertisements in high consideration purchase occasions may be vastly different. Thus, in the current paper, I use a high consideration purchase occasion (vacation stay) and additionally account for the position effect of advertisement.

Most of the papers above include advertising exposure as an element in the utility function. There are a few (Ackerberg (2003a), Erdem and Keane (1996),Anand and Shachar (2011)) which allow advertising to affect information set in their models. Doing so makes the model flexible enough to not only allowing advertising to provide direct information but also explicitly models the quality signal of the advertisement (Anand and Shachar (2011)). Such learning models in which consumer behavior during information acquisition through advertisements is explicitly specified are quite flexible (often enough to separately identify the direct and indirect (quality signal) roles of advertisements). Unfortunately, the limitations of the data at hand do not permit such a learning model in the current paper and hence I am unable to separately identify the different effects associated with sponsored listings. This is because, even though like Anand and Shachar (2011), advertisement in my case is assumed to provide information on the observable product characteristics, users in my data set do not necessarily have multiple exposures to the same product (advertised and non-advertised alike) and in majority of cases, observations on users are only for a single search occasion. This makes the separate identification of the different roles of sponsored listings impossible. If the data at hand had been longitudinal i.e. observing the same user react to listings on multiple search occasions and if the listings had been shown as organic as well sponsored, such a model might have been possible. To overcome this deficiency of the data, I include the sponsored - rank interaction term.[10] By doing this I am exogenously imposing a dependence between consumer response to a sponsored listing and the rank at which it is listed. This allows my model to capture the rank varying dynamics of consumer

[10]This is in the same spirit as Ackerberg (2001), where advertising is interacted with user experience dummy to account for prestige effects of advertising.

response to sponsored listings. A final point

2.2.2 Consumer Trust of sponsored listings

While the literature on the consumer trust of sponsored search was developed in relation to search engine results page (SERP) research, the concept is equally applicable in the case of sponsored listings which are intermingled with organic listings. As the internet penetration has increased globally, consumer usage has also matured. The mandatory legal disclosures of sponsored content has helped consumers become increasingly aware of the revenue implications associated with sponsored content (for the platform) (Ma, Liu, and Hossain (2013)). Prior research has identified that consumer perception of sponsored content tends to be negative. Consumers' negative emotional responses to sponsored links appear to be reflected as a result of the advertising nature of sponsored results (Marable (2003)). Jansen and Spink (2007) report that online searchers prefer to click links are perceived to be organic results rather than advertisements. Jansen and Resnick (2006) find a strong preference for organic links. This preference does not change, even as users gain a greater searching experience. Ma, Liu, and Hossain (2013) also find similar results as Jansen and Spink (2007), in that trust is lower for sponsored links compared with organic links, and consumers are less likely to buy from vendors in sponsored search results. However, the disclosure of information about vendors' reliability reduces this negative effect. Hotchkiss, Garrison, and Jensen (2005) find that consumers are suspicious of sponsored links and Hotchkiss (2004) finds that users consistently rated sponsored links as poor quality. Jansen (2007) finds that even if sponsored and organic links are comparable, consumers tend to perceive sponsored links as inferior. Edelman and Gilchrist (2012) show that consumers have a higher mistrust of the label "paid advertisement" as compared to "ads" or "sponsored links". Even in the case of a discerning audience like academia, labeling seems to be crucial. Beel, Langer, and Genzmehr (2013) show that organic recommendations are preferred over commercial recommendations, even when they point to the same freely downloadable research papers. Aribarg and Schwartz

(2018) document that increasing the prominence of an ad inserted in an email newsletter decreased clicks on it. As Jansen and Spink (2007) point out, when the search process becomes more focused, the likelihood that users will consider the sponsored listings increases. Understanding the nature of these effects will help to improve the online marketplace design.

2.3 Data and Reduced Form Analysis

2.3.1 Platform overview

The data has been provided by Expedia.com, one of the leading global travel companies headquartered in the U.S. The data is for online hotel booking for vacation travel. The search typically starts off with the consumers specifying the travel destination, check-in & check-out dates, number of children and adults along with the number of rooms. In response to the search query, Expedia.com returns a list of properties (see Figure 2.1) ranked as per the Expedia default algorithm (factoring in a variety of considerations which include among others the hotel availability, historical quality, occupancy and preference levels). The search page results give a snapshot of property characteristics such as its per night price, the star rating, average user rating, discount information, and included services (e.g. breakfast). The snapshot must also include sponsorship disclosure to be legally compliant. Since the list of the hotels can be very long, sponsored properties get a lift in the rank (for an auction determined price) and therefore can affect a change in the rank from the rank assigned by the algorithm. The organic listings for the corresponding sponsored property is shown at its naturally occurring rank, i.e. sponsored properties may potentially be seen twice by the user. To view further details about any hotel listed on the search page, consumers can click on any hotel and go to the property specific landing page which provides further information about the hotel including photographs, reviews and a detailed list of amenities. Consumers may make the purchase from the property specific landing page. For every purchase, Expedia.com receives a pre-determined commission.

2.3.2 Data overview

The data comprise of search and purchase impressions by US consumers, between 1-June-2017 and 15-January-2018, for popular international vacation destinations which include the Caribbean islands (including Cancun, Mexico), Rome (Italy), Paris (France) and Istanbul (Turkey). Planning accommodation for vacation usually requires involvement and high consideration due to the relatively higher emotional and financial involvement.[11]

The observations include *user id* which uniquely characterizes each consumer based on persistent cookies, *session id* which characterizes each session (i.e. till browser window is closed) and *search id* which characterizes every unique activity within a session (new destination search, any changes to the search results owing to sorting criteria or filtering criteria, etc). The cleaned data has approximately 2.5 million observations over 44,504 unique user ids. For every search id, I have data on all the impressions the user has seen, along with the rank ordering of the displayed properties. Additionally, the data includes information regarding the sponsored status of the listing, the user device and the date and time stamp at the search id level. For every property, I also have the click and purchase information at the 7 day aggregation level, i.e. for a specific search query (Location, dates, number of guests, etc), the click against a property cannot be discerned at the search id level but at the level of the past 7 day activity.

2.3.2.1 Consumer Search

Table 2.1 shows a quick overview of the data. The fact that the sponsored listings are applicable only for default sorting presented on Expedia.com needs to be emphasized. When any user customizes the sort order, sponsored listings are no longer displayed and only organic listings are displayed. 50,637 search ids out of 61,068 have default sort order. In the sample, a user on average performs 1.4 searches over 1.3 sessions. Table 2.2 shows the search related

[11] http://www.hotel-industry.co.uk/2015/10/the-luxury-way-marketing-for-hotels/

characteristics of the sample. Sufficient variation in the consumer search behavior seems to exist, implying heterogeneity among consumers. Table 2.3 shows evidence of the link between search and choice across attributes. The average for each attribute was computed for each consumer and then the averages across consumers. The column "Raw Results" includes all the results as displayed to the consumers 5. The averages for other attributes are in the expected direction, except for the discount percentage. It may be argued that consumers do not factor in the discount percentage (struck off price) while making their search and purchase decision. The click and purchase histograms in figure 2.2 confirm this.

Table 2.1: Impression Level Summary Statistics

Variable	Mean	Std. Dev	Min	Max
Price	179.51	219.81	50.00	5000
Star Rating	3.08	1.11	0	5
User Rating	3.37	1.51	0	5
No. of Reviews	613.47	2111.97	0	56372
Branded (Yes/No)	0.38	0.48	0	1
Sponsored (Yes/No)	0.03	0.18	0	1
Deal of the Day	0	0.06	0	1
Clicks	0.04	0.2	0	1
Transactions	0.02	0.15	0	1
Rank	27.94	24.23	1	400
Filtered (Yes/No)	0.04	0.19	0	1

Notes: Summary statistics for the data at Impression level

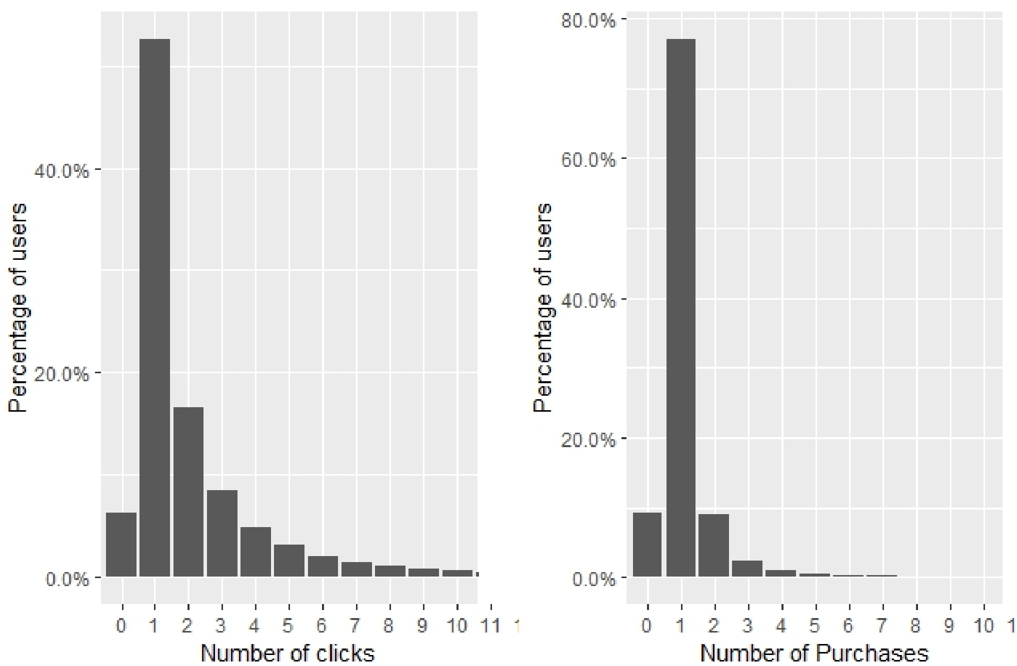
Table 2.2: Search Summary Data

Variable	Mean	Std. Dev.
Purchases	1.15	1.01
No. of Clicks	2.22	2.79
No. of Sessions	1.3	1.21
No. of Searches	1.37	1.44

Table 2.3: **Search Evidence**

Attribute	Raw Results	Searched (Clicked)	Purchased
Price (USD)	196.48	168.68	164.76
Star Rating	3.09	3.39	3.41
User Rating	3.28	3.71	3.74
Reviews	500.28	819.78	846.34
Branded (Part of Chain)	0.35	0.46	0.47
Discount Percentage	-0.59	-0.36	-0.33

Figure 2.2: Search Behavior - Clicks and Purchases



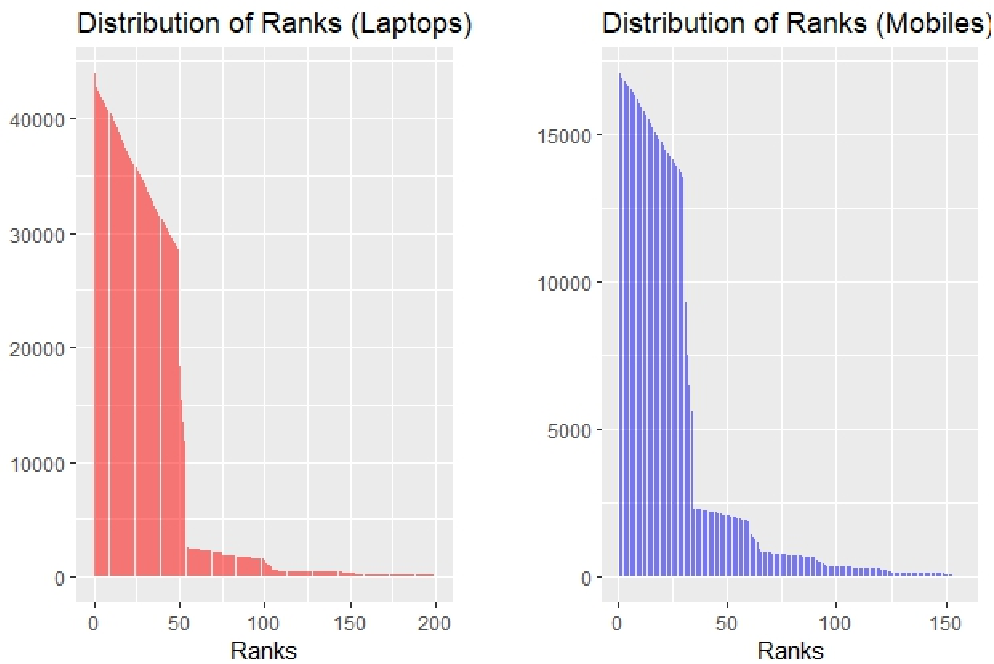
2.3.2.2 Rank Effect

To enter the consideration set typically marketers insist on visibility which in the current online context translates to having a better (lower) impression rank. So I draw attention to the importance of rank (position) as seen in the data. Consumers exhibit limited information search as seen in the browsing behavior shown in figure 2.3. The x-axis shows the rank and the y-axis shows the number of search ids in the data where that rank was browsed. For example, rank 1 was browsed in all search ids, rank was browsed in most but not all, etc. Clearly the browsing length is decreasing with the rank and very few consumers ever go beyond page 1 (across devices). Figure 2.4 shows that the click probability decreases exponentially with the rank. The x-axis is the rank and the y-axis is the average click through rate. These patterns for ranking (for the top ranked impressions) are consistent with those from Ghose, Goldfarb, and Han (2012). Besides the rank effects, the figure also shows the page effect for both devices. While the rank effect is prominent at the beginning and seems to be tapering off, this could be due to the significant destination heterogeneity (presence of smaller and larger destinations).

2.3.2.3 Sponsored Listings

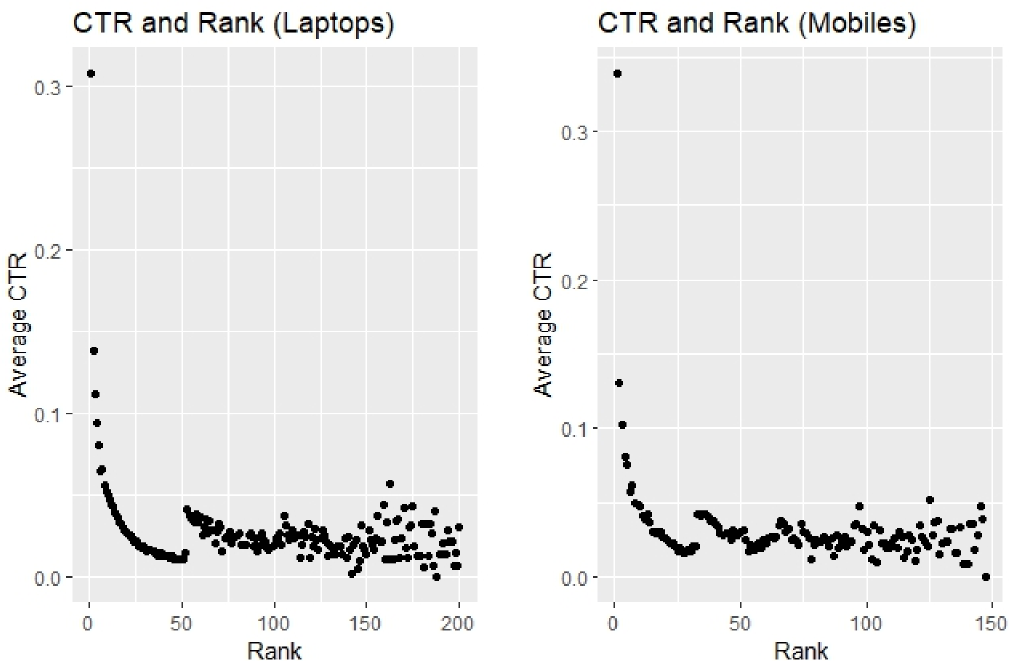
The idea behind sponsored listings is to promote visibility of properties (in return for a payment). The rank effect persists among sponsored listings implying that sponsored products at the top of the page perform better than sponsored products at the bottom of the page. Please note that in the dataset, sponsored listings have occurred at every possible rank. However for this analysis, only those ranks have been included at which a critical mass of properties (at least 50) were sponsored.

Figure 2.3: Browsing Behavior



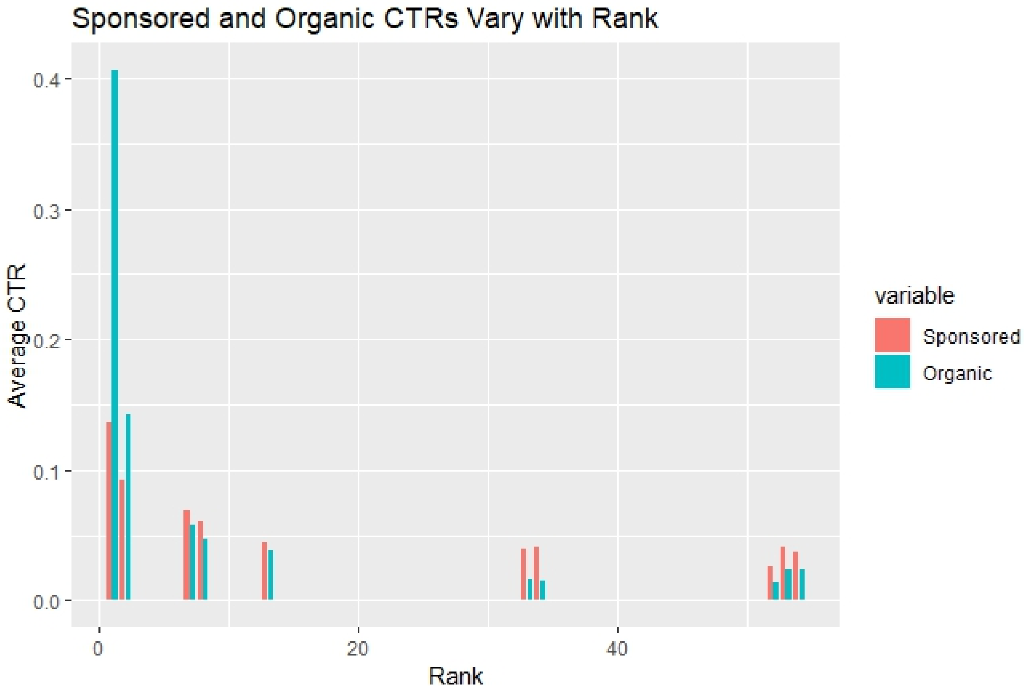
Notes: The figure shows for each rank/position the number of users who browsed up to it, in the data. For each rank/position further down the page, the number of users who browsed up to it is clearly decreasing

Figure 2.4: Rank/Position Effect on Clicks



Notes: Clicks are inversely proportional to the magnitude of the rank

Figure 2.5: Sponsored vs Organic listings - Average CTRs



Notes: The CTR for sponsored listings is lower than the CTR for organic listings at the top of the first page and vice versa at other ranks where sponsored listings are deployed.

2.3.3 Reduced form analysis

I first explore consumer perceptions of sponsored listings vis-à-vis organic listings. Is there any evidence of a "non-organic" effect or bias against sponsored listings?[12] Figure 2.5 shows that there is evidence for both in the data. It separately plots the average CTR for sponsored properties and for organic properties. Only those ranks for which sponsored properties were present above a threshold (greater than 50) have been included. At smaller ranks (top), organic listings seem to be preferred whereas at higher ranks (lower), the quality signal of sponsored listings seems to be making them attractive.

To ensure that there is no other confounding variable such as brand (chain affiliation of

[12]The signaling effect in Sahni and Nair (2020) is net of the positive quality signal derived effect a.k.a "ad-effect" and the negative "not-organic" effect

a hotel) that is resulting in the pattern seen above, I regress the click through rate (CTR) on sponsorship status of a listing while controlling for all observable characteristics of a property. From the regression estimates I obtain the difference in the predicted CTRs of sponsored and organic listings. The panel at the bottom of figure 2.6 shows these. The results from the regression confirm the focal pattern from figure 2.5.

2.3.3.1 Discussion

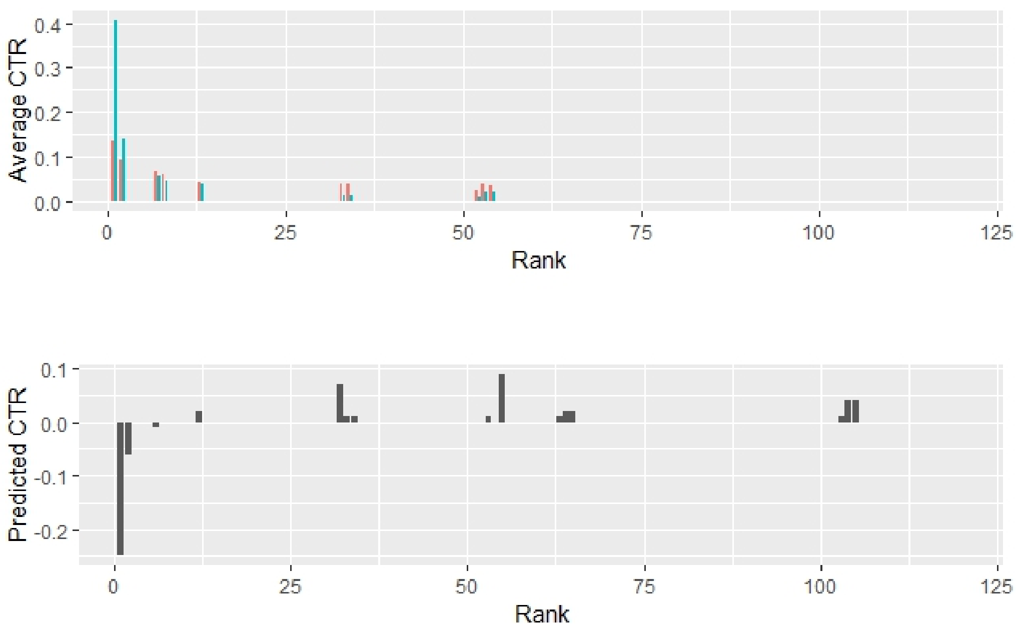
The regression predicted CTRs verify the pattern seen in figure 2.5 and show the evolution of net of the two effects which includes the "positive" quality signal of sponsored listings and the "negative" bias due to the "non-organic" nature of sponsored listings.[13] The theoretical direction of evolution of each of these two effects may be increasing, decreasing or constant with rank, but based on the empirical pattern I can rule out that both of them simultaneously are constant with rank. Extant literature may be of help in identifying the direction of the quality signal (perceived quality). Moorthy and Zhao (2000) find advertising expenditure and perceived quality are generally positively correlated even after accounting for objective quality, price, and market share. It may be argued that since Rank is perceived as a proxy for ad expenditure (higher rank implies greater expenditure), perceived quality is positively correlated with the rank implying that perceived quality is non-increasing in rank.

Thus, perceived quality is either decreasing or constant with rank and in either case the empirical pattern that we see would not be feasible with bias or "non-organic" effect that is increasing (becomes more negative) or stays constant with rank. So it is safe to deduce that the bias against sponsored listings is decreasing with rank.

Next, I explore for patterns in the data which are consistent with the predictions of advertising-quality signal theory. In order to do this I perform regression analysis from sup-

[13]The 2001 working paper version of Anand and Shachar (2011) demonstrated a "negative" consumption deterring effect of advertisements for products which do not fit well with consumer tastes. Repeated exposures of the same product are needed for such a negative ad effect to materialize and in my data I do not observe such repeat exposures at the individual user level for the same sponsored property.

Figure 2.6: Regression predicted CTRs



Notes: Event after using regression to control for all observable characteristics of hotels the CTR of sponsored listings is lower than the CTR for organic listings at the top of the first page and vice versa at other ranks where sponsored listings are deployed. This suggests that the pattern is robust to potential confounders.

ply perspective to check for evidence consistent with the advertising-quality signal (indirect information) hypothesis. From consumer perspective, for the advertising-quality signal hypothesis to be supported advertising must influence consumers' perceived quality, even after controlling for other characteristics such as objective quality and price. To show this, ideally one would need data regarding consumers' perception of quality (of the advertised product) before the advertising is shown and after the consumption of advertising. Unfortunately, such before and after measures of quality are not available in the current data.

From supply perspective, theory concerning signaling role of advertisement predicts that hotels of higher quality would advertise more (Sahni and Nair (2019)). Regressing the indicator for sponsored status on the underlying property's attributes, I get the results shown in table 2.4. The likelihood of a hotel opting for sponsored listing increases with user ratings and star rating, both of which are proxies for hotel quality and are as expected. Additionally, the likelihood to advertise is higher for hotels affiliated with a chain (a proxy for branding in this industry). Since hotels affiliated with a chain are (by contract) expected to maintain a higher minimal quality standards (Hollenbeck (2017), Hollenbeck (2018)), these results are also as per expectation. As expected advertising likelihood seems to decrease with the number of reviews, since well known and established hotels are unlikely to advertise. This is also consistent with Hollenbeck, Moorthy, and Proserpio (2019), who find that hotels (on TripAdvisor platform) with higher ratings spend less on advertising. The correlation patterns between advertising status and proxies for quality on supply side seem consistent with the quality signaling hypothesis of advertisement, where hotels of higher quality seem to be advertising more to signal higher quality.

2.4 Model

I model consumers' click (search) and purchase/no-purchase decision as a model of sequential search in which consumers' search for match value. The sequential search model is based on

Table 2.4: Regressing Advertising Decision on Hotel Attributes

	Estimate	Std. Error	p-value
Intercept	-0.0346	0.0005	0.0000
No. of reviews	-0.0000	0.0000	0.0000
Hotel is in chain	0.0594	0.0003	0.0000
Star rating	0.0098	0.0002	0.0000
Price	9.02e-06	0.0000	0.0000
24 hour front desk	0.0096	0.0004	0.0000
Beach nearby	-0.0080	0.0004	0.0000
Business hotel	0.0345	0.0004	0.0000
Free breakfast	-0.0388	0.0003	0.0000
Garden	-0.0003	0.0003	0.4204
Pet friendly	-0.0061	0.0004	0.0000
Indoor pool	0.0001	0.0007	0.8467
Outdoor Pool	0.0169	0.0003	0.0000
Ski area	0.0001	0.0026	0.9662
Spa	0.0062	0.0005	0.0000
Internet access	-0.0107	0.0004	0.0000
Guest rating	0.0048	0.0001	0.0000
Location fixed effects	Yes		
Observations	1711947		

Notes: Results from regressing an indicator for listing's sponsored status (i.e. decision to advertise) on hotel attributes seems to suggest that hotels with better quality tend to advertise more.

Weitzman (1979).

2.4.1 Utility and Search Cost

The utility for consumer i, for hotel j is given by:

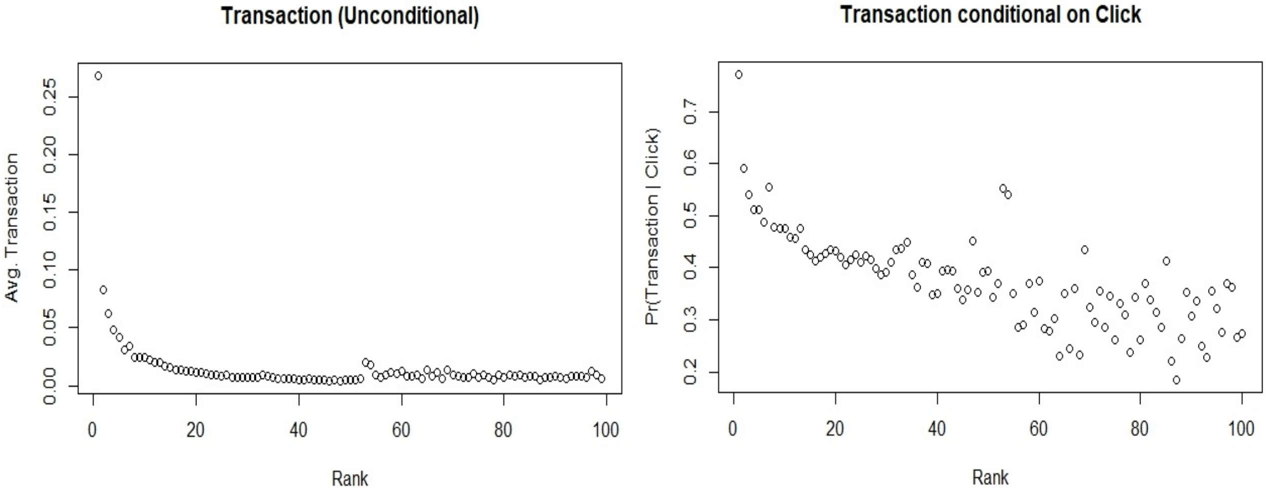
$$\begin{aligned} u_{ij} &= v_{ij} + \epsilon_{ij}; \\ v_{ij} &= x_{ij}\beta \\ \epsilon_{ij} &\sim N(0, \sigma_j^2) \end{aligned} \tag{2.1}$$

where I assume that the consumer is searching over ϵ_{ij}. Prior to the search, consumer knows v_{ij}, which is composed of consumers preference (β) for a vector of hotel attributes (x_{ij}). A more detailed representation of v_{ij} is given below. Note that the hotel attributes are consumer specific, i.e. attributes, such as price, number of reviews, etc, for the same hotel might vary based on the time at which a consumer views it. Due to healthy incidence of non-purchase (after search) in the data, I also include an outside option of no purchase. In the case of the outside option, only a fixed effect is estimated.

The mean utility component, represented below includes a rank-sponsorship interaction term to capture the net-impact (net of the positive quality signal effect due to sponsored status and the negative "non-organic" effect) due to the sponsorship status of a listing. This term is included to capture the different effect of sponsored listing at different ranks. Figure 2.7 provides empirical support for the inclusion of rank in utility specification.

In right panel of figure 2.7, the probability of transaction conditional on click seems to be dependent on the rank. It is not a clear exponential dependence as is in the case of the unconditional purchase probability. The dependence of the purchase (conditional on click) on the rank, implies that rank is a useful guide for search and purchase (Armstrong (2017)). This may be used to justify the inclusion of the rank in the utility specification. Accordingly,

Figure 2.7: Rank/Position Effect on Transactions



the mean utility component (v_{ij}) is represented as follows:

$$v_{ij} = X_{ij}\beta - \alpha P_{ij} + \mathcal{I}_{ij,ad}\, r_{ij} \tag{2.2}$$

where X_{ij} includes the non price characteristics, P_{ij} is the price, $\mathcal{I}_{ij,ad}$ is an indicator variable to indicate if hotel j has been sponsored and r_{ij} is the rank at which the sponsored property j has been displayed at.

After Expedia returns the list of hotels in the default sort order, consumers click on a hotel to find out more about it.[14] For every click, consumer has an associated cost to it. Following Jeziorski and Segal (2015), I endogenize the CTR decline with rank by including, rank fixed effects in the search cost specification. As mentioned previously, one of the mechanisms through which rank effect is assumed to manifest is through the impact rank has on the

[14]Consumers may also change the default sort order of the hotels. As stated in the data section, only an extremely small percentage of users ever use such sorting and filtering options, so I did not incorporate such actions from consumers in the model. I also did away with users who choose to move away from default sort order.

search cost. The search cost specification is similar to what is used in Ursu (2018):

$$c_{ij} = exp(c + \gamma r_{ij}) \tag{2.3}$$

where c is the mean search cost, r_{ij} is the position at which hotel is displayed and γ is the rank coefficient in cost specification. The exponential function is to ensure that the search cost is always positive and as stated in Ursu (2018) is standard way to handle search costs in the search literature.

2.4.2 Optimal Sequential Search Strategy

The rules which govern the optimal sequential search strategy are based on the consumer's reservation utility z_{ij}, defined as the utility that makes the consumer indifferent between the cost of carrying out an additional search at the marginal cost of c_{ij} and the expected marginal benefit from an additional search. The marginal benefit is in relation to a previously searched product with utility z_{ij}. Mathematically, as shown in Weitzman (1979), a unique z_{ij} solves the following:

$$c_{ij} = \int_{z_{ij}}^{\infty} (u_{ij} - z_{ij}) dF(u_{ij}) \tag{2.4}$$

where the integral on the right hand side is the marginal benefit from an additional search. (2.4) can be used to back out the reservation utility (of every hotel) based on the closed form equation derived in Kim, Albuquerque, and Bronnenberg (2010), under the assumption that ϵ_{ij} has normal distribution. The closed form equation is as shown below:

$$z_{ij} = v_{ij} + \zeta_{ij}\sigma_j \tag{2.5}$$

where ζ_{ij} is obtained from the implicit function defined in (2.4) and as shown in Kim, Albuquerque, and Bronnenberg (2010) simplifies to:

$$\frac{c_{ij}}{\sigma_j} = (1 - \Phi(\zeta_{ij}))(\frac{\phi(\zeta_{ij})}{1 - \Phi(\zeta_{ij})} - \zeta_{ij}) \tag{2.6}$$

Under the setting specified above (2.1 through 2.6), optimal sequential search strategy implies the set of three rules put forward by Weitzman (1979). These rules require each consumer to first sort the hotels in a decreasing order of their reservation utilities ($z_{i1} > z_{i2} > z_{i3} > ... > z_{iJ}$)

1. Selection rule: Consumers search the alternatives in the decreasing order of the reservation utility, i.e. the reservation utility of the first searched alternative has to be greater than the reservation utilities of all the other searched options and the reservation utilities of all non-searched options. Thus for the nth search:

$$z_{in} \geq \max_{k=n+1}^{J} z_{ik} \;\; \forall \, n \in 1, ..., M \tag{2.7}$$

   where M_i is the number of hotels searched by user i. Otherwise, the set of companies searched by the consumer would have been different.

2. Stopping rule: Consumers stop searching when the maximum realized utility from among the searched options exceeds the maximum reservation utility from among the un-searched option. For any searched hotel, this implies that its reservation utility must be greater than the maximum of the realized utilities of the hotels searched so far.[15] For an un-searched hotel, this implies that its reservation utility is lower than the maximum of the realized utilities searched hotels. Thus for the nth search:

$$\begin{aligned} z_{in} &\geq \max_{k=1}^{n-1} u_{ik} \;\; \forall \, n \in 1, ..., M \\ z_{iq} &\leq \max_{k=1}^{M} u_{ik} \;\; \forall \, q \in M+1, ...J \end{aligned} \tag{2.8}$$

[15] Honka and Chintagunta (2017) refer to this as "opposite of stopping rule"

Otherwise, consumer would end up searching all the hotels

3. Choice rule: From among the searched hotels, consumer picks the hotel with the highest realized utility. Note that this also includes the utility of the outside option.

$$\begin{aligned} j^* &= \arg\max_{k=1}^{K} u_{ij} \;\; \forall\, K \in M \cup \{0\} \\ u_{ij^*} &= \max_{k=1}^{K} u_{ij} \;\; \forall\, K \in M \cup \{0\} \end{aligned} \tag{2.9}$$

Thus, as long as the marginal benefit of an additional search is positive, consumer continues to search the option with the highest reservation utility. When the consumer decides to stop the search, the searched option with the highest utility will be purchased.

2.5 Estimation

In order for the estimation to be consistent with the optimal search strategy, information on the click order of the hotels is needed. This information is not available in the data set. It is possible to estimate using the Honka and Chintagunta (2017) approach which renders the click sequence redundant. However, I make the assumption that in cases with more than one click, consumer's click sequence is top down, i.e. the hotel listed at the top (lower numerical rank) gets clicked first. As can be seen in Ursu (2018), this assumption is not very far fetched. The rest of this section describes the empirical strategy employed for the model described in the previous section.

2.5.1 Likelihood

The probability of observing a consumer search a set of companies γ_i and purchase from company j (including outside option) under sequential search is given by the joint probability of the three conditions (equations 2.7, 2.8 and 2.9) in section 2.4.2.

$$\begin{aligned} L_i &= P_{ij\gamma_i} = Prob(z_{in} \geq \max_{k=n+1}^{J} z_{ik} \cap z_{in} \geq \max_{k=1}^{n-1} u_{ik} \cap u_{ij^*} \geq \max_{k=1}^{K} u_{ij}) \\ L_i &= \int_{-\infty}^{\infty} \mathcal{I}(z_{in} \geq \max_{k=n+1}^{J} z_{ik} \cap z_{in} \geq \max_{k=1}^{n-1} u_{ik} \cap u_{ij^*} \geq \max_{k=1}^{K} u_{ij}) f(\epsilon) d\epsilon \end{aligned} \tag{2.10}$$

The model likelihood is given by

$$L = \Pi_{i=1}^{N} L_i \tag{2.11}$$

The inter-relationship between search and purchase decisions, i.e. purchase conditioned on the consideration set, which in turn is endogenously determined by the search process, implies that the integral in equation 2.10 does not have a closed form solution. To overcome this challenge, simulated maximum likelihood estimation is used to estimate the sequential search model. In this approach, the random shocks (ϵ), which the consumers observe but which the researcher does not observe, need to be integrated out using simulations. These simulated probabilities would be non-smooth and would require non-gradient based optimization methods (McFadden (1989)). To avoid this, I use the scaled multivariate logistic CDF to smooth the probabilities. This logit smoothed accept reject simulation method has been used to estimate search models (Honka and Chintagunta (2017), Ursu (2018)). The details of the kernel-smoothed frequency simulator are as follows:

1. Take $d = 1, ..., D$ draws for ϵ_{ij} i.e. a total of $N * J * D$ draws
2. For each of the ϵ_{ij} draws (dth draw), compute indirect utility u_{ij}^d, search cost and the reservation utilities[16]
3. Compute the following entities (Weitzman's three rules):
   (a) $w_1^d = z_{ij} - \max_{k=n+1}^{J} z_{ik}$ (selection rule)

[16]Search cost and reservation utilities are not dependent on the epsilon draws

(b) $w_2^d = z_{in} - \max_{k=1}^{n-1} u_{ik}$ (stopping rule)

(c) $w_3^d = u_{ij^*} - \max_{k=1}^{K} u_{ij}$ (choice rule)

4. Use the logit smoothed Accept Reject probabilities $P_{ij}^d = \frac{1}{1+\sum_{m=1}^{3} -sw_m^q}$, where s is the scaling factor

5. Integrate over the ϵ_{ij} distribution by taking average of the probabilities $P_{ij} = \frac{1}{D}\sum_{d=1}^{D} P_{ij}^d$

The choice of the the scaling parameter is based on trial and error using Monte Carlo simulations (see 2.B).

2.5.2 Identification

2.5.2.1 Model parameters and search cost

The parameters in the model include the preference parameters in the utility specification, the search cost parameters (mean and position specific) and the uncertainty associated with the products (σ_j). I normalize σ_j to 1, which is a common practice in extant search literature while estimating models of sequential search (Kim, Albuquerque, and Bronnenberg (2010), Ursu (2018)). Parameters on user demographics cannot be identified as there is no variation by product. Further, consumer heterogeneity cannot be captured since the data has only a very small fraction of users (less than 5%) with multiple searches.

The preference parameters, which include β, α and the outside option indicator, are identified based on the correlation between the frequency of click & purchase and product characteristics for which we are estimating the parameter as well as the assumed click sequence. The identification strategy for these parameters is analogous to that of any typical discrete choice model. Within the optimal sequential search strategy equations 2.7 through 2.9 capture these correlations. These equations include conditions on reservation utilities and utilities i.e. all three of Weitzman's rules play a part in the identification process.

The point estimate of the mean search cost (intercept) parameter is identified based on 2.8 (stopping rule), the functional form of the utility and distribution of the utility function through the relationship between ζ_{ij} and v_{ij} in equation 2.5. The search and choice rules (Weitzman rules) play no part in the identification of the mean search cost parameters. This is because the search rule is based on the relative rank ordering of the reservation utilities. The relative ranking of the reservation utilities will not change due to the mean search cost (which is the same for all products). Similarly, the choice rule does not feature search costs so the rule plays no role in the identification process. Further the stopping rule only provides the upper and lower bounds of the mean search cost. The point estimate is identified due to the functional form of the utility and the distributional assumption on the utility (Honka and Chintagunta (2017)).

The position (product) specific search cost is identified based on the differences in the frequencies for search and purchase at each rank. For example, products clicked frequently but not purchased have low search costs, and are also low on the utility parameters as well.

2.5.2.2 Endogenous advertising

Hotels are likely to spend on advertising in a strategic manner, targeting (or attempting to target) travelers who are more likely to prefer the hotel. Such non-randomness in the assignment of advertising (as sponsored listing) is responsible for a strong possibility of advertising endogeneity due to selection bias. This is because the click (and purchase) data used to measure the returns to advertising are likely to be biased upwards since consumers more likely to click (and purchase) were targeted to start with (Rutz and Watson, 2019). From the perspective of the consumer utility specification, seen in equations 2.1 and 2.2, the presence of endogeneity would imply a correlation between exposure to a sponsored listing and the error term (ϵ), i.e. model unobservables. The econometrician is unable to see hotel characteristics (model unobservables) that are likely correlated with the decision to advertise. Advertising hotels are likely to target their sponsored listings, based on these model unobservables, to

Figure 2.8: Criteria Provided by Expedia to Target Consumers for Sponsored Listings

a specific set of users or geographies as is common in digital advertising. For example, figure 2.8 shows Expedia's promotion criteria for its online sponsored listings program called "TravelAds".[17] The arguments above imply that search ids which have sponsored listings may be systematically different from search ids which do not have any sponsored listings.

In the data, I observe sponsored listings with variation in clicks - some which were clicked and some which were not. There is also variation so that the different users who search for the same region, check-in, check-out dates and occupants, had different exposures to sponsored listings - a user may have been exposed to sponsored listing while another was not, or different properties were displayed as sponsored. Not all of this variation is likely exogenous. The institutional details which go into the process of allocating sponsored listings are necessary to handle the endogeneity. Hotels can place bids for the target customer segments based on the booking window, length of stay, origin country, etc as shown in the figure.[18] Hotels could be using such targeting practices for a number of reasons such as low demand, special events, seasonality or other hotel specific strategies. Except for the traveler origin, the data includes

[17] As the figure does not include any indication that user search history is used for Targeting, I assume that past user behavior does not influence exposure to sponsored listing. This assumption allows us to treat the data as cross sectional, which is ideal for the control function approach to treat endogeneity.

[18] https://searchsolutions.expedia.com/how-travelads-works

details of all the targeting criteria that may have been used by the hotels for targeting and these have been used as controls. Based on the bids (and the allocated budget) by hotels, Expedia allocates the properties to a set of pre-determined ranks.[19] It is interesting to note that hotels which get listed as sponsored (in the sponsored slots) continue to be displayed in their regular organic positions as well, i.e. sponsored hotels get listed twice. Hotels pay on a per click basis and continue to be listed as sponsored either till the allocated budget is exhausted or some other criterion specified by the hotel is fulfilled.[20] Between the bids and allocation, the source of randomness is due to the number and identity of hotels which participate in the bidding process. Not every hotel advertises daily and it is pertinent to note that many search ids do not have any sponsored listings at all. This could be either due to no bids or due to features of Expedia algorithm. The procedure described below takes this into consideration as well.

I handle the potential endogeneity bias using the control function approach (Petrin and Train (2010), Taylor and Hollenbeck (2021)).[21] The control-function approach derives a proxy variable that conditions on the part of advertising decision that is correlated with ϵ_{ij} so that the remaining variation in the endogenous variable becomes independent of the errors. Advertising decision is a function of X_{jt}, which are the product characteristics from the utility specification and Z, that do not enter the utility definition but affect the decision to advertise (exclusion restriction).

$$adv_{ijt} = \alpha + X_{jt}\beta_j + Z_{ijt}\delta_j + \nu_{ijt}$$

[19]Usually four ranks per page, two at the top and two at the bottom, are allocated for sponsored listings.

[20]Expedia does not advertise hotels which have reached their booking capacity

[21]In the absence of randomization in a micro data set, the correlation between model unobservables and the decision to advertise could ideally be handled by including product fixed effects. However, using hotel fixed effects is not feasible given the extremely large number of hotels in the data. Similarly, the correlation between unobserved consumer characteristics and the decision to advertise could be handled through a structural approach by including in the estimation the joint distribution of the unobserved consumer characteristics and the advertising exposure. This would have been possible if there were a direct measure of consumer browsing history in the data. Unfortunately, such information is lacking in the current data.

In the presence of endogeneity bias, ν_{ijt} from this equation will be correlated with ϵ_{ij} from the utility specification which results in the correlation of primary concern, i.e. between the decision to advertise (included in X_{jt}) and ϵ_{ij}. The choice of Z is such that both ν_{ijt} and ϵ_{ij} are independent of it. As discussed in Train (2009), we can decompose ϵ_{ijt} as $\epsilon_{ijt} = CF(\nu_{ijt}, \lambda) + \tilde{\epsilon}_{ijt}$, where CF stands for the control function and is the conditional expectation of ϵ_{ijt} given ν_{ijt} with parameters λ. By construction, $\tilde{\epsilon}_{ijt}$ are not correlated with ν_{ijt} and this does away with the motivating correlation. Equation 2.1 can thus be re-written as

$$u_{ij} = v_{ij} + CF(\nu_{ijt}, \lambda) + \tilde{\epsilon}_{ijt} \tag{2.12}$$

A linear regression to model the decision to advertise is used to obtain the control function ($CF(\nu_{ijt}, \lambda)$), i.e. the resulting residuals serve as a control for model unobserved characteristics while estimating the utility.[22] As noted in Petrin and Train (2010), we use an estimate of ν_{ijt} in the second stage, and not the true ν_{ijt}. This results in additional source of variation which the asymptotic variance needs to account for. A re-sampling method like bootstrap is often used to handle such complications in the asymptotic variance, which is what I propose to do as well.

For this method to work, instruments Z should be correlated with the decision to advertise but uncorrelated with consumer utility. Advertising cost could be one such instrument - it goes into the decision to advertise but is unlikely to be correlated with utility. However, since the data do not have the cost of sponsored listings, the lift in the rank ($\Delta rank_{ijt}$, i.e. difference in the rank at which a sponsored listings is displayed and the rank at which the listing occurs organically) is used as a proxy for it. The greater the lift in the rank, the more the hotel must have spent for advertising. In the data, for most of the sponsored properties we can also observe the organic listing (and the corresponding organic rank) which is used to

[22]Instead of a linear regression, a probit regression could also have been used. The choice of linear or probit regression could be treated as fungible. Please see Wooldridge (2010). Another point to note is that in equation 2.10, $\tilde{\epsilon}$ will replace ϵ.

compute the lift.[23] The number of hotels in the market (market size) is another instrument which is bound to influence the decision to advertise, but is unlikely to be correlated with utility. In larger markets, firms may be more compelled to advertise to gain visibility. The hotel availability in a market keeps varying based on occupancy rates and is dependent on the time of the year. Hotel management is likely to be aware of it, so I include the market size at the weekly level.[24] Thus lift in the rank and the number of hotels in the market serve as instruments needed for the exogenous variation in advertising decision.

2.5.2.3 Position endogeneity

The default rank order in which any hotel is shown by the OTA is based on a proprietary algorithm, details of which are usually not fully disclosed (2.A provides a few details regarding the factors that go into deciding the default rank ordering by a typical OTA). Such a default ordering is likely to be correlated with unobserved (by the econometrician) hotel quality as these rank orderings take into account the past hotel performance and preference by consumers. In the presence of such correlation, consumer preference for a position is confounded with unobserved product quality, i.e. we cannot be sure if the consumer's click or purchase is due to the rank or due to the unobserved quality of the hotel. There is likely to be random variation even in the default ranking algorithm of OTA (De los Santos and Koulayev, 2017). For example there could be random variation due to factors such as the absence of certain competing hotels due to non-availability of rooms (or of certain room types). This random portion of variation in default ranks is what I seek to extract and use in the utility specification. Since I do not have historical performance data on all the hotels in my data, I cannot follow the procedure followed by De los Santos and Koulayev (2017). Instead, I use the price sorted rank of a property. The price sorted rank of a property is

[23]There are also sponsored properties for which we do not see the corresponding organic rank. For such hotels the highest rank till which the consumer browses (i.e. the last rank that the consumer browses up to) is used to compute the lift (i.e. the minimum lift that advertising has enabled).

[24]I do this as the highest number of hotels that I see in the dataset for the destination for each week.

correlated with unobserved product quality through its correlation with price.[25] The price sorted rank of a property shares with the OTA assigned default rank the random portion of variation that we seek, e.g. the variation is the rank due to presence or absence of competing hotels. Using this information, a first regression of price sorted rank is carried out on price and other observed product characteristics.

$$Rank_{ijt}^{price} = X_{jt}\beta_j + P_{ijt}\delta_j + \mu_{ijt}$$

where X_{jt} are observed hotel characteristics and P_{ijt} is the price. The estimated residuals $\hat{\mu}_{ijt}$ are correlated with the randomness in the rank and will be used as instruments in the control function approach described next.

To obtain the controls to be used in utility specification of equation 2.1, I regress OTA assigned default rank on observed hotel characteristics and the instrument obtained from the previous regression, i.e. $\hat{\mu}_{ijt}$ and run the following regression

$$Rank_{ijt} = X_{jt}\beta_j + \hat{\mu}_{ijt}\delta_j + \nu_{ijt}$$

The residuals from this regression will be correlated with unobserved hotel characteristics that are responsible for the bias. By including the residuals in the utility specification in addition to the default rank, we should be able to control for unobserved hotel characteristics. The remaining variation in the default rank should be exogenous. An alternate modeling choice (ordinal logistic regression) was considered due to the ordered nature of the rank. However, the predicted probabilities from such an ordinal logistic regression are extremely difficult to work with as the expected odds (i.e. intercepts) would need to be considered at not only the different levels of the rank but also different values of the co-variates, which in this relatively high dimensional regression is significantly complex.

[25]In the next subsection on price endogeneity, it will become clear why price is correlated with unobserved product quality.

2.5.2.4 Price endogeneity

The price of a hotel may be endogenous due to some unobserved hotel characteristic being part of ϵ_{ij}, for example unobserved quality may be correlated with hotel price and could also be influencing consumer choices (due to an event at the hotel). Since price is correlated with unobserved product characteristics, price endogeneity is likely to bias the effect of price of click. Biased estimates of price may impact the policy simulations in unforeseen ways. To overcome this issue control function approach, along the lines of Chen and Yao (2016), is proposed. For this to work, valid instrumental variables are needed. I use BLP style instruments, i.e. using average prices in the market, average prices of hotels in the same market with identical star rating, average prices of hotels in the same market with identical user rating. These instruments are likely to be correlated with the pricing decision but unlikely to be correlated with the unobserved product heterogeneity that is suspected to be responsible for the price endogeneity. The control function approach is based on two stages. In the first stage, prices of hotels are regressed on the observed hotel characteristics and BLP style instruments. In the second stage, the residuals from the first stage price regression are used to control for unobserved product characteristics in the utility estimation.

2.6 Results

I estimate the model of sequential search for match value using simulated maximum likelihood estimation (SMLE) with logit smoothed accept-reject (AR) simulator. Separate estimations were carried out depending on whether consumers accessed the OTA using laptops and mobiles. This is to account for the different behavioral patterns across different device types. Additionally each device category serves as a robustness check for the presence of focal pattern of interest. I make 50 ϵ_{ij} draws for every consumer-hotel pair and set the scaling parameter to 3. The results are shown in table 2.5. Columns (1) & (2) show the results for laptops and column (3) & (4) are for mobiles. Further, columns (1) and (3) show the estimation results while not accounting for endogeneity, whereas columns (2) and (4) show the estimation results when endogeneity has been accounted for. Across device categories, the direction of the parameter estimates is along the expected lines. One key point of difference is the estimate of the outside option, which seems to be more preferred over laptops. This could potentially signal that consumers carry out more early stage search on laptops.

From column (1) to (2) and again from column (3) to (4), the change in the direction of the sponsored coefficient implies that the control function approach used to account for the selection bias leading to endogeneity is working. Note that the positive coefficient on the sponsored listings in columns (1) and (3) implies that when not accounting for endogeneity bias, we mistakenly overestimate the advertising elasticity. That the advertising coefficient is negative, implies that, in this empirical context, consumers are likely to be experiencing a dis-utility due to sponsored listings. However, as will be seen in the simulations, this dis-utility reverses based on the rank at which sponsored listings are shown.

Search cost is significant and has important implications for sponsored listings. Hotels with high ranks, i.e. those that appear lower in the ranking of slots have lower chances of being searched. These hotels thus have higher incentive to seek lift in position through sponsorship programs offered by OTA. This could also have interesting implications on ranking, i.e. rank hotels with high expected utilities in more lower ranked (better) positions to reduce the cost of search and high user satisfaction with OTA with potential adverse impact on revenue from sponsored listings. The mean search cost in \$ terms is \$ 0.38 for laptops and \$ 0.41 for mobiles. The average dollar cost per position is nearly identical at \$ 0.003. These results are in a similar ballpark as the search cost of 21 cents from Chen and Yao (2016).

Table 2.5: Sequential Search Model Results

	Laptops		*Mobiles*	
	(1)	(2)	(3)	(4)
Search Cost				
Constant	-1.021^{***}	-1.044^{***}	-0.947^{***}	-0.986^{***}
	(0.010)	(0.010)	(0.013)	(0.014)
Position	0.002^{***}	0.003^{***}	0.001^{*}	0.004^{***}
	(0.0004)	(0.0003)	(0.001)	(0.001)
Utility				
Price	-0.051^{***}	-0.048^{***}	-0.087^{***}	-0.080^{***}
	(0.002)	(0.003)	(0.004)	(0.005)
Star rating	0.082^{***}	0.080^{***}	0.075^{***}	0.071^{***}
	(0.003)	(0.004)	(0.006)	(0.006)
User rating	0.042^{***}	0.040^{***}	0.039^{***}	0.036^{***}
	(0.002)	(0.002)	(0.004)	(0.003)
Review count	0.001^{***}	0.001^{***}	0.001^{***}	0.001^{***}
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Chain	0.138^{***}	0.156^{***}	0.142^{***}	0.169^{***}
	(0.006)	(0.006)	(0.009)	(0.009)
Promotion flag	-0.394^{***}	-0.463^{***}	-0.401^{***}	-0.482^{***}
	(0.029)	(0.023)	(0.035)	(0.036)
Sponsored	0.077^{***}	-0.370^{***}	0.103^{*}	-0.391^{***}
	(0.027)	(0.019)	(0.058)	(0.060)
Outside option	0.169^{***}	0.297^{***}	-0.208^{***}	-0.087^{***}
	(0.039)	(0.022)	(0.038)	(0.033)
Advertising control		0.423^{***}		0.401^{***}
		(0.027)		(0.040)
Price control		0.001		-0.0002
		(0.003)		(0.005)
Position control		0.004^{***}		0.006^{***}
		(0.0003)		(0.001)
Fixed effects & interactions				
Position	Yes	Yes	Yes	Yes
Position X sponsorship	Yes	Yes	Yes	Yes
Log-likelihood	-116959.93	-116775.04	-50207.70	-50104.08
Observations	1,497,068	1,497,068	481,272	481,272
Note:				*p<0.1; **p<0.05; ***p<0.01

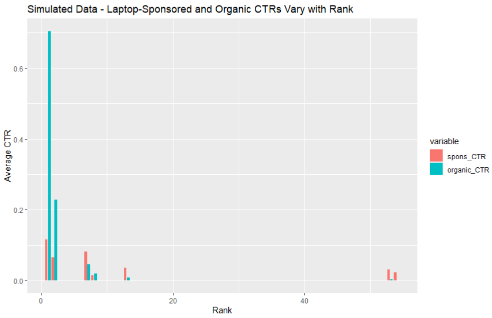
Notes: Columns 1 and 2 show the search model estimation results for laptops, where as columns 2 and 4 show the corresponding results for cases where consumers shopped using mobiles. Columns 1 and 3 show the estimation results while not accounting for endogeneity, whereas columns 2 and 4 show the estimation results when endogeneity has been accounted for. Std. errors are in parentheses. Coefficients (except sponsored) are in expected direction. Accounting for endogeneity in advertising results in directional changes to the sponsored coefficient.

2.6.1 Model Evaluation

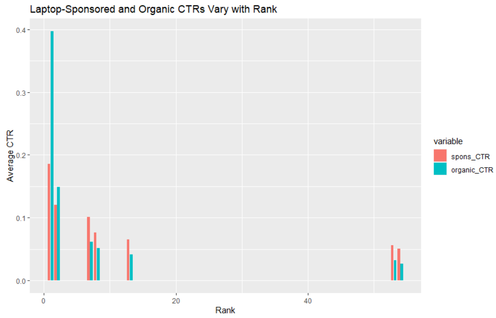
To evaluate the validity of the model, I use the estimation results, to simulate consumer search and purchase behavior. For each consumer, $N = 400$ (i.e. $46,400$ X 400 consumers) simulations are generated for mobiles and laptops separately to evaluate the in-sample predictive performance of the estimated search model. The evaluation criteria is the replication of the focal pattern of interest. For each consumer and the observed hotel data shown by the online marketplace, I use the model estimates to first compute the expected utility, search cost values, reservation utilities (based on search cost and expected utility values) and finally the indirect utility using the simulated value for ϵ_{ij} (of which I have 400 per consumer). Conditioned on the attributes of the hotels shown by the online marketplace, I infer the search (click) options and further conditioned on the search options the inference on the purchased option is made. This process is repeated for each one of the $46,400$ consumers. Based on the click sequences obtained from the simulated data, click through rates at each rank are computed separately for organic and sponsored listings. The results are shown in figure 2.9 below. To facilitate ready comparison, the figures in the column on the right show the patterns in the raw data. The images in the top row correspond to laptops, where as the bottom row is for mobiles.

The results show that the model is successful in replicating the focal pattern of interest, i.e. consumers' perception of sponsored listings vis-à-vis organic listings. Consistent with what extant literature has shown, the CTR continues to decline as the rank of the sponsored listings fall from the top to the bottom of the results page. However, the results confirm the existence of the two dual forces at play. Consumers' preference for organic listings (or lack of trust of the sponsored listings), which is evidenced at the top of the page where ceteris paribus organic listings have higher CTR compared to sponsored listings. Further down the page, this trend reverses as consumers pick up the signal of the inherent product quality.

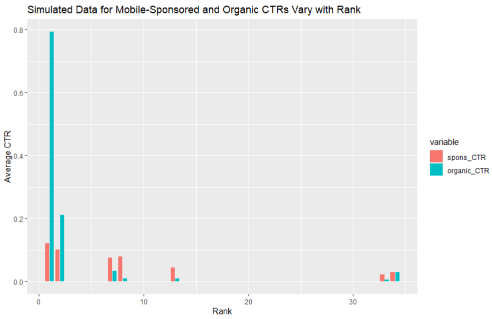
Figure 2.9: Replicating Pattern of Interest Using Estimated Model



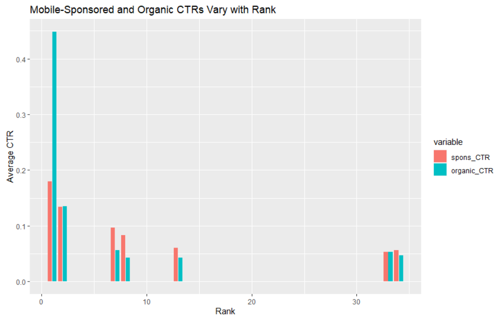
(a) Pattern with simulated (400 simulations) Laptop click data



(b) Pattern in raw Laptop data



(c) Pattern with simulated (400 simulations) Mobile click data



(d) Pattern in raw Mobile data

Notes: Patterns with simulated data in the left column and raw data in the right column. Top row is for Laptops data and bottom row is for Mobiles data. The overall patterns seem to have been replicated, although the relative proportions of organic and sponsored listings are not similar between the simulated and raw data across device categories.

2.7 Counterfactual Study

In this section I perform a few counterfactual analysis to better explore the model based insights into consumer perception of sponsored listings.

2.7.1 Cost of Sponsored Listings

The focal pattern of interest, as seen in both the raw data as well as through simulation study using model estimates, show that consumers' perception of sponsored listings varies along the position or rank ordering of properties from the top of the page to the bottom. The manifestation of such consumer perceptions can best be understood in terms of the customer engagement metrics with the online marketplace. The most common metrics of managerial relevance are the total number of clicks, *click-through rate* (CTR) and transactions (bookings made). Click through rate at any rank is obtained as the ratio of clicks at the rank to the total users who viewed up to that rank. The cost of consumer perceptions (regarding sponsored listings) to the platform can then be measured using these customer engagement metrics.

In order to do this, I first use the estimates from the sequential search model to obtain the simulated total clicks, click through rate and booking rate at each rank conditional on the observed hotel attributes. I do this by performing Monte Carlo integration over the distribution of ϵ_{ij}, by generating 400 random utility shocks per hotel per consumer. Using Weitzman's rules, I simulate the click and booking information for the raw data.

In the second step, I repeat the process under the no sponsored listings at a particular rank (e.g. rank 1) scenario. I do this for all the ranks that are prominently used for displaying sponsored listings for each device type. After simulating the search (click) and booking data, I compare how the consumer engagement varies over the two scenarios, i.e. with sponsored listings at rank xx and without sponsored listings at the same rank. Note that in the simulation process described above, I am comparing the simulated clicks and simulated booking under the two scenarios (with sponsored listings and without sponsored listings) instead of comparing the simulated data with the actual click and purchase data. This is consistent with what is done in existing literature.

While table 2.6 shows the detailed results for this counterfactual, it is worth noting that (a) the simulated behavior for laptops and mobiles is consistent with the focal pattern of interest (as was seen in the section 2.6.1), (b) at the top of the page (ranks 1 and 2), consumers prefer organic listings and the total number of clicks is likely to improve by 34.11% and 29 .57% on laptops and mobiles respectively if sponsored listings are replaced with organic listings. Additionally, at these ranks, conditional on clicking the booking rate is likely to improve by 44.38% and 39.33% on laptops and mobiles respectively if sponsored listings are replaced with organic listings and (c) consumer preference for organic listings vis-à-vis sponsored listings seems to reduce as we move down the page.

Table 2.6 shows, for each rank, the impact on *click-through rate* (CTR), of replacing the sponsored listings with identical organic listings. For example, the first row shows that for Laptops, at rank 1, under the as-is scenario, CTR for sponsored listings is 10.5%. For organic listings at rank 1, it is significantly high at 70.3%. Under the counterfactual, when no sponsored listing is allowed at rank 1, the organic CTR further improves to 74.9%. While returns to being an organic listing in terms of the click-through rate are higher at the top of the page, sponsored listings enjoy a higher click-through rate at larger ranks across devices. The difference in consumer perception towards sponsored listings as we move down the rank ordering, across device types, becomes stark in this table.

The results can be summarized as follows, (a) consumer engagement with the online marketplace is higher with organic listings at the top and with sponsored listings coming further down the page. (b) In fact consumer engagement is likely to be higher with sponsored listings at larger ranks. This is consistent with the expectation that the "non-organic" push back tends to reduce as we move down the page.

2.7.2 Search Cost and Sponsored Listing

I next investigate how search cost impacts consumer engagement with sponsored listings. Such an analysis helps the online marketplace identify conditions that make sponsored listings attractive to consumers. The presence of search frictions may make context relevant advertising such as sponsored listings very valuable for both the online marketplace as well as the advertising firm. This counterfactual allows us to study the trade-off between user experience (better user experience through low search frictions) and the monetization potential of sponsored listings (higher search frictions make sponsored listings more attractive to hotels). Thus in the context of the online marketplace's desire to optimize revenue by managing the

Table 2.6: Results for Counterfactual Click Through Rate at Each Rank/Position

Laptops				Mobiles			
Rank	Sponsored	Organic	Counterfactual	Rank	Sponsored	Organic	Counterfactual
1	0.105	0.703	0.749	1	0.119	0.792	0.824
2	0.067	0.229	0.219	2	0.107	0.216	0.194
7	0.066	0.046	0.171	7	0.068	0.036	0.153
8	0.027	0.021	0.058	8	0.086	0.011	0.040
13	0.031	0.009	0.028	13	0.040	0.012	0.037
53	0.040	0.009	0.038	33	0.048	0.011	0.039
54	0.033	0.0002	0.031	34	0.043	0.034	0.041

Notes: The table shows the click through rate at each rank. For each device type, the non-colored "Sponsored" & "Organic" columns (columns (2), (3), (6) & (7)), show the CTR at that rank under the as-is or non-counterfactual scenario. The colored column "Counteractual" (columns (4) & (8)) shows the CTR value at each rank, under the counterfactual scenario of "no sponsored" listing. The color coding is to segregate the results by rank - green implies ranks where replacing sponsored listings with organic is beneficial and red implies ranks where retaining sponsored listings is beneficial.

dynamics between its supply and demand sides, this counterfactual is important.

The search cost specification in my model has two components: intercept, which can be thought of as the fixed cost of clicking, and the slope coefficient of rank, which can be thought of as the variable cost of browsing. In this counterfactual, I vary only the slope coefficient. Since the intercept component of the search cost impacts all the hotels in the data equally, varying it is unlikely to show any interesting phenomenon. To perform this counterfactual, I use the estimates from the sequential search model to obtain the simulated click through rate and booking rate at each rank conditional on the observed hotel attributes. I generate 400 random utility shocks per hotel per consumer and simulate the click and booking data using Weitzman's rules. Repeating the process by varying the position coefficient of the search cost, I obtain (for each search cost value) the simulated search (click) and booking data and compare how the consumer engagement varies over the different search cost values.

Table 2.7 shows the results for both laptops and mobiles. It shows the contribution of sponsored listings, in percentage terms, to clicks and purchases as the search cost increases. For example, the first row shows that for laptops, if the search cost were zero, 11.59% of the total clicks would be for sponsored listings and 10.86% of the total purchases would be for

Table 2.7: **Results for Counterfactual on Search Cost**

			Laptops			
Search cost	Total Clicks	Sponsored Clicks	Percentage	Total Purchases	Sponsored Purchases	Percentage
0	41855	4852	11.59%	22014	2392	10.86%
0.1	31892	3738	11.72%	20016	2078	10.38%
0.2	27026	4446	16.45%	18654	2719	14.58%
0.4	20990	4523	21.55%	16348	3208	19.62%
0.5	18836	4212	22.36%	15306	3180	20.78%
			Mobiles			
Search cost	Total Clicks	Sponsored Clicks	Percentage	Total Purchases	Sponsored Purchases	Percentage
0	19661	2321	11.81%	10908	1175	10.77%
0.1	15229	1988	13.06%	10065	1146	11.38%
0.2	13174	2304	17.49%	9523	1458	15.31%
0.4	10588	2292	21.65%	8569	1681	19.62%
0.5	9648	2141	22.20%	8117	1670	20.57%

Notes: The table shows, for each device type, the share of clicks across all the ranks that sponsored listings get, as the search cost increases. The first five rows are for laptops and the last five are for mobiles. Column (2) is the average (across 400 simulations) total clicks across all the ranks and Column (3) shows the corresponding figure for only the sponsored listings. Column (5) shows the total purchases across all ranks, averaged across 400 simulations. Column (6) shows the corresponding value for only the sponsored listings in the data. Columns (4) and (7) show the percentage of sponsored listings, of the total, for clicks and purchases respectively.

sponsored hotels.

The results show that across device types as the search cost increases, sponsored listings become more effective. The share of sponsored listings in clicks and purchases increases. The result is highly consistent with the economic theory which predicts that advertising as a source of information can be used to counter the market inefficiencies such as high search costs (Bagwell, 2007). As search cost increases, sponsored listings can be seen to provide low cost direct information to consumers.

It is important to note that while high search costs may improve consumer engagement with sponsored listings, the overall engagement of the consumer with the platform is likely to decrease. This can be seen in results; as the search cost increases the total clicks and total purchases decrease. The platform thus needs to identify the sweet spot that maximizes revenue from sponsored listings without excess loss of consumer engagement with the platform.

2.7.3 Welfare Effects

In this counterfactual, I analyze the welfare effects of a "no-sponsored" ranking scenario. To perform this counterfactual, I re-create the "default" ranks that would have manifested in the absence of any sponsored listings. This entails two steps. First, I delete sponsored listings in every search id. Second, for each search id, the listings that remain after deleting sponsored listings are ordered in the ascending order of the default rank, and new rank are assigned based on this ordering. I then carry out simulations to obtain consumer clicks and choices using the estimates of the parameters obtained from the sequential search model. Simulations are carried out separately using the list of hotels displayed under Expedia's default ranking and under the ranks assigned under the "no-sponsored" ranking counterfactual. Again, by comparing simulated clicks and choices under the two scenarios, I am able to avoid a potential bias. This counterfactual involves changing not just the sponsored status but also the positions of the listings. The analysis is carried out separately for each device type.

Table 2.8: Results for Counterfactual on Consumer Welfare

	D-NS	
	Laptops	**Mobiles**
Utility	3.17%	4.05%
Total search cost	-3.04%	-3.56%
Welfare	15.70%	28.26%
Transaction Rank	-1.85	-2.66
Number of Clicks	-0.80%	0.36%
Total number of Transactions	3.43%	3.28%
Cumulative Revenue	3.49%	5.96%

Note: D = Default ranking; NS = Ranking with no sponsored listings

Notes: Table shows the difference in the simulated average values under Expedia's default ranking and under the counterfactual ranking with no sponsored listing.

I define consumer welfare along the lines of Ursu (2018), i.e. the difference of the consumer's utility from the chosen hotel and the total search cost (from all the clicked hotels) is used as a measure of consumer welfare. I repeated the simulation 400 times for each device and obtained the average welfare for each set of draws. I report the average of the 400 simulations for only those consumers who purchase an "inside" good.

The results, in table 2.8, show that when compared with the default ranking, which includes sponsored listings, the consumer welfare under the "no-sponsored" ranking scenario is on average 15.7% higher on laptops and on average 28.26% higher on mobiles. The results confirm that consumers are better off under the platform assigned ranks, but with no scope for listing hotels to seek "non-organic" lift. These results are broadly in agreement with the predictions of the behavioral literature concerning the "non-organic" effect of sponsored listings. It appears that the welfare improvement is due in equal parts to better matches (utility) as well as lower search costs. Search cost, composed of fixed clicking cost and rank based browsing cost,

2.8 Conclusion and Future Research

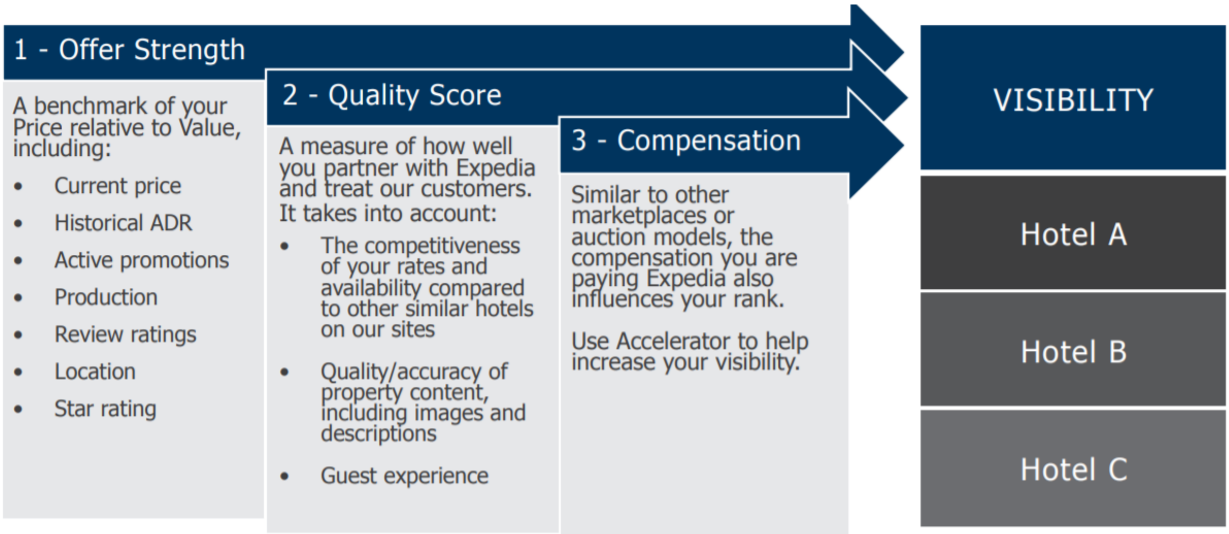
In this paper, I document the rank varying nature of consumer perceptions regarding sponsored listings. I show the empirical existence of multiple mechanisms, which are theoretically at loggerheads with each other. To this end, I use a rich micro level data set with information on an online travel agent's marketplace rank listings & their attributes, and details of the user engagement with the said listings. I rationalize the focal patterns of interest using a sequential model of consumer search, while accounting for potential endogeneity concerns using control function approach with innovative instruments that help extract exogenous variation.

I make the following contributions. First, I find that consumers have a natural preference for organic listings at the top of the page ranks and replacing these with sponsored listings deteriorates consumer engagement with the online marketplace. Second, I find that the consumers' push back against sponsored listings weakens further along the page, so much so that sponsored listings may even outperform organic listings in the middle & lower ranks of a page. I next find empirical evidence that in an environment characterized by high search costs, sponsored listings, as bearers of low cost & direct information, help the listing hotels and encourage competition among competing hotels as predicted by theory. This supports the conclusion that an ideal strategy for the online marketplace is to identify the optimal search cost of the online marketplace, where it is neither too high to dissuade consumers from engaging with the platform nor is it too low to render sponsored listings ineffectual. Finally, measures of the consumer welfare based on the model estimates show that consumers are better off under the default ranking with only organic listings.

Future work on this topic could explore the exact mechanisms that govern the two forces and identify precisely the inflection point where preference for organic listings is supplanted by preference for sponsored listings. Another aspect of the research would be identify, in a structural sense, the device (laptops and mobiles) specific differences concerning the two mechanisms at play.

2.A Appendix: OTA Ranking Basis

Figure 2.10: Expedia Ranking Algorithm Drivers



Adapted from "Understanding the science behind Expedia's marketplace: What drives hotel visibility online." by Melissa Maher.

2.B Appendix: Monte Carlo Simulations

In this section, I describe simulation results that show that the simulated maximum likelihood estimation method with logit smoothed Accept-Reject simulator recovers the utility and search cost parameters in this model. I simulate data for $N = 10000$ consumers, each seaching among $J = 5$ hotels including the outside option. The hotel characteristics used are:

- Price $\sim N(2, 0.2^2)$
- Star Rating $\sim N(3, 0.5^2)$
- ϵ (match value) $\sim N(0, 1)$

I also include user id and rank of the hotel as other necessary characteristics. Simulatoion study results and the details of the simulation process are shown below:

Table 2.9: Results from Simulation Study

Parameter	**True**	**Estimate**	**se**
Price	-1.5	-1.42	1.38
Star Rating	1	0.92	0.55
Outside Option	1	1.16	1.10
Mean Cost	-5.5	-4.99	0.19
Rank	1	0.89	0.05
Log Likelihood		-2242.16	

Notes: Estimates for N=50 simulations

1. Utility is specified as $u_{ij} = V_{ij} + e_{ij}$, for consumer i and hotel j and with $V_{ij} = X_{ij}\beta$ representing the expected Indirect Utility (X_{ij} is the matrix of hotel attributes including price)

2. The true utility parameters are: $\beta_{price} = -1.5$, $\beta_{starrating} = 1$, $\beta_{outsideoption} = 1$, where as the mean search cost parameter, $\beta_{meancost} = -5.5$ and slope parameter for position is $\beta_{position} = 1$ ($c_{ij} = exp(\beta_{meancost} + \beta_{position} * rank)$)

3. I make $N * J$ random std. normal draws since $e_{ij} \sim N(0,1)$ and compute the indirect utility values

4. I compute the search values cost using the above mentioned functional form and use the pre-computed reservation utility grid to look up the reservation utility corresponding to the cost

   (a) I obtain the reservation utility grid using the implicit value function defined in Kim, Albuquerque and Bronnenberg (2010)

   (b) For a fine sequence of ζ values, I use the implicit value function to obtain the corresponding cost value. The implicit value function is defined as: $\frac{c_{ij}}{\sigma_{ij}} = (1 - \Phi(\zeta))(\frac{\phi(\zeta)}{1-\Phi(\zeta)} - \zeta)$, where we take $\sigma_{ij} = 1$

5. After obtaining the reservation utilities for each hotel for every user, I re-arrange the data for every user so that the outside option is the first option followed by the hotels in descending order of reservation utilities. This is to be consistent with Weitzman's (1979) Selection rule

6. Follow Weitzman's (1979) Stopping rule, For every user I mark as clicked each hotel whose reservation utility is greater than the maximum utility of the hotels strictly before it

7. Finally, following Weitzman's (1979) Choice rule, I mark as chosen that hotel which has the highest utility from among the clicked hotels

From the data generation I receive the reservation utilities grid and data which includes the user id, rank, price, star rating, outside option flag, click flag and transaction flag.

1. I make $d = 1, ..., D$ draws for ϵ_{ij} i.e. a total of $N * J * D$ draws from $N(0,1)$

2. For each of d compute indirect utility u_{ij}^d and compute the search cost and obtain the reservation utilities from the look up grid (note that search cost and reservation utilities are not dependent on the epsilon draws)

3. I then compute the following

   (a) $w_1^d = \min_{j \in S_i} z_{ij} - \max j' \notin S_i z_{ij'}$, i.e. minimum reservation utility among the clicked options is greater than or equal to the maximum reservation utility among the non-clicked (Selection Rule)

   (b) $w_2^d = \max_{j \in S_i} u_{ij} - u_{ij''}, \max j' \notin S_i z_{ij'}$, i.e. utility of the chosen hotel is greater than the utility of the non-chosen (but clicked) hotels and greater than the maximum reservation utility among the non-clicked hotels (Stopping and Choice rules)

   (c) $w_3^d = \cap_{t=2}^{T} \max_{t<L} u_{it} - z_{it=l}$, i.e. the maximum utility from among the so far searched hotels is less than the reservation utility of next clicked hotel. This condition is the opposite of the stopping rule

4. Use the logit smoothed Accept Reject simulator $P_{ij}^d = \frac{1}{1+\sum_{m=1}^{3} -sw_m^q}$, where $s = 4$ is the scaling factor

5. Integrate over the ϵ_{ij} distribution $P_{ij} = \frac{1}{D} \sum_{d=1}^{D} P_{ij}^d$

CHAPTER 3

Impact of Sponsored Listings on Awareness, Consideration and Choice

3.1 Introduction

Online marketplaces such as Amazon, Expedia, UberEats, etc present consumers a rank ordered list of available choices. In the presence of a large number of choice alternatives, consumers face uncertainty, which is usually resolved through a costly search. The prevalence of search frictions implies that the consumers are unlikely to consider all the choices and those listed at the top (and readily visible) are more likely to be considered than those at the bottom. The rank ordering of choices by the search intermediary, usually done based on some confidential proprietary algorithm, plays a critical role in this process.

The Online Travel Agents (online marketplaces such as Expedia, Tripadvisor, etc) allow hotels to pay for improving visibility by allowing sponsored (or advertised) listings.[1] Such a sponsored listing improves a hotel's visibility by allowing the listing to appear at a better rank as compared to the OTA's default (or organic) rank for the property.[2] Literature exploring sponsored listings is nascent and has focused on its implications on the online marketplace (e.g. Choi and Mela (2019), Sharma and Abhishek (2017), Chen, Wang, and

[1] usually through a bidding process

[2] Note that many online marketplaces allow sponsored (advertising) hotels to be double listed, i.e. list at their organic rank slot as well as at the advertised rank slot.

Webster (2021)).[3] The mechanism, for sponsored listings, that these papers have exclusively focused on is the lift in the rank due to being sponsored (which brings the product to a better rank/position and placed it within consumer's attention/visibility zone) and the impact of rank on consumer demand. While the improved rank is definitely an important mechanism behind the effectiveness of sponsored listings, it is possible that the sponsored status (or rather the disclosure of the advertised status) itself is prompting the consumers to act differently due to a signaling effect (Sahni and Nair, 2020). This alternate mechanism implies that the consumer response to an advertised listing differs from its organic counterpart because the *disclosure* of the sponsored status of a listing has an impact on the utility that consumer draws from the listing.[4] The utility impact of the disclosure could be due to the effect on perceptions of consumers towards the unobserved quality of the listing (Nelson (1974)). From a marketing view point, impact of sponsored status could thus spill across stages of the purchase funnel (beyond awareness/visibility). At the consumer level, the impact of the sponsored status on the different stages of the shopping and purchase process remains un-investigated. The goal of this paper is to better understand the mechanism(s) behind sponsored listings and explore if the sponsored status of a listing has an impact on the consideration and the choices stages? And to, understand what the impact of rank (position) is on this mechanism? I do this in the empirical context of online hotel booking for vacation stay.

Understanding the mechanisms behind sponsored listings is important for managers and academics alike. From a managerial viewpoint these questions are worth asking due to the impact of the answers on the marketing spend and effort of the advertising firms. Understanding the implications beyond the visibility gain may help hotels reconsider their marketing strategy and reallocate their marketing spends. In the short run, this is useful to

[3]There is a large body of literature which has explored sponsored search advertising (and continues to do so). Sponsored listings however are distinct, in that organic and sponsored listings are interspersed.

[4]The "disclosure" that I refer to here has also been referred to as prominence in existing literature, e.g. (Joo, Shi, and Abhishek (2021)).

improve the advertising returns. A more strategic implication of this research would be to verify the possibility of improving the organic rank of a hotel itself through advertising. If the hotels which use sponsored advertising get higher consideration (& purchases) not just due to lift in the rank but also due to the disclosure effect, then the long run implication of sponsored advertisement would be a better organic rank/position because one of the determinants of the organic rank is historical performance of a listing.[5] Depending on the impact of the sponsored status on consideration and purchase, an advertising hotel may be able to improve upon its profitability metrics vis-à-vis the OTA. Profitability potential of a hotel is among the factors that go into the organic ranking of OTAs. Appendix 3.A provides a few details about the criterion used by a couple of OTAs to determine hotel competitiveness and rank ordering. It is safe to assume that similar criterion are used by other online marketplaces as well. From an academic viewpoint, this research helps identify the mechanisms at play for sponsored advertisements. Specifically, it helps determine if there is a utility impact of sponsored listings.

3.1.1 Model Overview & Main Findings

For the empirical analysis, I develop a structural model of sequential consumer search. Consumers are assumed to be searching for a match from among the listed hotels. The key feature of search models is that consumers make a trade-off between an additional costly search (incurring search cost) which results in higher expected utility, i.e. consumers search up to the point that they are indifferent between the marginal cost and the marginal benefit from an additional search. The model includes two components, search cost and utility, each of which is designed to capture the impact of sponsored listings on consumer attention (or awareness or visibility), consideration and choice. An important point to note is that in the rank ordered, online market platform's context, I am treating user awareness, attention and visibility as fungible concepts. In the proposed model, search cost is being conceptualized

[5]Please see appendix 3.A.

as a cognitive cost, which can be labeled as "cost of user attention". The presence of search cost prevents consumers from sampling all rank ordered hotels (from top of the page to bottom, across many such pages). User attention is modeled, via search cost specification, as a function of rank & page. The reason this specification has been adopted is made clear in the discussion on figure 3.2 later in the chapter. The strategy to estimate the impact of sponsored listing on user awareness is based on a novel aspect of the data, where I observe the organic rank of each sponsored listing. This allows me to estimate the difference in the search costs of the organic and sponsored listings, i.e. difference in "user attention" at the sponsored position and the organic position which serves as the estimate of the impact of sponsored listing on user awareness.

The impact of advertising on consideration and choice stages is captured through its impact on the perceived quality. Similar to Mehta, Rajiv, and Srinivasan (2003) and Jang, Prasad, and Ratchford (2012), the model uses Bayesian updating within a consumer search context. In my model, consumers update in a Bayesian fashion their prior beliefs about quality at each position, based on the noisy signal of quality conveyed by the disclosure of sponsored status of the listing. The paper thus models consumer's learning from advertising signal. The impact of sponsored listings on the consideration (and choice) stages is measured by the extent to which posterior quality beliefs are influenced by the advertising disclosure signal. Posterior quality beliefs due to the organic/sponsored nature of a listing are included in model's utility specification.[6] In order to test for the impact of disclosure of sponsored status on consumer's posterior quality beliefs, the model parametrizes the strength of the signal (of quality) from sponsored listing with a parameter α. This parameter, α, is defined as the "inverse of signal to noise" and is the ratio of the sponsored signal's variance (which is the random noise in the signal) and the variance of prior quality.[7] Thus, a larger variance

[6] For an organic listing, the posterior quality beliefs are the same as the prior quality beliefs.

[7] I fix the random noise in the signal to a constant (= 1) for identification purposes. This is consistent with extant literature.

of the prior quality beliefs would imply that consumer is uncertain of his/her quality beliefs. This would then result in a small value of α. In such a scenario, the quality (posterior) perception associated with a position is influenced by the disclosure of advertisement at that rank. Where as a high value of α would suggest that the users prior beliefs are more or less certain (low variance) and the advertising signal is not very effective, i.e. posterior quality beliefs are not highly influenced by the advertising signal. To make the model further managerially impactful, I allow the α to be a consequence of the rank/position (within a page) at which sponsored listing is shown. I do this by categorizing the ranks within a page as top of the page rank, i.e. top 5 ranks within a page and non-top rank, i.e. any of the remaining 45 ranks within a page.

The empirical analysis in this paper contributes by modeling consumer awareness in the online marketplace's context through the search cost. Based on this, I find that advertisers use sponsored listings for not just lift but also for prominent locations where consumer attention experiences spikes. Further, I find that the advertising disclosure of sponsored listings, through noisy advertising signal, plays a strategic role in enhancing the consideration and choice probabilities at the top of the page. At the rest of the positions (ranks), advertising disclosure, does not seem to have a big impact on consideration and choice stages. These results are important from a managerial perspective, as improved consideration and choice probabilities serve to enhance the organic rank of the listing via advertising.

The remainder of the article is organized as follows. I discuss the relevant literature in the next section. In section 3.3, I present the model. Section 3.4 presents the data and patterns in the data that motivate the model, this is followed by the estimation approach and a brief discussion on identification. Finally, section 3.5 shows the results and presents a discussion of the same.

3.2 Literature Review

This paper is related to and contributes to at least three streams of literature.

This paper is strongly related to the literature on sponsored listings and helps in furthering our understanding of it. While Abhishek, Jerath, and Sharma (2019), Choi and Mela (2019), Sharma and Abhishek (2017) have helped start the discussion of sponsored listings as entities different from search advertisement and helping improve the understand of aspects peculiar to this form of advertising. For instance Choi and Mela (2019) consider the trade-off between the supply side (vying for visibility) & the demand side (efficiency loss) and explore the impact on marketplace revenues. In this regard, Joo, Shi, and Abhishek (2021) explore a similar premise (as I do in this paper) to understand the advertiser's incentive (lift or visual prominence). In the current paper I seek to understand the roles that lift and disclosure, two aspects of sponsored listings, play towards its impact. This is a yet unexplored topic related to sponsored advertising.

My work in this paper is also related to the literature on understanding the effects of advertising. I have tried to define the information role associated with advertising, as proposed by Chamberlin (1949), in a novel way so that it is relevant to the context of digital advertising. I consider consumer's awareness (information) as closely related to his/her attention span. In fact, the same may also be extended to consumer's visibility as the three are all closely related in a rank ordered context. This is because digital advertisements, especially sponsored listings, are designed to keep the product at consumer's eye-level i.e. to catch consumer eyeball. Consistent with literature in this area, e.g. Ackerberg (2001), Ackerberg (2003b), Anand and Shachar (2009), Narayanan, Manchanda, and Chintagunta (2005), etc, I allow for the possibility of sponsored advertising to have multiple mechanisms through which it is effective. While the information role of sponsored listings is related to its lift enhancement capacity, the role of advertising which could alter user's utility (persuasive role) is assumed to be related to the disclosure of advertising status.

Finally, this paper is related to the literature on consumer search behavior modeling, e.g. Bronnenberg, Kim, and Mela (2016), Kim, Albuquerque, and Bronnenberg (2010), etc. Two papers in this area that have investigated the effect of advertising along purchase funner, i.e. Seiler and Yao (2017) and Honka, Hortaçsu, and Vitorino (ming). The analysis in both these papers is however set in offline contexts of brick and mortar store advertising and financial industry advertising. In the online context, this is perhaps the first paper to use search model to explore the impact of advertising along various stages of consumer's decision making. Two other papers in this area, which are relevant due to their emphasis on learning are Koulayev (2013) and De los Santos, Hortaçsu, and Wildenbeest (2017). These are two of the earliest empirical papers in consumer search behavior which modeled for consumers updating (learning) the prior beliefs that they have. Just like this paper, both these papers do not explicitly model for the priors (non-parametric priors) hence they are search models of partial information (about distribution of product offerings).

3.3 Model

I develop a model with two stages, consideration and choice, where consumer's choice is modeled conditional on the consideration set. A common utility specification is used across the two stages, which is consistent with the findings of Bronnenberg, Kim, and Mela (2016) who state that "preferences revealed during search are highly similar to those revealed by choice". Consumers are modeled to have prior perceived quality for organic listings at each position. Consumers update their beliefs about the quality of a hotel based on the sponsored status of the hotel, i.e. sponsored status serves as a signal of quality.

3.3.1 Bayesian Updating

Consumer wants to learn the true quality of the hotel (state of nature), however the true quality will never be known with certainty. Consumers have prior perceptions of quality of the organic listing at each rank. It can be assumed that these priors are formed based on user's prior online shopping and browsing occasions.[8] While in this paper, I do not model the explicit learning of priors, I do assume data based priors, i.e. priors are assumed to be normal and are based on the actual distribution of the quality of organic listings at each rank (separate for every position).

$$q_{ij(r)} \sim \mathcal{N}(Q_{r0}, \sigma^2_{ir0}) \tag{3.1}$$

where $q_{ij(r)}$ denotes consumer i's prior perceived quality of hotel j at rank r. The prior in equation 3.13 accounts for the quality heterogeneity at the rank level through the rank specific mean quality. Further the priors are assumed to be unbiased i.e. the mean of the consumers' prior at rank r is the true mean quality at rank r. The consumer level heterogeneity over the priors is due to the variance σ^2_{ir0} and following Jeziorski and Segal (2015), I set it as $\sigma_{ir0} = \bar{\sigma} + \sigma_{\bar{\sigma}}\gamma_i$, where $\gamma_i \sim \mathcal{N}(0, 1)$.

Let $\omega_{ij(r)}$ be the noisy signal of quality associated with seeing a sponsored listing for property j at rank r.

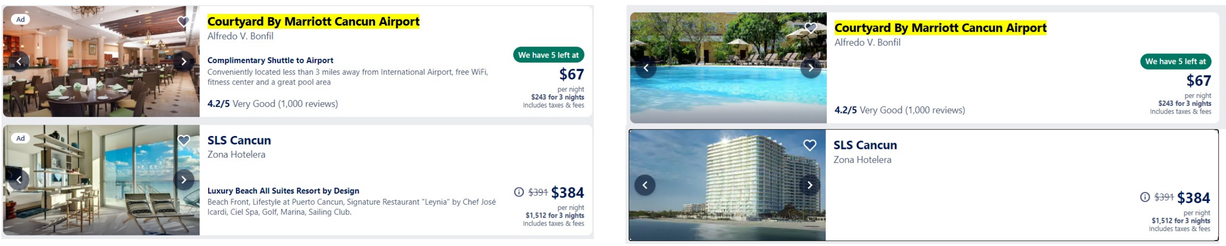
$$\omega_{ij} = Q_j + \nu_{ij} \tag{3.2}$$

Equation 3.2 shows that the signal is composed of hotel j's true quality, Q_j and noise $\nu_{ij} \sim \mathcal{N}(0, \sigma^2_\nu)$ which prevents consumer's learning to be perfect (and hence the consumer does not, in expectation, learn the true quality Q_j). Following Mehta, Rajiv, and Srinivasan (2003), I assume σ^2_ν to be the same across rank and hotels to ensure analytical tractability. Thus $\omega_{ij} \sim \mathcal{N}(Q_j, \sigma^2_\nu)$. Figure 3.1 shows the sponsored listings and the organic listings. Sponsored listings include all the details of organic listings and in addition they also display the product description associated, a caption to sum up the core hotel offering and choice of cover photograph (which for an organic listing is picked at random from the hotel's col-

[8]This is a reasonable assumption to make given that 70% all consumers do their research online (source: https://www.stratosjets.com/blog/online-travel-statistics/)

lection of photographs). As can be seen in the figure, the product description highlights the amenities offered by the hotel, amenities which are usually seen in the inner page (landing page after clicking). In the data, I see only the sponsored listing identifier. I do not see the product description, the ad caption or the photographs. However, I do have the data on the amenities associated with the hotel, which allows me to create a quality index (details of which are in the appendix 3.B.

Figure 3.1: Difference Between Sponsored and Organic Listings



Notes: Sponsored listings (left column) and the corresponding organic listings (right column). When viewed side by side we can see that sponsored listings permit additional description and the choice of a photograph.

Let $s_{ij(r)}$ be the indicator for the sponsored status, i.e. $s_{ij(r)} = 1$ implies listing j at rank r is sponsored. Using the results for conjugate prior distribution for normal distribution, the posterior perceived quality and variance are given as follows.

$$
\begin{aligned}
Q_{r1} &= \frac{Q_{r0}\alpha_{r0} + s_{ij(r)}\widehat{\omega}_{ij(r)}}{\alpha + s_{ij(r)}} \\
\alpha_{r1} &= \alpha_{r0} + s_{ij(r)} \\
\alpha_{r0} &= \frac{\sigma_\nu^2}{\bar{\sigma}^2}
\end{aligned}
\tag{3.3}
$$

where $\widehat{\omega}_{ij(r)}$ is the realized value of the noisy quality signal. α_{r1} can be thought of as the inverse of signal to noise ratio (Mehta, Rajiv, and Srinivasan, 2003) and hence a larger value implies that sponsored status is not very informative of the hotel quality. Unlike Mehta, Rajiv, and Srinivasan (2003), who do not observe the true quality, I am able to obtain a measure of the true quality of a listing a described in appendix 3.B.

The sequence of events is thus as follows:

(i) Consumers have prior perceptions of quality associated with the organic listings at various positions (or ranks) based on historical browsing,

(ii) Consumers observe the advertising status of listings, reads the product description and updates the prior beliefs about quality resulting in the posterior quality beliefs,

(iii) Consumer decides to either click or not as per Weitzman (1979) rules, and

(iv) If clicked, consumer observes the landing page of the listing. Landing page of the hotel listing does not provide the true hotel quality. If it were to do so, then consumer would know the quality of hotel exactly & unambiguously even before the stay experience. From a research question perspective this is important as otherwise the impact of sponsored listing cannot be tested at the choice stage.

3.3.2 Utility and Search Cost

The utility for consumer i, for hotel j is given by:

$$\begin{aligned} u_{ij} &= v_{ij} + \epsilon_{ij}; \\ v_{ij} &= x_{ij}\beta \\ \epsilon_{ij} &\sim N(0, \sigma_j^2) \end{aligned} \tag{3.4}$$

where I assume that the consumer is searching over ϵ_{ij}. Prior to the search, consumer knows v_{ij}, which is composed of consumers preference (β) for a vector of hotel attributes (x_{ij}). A more detailed representation of v_{ij} is given below. Note that the hotel attributes are consumer specific, i.e. attributes, such as price, number of reviews, etc, for the same hotel might vary based on the time at which a consumer views it. Due to healthy incidence of non-purchase (after search) in the data, I also include an outside option of no purchase. In the case of the outside option, only a fixed effect is estimated.

The mean utility component (v_{ij}) is represented as follows:

$$v_{ij} = X_{ij}\beta - \beta_p P_{ij} + \gamma q_{ij} \tag{3.5}$$

where X_{ij} includes the non price characteristics, P_{ij} is the price, q_{ij} is the perceived quality of the hotel j.

After Expedia returns the list of hotels in the default sort order, consumers click on a hotel to find out more about it [9]. For every click, consumer has an associated cost to it. Following Jeziorski and Segal (2015), I endogenize the CTR decline with rank by including, rank fixed effects in the search cost specification. As mentioned previously, one of the mechanisms through which rank effect is assumed to manifest is through the impact rank has on the search cost. The search cost specification is similar to what is used in Ursu (2018):

$$c_{ij} = exp(\gamma_r r_{ij} + \gamma_p pg_{ij}) \tag{3.6}$$

where r_{ij} is the position at which hotel is displayed, pg_{ij} is the page at which hotel is displayed, γ_r and γ_p are the rank and page coefficients respectively in cost specification. The exponential function is to ensure that the search cost is always positive and as stated in Ursu (2018) is standard way to handle search costs in the search literature.

[9] Consumers may also change the default sort order of the hotels. Only an extremely small percentage of users ever use such sorting and filtering options, so I did not incorporate such actions from consumers in the model. I also did away with users who choose to move away from default sort order.

3.3.3 Optimal Sequential Search Strategy

The rules which govern the optimal sequential search strategy are based on the reservation utility z_{ij}, defined as the utility that makes the consumer indifferent between the cost of carrying out an additional search (marginal cost) and the expected marginal benefit from an additional search (marginal benefit). Mathematically, it can be shown that:

$$c_{ij} = \int_{r_{ij}}^{\infty} (u_{ij} - r_{ij}) dF(u_{ij}) \tag{3.7}$$

where the integral on the right hand side is the marginal benefit from an additional search. The implicit value function in (3.7) can be used to back out the reservation utility (of every hotel) based on the closed form equation derived in Kim, Albuquerque, and Bronnenberg (2010) under the assumption that ϵ_{ij} has normal distribution. The closed form equation is as shown below:

$$z_{ij} = v_{ij} + \zeta_{ij}\sigma_j \tag{3.8}$$

where ζ_{ij} is obtained from the implicit function defined in 3.7:

$$\frac{c_{ij}}{\sigma_j} = (1 - \Phi(\zeta_{ij}))(\frac{\phi(\zeta_{ij})}{1 - \Phi(\zeta_{ij})} - \zeta_{ij}) \tag{3.9}$$

Under the setting specified above (3.4 through 3.9), optimal sequential search strategy implies the set of three rules put forward by Weitzman (1979). These rules require each consumer to first sort the hotels in a decreasing order of their reservation utilities ($z_{i1} > z_{i2} > z_{i3} > ... > z_{iJ}$)

1. Selection rule: Consumers search the alternatives in the decreasing order of the reservation utility, i.e. the reservation utility of the first searched alternative has to be greater than the reservation utilities of all the other searched options and the reserva-

tion utilities of all non-searched options. Thus for the nth search:

$$z_{in} \geq \max_{k=n+1}^{J} z_{ik} \;\; \forall\, n \in 1, ..., M \tag{3.10}$$

where M_i is the number of hotels searched by user i. Otherwise, the set of companies searched by the consumer would have been different.

2. Stopping rule: Consumers stop searching when the maximum realized utility from among the searched options exceeds the maximum reservation utility from among the un-searched option. For any searched hotel, this implies that its reservation utility must be greater than the maximum of the realized utilities of the hotels searched so far. For an un-searched hotel, this implies that its reservation utility is lower than the maximum of the realized utilities searched hotels. Thus for the nth search:

$$\begin{aligned} z_{in} &\geq \max_{k=1}^{n-1} u_{ik} \;\; \forall\, n \in 1, ..., M \\ z_{iq} &\leq \max_{k=1}^{M} u_{ik} \;\; \forall\, q \in M+1, ...J \end{aligned} \tag{3.11}$$

Otherwise, consumer would end up searching all the hotels

3. Choice rule: From among the searched hotels, consumer picks the hotel with the highest realized utility. Note that this also includes the utility of the outside option.

$$\begin{aligned} j^* &= \arg\max_{k=1}^{K} u_{ij} \;\; \forall\, K \in M \cup \{0\} \\ u_{ij^*} &= \max_{k=1}^{K} u_{ij} \;\; \forall\, K \in M \cup \{0\} \end{aligned} \tag{3.12}$$

Thus, as long as the marginal benefit of an additional search is positive, consumer continues to search the option with the highest reservation utility. When the consumer decides to stop the search, the searched option with the highest utility will be purchased.

3.4 Data and Estimation

3.4.1 Data & Analysis

For modeling and analysis, I use data from Expedia.com. The cross-sectional data comprises of 20,177 random user searches for hotels to stay, at specific destinations and includes 1.08 million observations. These searches include sponsored advertisements (listings) at a few ranks. While sponsored listings in the data appear at a variety of ranks, only those ranks with at least 50 sponsored listings across searches (cumulative in the data) have been considered for the analysis. Table 3.1 shows the details of these positions. Each page has 50 positions (when consumers use laptop or desktop to browse Expedia.com) and for the purpose of the rest of analysis, I have classified the ranks at which sponsored listings are deployed as those occurring at the top of the page (first 5 ranks on each page) and those that are not at the top (all other ranks). Such a classification allows the analysis to account for any position specific effects on quality updating due to sponsored listings. This was necessitated since interactions of advertising and rank are not feasible in the model. Such a classification of ranks within a page implies that the parameters of the model that govern the Bayesian updating of perceived quality need to be separately estimated for the sponsored listings at the top of the page and for those shown at other positions.

Another aspect of sponsored listings is that they come up at least twice in the rank ordered list (once as Sponsored and Organic). For example, the data has 55,813 properties which have been listed twice (dual listed) in the same search, once as sponsored and once as organic. In fact, every property that is listed as sponsored has a corresponding organic listing in the same search irrespective of whether the consumer browses up to it or not. This aspect of the data makes it unique and is at the heart of modeling the awareness/attention impact of sponsored listings in this paper.

Figure 3.2 compares the ranks of the dual listed properties (i.e. those that appear as sponsored and organic within the same search). The y-axis is the rank at which a property

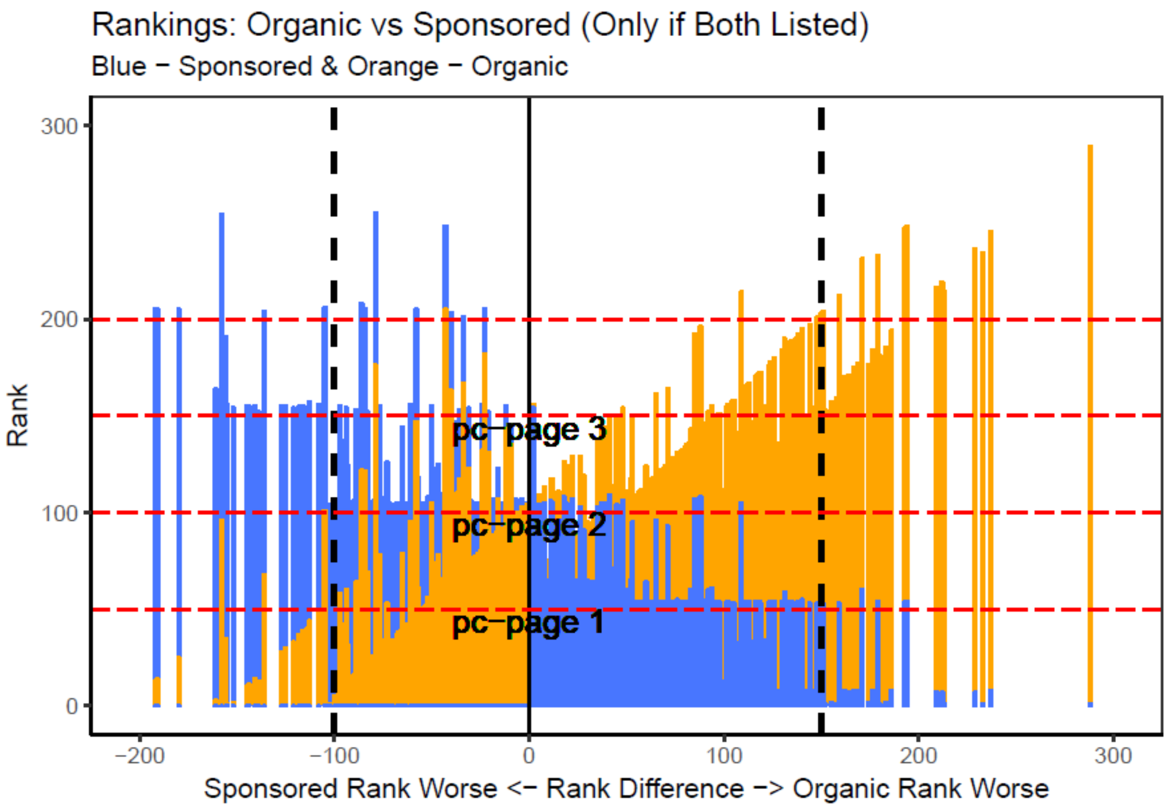
Table 3.1: **Count of Sponsored Listings at Each Rank/Position**

Position	Cumulative Sponsored Listings
1	9613
2	5460
6	51
7	11447
8	4512
12	60
13	3625
52	130
53	11150
54	9289
103	51
104	66
105	86

Notes: Table shows details of positions (or ranks) at which there are at least 50 sponsored listings in cumulative across searches. These ranks, based on the frequency of sponsored listings, have been identified as the ranks which Expedia has earmarked for sponsored ads

is displayed and the x-axis is the difference between the organic and sponsored rank of a listing. A negative value implies that organic listing is ranked better (higher on the page) than sponsored listing and a positive value implies that sponsored listing is ranked better than organic listing. For example, if the organic rank of a listing is 35 and the rank at which it shows up as sponsored is 2, then the difference is 33; however if the organic and sponsored ranks are 20 and 50 respectively, then the difference is -30. In the figure below, the properties to the left of the solid black line at $x = 0$ have negative rank difference (better rank for organic listings). For each of the 55,813 properties, the height of blue bar shows the rank of sponsored listing and the height of orange bar shows the rank for the organic listing corresponding to the sponsored listing, i.e. the sponsored blue bar and organic orange bar are pairs. The taller blue bars and shorter orange bars are to the left and vice-versa to the right. The dashed horizontal red lines in the figure are to demarcate pages (on Laptops). Each page on a laptop browser has 50 properties listed on it.

Figure 3.2: Sponsored Listings Provide Lift & Prominent Ranks



Notes: Figure compares the ranks of the dual listed properties (sponsored and organic). The y-axis is the rank at which a property is displayed and the x-axis is the difference between the rank of organic listing and sponsored listing. A negative value implies that organic listing is ranked better (small absolute value of rank) than sponsored listing and a positive value implies that sponsored listing is ranked better than organic listing. The properties to the left of the solid black line at $x = 0$ have negative rank difference (better rank for organic listings). For every property, the height of blue bar shows the rank of sponsored listing and the height of orange bar shows the corresponding value for the organic listing (taller blue bars and shorter orange bars to the left and vice versa to the right). The dashed horizontal red lines in the figure are to demarcate pages (on Laptops). Each page on a laptop browser has 50 properties listed on it.

Table 3.2: Summary Statistics of the Pattern of Key Interest

Rank Difference	**<-100**	**[-100,0)**	**(0,150]**	**>150**	**Overall**
CTR	.179	.100	.106	.070	.104
Transactions	.036	.064	.058	0	.060
Branded	.786	.772	.795	.953	.786
Price	226.91	248.75	280.73	212.52	268.39

Notes: Tables shows the summary statistics in each of the 4 regions in figure 3.2

This figure shows that the lift can be positive, e.g. organic rank 70 to sponsored rank 2 implies a lift of 68, or it can be negative, e.g. organic rank 45 to sponsored rank 52 implies a lift of -7. Properties with better organic rank, i.e. with negative rank difference value, may at first seem like a mistake by the advertiser or a loss in the bidding process. However, this is not the case as has been confirmed by the data.[10]. A quick look at the rank of the sponsored listing (along the dashed red lines) shows that such sponsored listings are strategically placed (at the top of the page) to target consumers with *low-search costs*, i.e. those who search beyond page 1. The attention span of consumers is likely to be higher at the top of the subsequent page rather than the bottom of the current page.

The figure also has two black vertical dashed lines used to segregate properties which may be outliers. There are only 28 properties to the left of the first one (from left) and 43 to the right of the second dashed line. Table 3.2 shows a few key summary statistics for each of the four regions (vertical blocks) of the figure.[11] Using a structural model of consumer search, makes it ideal to study the the impact of sponsored listings on consumer attention/visibility, through the search cost specification. Making the search cost a function of both page and position (within a page), allows the model to capture the precise impact of sponsored listings on awareness/attention stage.

3.4.2 Estimation

I use the simulated maximum likelihood, described below, to estimate the parameters of the model. The coding was done using R programming language and the following parameters were estimated:

1. Preference parameters (β's) for hotel attributes (in the utility)

[10]This has also been confirmed independently by industry practitioners.

[11]The higher than average CTR in the first block could either be due to the low number of properties or it could be due consumer trust in better ranked organic listings.

(a) Price

(b) Star rating

(c) Average user rating

(d) Number of reviews

(e) Chain affiliation of a hotel (Yes/No)

(f) Discount (if any)

(g) Perceived quality

(h) In the data, there are instances of consumers making no purchase even after clicking. To account for this preference parameter for outside option is also estimated.

2. Parameters for the inverse of signal to noise ratio in the sponsorship signal

   (a) α_t, for sponsored listings shown in the positions (ranks) at the top of the page

   (b) α_{nt}, for sponsored listings shown in the rest of the positions (ranks) on the page

3. Search cost parameters

   (a) Effect of position (rank) on search cost

   (b) Effect of page on search cost

The probability of observing a consumer search a set of companies γ_i and purchase from company j (including outside option) under sequential search is given by the joint probability of the three conditions (equations 3.10, 3.11 and 3.12) in section 3.3.3.

$$
\begin{aligned}
L_i &= P_{ij\gamma_i} = Prob(z_{in} \geq \max_{k=n+1}^{J} z_{ik} \cap z_{in} \geq \max_{k=1}^{n-1} u_{ik} \cap u_{ij^*} \geq \max_{k=1}^{K} u_{ij}) \\
L_i &= \int_{-\infty}^{\infty} \mathcal{I}(z_{in} \geq \max_{k=n+1}^{J} z_{ik} \cap z_{in} \geq \max_{k=1}^{n-1} u_{ik} \cap u_{ij^*} \geq \max_{k=1}^{K} u_{ij}) f(\epsilon) d\epsilon
\end{aligned} \tag{3.13}
$$

The model likelihood is given by

$$L = \Pi_{i=1}^{N} L_i \tag{3.14}$$

The inter-relationship between search and purchase decisions, i.e. purchase conditioned on the consideration set, which in turn is endogenously determined by the search process, implies that the integral in equation 3.13 does not have a closed form solution. To overcome this challenge, simulated maximum likelihood estimation is used to estimate the sequential search model. In this approach, the random shocks (ϵ), which the consumers observe but which the researcher does not observe, need to be integrated out using simulations. These simulated probabilities would be non-smooth and would require non-gradient based optimization methods (McFadden (1989)). To avoid this, I use the scaled multivariate logistic CDF to smooth the probabilities. This logit smoothed accept reject simulation method has been used to estimate search models (Honka and Chintagunta (2017), Ursu (2018)). The details of the kernel-smoothed frequency simulator are as follows:

1. Take $d = 1, ..., D$ draws for ϵ_{ij} i.e. a total of $N * J * D$ draws
2. For each of the ϵ_{ij} draws (dth draw), compute indirect utility u_{ij}^d, search cost and the reservation utilities[12]
3. Compute the following entities (Weitzman's three rules):
   (a) $w_1^d = z_{ij} - \max_{k=n+1}^{J} z_{ik}$ (selection rule)
   (b) $w_2^d = z_{in} - \max_{k=1}^{n-1} u_{ik}$ (stopping rule)
   (c) $w_3^d = u_{ij^*} - \max_{k=1}^{K} u_{ij}$ (choice rule)
4. Use the logit smoothed Accept Reject probabilities $P_{ij}^d = \frac{1}{1+\sum_{m=1}^{3} -sw_m^q}$, where s is the scaling factor

[12]Search cost and reservation utilities are not dependent on the epsilon draws

5. Integrate over the ϵ_{ij} distribution by taking average of the probabilities $P_{ij} = \frac{1}{D}\sum_{d=1}^{D} P_{ij}^{d}$

To get a consistent estimate of the parameters, I generated 10 sets of the posterior quality draws in the above process. The choice of the the scaling parameter is based on trial and error using Monte Carlo simulations.

3.4.3 Identification

The preference parameters, which include β's and the outside option indicator, are identified based on the correlation between the frequency of click & purchase and product characteristics for which we are estimating the parameter as well as the assumed click sequence. The identification strategy for these parameters is analogous to that of any typical discrete choice model. Within the optimal sequential search strategy equations 3.10 through 3.12 capture these correlations. These equations include conditions on reservation utilities and utilities i.e. all three of Weitzman's rules play a part in the identification process.

The point estimate of the mean search cost (intercept) parameter cannot be identified in the presence of the "statistical error term" (e.g. Mehta, Rajiv, and Srinivasan (2003)). Giving up the estimation of this time invariant search cost, does not affect the answers to the research questions of interest. This is because, to get an estimate of the impact of sponsored listings on awareness/attention phase, we only need the difference of search costs (between the sponsored position and the organic position) and the fixed cost term would not matter. The other parameters of interest are the inverse signal to noise parameters (α's), which are also not impacted by the mean search cost. The position (rank) and page specific search costs are identified based on the differences in the frequencies for search and purchase at each rank and page respectively. For example, products clicked frequently but not purchased have low search costs, and are also low on the utility parameters as well.

3.5 Results and Discussion

The search model parameter estimates are shown in table 3.3. All the parameters are statistically significant and along the expected direction. In the model specification, the search cost is affected only by the relative position of a listing and not by any product attribute. Attributes (except the sponsored status) affect consumers utility, and therefore the consumer's decision regarding the size & composition of the consideration set and choice probabilities.

A listing's sponsored status (and the position/rank at which it occurs) provides the consumer a noisy signal of the product's quality which is used to update consumer's priors in a Bayesian fashion.

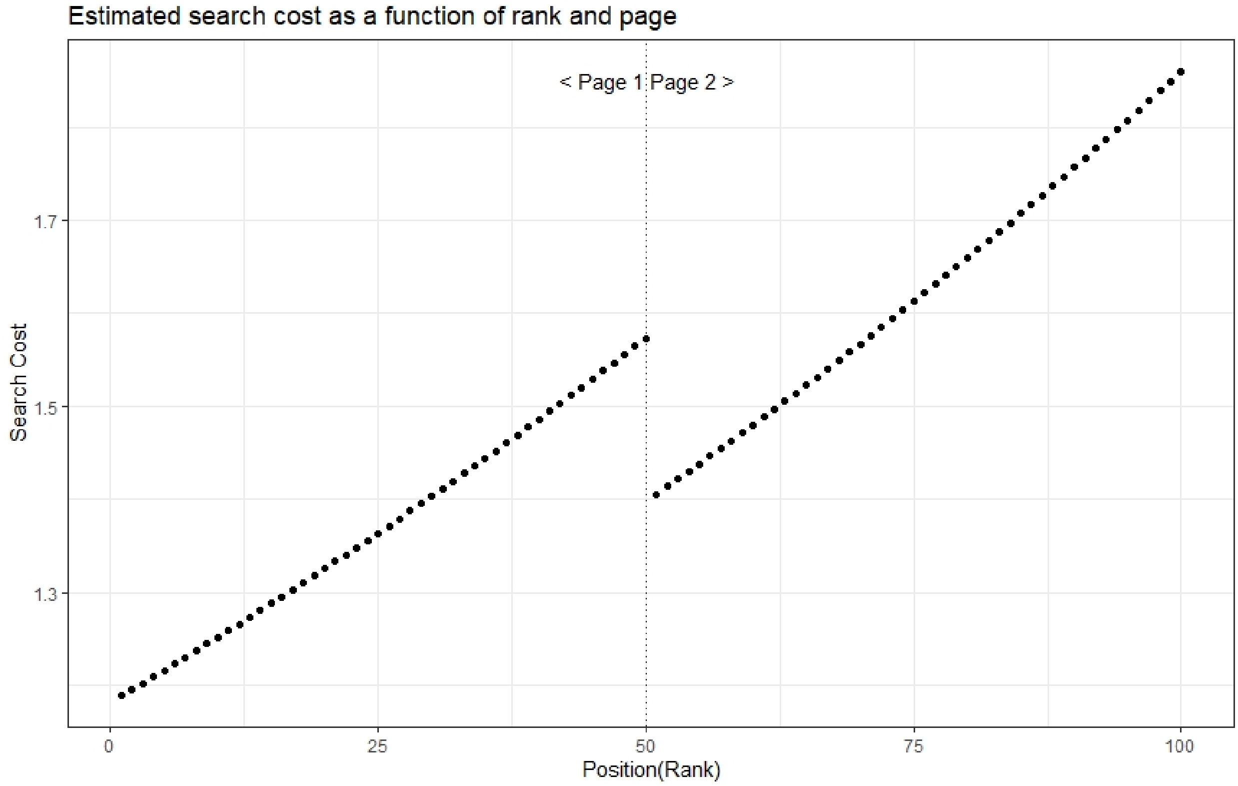
The model's estimate of the search cost is depicted in figure 3.3. The estimates capture consumer's attention span quite well. The search cost is increasing in the position/rank along a page, i.e. search cost is lowest at the top of a page where consumer attention is the highest. As the consumer browses down the page, the attention span decreases & hence the search cost is increasing. Further, as consumer end the browsing on a page and proceeds to the next page, the attention improves which implies that the search cost is lower at the top of the next page. This is consistent with the *primacy* effect in behavioral literature. The average difference in the search cost of the sponsored listing and its corresponding organic listing is 0.0973, which is the equivalent of an average lift of 17.05 positions (ranks). It is interesting to note that the corresponding average lift in the raw data is only 15.25 positions (ranks). Accounting for the improved consumer attention span, on the subsequent page, is clearly important.

Table 3.3: Model Estimation Results

Parameter	*Estimate*
Search Cost	
Position	0.0057***
	(0.0001)
Page	0.1674***
	(0.0004)
Sponsored Signal	
α_t	1.8367***
	(0.0002)
α_{nt}	4.1171***
	(0.0001)
Utility	
Price	-0.0398***
	(0.0002)
Star Rating	0.0765***
	(0.0001)
User Rating	0.0610***
	(0.0014)
Review Count	0.0015***
	(0.0001)
Chain Affiliation	0.0803***
	(0.0023)
Promotion Flag	-0.2593***
	(0.0015)
Outside Option	0.0578***
	(0.0016)
Perceived Quality	0.0356***
	(0.0034)
Log-likelihood	-82143.71
Observations	1,076,821

Notes: Parameter estimates for the search model. Standard errors are in parentheses. Note: *p<0.1; **p<0.05; ***p<0.01

Figure 3.3: Search Cost Plot



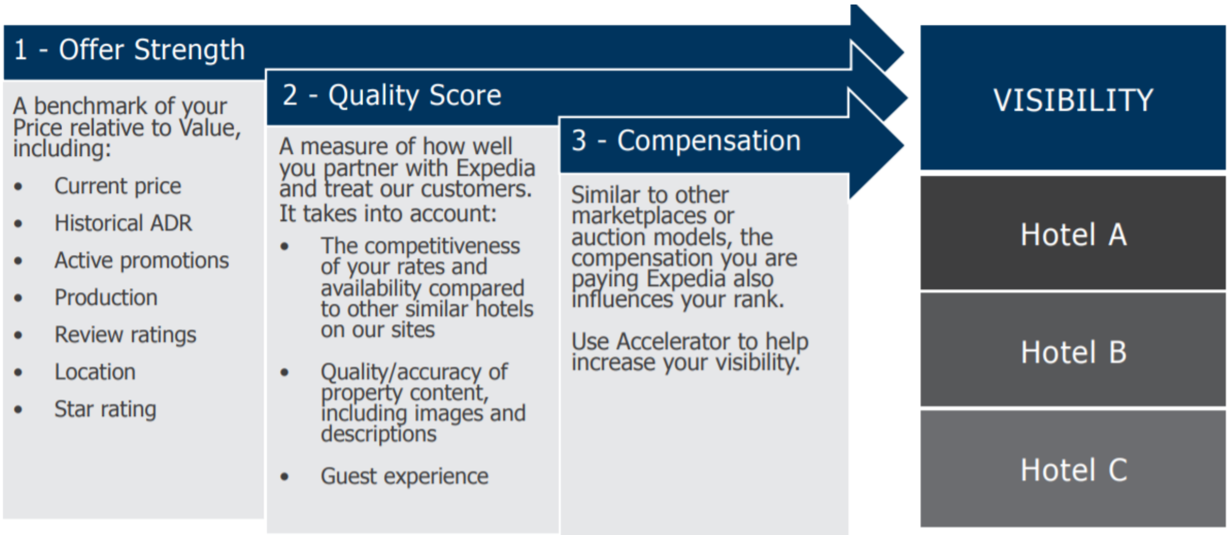
Notes: Figure shows the plot of the search cost. Search cost is increasing with the rank of a page and consistent with primacy effect, it falls at the top of the next page. The dashed line at rank 50 represents the page demarcation.

The impact of sponsored status on consumer learning of quality, which affects the consideration and choice is captured through the α_t and α_{nt} parameter estimates. A high value implies a low variance in prior quality beliefs of the consumers. Thus, a high value implies that quality learning through the sponsored signal will be low Mehta, Rajiv, and Srinivasan (2003). The estimates are $\alpha_t = 1.83$ and $\alpha_{nt} = 4.11$. This means that at the top ranks the α value is low and the extent of quality learning due to sponsored status is high. However, as consumers are exposed to the sponsored signal further down the rank ordering on a page, the the extent of quality learning due to sponsored status is low, not only in absolute terms but also relative to the sponsored listings at the top of the page. This leads to the conclusion that sponsorship disclosure at the top of the page has the most impact on the consideration and the choice probability of the listing.

From a managerial point of view, it is evident that the sponsored status of a listing has an impact on the lift. The search cost function and the corresponding search cost differences between organic & sponsored listings attest to these facts. However, the disclosure of the sponsored nature of a listing, at the top of the page, makes advertising strategically beneficial by improving the listings consideration & choice probability. Sponsored advertising down the page may help listings enter consumer's visibility/awareness zone, which is tactically useful, but the impact on consideration and choice probability is muted.

3.A OTA Ranking Basis

Figure 3.4: Expedia Ranking Algorithm Drivers



Adapted from "Understanding the science behind Expedia's marketplace: What drives hotel visibility online." by Melissa Maher.

Figure 3.5: Factors that Most Impact Rankings on Booking.com.

"Even though there's no single determining factor, conversion (i.e. turning lookers into bookers) is important, and improving your conversion rate starts with being seen by more potential guests. You can learn more about the basics of visibility and find out which tools boost your visibility."

Source: https://partner.booking.com/en-us/help/growing-your-business/all-you-need-know-about-ranking-search-results-and-visibility

3.B Estimating the true quality index

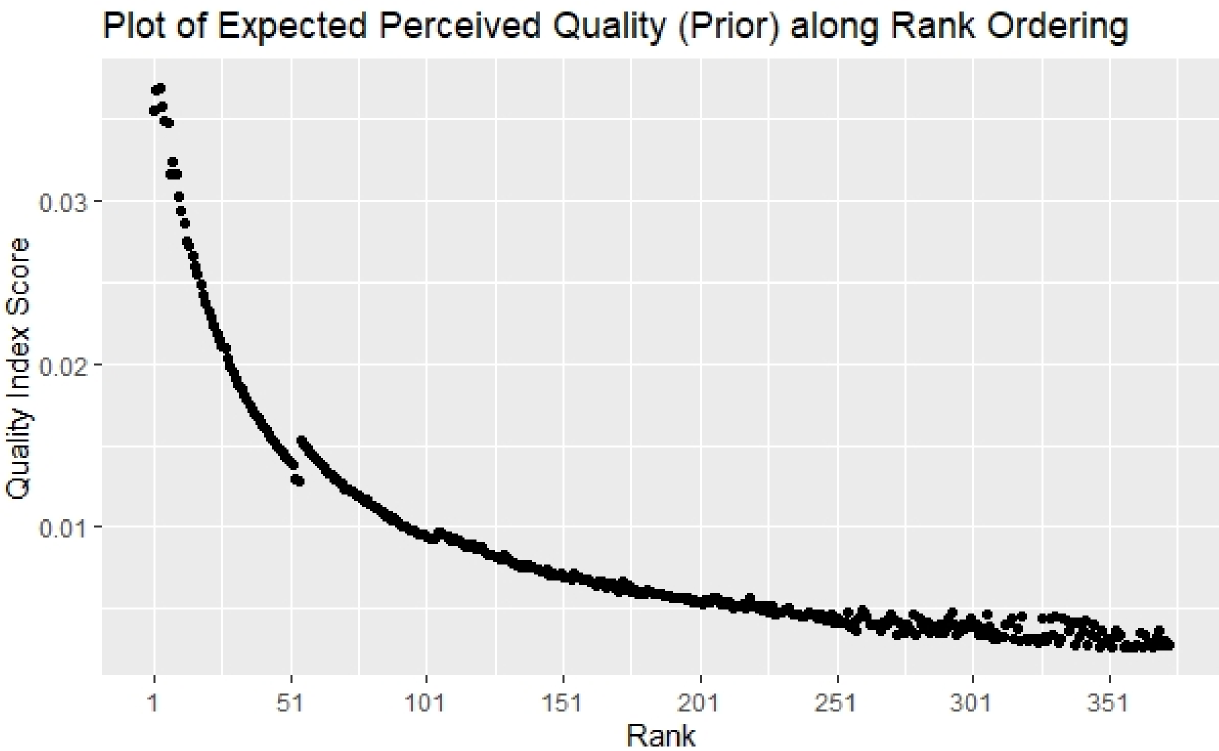
The default positions assigned by the OTA, also referred to as organic ranks, are based on broadly two measures - quality of the hotel and business considerations of the OTA. Business considerations include quality correlated attributes, e.g. profitability potential of the hotel[13] and non-quality correlated attributes, e.g. OTA power enhancing considerations (Hunold, Kesler, and Laitenberger (2020)). Thus the default rank order in which any hotel is shown by the OTA is likely to be correlated on unobserved (by the econometrician) hotel quality, as these rank orderings take into account past hotel performance, preference by consumers, etc. If I were to model the rank of hotel for each search id as:

$$R_{ijt} = X_{jt}\beta_j + Z_{ijt}\delta_j + \nu_{ijt}$$

Where X_{jt} are observed hotel characteristics. Z_{ijt} is price sorted rank of the hotel, which will be correlated with unobserved hotel characteristics since price is correlated with unobserved hotel characteristics. The residuals from this model, ν_{ijt}, include the variation in the rank due to unobserved (to the researcher) hotel quality and due to business considerations. These residuals can thus serve as a proxy for the hotel quality.

[13]A higher quality hotel is likely to have a higher profitability potential.

Figure 3.6: Quality index.



Since we expect that Expedia.com would rank based on the quality of the listing hotels, we expect the quality be highest at low ranks and that the quality would decrease with rank. Figure 3.6 shows the quality index that is obtained as per the steps described previously. The quality index is along the expected lines, i.e. the listings at lower ranks have better quality.

Bibliography

Abhishek, Vibhanshu, Kinshuk Jerath, and Siddhartha Sharma (2019), "Advertising on online marketplaces: Information asymmetry and the relevance of sponsored listings," *Available at SSRN 3013468*.

Ackerberg, Daniel A (2001), "Empirically Distinguishing Informative and Prestige Effects of Advertising," *RAND Journal of Economics*, 32 (2), 316–333.

— (2003a), "Advertising, learning, and consumer choice in experience good markets: an empirical examination," *International Economic Review*, 44 (3), 1007–1040.

— (2003b), "Advertising, Learning, and Consumer Choice in Experience Good Markets: An Empirical Examination," *International Economic Review*, 44 (3), 1007–1040.

Anand, B. and Ron Shachar (2009), "Targeted Advertising as Signal," *Quantitative Marketing and Economics*, 7 (3), 237–266.

Anand, Bharat N and Ron Shachar (2011), "Advertising, the matchmaker," *The RAND Journal of Economics*, 42 (2), 205–245.

Anderson, Simon P and R 'e gis Renault (2005), "Descriminating pricing," *Transport pricing: challenges and challenges, Economica*, 107–151.

Animesh, Animesh, Vandana Ramachandran, and Siva Viswanathan (2007), "An empirical investigation of the performance of online sponsored search markets," in "Proceedings of the ninth international conference on Electronic commerce," ACM, 153–160.

Arbatskaya, Maria (2007), "Ordered search," *The RAND Journal of Economics*, 38 (1), 119–126.

Aribarg, Anocha and Eric M Schwartz (2018), "Native advertising in online news: Tradeoffs between clicks and brand recognition," *Available at SSRN 2995467*.

Armstrong, Mark (2017), "Ordered consumer search," *Journal of the European Economic Association*, 15 (5), 989–1024.

Bagwell, Kyle (2007), "The Economic Analysis of Advertising," in Mark Armstrong and Robert H. Porter (Editors), "The Handbook of Industrial Organization," New York: North Holland, volume 3, 1701–1844.

Beel, Joeran, Stefan Langer, and Marcel Genzmehr (2013), "Sponsored vs. organic (research paper) recommendations and the impact of labeling," in "International Conference on Theory and Practice of Digital Libraries," Springer, 391–395.

Blake, Thomas, Chris Nosko, and Steven Tadelis (2015), "Consumer heterogeneity and paid search effectiveness: A large-scale field experiment," *Econometrica*, 83 (1), 155–174.

Bronnenberg, Bart J., Jun B. Kim, and Carl F. Mela (2016), "Zooming in on Choice: How Do Consumers Search for Cameras Online?" *Marketing Science*, 35 (5), 693–712.

Chamberlin, Edward Hastings (1949), *Theory of monopolistic competition: A re-orientation of the theory of value*, Oxford University Press, London.

Chen, Xiangjing Olivia, Yimin Wang, and Scott Webster (2021), "Open Slots: An Economics Analysis of Sponsored Listings on E-tailing Platforms," *Available at SSRN 3785034*.

Chen, Yuxin and Song Yao (2016), "Sequential search with refinement: Model and application with click-stream data," *Management Science*.

Choi, Hana and Carl F Mela (2019), "Monetizing online marketplaces," *Marketing Science*, 38 (6), 948–972.

Dai, Daisy and Michael Luca (2016), "Effectiveness of paid search advertising: Experimental evidence," *Harvard Business School NOM Unit Working Paper*, (17-025).

De los Santos, Babur, Ali Hortaçsu, and Matthijs R Wildenbeest (2017), "Search with learning for differentiated products: Evidence from e-commerce," *Journal of Business & Economic Statistics*, 35 (4), 626–641.

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

Edelman, Benjamin and Duncan S Gilchrist (2012), "Advertising disclosures: Measuring labeling alternatives in internet search engines," *Information Economics and Policy*, 24 (1), 75–89.

Erdem, Tülin and Michael P Keane (1996), "Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets," *Marketing science*, 15 (1), 1–20.

Ghose, Anindya, Avi Goldfarb, and Sang Pil Han (2012), "How is the mobile Internet different? Search costs and local activities," *Information Systems Research*, 24 (3), 613–631.

Ghose, Anindya and Sha Yang (2009), "An empirical analysis of search engine advertising: Sponsored search in electronic markets," *Management science*, 55 (10), 1605–1622.

Hollenbeck, Brett (2017), "The economic advantages of chain organization," *The RAND Journal of Economics*, 48 (4), 1103–1135.

— (2018), "Online reputation mechanisms and the decreasing value of chain affiliation," *Journal of Marketing Research*, 55 (5), 636–654.

Hollenbeck, Brett, Sridhar Moorthy, and Davide Proserpio (2019), "Advertising strategy in the presence of reviews: An empirical analysis," *Marketing Science*, 38 (5), 793–811.

Hong, Han and Matthew Shum (2006), "Using Price Distributions to Estimate Search Costs," *RAND Journal of Economics*, 37 (2), 257–275.

Honka, Elisabeth and Pradeep K. Chintagunta (2017), “Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry,” *Marketing Science*, 36 (1), 21–40.

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino (forthcoming), “Advertising, Consumer Awareness, and Choice: Evidence from the U.S. Banking Industry,” *RAND Journal of Economics*.

Hotchkiss, Gord (2004), “Inside the mind of the searcher,” *Enquiro Search Solutions Inc, Kelowna, available at: www. enquiro. com/Downloads/FileDownload. aspx*.

Hotchkiss, Gord, M Garrison, and S Jensen (2005), “Search engine usage in North America,” *White paper, Enquiro, Kelowna, BC, Canada*.

Hunold, Matthias, Reinhold Kesler, and Ulrich Laitenberger (2020), “Rankings of online travel agents, channel pricing, and consumer protection,” *Marketing Science*, 39 (1), 92–116.

Jang, Sungha, Ashutosh Prasad, and Brian T Ratchford (2012), “How consumers use product reviews in the purchase decision process,” *Marketing letters*, 23 (3), 825–838.

Jansen, Bernard and Amanda Spink (2007), “The effect on click-through of combining sponsored and non-sponsored search engine results in a single listing,” in “Proceedings of the 2007 Workshop on Sponsored Search Auctions, WWW Conference,” .

Jansen, Bernard J (2007), “The comparative effectiveness of sponsored and nonsponsored links for Web e-commerce queries,” *ACM Transactions on the Web (TWEB)*, 1 (1), 3.

Jansen, Bernard J and Marc Resnick (2006), “An examination of searcher’s perceptions of nonsponsored and sponsored links during ecommerce Web searching,” *Journal of the American Society for information Science and Technology*, 57 (14), 1949–1961.

Jeziorski, Przemyslaw and Ilya Segal (2015), "What makes them click: Empirical analysis of consumer demand for search advertising," *American Economic Journal: Microeconomics*, 7 (3), 24–53.

Joo, Mingyu, Jiaqi Shi, and Vibhanshu Abhishek (2021), "Seller Incentives in Sponsored Product Listings on Online Marketplaces," *Available at SSRN 3896716*.

Kihlstrom, Richard E and Michael H Riordan (1984), "Advertising as a Signal," *journal of Political Economy*, 92 (3), 427–450.

Kim, Jun B., Paulo Albuquerque, and Bart J. Bronnenberg (2010), "Online Demand under Limited Consumer Search," *Marketing Science*, 29 (6), 1001–1023.

Kirmani, Amna (1990), "The effect of perceived advertising costs on brand perceptions," *Journal of consumer research*, 17 (2), 160–171.

Kirmani, Amna and Peter Wright (1989), "Money talks: Perceived advertising expense and expected product quality," *Journal of consumer research*, 16 (3), 344–353.

Koulayev, Sergei (2013), "Search with dirichlet priors: estimation and implications for consumer demand," *Journal of Business & Economic Statistics*, 31 (2), 226–239.

Long, Fei, Kinshuk Jerath, and Miklos Sarvary (2018), "The Informational Role of Sponsored Advertising on Online Retail Marketplaces."," .

Ma, Zhongming, Xin Liu, and Tarique Hossain (2013), "Effect of sponsored search on consumer trust and choice," *International Journal of Electronic Business Management*, 11 (4), 227–237.

Marable, Leslie (2003), "False oracles: consumer reaction to learning the truth about how search engines work," *Results of an Ethnographic Study*.

McFadden, Daniel (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57 (5), pp. 995–1026.

Mehta, Nitin, Surendra Rajiv, and Kannan Srinivasan (2003), "Price Uncertainty and Consumer Search: A Structural Model of Consideration Set Formation," *Marketing Science*, 22 (1), 58–84.

Moorthy, Sridhar and Scott A Hawkins (2005), "Advertising repetition and quality perception," *Journal of Business Research*, 58 (3), 354–360.

Moorthy, Sridhar and Hao Dong Zhao (2000), "Advertising Spending and Perceived Quality," *Marketing Letters*, 11, 221–233.

Narayanan, Sridhar, Puneet Manchanda, and Pradeep K. Chintagunta (2005), "Temporal Differences in the Role of Marketing Communication in New Product Categories," *Journal of Marketing Research*, 42 (3), 278–290.

Nelson, Phillip (1970), "Information and consumer behavior," *Journal of political economy*, 78 (2), 311–329.

— (1974), "Advertising as information," *Journal of political economy*, 82 (4), 729–754.

Petrin, Amil and Kenneth Train (2010), "A Control Function Approach to Endogeneity in Consumer Choice Models," *Journal of Marketing Research*, 47 (1), 3–13.

Rutz, Oliver J and George F Watson (2019), "Endogeneity and marketing strategy research: An overview," *Journal of the Academy of Marketing Science*, 47 (3), 479–498.

Sahni, Navdeep S and Harikesh S Nair (2019), "Sponsorship disclosure and consumer deception: Experimental evidence from native advertising in mobile search," *Marketing Science*.

— (2020), "Does advertising serve as a signal? Evidence from a field experiment in mobile search," *The Review of Economic Studies*, 87 (3), 1529–1564.

Seiler, Stephan and Song Yao (2017), "The impact of advertising along the conversion funnel," *Quantitative Marketing and Economics*, 15 (3), 241–278.

Sharma, Siddhartha and Vibhanshu Abhishek (2017), "Effect of Sponsored Listings on Online Marketplaces: The Role of Information Asymmetry," *Available at SSRN 3013468*.

Taylor, Wayne and Brett Hollenbeck (2021), "Leveraging loyalty programs using competitor based targeting," *Quantitative Marketing and Economics*, 19 (3), 417–455.

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 (4), 530–552.

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

Wooldridge, Jeffrey M (2010), *Econometric analysis of cross section and panel data*, MIT press.

Yang, Sha and Anindya Ghose (2010), "Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence?" *Marketing science*, 29 (4), 602–623.