



Marketing Science

Publication details, including instructions for authors and subscription information:
<http://pubsonline.informs.org>

Monetizing Online Marketplaces

Hana Choi, Carl F. Mela

To cite this article:

Hana Choi, Carl F. Mela (2019) Monetizing Online Marketplaces. Marketing Science 38(6):948-972. <https://doi.org/10.1287/mksc.2019.1197>

Full terms and conditions of use: <https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions>

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2019, INFORMS

Please scroll down for article—it is on subsequent pages



With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.

For more information on INFORMS, its publications, membership, or meetings visit <http://www.informs.org>

Monetizing Online Marketplaces

Hana Choi,^a Carl F. Mela^b

^a Simon School of Business, University of Rochester, Rochester, New York 14627; ^b Fuqua School of Business, Duke University, Durham, North Carolina 27708

Contact: hana.choi@rochester.edu,  <https://orcid.org/0000-0001-6632-6781> (HC); mela@duke.edu,

 <https://orcid.org/0000-0001-7893-6966> (CFM)

Received: August 11, 2016

Revised: August 22, 2018; April 26, 2019

Accepted: July 10, 2019

Published Online in Articles in Advance:
November 14, 2019

<https://doi.org/10.1287/mksc.2019.1197>

Copyright: © 2019 INFORMS

Abstract. This paper considers the monetization of online marketplaces. These platforms trade off fees from advertising with commissions from product sales. Although featuring advertised products can make search less efficient (lowering transaction commissions), it incentivizes sellers to compete for better placements via advertising (increasing advertising fees). We consider this trade-off by modeling both sides of the platform. On the demand side, we develop a joint model of browsing (impressions), clicking, and purchase. On the supply side, we consider sellers' valuations and advertising competition under various fee structures (cost-per-mille, cost-per-click (CPC), and cost-per-action) and ranking algorithms. Using buyer, seller, and platform data from an online marketplace where advertising dollars affect the order of seller items listed, we explore various product-ranking and ad-pricing mechanisms. We find that sorting items below the fifth position by expected sales revenue while conducting a CPC auction in the top 5 positions yields the greatest improvement in profits (181%) because this approach balances the highest valuations from advertising in the top positions with the transaction revenues in the lower positions.

History: K. Sudhir served as the senior editor and Tat Chan served as associate editor for this article.

Supplemental Material: Data and the online appendix are available at <https://doi.org/10.1287/mksc.2019.1197>.

Keywords: online marketplaces • e-commerce • online advertising • sequential search model • dynamic discrete choice model

1. Introduction

1.1. Overview

With buyers on one side and third-party merchants on the other, online marketplaces are a two-sided platform of substantial economic importance. The market capitalization of Alibaba, the world's largest online marketplace, was around \$481 billion in the first quarter of 2019, and the market capitalization of Amazon, the largest online retailer in the United States, was over \$910 billion.¹ In 2018, 52% of units on Amazon were sold by third-party sellers, generating \$42.75 billion, up from \$31.88 billion in the previous years.² An estimated \$1.86 trillion was transacted on the top 100 online marketplaces around the world in 2018.³ With the rise in mobile shopping, online marketplaces are expected to continue this rapid growth in coming years.⁴

Online marketplaces' revenue models are built upon several different fee types, including fees charged to merchants (i) for impressions delivered to the consumers by the platform (or cost-per-mille; CPM), (ii) for clicks made by the consumers (or cost-per-click; CPC), or (iii) per completed merchant transaction (or cost-per-action; CPA). For example, commissions on sales are a common form of CPA wherein merchants are usually charged fees per item sold as a percentage of the total sale amount, and these fees vary between

6% ~ 25% depending on the platform and categories.⁵ Advertising fees are commonly charged based on CPM or CPC pricing.⁶ Marketplace platforms commonly consider their product display ranking algorithm in conjunction with the fee types because listing order affects both advertiser and consumer sources of revenue. For example, ranking items from low to high price can lower CPA fees if consumers substitute lower-price goods but can raise CPC fees if more clicks are generated.

In spite of the growth in online marketplace platforms, research is limited regarding their fee structure and ranking strategies. Accordingly, this paper considers the monetization of advertising and sales in the context of online marketplaces by considering (i) how product-ranking decisions affect consumers' browsing (i.e., the impressions that can be monetized via CPM), consideration (i.e., the clicks that can be monetized by CPC), and choice (which affects monetization via CPA); and (ii) how ranking algorithms as well as fee structure (i.e., CPM, CPC, and CPA) affect sellers' advertising decisions and platform profits.

Toward answering these two questions, we develop a joint model of (i) consumer impressions, clicks, and purchases; and (ii) sellers' advertising competition, where advertising behaviors take consumers' search (browsing to become aware of items and clicking to

consider them) and choice (purchase) into account. Because of the interdependency across both sides of the platform (advertising can make search inefficient, thereby lowering consumer sales), a complete accounting of platform monetization requires the joint consideration of both consumer and advertiser behaviors. Thus, to address our research objectives, we develop a joint model of consumer and advertiser behaviors on online marketplaces. Next, we discuss relevant research pertaining to both sides of the platform and how our model builds on those foundations.

1.2. Relevant Research

1.2.1. Consumer Behavior. Although there is a prolific literature on consumer search in marketing and economics (e.g., Stigler 1961, Weitzman 1979, Mehta et al. 2003, Hong and Shum 2006, Kim et al. 2010, De los Santos et al. 2012, Seiler 2013, Honka 2014, Koulayev 2014, Bronnenberg et al. 2016, Honka and Chintagunta 2016, Chen and Yao 2017, Honka et al. 2017, Moraga-González et al. 2018, Ursu 2018), our paper builds on ordered search theory (e.g., Arbatskaya 2007, Armstrong et al. 2009, Wilson 2010, Armstrong and Zhou 2011, Zhou 2011) by considering the case when the search order is influenced by sellers' advertising decisions and the order of items presented to the consumer is predetermined by an intermediary platform. We extend the optimal stopping problem framework therein (e.g., Zhou 2011) to our empirical context by accommodating selective clicking decisions, differential information revealed at browsing and clicking stages, and consumers' expectations about the ordering of items arising from the platform's ranking algorithm.

Like Mojir and Sudhir (2019) and Chan and Park (2015), we consider the joint demand-side problem of consumer search and product choice. Our emphasis on monetization of advertising in the online marketplace context (as opposed to spatio-temporal price search in Mojir and Sudhir 2019 and sponsored search in Chan and Park 2015) motivates several differences in modeling choices. Extending Mojir and Sudhir (2019), we decouple the store-visit decision (browsing in our context) and category consideration (clicking in our context), as these are coincident decisions in Mojir and Sudhir (2019). In our context, we often find the absence of clicking after browsing and/or extensive browsing after the terminal click, suggesting that browsing and clicking decisions are not coincident in the online marketplace context. Moreover, as our goal is to consider the monetization of each step, it is useful to decouple them. Our model more closely hews to Chan and Park (2015), who consider sponsored search. Unlike our context, purchase is rarely observed in search advertising, so the terminal click is often proxied for purchase. In the online marketplace

context, wherein purchase is observed, we find that purchase rarely occurs at the last click. Hence, our consumer model decouples click and purchase decisions and accommodates the possibility that consumers purchase even the nonterminal clicked items.

1.2.2. Advertiser Behavior. The marketplace context we consider also has implications for the supply side. First, it is common that online marketplaces' revenues come from both advertising and transactions. As such, we build on Chan and Park's (2015) specification for advertiser valuation by including the observed dollar value of purchases. Second, there are typically a vastly greater number of advertisers observed in online marketplaces (sometimes thousands of advertisers competing for a limited number of slots). Hence, (i) it can be difficult to scale the inequality constraint approach in Chan and Park (2015); and (ii) the common-knowledge assumption on valuations of other advertisers is more difficult to ensure when the number of them becomes large. To address a similar challenge in display advertising markets, Balseiro et al. (2015) and Lu and Yang (2016) use a Mean Field Equilibrium (MFE). In the MFE, advertisers condition on the aggregate stationary distribution of states rather than each competitor's, an approach that obviates the need to invoke a common-knowledge assumption. By characterizing advertiser competition, we provide insights on the platform's profits and equilibrium outcomes under different fee structures and ranking algorithms. Third, because we observe both paid and nonpaid listings, we consider the trade-off that advertisers face in advertising and not advertising (i.e., organic ranking) (Blake et al. 2015, Sharma and Abhishek 2017, Simonov et al. 2018).

1.2.3. Platform Behavior. An area of enhanced interest in marketing, economics, and computer science concerns the use of ranking/scoring algorithms and/or recommendation systems to improve consumer search (e.g., Resnick and Varian 1997, Schafer et al. 1999, Bodapati 2008, Varian 2010, Ghose et al. 2012, De los Santos and Koulayev 2017, Smith and Linden 2017, Yoganarasimhan 2019, Zhang et al. 2019). Our focus lies on consumer search in two-sided markets, where strategic platforms face trade-offs between the needs of both sides of the market (e.g., Damiano and Li 2007, Horton 2017, Huang 2018, Fradkin 2019). In the context of e-commerce platforms, Hagiu and Jullien (2011) show that profit-maximizing platforms' incentives to optimize consumers' search process depend on the structure of the revenues they derive from the parties they serve. Long et al. (2018) consider leveraging sellers' bids for advertising in ranking organic listings to alleviate the information asymmetry between sellers and the platform. Dinerstein et al. (2018)

find that recommending products first, then comparing prices across sellers second, increases consumers' search efficiency while intensifying sellers' price competition. Our work builds on these papers by structurally modeling both consumer search and sellers' advertising competition and jointly considering fee structures (CPM, CPC, CPA, and fixed rate versus auctions) and ranking algorithms.

1.3. Key Findings

The consumer search and choice model results indicate that price and the number of pictures affect consumer preferences the most. The consumers' average marginal cost of browsing and clicking are \$0.89 and \$3.90, respectively, although there exists considerable heterogeneity in search costs across consumers. The model of advertiser behavior indicates that the typical seller's valuation from demand is negative (−4% of the transaction amount) when the seller opts in for advertising under the current fee structure. In other words, sellers are worse off on each advertised sale. In contrast, the median valuation from a click is estimated to be \$0.13, possibly because clicks generate awareness for items that can also be sold in other channels. Together, these results could suggest that clicks have a branding value because they can generate future demand for the advertiser's goods. Of note, impressions generate little value beyond clicks in our data.

Owing to consumers' high level of price sensitivity, we further find that a policy wherein the platform orders products by consumer utility or by ascending price lowers the platform's profits. Although more items are sold by reordering the product list, those that are sold are lower-price items. Sorting items by past sales volume also reduces the platform's profits because the increase in transaction commissions does not offset the decrease in advertising commissions. On the other hand, listing items by expected transaction revenue enhances the platform's profits, although it reduces consumer's consumption utility.

Because of the advertisers' high value for clicks and low value for sales (due to negative estimated margins for advertised goods caused by high levels of commissions imposed by the platform), the policy that lowers the cost-per-action and increases the cost-per-click more than doubles the platform's profits. Moreover, this policy improves advertiser welfare because their payments are better aligned with their valuations.

Finally, the platform's profits can be nearly tripled by changing both the pricing mechanism and the product-ranking algorithm. Specifically, (i) using a second-price CPC auction on the top 5 positions (i.e., thereby limiting the advertising slots) and (ii) ordering the remaining positions (6 and lower) by expected transaction revenue generates the highest

platform's profits. The intuition behind this result is that rationing the top positions monetizes the highest-sellers' valuations for advertising, whereas the transaction revenues are enhanced by ranking slots 6 and lower by the expected revenue.⁷ This outcome is illustrative of the value that accrues from considering the motivation of agents on both sides of the platform.

This paper is organized as follows. Section 2 describes our data and highlights key features pertinent to online marketplaces. Next, we present the model of buyers' purchase funnel decisions and sellers' advertising decisions. Section 4 discusses estimation method and identification argument, and Section 5 describes the estimation results. In Section 6, policy simulations are conducted to address questions that are of interest to practitioners.

2. Data

In order to better motivate the model assumptions and development, this section overviews our data context—first discussing the platform, then the buyers, and finally the advertisers.

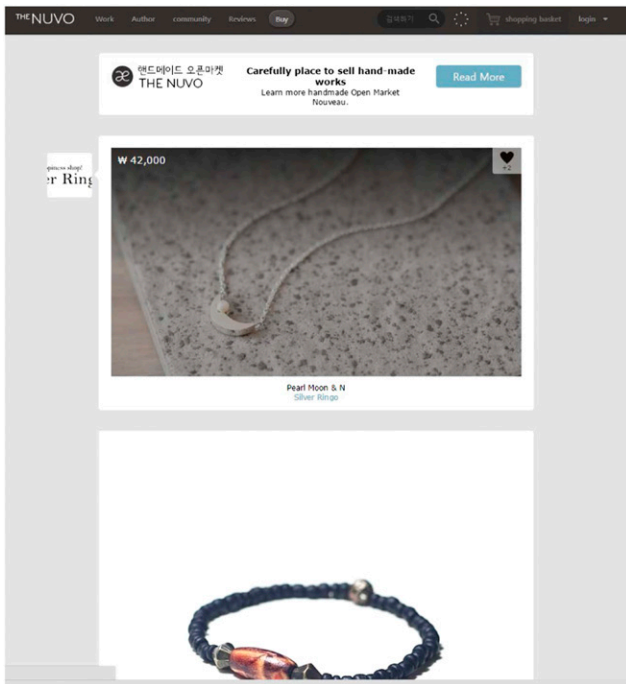
2.1. The Platform

The data we use are furnished by a Korean online marketplace (the-nuvo.com) specializing in hand-made goods. A unique aspect of the data is the depth of information provided by the platform on *both* buyers (browse, click, and purchase behaviors) and sellers (advertising decisions), along with their operational details, including product-display ranking algorithm. Moreover, owing to the unique nature of the handcrafted items in the data, browsing lengths and clicks are extensive and advertising is frequent, making it an ideal context to assess the consumer purchase funnel and how it is affected by advertising. The data include several files, each discussed below.

2.1.1. Platform Structure. We consider three aspects of the platform structure: the design of its pages (i.e., how attributes are allocated across the product-listing and product-detail pages), the ranking algorithm used to display products to consumers, and the fees charged for advertising.

Website Design. When a consumer first visits the site, she arrives on the main landing page. On this page, the platform displays products in a sequential product-feed format.⁸ Figure 1 provides a screenshot of the product feed in the main page, where typically one product is fully visible at a time on a regular-size browser. Consumers can scroll down to view more items or can interrupt browsing by clicking upon a specific product to access its product-detail page and to gather additional information. Upon continuation of browsing, the platform loads more products in

Figure 1. (Color online) Website Design



response to scroll-down requests, and the main-page product feed continues until the consumer stops browsing.

Although a consumer examines one product at a time as she scrolls down the product feed, the platform's server loads 10 items at a time to the back-end browser queue in response to a consumer's scroll-down request (e.g., position 11–20 items are loaded into the browser queue by the server as the consumer scrolls past the 10th positioned item). From an estimation perspective, the researcher observes the 10 products *loaded* last, not necessarily the last item *browsed* (amongst the 10 loaded last). Thus, in the empirical analysis, consumers are assumed to have browsed all items that are loaded from the requests.⁹

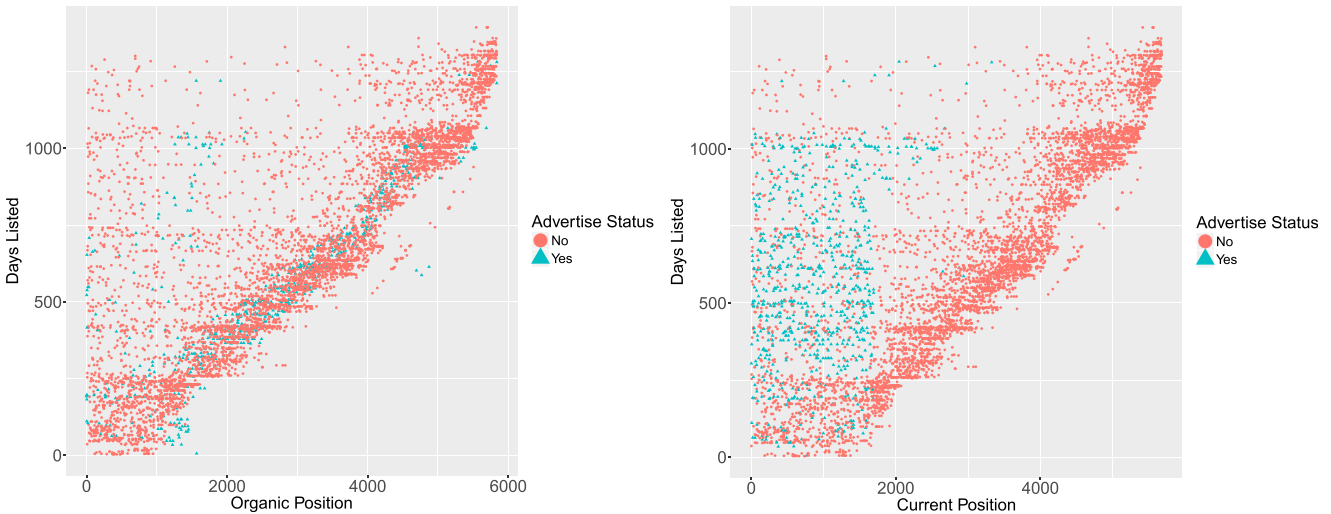
Information included in the “*product-listing page*” (defined as the product's information presented on the main landing page) includes the item's name, seller's (brand) name, price, number of likes, and discount percentage if the product is on sale. All other product-specific information is revealed in the “*product-detail page*” (defined as the page returned after a click upon an item), including a detailed product description, additional pictures, questions and answers, user reviews, size/color/material options, customizability (e.g., personal engraving), quantities remaining, shipping methods, the exchange and return policy, and the seller contact information.

Although the transactional site we consider has several categories, analogous to a retailer with many categories such as a department store, we focus our attention on items listed on the main landing page and

subsequent listings returned as consumers scroll down the main landing page. This focal category selection arises from the institutional details of our setting, where advertising works via the main-page product-feed ranking algorithm, whereas other (sub)categories are sorted purely from the newest to the oldest. Exits from this main-page product feed imply that consumers either leave the site (like leaving a store) or shop in another set of categories, which are captured as the outside option in our model.¹⁰

Product-Display Ranking Algorithm. The products displayed to consumers are ordered by using an algorithm determined by the platform, and the product list is updated daily using this algorithm. Although this algorithm is known to the researchers, it is not known to the sellers. The site presents the same list to all consumers and does not present sponsorship tags (so the consumers cannot distinguish between advertised and nonadvertised listings).¹¹ Key inputs to the algorithm include an item's (i) popularity score, (ii) slot-adjustment score, (iii) days listed, and (iv) advertising score. The popularity score includes the cumulative total number of purchases, clicks, likes, comments, reviews, SMS shares, and seller activities. The popularity score is measured in cumulative (running) totals, so popular items ranked high are likely to acquire higher popularity scores via more exposures, clicks, and likes. To offset this positive loop and to present more of a variety of items, the site applies a cumulative negative weight (slot-adjustment score) to the items previously shown in top 30 positions. Further, to offset the effect of older items acquiring higher popularity scores, the site applies a negative weight to the total number of days listed. Last, the advertising score mitigates the negative weighting on days listed, so older products can substantially increase their rank order in the listed items via advertising. The advertising advantage is attenuated as more sellers advertise because the gains in position are offsetting.¹²

To visualize the role of advertising in determining a product's position on the site, the left side of Figure 2 plots each product's organic position in the absence of advertising score against days listed. Each point represents an advertiser-product-day, and points marked in blue represent advertised goods. On the y-axis, smaller numbers indicate newer products. On the x-axis, smaller numbers indicate higher display positions. We find a strong relationship between the organic position and the days listed. Older products are pushed down to a lower rank, making it harder for consumers to find them (note that some older products attain higher position owing to higher popularity scores). On the right side, we plot each product's displayed position, and those that do advertise are moved to upper

Figure 2. Ranking Algorithm

positions in the product feed. Contrasting the two plots, we see that the positions can improve substantially with advertising.

Advertising Fee Structure. The website imposes zero listing fees and 13% transaction fees (f^T). The platform also receives an additional 17% of the transaction price (f^A) if the product sold is an advertised product at the time of transaction. When listing an item, a seller has an option to opt-in for advertising and can change its advertising decision at any time. Currently, the website does not impose fees based on clicks or impressions.

2.2. The Buyers

The buyer-side data include every visit, scroll, click, or purchase the website receives from its visitors. These data yield the number of times users visit the website, the products they browse, the product-detail pages they click, and what items they purchase. Registration to the website is optional for the buyers, and non-logged-in users are tracked by their cookie IDs.

2.2.1. Data Sample. The data were collected from mid-May 2014 to mid-February 2016. During the estimation sample period, 74,224 users made 238,646 visits to the platform.¹³ We focus our attention on the main landing-page visits. Excluding other-category visits and main-page visits that were followed by immediate visits to another area on the site leaves us 72,030 users with 85,632 visits. We further restrict our attention to the users with at least one purchase (within the estimation period, across all categories), giving us 263 users with 956 visits. This approach is analogous to research using scanner data that filter customer based on a minimum number of category purchases (Guadagni and Little 1983, Gupta 1988).¹⁴ As the website imposes zero listing fees, many sellers do not

unlist items when they become unavailable (e.g., temporarily sold out); instead, sellers change the price to zero and mention in the product-detail description that the product cannot be purchased. Hence, of the 74,969 total browsing instances, we exclude 569 with zero prices.

2.2.2. Buyer-Side Statistics. Consumers make sequential decisions regarding visit, product browse (impression), click, and purchase. Below, we discuss each in the order of consumer decision process.

Visit (Search Session). An individual makes 3.6 visits on average (median 2) during the sample period. These consumers browse 74,400 instances in total, among which 795 are considered, and 40 are purchased within the main-page product feed.¹⁵

Product Search. Summary statistics of consumer browsing and clicking behavior are presented in Table 1. The table indicates a mean level of 78 products browsed and 0.8 product-detail pages clicked, but there is a large standard deviation associated with each. The cumulative distribution of each behavior is present in Figure 3. Consumers in our sample generally search extensively, and there exists significant heterogeneity in search across individuals. These consumers differ in length and depth of their search processes. Some browse longer and make few clicks, whereas others browse shorter and click relatively many. All point to heterogeneity present in valuations and/or search costs and that consumers might possess click costs that are different from browsing costs.

Position Effect. In this subsection, we draw attention to the importance of an integrated model of browse, click, and purchase. Specifically we consider the role

Table 1. Summary Statistics of Consumers Behavior

No. per visit	Mean	Median	Standard deviation	Minimum	Maximum
Items browsed per visit	78	20	277	7	4,867
Clicks per visit	0.8	0	3.0	0	44
Purchases per visit	0.04	0	0.2	0	1
#Clicks/#Browses (%)	1.1	0	2.9	0	25
#Purchases/#Clicks (%)	7.4	0	22.2	0	100

of position effects as advertisers seek to obtain better positions in the product display list.

The product ranking and placement of advertised goods can have a considerable impact on items browsed and clicked. Such effects can be amplified for consumers with larger browsing and clicking costs. To explore this potential, Figure 3 displays how products placed in different positions are browsed, clicked, and purchased. The product position in the display list is plotted on the x-axis (larger number means lower position in the display feed). The cumulative probability of browsing, clicking, and purchase attained by the position is plotted on the y-axis. The position effect is strongest for the browsing length, and the number of browsing instances decreases exponentially with position, similar to the findings in Ansari and Mela (2003).

Conditional on browsing, however, the click likelihood does not exhibit an exponential decrease with the listing position, indicating that the magnitude of browsing costs and click costs may differ. This is shown on the left side of Figure 4, where product position is plotted on the x-axis, and the probability of click conditional on browsing is plotted on the y-axis. On the right side of Figure 4, the x-axis is again product position, and the y-axis represents the probability of purchase conditional on click. Here, the decrease in purchase with position is even smaller conditional on

click, suggesting that preference plays a bigger role at this decision stage relative to the sunk browsing and click costs.¹⁶ These plots are consistent with our modeling approach in that consumers first form a consideration set, taking into account their preference and the costs of browsing and clicking, but then make a purchase decision at the end based on preference alone. In sum, all above findings suggest the desirability of explicitly modeling the purchase (demand) as well as browsing and click behaviors separately.

Top-Down Search Assumption. An important assumption in our search model is that consumers browse and click products sequentially from top to bottom (scroll down the product feed).¹⁷ We begin by noting that browsing must be top-down when products are encountered for the first time because there is no way to be exposed to a later item before being exposed to an earlier one. However, this is not necessarily true for clicks. To explore this assumption further, we count the total number of occurrences in which the consumers click in the reverse rank order within each visit. Table 2 suggests that our assumption of top-down clicking is not violated for 98.2% (91.6 + 6.6) of all visits.¹⁸ In those instances in which consumers deviate from the top-down search pattern, we presume that the observed browse/click sequences follow the order in which products are first encountered (that is, as exogenously determined by the firm’s ranking algorithm).

2.3. The Advertisers

On the seller side, the site’s log file includes advertisers’ product listing, pricing, and advertising decisions. These include details of listed items, when they are listed, and at what price. If sellers update their pricing and advertising decisions after the initial listing is created, these changes are also recorded.

2.3.1. Data Sample. The data were collected from mid-May 2014 to mid-February 2016, but the key inputs to the ranking algorithm (popularity score and slot-adjustment score) were only available after mid-November 2015. As such, we use the shorter span when estimating the advertiser model. During this sample period, a total of 6,235 products from 595 sellers were exposed to the consumers. On a given day, on

Figure 3. (Color online) Browsing Length, Click, and Purchase by Position

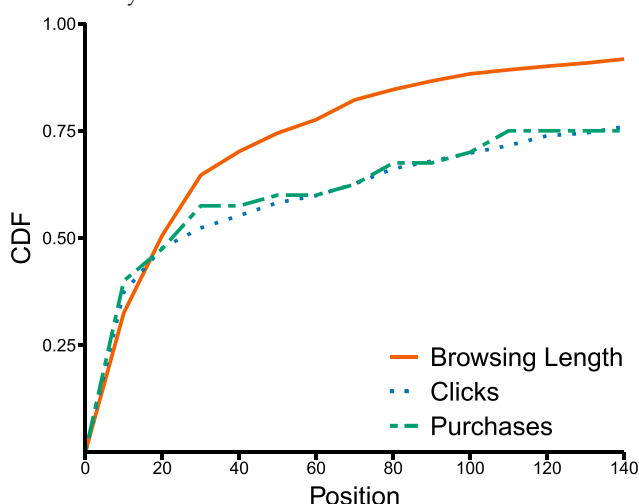
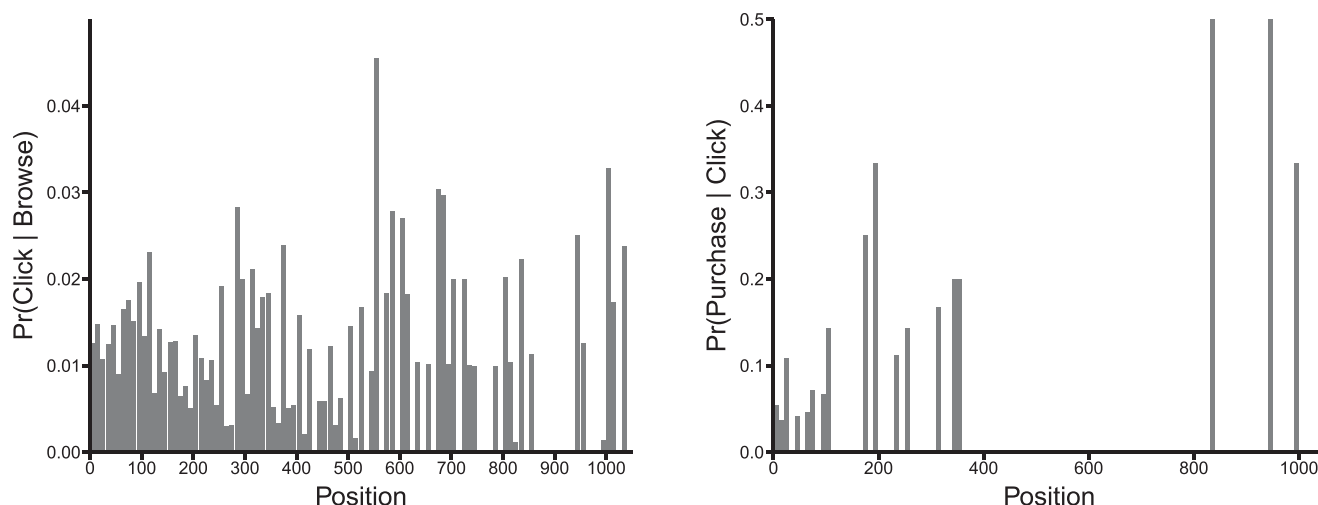


Figure 4. Position Effect on Click and Purchase

average, 5,847 products were available and displayed as product feed, among which 754 were advertised products. We omit products whose ranks are so low that they are never seen by the consumers even with advertising. Excluding products whose positions were never above 3,000 during our sample period yields a sample of 3,466 products. We then restrict our sample to the product listings initially created after March 2014, when the website went through a major renewal in its design and ranking algorithm. Lastly, we exclude products with zero prices and one product with an extreme price point (\$6,500), leaving us a final sample of 2,853 products.

2.3.2. Supply-Side Statistics. To obtain a better sense of seller listing strategies, we provide several summary statistics for the final sample of products ($N = 2,853$).

Product Attributes. Table 3 reports summary statistics of product attributes. The products have an average price of \$19.50, with a large variation across products. The products also vary in their promotion percentage (discount percentage), number of likes, and pictures.

Product Listing and Advertising Decisions. A seller lists 9.3 items on average (median 4) with a standard deviation of 16. Although there are a couple of sellers with more than 50 items, most are casual sellers with

few listings. This implies that most sellers are sufficiently atomistic, and none are likely to have undue influence on consumers, the platform, or other listing firms (see Figure A.1 in Online Appendix A.2.1).

We found that 35.8% of the sellers advertise at least one item, and advertised products constitute 19.5% of the total listed items. We found that 76.5% of sellers adopt a simple binary strategy in their advertising decision in that they either list all their items as advertised products or vice versa (Table 4). Although sellers can change their advertising decisions at any time on the website, we find that these changes rarely occur, suggesting that sellers play a static, binary opt-in or opt-out strategy at the time of listing an item. The phenomenon is even more pronounced at the seller-item level. Only 1.1% of advertising decisions changed across the products listed in our sample period (32 products from 7 sellers). As there is minimal longitudinal variation in advertising decisions, we aggregate data to the product level and treat advertising decisions at the product level as an observation unit instead of treating advertising decisions at the product-day level as an observation unit.¹⁹

Organic Strength and Advertising Decisions. To further illuminate the rationale underpinning sellers' advertising decisions, we compute products' "organic strength" as the mean residuals of the popularity score on days listed and feed position (see Online Appendix C.2.1). In the absence of an advertising effect, a higher organic strength implies that a product is more likely to attain a higher organic position in the product list. In Figure 5, we consider the relationship between a listing's organic strength and the seller's likelihood of advertising conditioned on that organic strength; the organic strength percentile is plotted on the x-axis (bigger percentile means higher strength), and the percentage of products advertised within

Table 2. Deviations from Top-Down Search Process

Deviations in clicks	% of visits
#Clicks < 2	91.6%
#Clicks ≥ 2	
No deviation	6.6%
1	1.6%
2 or more	0.2%

Table 3. Summary Statistics of Product Attributes

Attribute	Mean	Median	Standard deviation	Minimum	Maximum
Listing price (\$)	19.5	14.0	23.3	0.1	430.0
Discount (%)	0.9	0	4.6	0	50.0
No. of likes	1.6	1	2.2	0	31
No. of pictures	3.6	4	1.7	0	27

each bin is plotted on the y-axis. The figure shows that products that can organically appear early in the product list advertise less, suggesting strategic behavior on the part of the advertiser.

The observed pattern that organically highly ranked products advertise less than those organically ranked lowly suggests diminishing marginal returns to clicks/impressions. To the extent that diminishing marginal returns exist, one might expect strategic sellers at the bottom of the queue to be more disposed to advertise in order to be bumped up into the range of searched goods and gain the first impressions. In other words, the marginal benefit of being exposed via advertising is greater for those organically ranked low products. Hence, we accommodate diminishing marginal returns in our advertiser valuation model.²⁰

3. Model

In this section, we present a structural model encompassing the online marketplace. This model contains two components: (i) a model of consumer browsing (impressions), clicking (selection of product-detail pages), and purchases (choice); and (ii) a model of sellers' advertising decisions wherein sellers compete for positions in order to maximize their valuations from consumer impressions, clicks, and purchases. The platform moves first by setting the rules of the advertising game (i.e., the ranking algorithm and the fee structure). The advertisers move second by responding to the rules of the game, and the consumers move last conditioned on platform and advertiser decisions. Thus, we solve the problem via backward induction. Figure 6 depicts the agents and their interactions as well as the respective sections that discuss how we model each agent's problem.

3.1. The Consumer Model

Figure 7 summarizes the series of conditional decisions described below.

(1) **Visit:** A consumer first decides whether to visit the e-commerce website (start search session). We take

the consumer's visit decision as exogenously given.²¹ That is, the consumer's visit decision is independent of other consumers' behavior, sellers' advertising behavior, and the platform's ranking algorithm.²²

(2) **Product Search:** Product search consists of two stages: browsing (which generates impressions) and clicking (on a product-detail page yielding additional information about the items). Upon first visiting the website, the consumer is presented with an ordered list of items, one product at a time, where the arrival order of the products is exogenously determined by the platform's ranking algorithm. Faced with this list, a consumer can either click on the first item, incurring a click cost, or browse the next product on the list while incurring a browsing cost; that is, we presume a sequential search process. This leads to the following sequence of steps:

• Clicking Decision

The consumer is presented with the t -th positioned product (starting at $t = 1$), with some subset of the t -th item's attributes Z_t available on the product listing page (denoted "external" attributes). Having this partial information about the product's attributes, the consumer decides whether to add the t -th product into the consideration set by accessing (clicking) its product-detail page. Once clicked, the consumer gathers all information on the product-detail page's "internal" attributes X_t (possibly correlated with the Z_t) and the matching value ϵ_t and fully resolves any

Figure 5. Mean Organic Position and Advertising Percentage

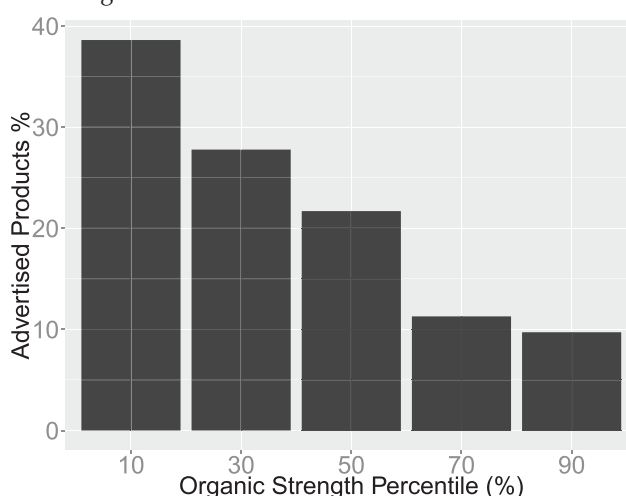
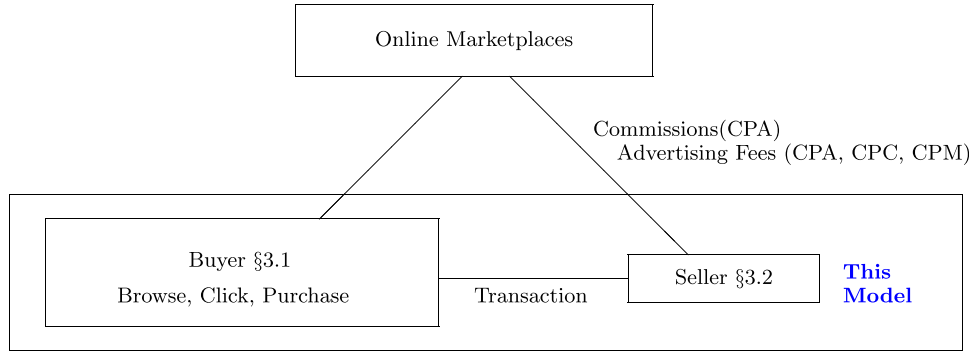


Table 4. Advertising Strategies

Pr(Advertise)	0	$0 < \text{Pr} < 1$	1	Total
No. of sellers	197	72	38	307
No. of products	2,298	32	523	2,853

Figure 6. Model Overview



product uncertainty with regard to its utility.²³ Once the click decision is made, the consumer decides whether to browse the $(t + 1)$ th position product. We present the click model in Section 3.1.2.

• Browsing or Exit Decision

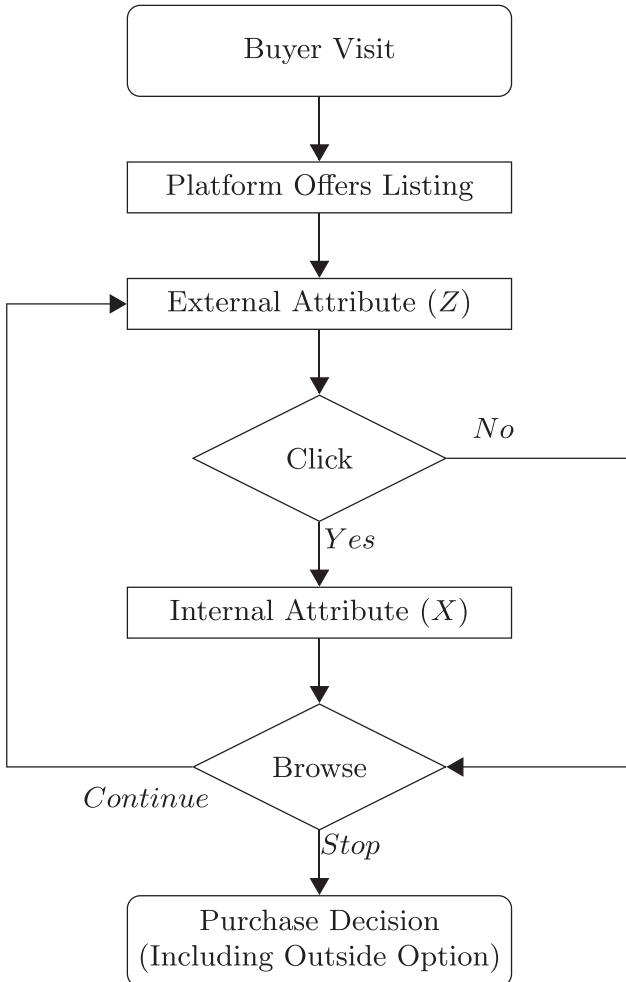
Conditioned on the information obtained in searching so far (the set $\{Z_1, Z_2, \dots, Z_t, d_1^c \cdot (X_1, \epsilon_1), d_2^c \cdot (X_2, \epsilon_2), \dots, d_t^c \cdot (X_t, \epsilon_t)\}$, where $d_t^c = 1$ if an item is clicked in step t , else 0), the consumer decides whether

to continue browsing the $(t + 1)$ th product. If the consumer decides to continue browsing, partial information on $(t + 1)$ th product is revealed, Z_{t+1} , the consumer incurs a browsing cost, and the consumer moves to the first stage of $(t + 1)$ th step. If the consumer decides not to continue, the entire search process terminates. We present the browsing model in Section 3.1.3.

(3) **Purchase (Choice)**: Once the search process terminates, the consumer has a final consideration set that consists of the items whose product-detail pages have been clicked and the outside option of nonpurchase. The consumer rationally chooses the highest-utility alternative in the consideration set. See Section 3.1.1.

We explicate each step of the purchase funnel—first, the utility function related to choice (purchase) is specified, and, then, clicking and browsing decisions are explained. The existence and the uniqueness of the consumer model are detailed in Online Appendix B.2.

Figure 7. Consumer Model



3.1.1. Purchase (Choice). Let consumer i 's indirect utility from purchasing a product j be

$$u_{ij} = \mathbf{X}_j \alpha + \mathbf{Z}_j \beta + \epsilon_{ij} \quad (1)$$

$$u_{i0} = \epsilon_{i0},$$

where $\{\mathbf{X}_j, \mathbf{Z}_j\}$ are row vectors of product attributes. ϵ_{ij} s follow $N(0, \sigma_\epsilon^2)$, and these are independent and identically distributed (iid) across consumers and products. When a consumer browses through the product list, some product characteristics are accessible without retrieving the product-detail page. These external attributes presented on the product-listing page are defined as \mathbf{Z}_j . Other product attributes revealed inside the product-detail page (which is accessed after clicking on an item in the product listing page) are denoted as \mathbf{X}_j . The last term ϵ_{ij} captures consumer i 's idiosyncratic taste about product j , and this match value is also inferred together with \mathbf{X}_j when the product is clicked. For example, a consumer looking for a handmade item finds a product from certain brand (\mathbf{Z}_j) on a product-listing page, clicks the link and

considers its product-detail page, and then finds that it is not adorned with a particular gemstone (X_j) although he likes the design detail (ϵ_{ij}). Consumers do not know the specific values of $\{X_j, \epsilon_{ij}\}$ before accessing the product-detail page, but they know the distribution of $\{X_j, \epsilon_{ij}\}$ conditional on the information in hand $\{Z_j\}$. This conditioning becomes material when there is a correlation between the external attributes (Z_j) and the internal attributes (X_j), enabling consumers to better forecast the attributes on the detail page (in the extreme case of a perfect correlation, the $\{X_j\}$ provide no additional information, and the only uncertainty is given by the $\{\epsilon_{ij}\}$). The outside good (not purchasing) does not require a search and is available in the consideration set from the beginning at no cost. The outside good can be construed to capture the option value of shopping in other stores on the site or leaving the site without a purchase (e.g., the option value of shopping for the good elsewhere).

The consumer's choice probability conditional on the consideration set Γ_i is

$$\Pr(d_i^p = j \mid \Gamma_i; \theta^p = \{\alpha, \beta\}) = \Pr(u_{ij} \geq u_{ij'} \forall j' \neq j, j' \in \Gamma_i),$$

where superscript p stands for purchase, and d_i^p indicates whether item j is chosen by consumer i .

3.1.2. Clicking Decision. Clicking an item involves reviewing its product-detail page and adding it to one's consideration set. This is a necessary step prior to purchase. Clicking a product-detail page does not afford any current period utility, although it is costly; rather, the benefit from clicking accrues in future periods via adding an item to a consideration set for purchase. As $\{X_j, \epsilon_{ij}\}$ are not known prior to clicking the detail page, the decision to click involves a trade-off between the cost of clicking and the likelihood that the clicked product's utility will be higher than any other item currently in the consideration set. Stated differently, consumers will click if the expected benefit of doing so exceeds the costs.

Clicking Costs. We proceed under the assumption that click costs are constant and specify the cost of click, c^c , as

$$c^c = \exp(\gamma_1).$$

Because there is no immediate period benefit from clicking, the current period payoff of click decision, U^c , is given by its costs,

$$U_{d^{c:ij}(t)}^c = \begin{cases} \eta_{0ij(t)}^c & \text{if not click, } d^c = 0 \\ -c^c + \eta_{1ij(t)}^c & \text{if click, } d^c = 1, \end{cases}$$

where $j(t)$ represents product j encountered by consumer at position t , and $\eta_{d^{c:ij}(t)}^c$ is assumed to follow iid Type I Extreme Value (TIEV) distribution. The

alternative-specific shock can be interpreted as a classic structural error term related to click preference that is known to the consumer but not observed by the researcher. Examples of $\eta_{d^{c:it}}^c$ might include an image-specific content revealed upon browsing but before clicking, the current position of the mouse pointer, or unobserved ongoing activities that may affect consumers' clicking decisions. This error term is distinguished from the match value $\epsilon_{ij(t)}$, which is revealed inside the product-detail page after clicking.

Clicking Benefits. Recall that the benefit from visiting a product-detail page accrues in future periods via its addition to the consideration set. Given that the utility of this item is not fully revealed until clicked, the consumer makes the click decision based on beliefs about whether adding the current item to the consideration set will yield higher utility than previously clicked items. The maximal utility, u_{it}^* , among the products in the consideration set Γ_{it} can be expressed as

$$u_{it}^* = \begin{cases} \max\{u_{ij(t)}, u_{it-1}^*\} & \text{if } j(t) \in \Gamma_{it} \\ u_{it-1}^* & \text{if } j(t) \notin \Gamma_{it}, \end{cases}$$

where $j(t)$ represents product j encountered by consumer at position t . In other words, if there is no additional click, there can be no increase in the maximum utility in the consideration set. The outside good option of not purchasing is included in the consideration set from the beginning, so $u_{i0}^* = u_{i0}$. Using this notation, the conditional value function of the click decision is given by the sum of the current period utility ($-c^c$ if click and 0 if not click) and the future utility flows accruing from the click decision, net of the choice specific error $\eta_{d^{c:it}}^c$; that is:

$$\begin{aligned} v_0^c(u_{it-1}^*, Z_{j(t)}) &= \int_{u_{it}^*} Emax^{browse}(u_{it}^*, Z_{j(t)}) f_0^c(u_{it}^* \mid u_{it-1}^*, Z_{j(t)}) \\ v_1^c(u_{it-1}^*, Z_{j(t)}) &= -c^c + \int_{u_{it}^*} Emax^{browse}(u_{it}^*, Z_{j(t)}) \\ &\quad \cdot f_1^c(u_{it}^* \mid u_{it-1}^*, Z_{j(t)}), \end{aligned} \quad (2)$$

where the future utility flows after clicking an item involve the expected value that will accrue from the next browsing decision (see Figure 7), or

$$\begin{aligned} Emax^{browse}(u_{it}^*, Z_{j(t)}) &= E_{\eta_0, \eta_1} [\max\{v_0^s(u_{it}^*, Z_{j(t)}) + \eta_{0it}^s, v_1^s(u_{it}^*, Z_{j(t)}) + \eta_{1it}^s\}] \\ &= \ln[\exp(v_0^s(u_{it}^*, Z_{j(t)})) + \exp(v_1^s(u_{it}^*, Z_{j(t)}))] + \kappa, \end{aligned} \quad (3)$$

where κ is the Euler constant, and v_0^s, v_1^s are conditional value functions of the browse decision, which are later defined in Equation (6).

The first line in Equation (2) captures the value of not clicking, v_0^c , and the second line captures the value of clicking, v_1^c . Functions f_0^c and f_1^c are the state transitions for consumers' beliefs about the future highest utility achievable, u^* , based on current state, u_{it-1}^* , in the consideration set and the partial attributes, $\mathbf{Z}_{j(t)}$, presented at position t on the product listing page. Online Appendix B.1 derives the state transitions, f^c , on these beliefs.

$Emax^{browse}(u_{it}^*, \mathbf{Z}_{j(t)})$ represents the expected value of browsing (which immediately follows the click decision per Figure 7). This expectation is taken over the browsing alternative-specific shocks, η_{it}^s s, as they are not observed at the time of click (that is, they are revealed to the consumer in the subsequent browsing step). Under the logit error assumption, this is the inclusive value of the browse decision. We don't specify discount factor in front of the future values, as the time interval between click and browse decision is short.

Clicking Decision. Given the choice specific value functions above, the conditional choice probability of no click, $d^c = 0$, can be expressed as

$$p_0^c(u_{it-1}^*, \mathbf{Z}_{j(t)}) = \frac{1}{1 + \exp(v_1^c(u_{it-1}^*, \mathbf{Z}_{j(t)}) - v_0^c(u_{it-1}^*, \mathbf{Z}_{j(t)}))}. \quad (4)$$

This is the popular dynamic logit model where the choice probabilities depend on differences in choice specific value functions (Arcidiacono and Miller 2011). Once the click decision is made at position t , a consumer proceeds to browse decision and decides whether they want to terminate or continue to browse the $(t+1)$ th item.

3.1.3. Browsing Decision. Analogous to click, consumers will browse as long as the expected benefit of doing so exceeds the cost.

Browsing Costs. Browsing costs c^s are specified as

$$c^s = \exp(\gamma_2).$$

Browsing Benefits. Once u_{it}^* is revealed based on the click decision ($u_{it}^* = u_{it-1}^*$ if t -th position product is not clicked), the consumer must then decide whether to browse the $(t+1)$ th product in order to obtain information about the $\mathbf{Z}_{j(t+1)}$ (see Figure 7). The current period payoff of browse decision is

$$U_{d^s it}^s = \begin{cases} u_{it}^* + \eta_{0it}^s & \text{if browsing stops, } d^s = 0 \\ -c^s + \eta_{1it}^s & \text{if browsing continues, } d^s = 1, \end{cases} \quad (5)$$

where $\eta_{d^s it}^s$ is assumed to follow iid Type I Extreme Value (Gumbel) distribution. The first line in Equation (5) indicates that a consumer who stops browsing at step t will receive utility u_{it}^* (reflective of the best alternative found prior to stopping browsing), plus a random shock observed by the consumer but not the researcher. This alternative-specific shock might include unobserved factors such as Internet connectivity, incoming online messages from a friend, or general time constraints that affect browsing behavior. Alternatively, if a consumer continues to browse, he will pay a browsing cost now but accrues no benefit until after the entire search process is completed. This benefit represents the expected future value arising from potentially finding a better alternative to add to the consideration set and purchase by continuing browsing. The conditional value function for the browse decision at position t —that is, the sum of the current period utility ($-c^s$ if browsing is continued, and u_{it}^* if browsing is stopped) and the future utility flows accruing from the browse decision, net of the browsing choice specific error $\eta_{d^s it}^s$, can be written as

$$\begin{aligned} v_0^s(u_{it}^*, \mathbf{Z}_{j(t)}) &= u_{it}^* \\ v_1^s(u_{it}^*, \mathbf{Z}_{j(t)}) &= -c^s + \int_{\mathbf{Z}_{j(t+1)}} Emax^{click}(u_{it}^*, \mathbf{Z}_{j(t+1)}) \\ &\quad \cdot f_1^s(\mathbf{Z}_{j(t+1)} | \mathbf{Z}_{j(t)}), \end{aligned} \quad (6)$$

where

$$\begin{aligned} Emax^{click}(u_{it}^*, \mathbf{Z}_{j(t+1)}) &= E\left[\max\left\{v_0^c(u_{it}^*, \mathbf{Z}_{j(t+1)}) + \eta_{0it(t+1)}^c, v_1^c(u_{it}^*, \mathbf{Z}_{j(t+1)})\right.\right. \\ &\quad \left.\left.+ \eta_{1it(t+1)}^c\right\}\right] \\ &= \ln(\exp(v_0^c(u_{it}^*, \mathbf{Z}_{j(t+1)})) + \exp(v_1^c(u_{it}^*, \mathbf{Z}_{j(t+1)}))) + \kappa; \end{aligned} \quad (7)$$

$f_1^s(\mathbf{Z}_{j(t+1)} | \mathbf{Z}_{j(t)})$ is the distribution of consumers' beliefs on future $\mathbf{Z}_{j(t+1)}$ conditional on the decision to continue browsing (see Online Appendix B.1). The continuation value of browsing is given in the second line of Equation (6) and corresponds to the expected maximum of the utility of the ensuing click decision because the continuation of browsing affords the option of potentially adding another item to the consideration set. This expected future value is given in Equation (7). The discount factor is again assumed to be 1, as the time interval between browse decision and following click decision for the $(t+1)$ th product is short. Although we discuss product attribute state transitions $f_1^s(\mathbf{Z}_{j(t+1)} | \mathbf{Z}_{j(t)})$ in Online Appendix B.1, it is worth noting that the browse decision can be informative about click if the attributes on the

product-listing page \mathbf{Z}_t , are correlated with the attributes inside the product-detail page \mathbf{X}_t .

Browsing Decision. Note that *stop browsing* is a terminal decision that ends the search process altogether. With the double-exponential parametric assumption on $\eta_{d^s|it}^s$, the conditional choice probability of ending browsing, $d^s = 0$, is given by

$$p_0^s(u_{it}^*, \mathbf{Z}_{j(t)}) = \frac{1}{1 + \exp(v_1^s(u_{it}^*, \mathbf{Z}_{j(t)}) - v_0^s(u_{it}^*, \mathbf{Z}_{j(t)}))}. \quad (8)$$

If we denote *stop browsing* position as $t = T^s$, the consumer's optimal purchasing decision is to choose the alternative (including the outside option of not purchasing) that delivers the highest utility $u_{iT^s}^*$ within the consideration set Γ_{iT^s} . This payoff related to purchase is embedded in the browse decision as we model $v_{0T^s}^s = u_{iT^s}^*$.

3.2. The Advertiser Model

Upon deciding to list an item on the platform, sellers are faced with the decision of whether to advertise. Advertising on the site has two offsetting consequences. On the positive side, advertised goods are listed in more favorable positions, thereby increasing exposures and potentially clicks and purchases, which, in turn, increase advertiser revenue. On the negative side, sellers pay fees for advertising. We presume that sellers advertise if the expected valuation gains from advertising surpass the expected cost of advertising. These expected valuation gains depend on (i) how advertising affects *consumer browsing*, *clicking*, and *purchasing*; (ii) the *competition for advertised slots* improving an advertised product's position necessarily entails lowering those of other products; and (iii) the *cost of advertising* arising from fees charged by the platform. As the solution to the advertiser problem requires firms to form beliefs about consumer response, product position, the cost of advertising, and the competitive landscape, we detail these points in Section 3.2.1 before formalizing the advertiser problem in Section 3.2.2.

3.2.1. Key Assumptions. The advertiser problem conditions upon the consumer behavior, competitor behavior, and the platform's behavior in terms of fee structure and ranking algorithm. We detail our assumptions pertaining to each.

Consumer Behavior. We assume that sellers form rational expectations about demand, clicks, and browses based on their beliefs about increase in product placement via advertising and that strategic interactions (competitive effects) work through the changes in product placement. Specifically, given the belief on product

position from the advertising strategy, the seller is assumed to form rational beliefs on consumer demand, click, and browse (impression) responses based on the distribution of consumer preferences and costs from consumer model:

$$\begin{aligned} \widehat{D_{j,d^d|j}} &= D(\widehat{Rank_{j,d^d|j}}, \mathbf{X}, \mathbf{Z}); \\ \widehat{C_{j,d^d|j}} &= C(\widehat{Rank_{j,d^d|j}}, \mathbf{X}, \mathbf{Z}); \\ \widehat{I_{j,d^d|j}} &= I(\widehat{Rank_{j,d^d|j}}, \mathbf{X}, \mathbf{Z}), \end{aligned} \quad (9)$$

where $\widehat{Rank_{j,d^d|j}}$ is the belief on product j 's position when the competing advertising strategies are given by \mathbf{d}_{-j}^a , which is a vector of beliefs regarding competing advertiser advertising decisions.

Competitive Behavior. Consistent with the lack of evidence for dynamics in the data, we presume that the seller's advertising decision is a static, binary, discrete choice at the product level. That is, the seller opts in for advertising when listing an item if it is profitable to do so to compete for better placement. We presume that sellers form bounded rational beliefs about others' advertising decisions. Under the rational expectations assumption, solving optimal advertising decisions in our context of an online marketplace requires forming beliefs about many thousands of other sellers' (products') advertising strategies. This is not only computationally intractable due to the curse of dimensionality, but also implies that small firms (who carry a median of 4 products in our data) know the valuations of thousands of other small firms. This assumption strikes us as implausible given the effort that such a task would entail. Moreover, in the limit, an advertiser's rank does not explicitly depend on what other specific firms do but, instead, the aggregate number of firms that advertise. Accordingly, we assume that each seller (product) is sufficiently atomistic that each seller (product) conditions on the advertising probability distribution moments (aggregate states) rather than each other seller's actual advertising probability (individual states) when forming beliefs about their own ranking. Finally, we presume that the aggregate beliefs are consistent with the underlying advertisers' decisions at equilibrium. For example, we presume an advertiser's beliefs about the expected number of competing advertisers is simply the sum of individual advertising decisions across competing firms.²⁴

Platform Behavior. We consider two aspects of platform behavior: search rankings wherein the platform determines the order of items presented to consumers and the fees charged to sellers. Although the cost of advertising could involve a variety of potential pricing mechanisms available to the platform

(fixed-fee-per-ad-slot, auction-mechanism-per-ad-slot, cost-per-click, cost-per-mille, and/or cost-per-action), our inference regarding the advertiser model reflects the institutional details of our setting wherein the e-commerce platform charges a percentage commission as advertising fees based on sales. We will further incorporate cost-per-click, cost-per-mille, and auction mechanisms in the advertiser model as part of our policy simulations.

3.2.2. Valuations. A seller k chooses an optimal advertising strategy for product j as defined by an indicator variable d_j^a ($d_j^a = 1$ advertises, and $d_j^a = 0$ does not). Building on Chan and Park (2015), we model sellers as gaining valuations from three sources: (i) demand, (ii) clicks, and (iii) impressions (browsing). Impressions and clicks can generate value from, for example, creating value via branding. Specifically, seller k 's valuation for product j from advertising decision d_j^a is

$$\pi_{jk d^a} = \theta \cdot \mathbf{w}_{jk} \cdot \mathbf{1}(d^a = 1) + \pi_{j d^a}^D + \pi_{j d^a}^C + \pi_{j d^a}^I,$$

where $\pi_{j d^a}^D$, $\pi_{j d^a}^C$, and $\pi_{j d^a}^I$ are valuations from demand, clicks, and impressions, respectively. To accommodate product–seller level heterogeneity, a vector \mathbf{w}_{jk} is introduced as an additive term and includes the following: seller fixed effects, material fixed effects, category fixed effects, whether a product is refundable, and whether a URL is specified in the product description (see Online Appendix A.2.2). The unobserved heterogeneity at the product level is captured by the structural error term ξ_j ($\equiv \xi_{j1} - \xi_{j0}$), which is assumed to follow a normal distribution. Examples of ξ_j include product-related nonsite activities or promotional/marketing strategies that might affect sellers' advertising decisions. The seller advertises product j if doing so is profitable—that is, if the below condition is satisfied:

$$\left(\theta \cdot \mathbf{w}_{jk} + \pi_{j1}^D + \pi_{j1}^C + \pi_{j1}^I \right) + \xi_j \geq \left(\pi_{j0}^D + \pi_{j0}^C + \pi_{j0}^I \right) \\ \xi_j \sim N(0, \sigma_\xi^2). \quad (10)$$

Valuations from Demand. The first component of the advertiser's valuation comes from profit earned when a product is sold on the website. The sale of a product accrues revenue, and, at the same time, the seller pays a fixed transaction fee as a percentage of the transaction amount, f^T . In addition, the seller also pays an additional fixed percentage as a commission, f^A , when the product is advertised and sold. The valuation from demand is represented as

$$\pi_{j d^a}^D = \theta^D \left(1 - f^T - f^A \mathbf{1}(d_j^a = 1) - \delta_j \right) D_{j d^a} p_j, \quad (11)$$

where $D_{j d^a}$ and p_j are demand and price for product j , respectively. The δ represents the underlying marginal cost, and θ^D is a scale parameter that maps the seller's short-term profit valuations.²⁵ For the same marginal costs, higher f^T or f^A implies that the seller has a greater incentive to redirect consumer to the outside channels for purchase.

Valuations from Clicking and Browsing (Impressions).

The other two components of the advertiser's valuation come from clicks and impressions. The seller gains benefit from clicks and impressions, but the seller also pays potential cost-per-click fees, f^C (a fixed fee per click made by the consumer), and/or potential cost-per-mille fees, f^I (a fixed fee per thousand impressions delivered to the consumer). These potential fees are charged to the sellers regardless of whether an item is sold on the website (recall that a consumer must click on an item for it to enter the consideration set for potential subsequent sale). These valuations reflect the standard concept that exposures and clicks have advertising value to the seller over and above an immediate sale, either through branding or future sales.²⁶

We assume that the seller's valuation from clicks and impressions exhibit diminishing marginal returns. This assumption is motivated by the findings in our data (see Section 2.3.2) and the widely used practice of "frequency capping" in display advertising market. Many experts believe that repeated exposures past a certain threshold will not increase conversion rate or brand equity; thus, the number of impressions served needs to be capped to avoid overexposure.²⁷ The valuation from clicks and impressions is given by

$$\pi_{j d^a}^C = \theta^C \log(C_{j d^a}) - f^C C_{j d^a} \\ \pi_{j d^a}^I = \theta^I \log(I_{j d^a}) - f^I I_{j d^a}, \quad (12)$$

where $C_{j d^a}$ and $I_{j d^a}$ are clicks and impressions (in thousands), respectively.

Advertiser Decision. Given the underlying parameters of the model (θ , δ , θ^D , θ^C , θ^I) and with the parametric assumption on ξ_j , the probability of advertising in equilibrium is given by

$$p_{jk1}^a = \Phi \left[\frac{\left(\theta \cdot \mathbf{w}_{jk} + \pi_{j1}^D + \pi_{j1}^C + \pi_{j1}^I \right) - \left(\pi_{j0}^D + \pi_{j0}^C + \pi_{j0}^I \right)}{\sigma_\xi} \right]. \quad (13)$$

4. Estimation

In this section, we outline our estimation approach to the consumer model and the advertiser model. The goal of the consumer model is to infer preferences and browsing/clicking costs, and the goal of the

advertiser model is to infer advertiser valuations for impressions, clicks, and purchases.

4.1. The Consumer Model

In this subsection, we develop the consumer model likelihood and overview identification.

4.1.1. Consumer Utility. We specify consumer i 's utility from purchasing product j from category-seller k to be

$$u_{ijk} = \mu_k - \beta_p \log(P_j) + \beta_z Z_j + \alpha X_j + \epsilon_{ij}$$

$$u_{i0} = \epsilon_{i0}.$$

The information depicted on the product-listing page and known to consumers upon browsing an item (but before clicking it) includes seller identity, price, and the number of likes (μ_k, P_j, Z_j) . The number of pictures X_j and the match value ϵ_{ij} are revealed inside the product-detail page.

We abstract away from product-level unobservables μ_j and include category-seller level fixed effects μ_k to capture preferences for certain categories and brands. Many products that are browsed have zero demand and zero clicks in our data, making it difficult to recover product-level unobservables. Second, seller-level unobservables capture unobserved vertical differentiation in this market, where authorship and craftsmanship create uniqueness and distinguishable features at the seller level. Products nested within seller share these unobservables.²⁸

4.1.2. Likelihood and Heterogeneity. The log-likelihood of browsing, clicking, and purchase is denoted as

$$L(\Theta_1) = L(\alpha^g, \beta^g, \gamma_1^g, \gamma_2^g, \lambda^g) \quad g = 1, \dots, G,$$

where $\lambda^1, \dots, \lambda^G$ represents the type probability of each segment when there are G latent classes (Kamakura and Russell 1989).

Let T_i^g reference the position where individual i chooses to stop browsing such that $d_{iT_i^g}^s = 0$. The likelihood of observing $\mathbf{d}_i = \{d_{i1}^c, \dots, d_{iT_i^g}^c, d_{i1}^s, \dots, d_{iT_i^g}^s, d_{i1}^p, \dots, d_{iT_i^g}^p\}$ for individual i in latent class g is defined as

$$L_i(\Theta_1^g) = \int_{u_{iT_i^g}^*} \dots \int_{u_{i1}^*} \int_{u_{i0}^*} f^u(u_{i0}^*)$$

$$\cdot \mathcal{L}(\mathbf{d}_i | u_{i0}^*, \dots, u_{iT_i^g}^*, \mathbf{Z}, \mathbf{X}; \Theta_1^g)$$

$$= \int_{u_{iT_i^g}^*} \dots \int_{u_{i1}^*} \int_{u_{i0}^*} f^u(u_{i0}^*) \prod_{t=1}^{T_i^g} \mathcal{L}_t^{\text{browse}} \mathcal{L}_t^{\text{click|browse}}$$

$$\cdot \mathcal{L}_t^{\text{purchase|browse,click}},$$
(14)

where the initial probability $f^u(u_{i0}^*)$ is the distribution of outside option value $f^u(\epsilon_{i0}) = \phi(\epsilon_{i0})$, and

superscripts c , s , and p represent click, browse, and purchase, respectively. The information used to infer the consumer primitives comes from these three observed decisions, and the joint log-likelihood of the sample data can be written as

$$L(\Theta_1) = \sum_{i=1}^I \ln \left(\sum_{g=1}^G \lambda^g L_i(\Theta_1^g) \right), \quad (15)$$

where we integrate out latent class consumer heterogeneity. In Online Appendix C.1.1, we derive the joint likelihood of browsing, clicking, and purchase. Of note, the likelihood function is not separable, and the state transition, $f^u(u_{it}^* | u_{it-1}^*, \mathbf{Z}_{j(t)}, \mathbf{X}_{j(t)})$, links \mathcal{L}_t and \mathcal{L}_{t-1} .

4.1.3. Solving the Dynamic Problem. We formulate the consumer model as an infinite horizon problem and maximize the joint likelihood using maximum-likelihood estimation in the outer loop (parameter estimation) and value function iteration for the inner loop (future value terms and resulting choice probabilities conditioned on those parameters). The rationale for using an infinite-horizon formulation and the estimation approach can be found in Online Appendix C.1.2.

4.1.4. Identification. Our identification discussion covers four domains—the identification of costs, preferences, heterogeneity, and the discussion on the error terms.

Search Costs. Browsing costs are identified from the variation in the number of items browsed with respect to the (exogenous) variation on product positions conditioned on the product characteristics. For example, the product positions provide an exclusion restriction for the observed browsing length because they affect search costs, but not consumers' valuation of a good. The clicking cost is separately identified from the browsing cost based on variation in how many products consumers click, conditioned on the browsing length and product characteristics.

Preferences. The identification of the utility parameters comes from the consumers' browsing/clicking decisions and the purchase decisions. As we observe purchases directly, identification of the preference parameters is standard as in a conditional choice model. Additionally, observing consumers' browsing and clicking behaviors strengthens identifiability of the preference parameters because the selection of which product characteristics to click (in addition to how many) helps to pin down the preference parameters (see Kim et al. 2016, Chen and Yao 2017, Honka et al. 2017). For example, given a fixed browsing length, clicking on more low-price items implies greater price sensitivity.

Online Appendix C.1.3 reports simulation results suggesting that the inclusion of purchase data significantly reduces the standard errors of the parameter estimates over browsing/clicking data alone, especially for the preference parameters.

Heterogeneity. In our empirical application, most consumers' visits are highly episodic (with median 2 visits per individual); thus, we use a finite mixture model assumption to help identify heterogeneity in costs.

Structural Errors.

(1) **Match Value:** The normal distribution on the match value, ε , follows the commonly adopted distributional assumption in existing Weitzman-type search models (e.g., Kim et al. 2010, Chen and Yao 2017). The variance of match error term (ε) is normalized to be $\sigma_\varepsilon^2 = 1$ for identification purposes.

(2) **Structural Error for Clicking:** The introduction of the structural error term for clicking ($\eta_{d^{c:it}}^c$, known to consumer *prior* to clicking), separately from the match value (ε , revealed *after* clicking an item), accommodates the possibility that consumers' clicking decisions may vary even after controlling for the observed external attributes and u^* .²⁹ Following the dynamic discrete choice model literature (Arcidiacono and Miller 2011), the structural error term for clicks ($\eta_{d^{c:it}}^c$) is assumed to follow TIEV distribution, and the scale is normalized to 1 for identification. With this TIEV distributional assumption, the value functions in Equation (7) has a closed form, which greatly reduces the computational complexity associated with estimation.

(3) **Structural Error for Browsing:** A separate error term for browsing ($\eta_{d^{b:it}}^b$) is required to accommodate the possibility consumers browse extensively after the last click.³⁰ Similarly to the structural error term for clicking, the structural error term for browsing is assumed to follow TIEV with scale 1.³¹

4.2. The Advertiser Model

4.2.1. Constructing Advertisers' Beliefs. As the platform's ranking algorithm and the underlying scores are not shared with the sellers, they must form beliefs regarding their relative product rank with and without advertising in order to assess the attendant impact on impressions, clicks, and purchase. Following the discussion in Section 3.2.1, we assume that each seller (product) is sufficiently atomistic and forms bounded rational beliefs about others' advertising decisions in predicting his own product rank. Specifically, we assume that advertisers' beliefs on the product placement for a given day t depend on their own advertising strategy d_j^a , the aggregate states of others' advertising strategies $E_t(\mathbf{d}_{-j}^a)$, the total number

of products available J_t , and own product j 's attributes that affect the rank score:

$$\widehat{Rank}_{j,t,d_j^a,d_{-j}^a} = g\left(d_j^a, E_t(\mathbf{d}_{-j}^a), J_t, \text{Days Listed}_{jt}, \text{Organic Strength}_j\right), \quad (16)$$

where "organic strength" is the mean residuals of the popularity score on days listed and product position. We specify the function $g(\cdot)$ to be a generalized additive model with interaction terms included (see Online Appendix C.2.1).³² Note that the effect of competition manifests via $E(\mathbf{d}_{-j}^a)$. As competing firms advertise more, one's own rank (and thus impressions, clicks, and sales) decreases. Because each advertiser faces a similar problem, to find the equilibrium behavior, we solve each advertiser's respective problem conditioned on $E(\mathbf{d}_{-j}^a)$, recompute $E(\mathbf{d}_{-j}^a)$ using these collective decisions, and iterate until convergence for policy simulations. For more detail, see Section 4.2.3 and Online Appendix C.2.3.

In addition to beliefs about competing firms' behaviors, advertisers form beliefs about consumer behavior as well. Equipped with beliefs about their own product placement in the search queue, $\widehat{Rank}_{j,t,d_j^a,d_{-j}^a}$, sellers form beliefs about consumer behavior in terms of demand, click, and impression responses (Equation (9)). That is, sellers form expectations by integrating out over the belief distribution of product ranks and consumer behaviors. As we formulate the advertiser model in a static framework, expected impressions, clicks, and demand are imputed over the duration of the product listing (i.e., the net present value of impressions, clicks, and purchases). By using the consumer demand model, consumer responses are simulated for each day based on sellers' product position beliefs $\widehat{Rank}_{j,t,d_j^a,d_{-j}^a}$ and aggregated across time periods.

4.2.2. Likelihood. The advertising model parameters are $\Theta_2 = (\theta, \theta^D, \theta^C, \theta^I, \delta)$. The likelihood of observing seller k 's advertising decision on product j , d_{jk}^a , is given by

$$\mathcal{L}_{jk}^a(d_{jk}^a; \Theta_2) = p_{jk1}^a 1^{(d_{jk}^a=1)} \times [1 - p_{jk1}^a] 1^{(d_{jk}^a=0)},$$

where p_{jk1}^a is the advertising probability defined in Equation (13). Further, the log-likelihood of the sample data for the advertiser probit model is given by

$$L^a(\Theta_2) = \sum_{j=1}^J \ln \left(\mathcal{L}_{jk}^a(d_{jk}^a; \Theta_2) \right). \quad (17)$$

4.2.3. Solving the Advertiser Problem. We estimate the advertiser model in three stages. In stage 1, we estimate the function governing sellers' beliefs on

product rank, Equation (16). In stage 2, sellers' beliefs on product placement and consumer responses with respect to advertising are constructed. By contrasting the valuation from demand, click, and impression responses when advertising and when not advertising, the seller's advertising probability is imputed. The parameters of interest, $\Theta_2 = (\theta, \theta^D, \theta^C, \theta^I, \delta)$, are then recovered in stage 3 by using a maximum-likelihood estimation method based on the likelihood function in Equation (17). In Online Appendix C.2.2, we describe these estimation stages in detail and discuss how the equilibrium advertising strategies are computed for the policy simulation.

4.2.4. Identification. As in the standard probit model, the variance of the structural error term is normalized to $\sigma_\xi = 1$. Under the functional specification assumed in the advertiser model, the advertiser valuations for demand, clicks, and impressions are identified from the observed likelihood of advertising with respect to variation in rank and resulting changes in consumer responses due to advertising.³³ More specifically,

rewriting the difference in seller k 's valuation for product j from opting in and opting out of advertising yields:

$$\begin{aligned}\pi_{jk1} - \pi_{jk0} = & \theta \cdot \mathbf{w}_{jk} + \theta^D(1 - f^T - \delta)p_j(D_{j1} - D_{j0}) \\ & + \theta^C(\log(C_{j1}) - \log(C_{j0})) - f^C(C_{j1} - C_{j0}) \\ & + \theta^I(\log(I_{j1}) - \log(I_{j0})) - f^I(I_{j1} - I_{j0}) \\ & - \theta^D f^A p_j D_{j1}.\end{aligned}\quad (18)$$

Note that in our empirical setting, the sellers pay advertising fees only when opting in for advertising and when the sales are realized. Thus, the valuation from demand, θ^D , can be identified from the sensitivity of advertising decision with respect to the variation in expected advertising commissions incurred ($f^A p_j D_{j1}$). Second, the valuations from clicks and impressions are recovered from the increase in clicks and impressions via advertising. If an increase in clicks (impressions) is correlated with advertising, valuations will be positive. Finally, δ is identified from the revenue increase due to advertising. Given

Table 5. The Consumer Model Estimates

Parameter	Number of segments				
	1	2	2 on Cost	3 on Cost	4 on Cost
Preference					
Type1					
# Pictures (X)	0.20 (0.11)	0.09(0.15)	0.18(0.15)	0.18 (0.09)	0.18 (0.10)
Log (Price) (Z)	-0.25 (0.15)	-0.11(0.22)	-0.24 (0.14)	-0.25 (0.15)	-0.29 (0.15)
# Likes (Z)	-0.00(0.01)	-0.00(0.01)	-0.00(0.01)	-0.00(0.01)	-0.00(0.01)
Constant	-3.05 (0.61)	-2.86 (0.80)	-2.84 (0.87)	-2.77 (0.24)	-2.67 (0.61)
Type2					
# Pictures (X)		0.27 (0.15)			
Log (Price) (Z)		-0.37 (0.22)			
# Likes (Z)		-0.01(0.02)			
Constant		-2.79 (0.84)			
Cost					
Type1					
Clicking	1.52 (0.01)	1.48 (0.01)	1.48 (0.01)	1.37 (0.02)	1.42 (0.02)
Browsing	0.16 (0.00)	0.18 (0.00)	0.18 (0.00)	0.16 (0.00)	0.16 (0.00)
Type2					
Clicking		1.55 (0.01)	1.55 (0.01)	1.75 (0.03)	1.77 (0.03)
Browsing		0.15 (0.00)	0.15 (0.00)	0.15 (0.00)	0.15 (0.00)
Type3					
Clicking				1.52 (0.02)	1.16 (0.04)
Browsing				0.18 (0.00)	0.20 (0.00)
Type4					
Clicking					1.76 (0.05)
Browsing					0.17 (0.00)
Type probability					
Type1		0.91 (0.04)	0.91 (0.04)	0.05 (0.01)	0.05 (0.01)
Type2				0.06 (0.01)	0.04 (0.01)
Type3					0.20 (0.04)
Log-likelihood (N = 74,400)	-9,657.56	-9,244.66	-9,245.83	-9,147.45	-9,076.56
BIC	19,404.9	18,657.6	18,615.0	18,452.0	18,343.8
Akaike information criterion	19,331.1	18,519.3	18,513.7	18,322.9	18,187.1

Note. Bold indicates $p < 0.1$.

θ^D , if firms are less likely to advertise when there is an increase in demand, this implies a higher δ .

5. Results

5.1. The Consumer Model

Table 5 presents the consumer model results. The first column reports the parameter estimates of the homogeneous model. Estimates from the preference utility model indicate that the price (an external attribute) and the number of pictures (an internal attribute) affect consumers' preferences, and thus consumers' browsing, clicking, and purchase behaviors. Both the browsing and clicking costs significantly affect the length and depth of search and the formation of the consideration set. The second column in Table 5 reports the results from a two-segment model where the heterogeneity is imposed on both preference and cost. The third to fifth columns report results from the model with two to four segments where the heterogeneity is imposed only on the cost parameters. The four-segment model with heterogeneity on the cost parameters yields the best result in terms of the Bayesian information criterion (BIC).³⁴ About 71% of the consumers belong to the group with the browsing-cost estimate of 0.17 and the clicking-cost estimate of 1.76. About 20% (4%) of the consumers browse considerably more (less) but click less (more), and about 5% of the consumers browse and click more than the majority. The average marginal cost of browsing and clicking are \$0.89 and \$3.90,³⁵ respectively, but there exists considerable heterogeneity (ranging from \$0.87–\$0.92 for browsing and \$2.39–\$4.41 for clicking costs).³⁶ In-sample and out-of-sample model fits are reported in Table C.6 in the online appendix.³⁷

5.2. The Advertiser Model

Table 6 details the estimates from the advertiser model. To accommodate diminishing marginal returns for

clicks and impressions as discussed in Section 2.3, these variables are log-transformed. Additionally, a number of covariates control for various product types' observed differences in advertising rates, apart from their impact on consumers' browsing, clicking, and demand responses. For example, consumers' behavior is not responsive to different product materials, conditional on the variables entering the consumer model. However, the sellers systematically advertise stone-made products more frequently in our data, which suggests that the competition might be more intense with this type of product.

Of note, advertisers in this online marketplace face negative valuations from demand when opting in to advertise, owing to (i) high commissions from transactions and advertising (f^T, f^A); and (ii) the high value for δ , which captures the marginal cost. As the commissions from transaction and advertising constitute a large portion of the cost, with $f^T + f^A = 17\% + 13\% = 30\%$, the resulting valuation from demand is negative when sellers advertise ($100 - 17\% - 13\% - 74\% = -4\%$ of the transaction amount). This loss presumably motivates sellers to redirect consumers' purchases to outside channels (to their own websites or stores) to avoid paying high commissions on sales or promote buyers' web-rooming behavior.

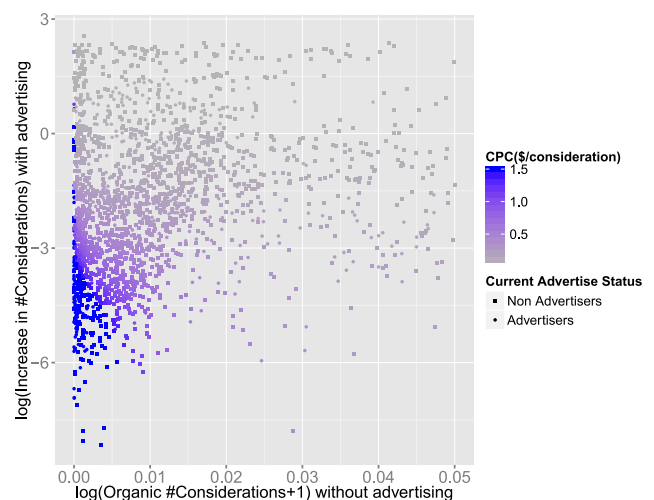
To assess when the valuations from clicks are highest, Figure 8 plots the increase in logged clicks from advertising on the y-axis and the number of logged clicks conditioned on not advertising on the x-axis (holding others' advertising decisions fixed). Each dot represents a listed product in the data. The color of the dots indicates the valuation per consideration (click) calculated based on the estimate θ^C and adjusted to be in dollar metric. The shape of the dots indicates the observed advertising decisions

Table 6. The Advertiser Model Estimates

Parameter	Estimates
θ^D Demand	0.620 (0.37)
θ^C log(Clicks)	0.017 (0.005)
θ^I log(Impressions (in thousand))	0.000(0.000)
δ Marginal cost	0.740 (0.03)
θ Constant	-1.010(0.06)
Include URL	0.280 (0.06)
Silver	0.440 (0.15)
Stone	0.200 (0.09)
Bracelet	0.290 (0.07)
Refundable	0.260 (0.06)
Seller fixed effects	Y
LL (N = 2,853)	-850.0

Notes. Negative valuation from demand if advertised: $(1 - f^T - f^A - \delta) = -0.04$. Bold indicates $p < 0.1$.

Figure 8. Valuations from Consideration (Click) Avg value per consideration quantiles: \$0.04 (25%), \$0.13 (50%), \$0.48 (75%)



in the data, where the squares (rounds) represent currently “nonadvertising” (“advertising”) products. The product observations with close to zero clicks in the absence of advertising have higher valuations from a unit increase in click (darker-colored dots) and are more likely to advertise. In other words, the first few clicks generate the largest valuations to advertisers. The quantiles for average value per click are \$0.04 (25%), \$0.13 (50%), and \$0.48 (75%). The average conversion rate ($\# \text{total demand} / \# \text{total clicks}$) in our data is 5%, so the cost per conversion is calculated to be \$2.60. As the median price is \$14, the total willingness to pay for clicks is about 18.6% of the transaction amount.³⁸ Although the results suggest that advertisers accrue valuations from clicks beyond valuations from purchases, we find that advertisers rarely gain valuations from impressions. This is consistent with the findings in Chan and Park (2015), where the value per impression is found to be zero in the context of a leading search-engine firm in Korea.

6. Policy Simulation

Owing to the structural underpinning of the models of consumer and advertiser behavior, it is possible to explore options by which the platform can improve its revenue and/or welfare of consumers and advertisers. On the consumer side, we explore how product-ranking decisions (e.g., sorting by consumers’ utility, price, past sales, or expected revenue) affect consumers’ browsing (impressions), consideration (clicking), and choice (purchase) of merchant goods. On the supply side, we explore how payment mechanisms (CPM, CPC, and CPA) and ranking rules together affect consumer and advertiser behaviors and welfare. We detail these policy analyses below.

6.1. Simulation Procedures

Details on the policy-simulation procedures are included in Online Appendix C.2.3. Of note here, we update consumers’ beliefs (state transitions $f_1^s(\mathbf{Z}_{j(t+1)} | \mathbf{Z}_{j(t)})$) in simulations to account for the changes in either the platform’s ranking algorithm or the aggregate consumers’ behaviors with respect to the changes in sellers’ advertising decisions. For example,

if the ranking algorithm changes, consumers’ beliefs about the characteristics of the next product to be potentially considered should also change. Second, as the ranking algorithm changes, sellers’ beliefs on their own and others’ product positions will also change. Thus, to account for the competitive responses, we construct sellers’ counterfactual beliefs per Equation (16) under the new ranking algorithm. One advantage of the structural approach over a simpler model is that it explicitly captures changes in consumer and advertiser beliefs. Lastly, changing the ranking or the fee structure may affect sellers’ listing behavior (i.e., a seller will unlist an item if the expected fees are higher than the expected gains). To account for the change in the seller’s listing behavior, we impose a participation constraint for the advertiser model simulations that each seller’s utility is greater than the minimum of the seller utilities estimated in the actual fee-structure setting in each iteration step. Those who gain lower than this threshold are assumed to drop out (delist items).

6.2. Consumer Model Simulations

6.2.1. The Effect of the Marketplace Ranking Algorithm on Browsing, Clicking, and Purchase. Although featuring advertised products generates advertising revenue, it can also impede search, thereby reducing transaction commissions. This leads to a trade-off between advertising revenue and sales commissions that can be considered using our model. Hence, we contrast the current ranking scheme with one that orders products by (i) utility level, (ii) price (from lowest to highest), (iii) past sales (volume), and (iv) expected revenue (i.e., expected item demand \times item price).³⁹ In each simulation, we first measure consumer response, then revenue implications for the platform, holding seller response fixed.

Table 7 suggests that ranking goods by consumer preference, price, or past sales volume generates increased consumption utility, u_{ij} , relative to the current ranking algorithm that favors advertised goods in the rankings. Specifically, consumers’ choice utility increases by (165%, 57%, 15%) and the number of items sold by (120%, 35%, 14%) when sorting by utility,

Table 7. Consumer Response

Consumer response (% change)	Ranking rule			
	Utility	Price	Past sales (volume)	Revenue
Browsing	−1.98	0.68	−0.53	−0.42
Clicking	−1.4	1.07	−0.21	−0.89
Purchase probability	119.8	35.4	13.8	−2.49
Average price purchased	−58.4	−86.3	−4.66	207
Total search costs (browsing + clicking)	−1.97	0.71	−0.52	−0.44
Choice utility	164.9	56.6	14.6	−1.51
Overall utility	2.04	−0.68	0.53	0.44

price, and past sales volume, respectively. On the other hand, sorting by expected revenue decreases both the number of items sold and the consumers' choice utility.

Sorting products by preferences has two countervailing effects on search behavior. On the one hand, consumers may browse/click less if they find the best item early in the search process. On the other hand, consumers may browse/click more if the expected future benefit is high. When products are sorted by consumers' utility, the former effect dominates, as consumers' browsing (clicking) decreases by 2% (1%). Combined, the effect of decreased search costs (browsing and clicking costs) from finding the preferred item sooner and the increase in choice utility from finding a better item leads to an overall utility increase of 2% when sorting by utility.⁴⁰

Although sorting products by consumer preferences can increase consumer welfare (and potentially transaction commissions), it can also lower revenue from advertising (i.e., sellers have no incentive to pay for advertising because there is no increase in rank position from advertising). Table 8 highlights this trade-off. Reordering items by consumers' utility or price decreases the commissions from transactions. This result is mainly driven by the fact that consumers are price-sensitive and purchase lower-price items displayed earlier in the product list. The increase in sales volume is not large enough to offset the decrease in the transaction commissions. As such, sorting by consumer's utility or price neither increases transaction revenues nor advertising revenues.

When sorting products by past sales volume, the increase in transaction commissions also does not offset the decrease in advertising commissions.⁴¹ Accordingly, the platform's profits decrease by 10% when sorting by past sales. Our analysis provides one insight regarding why many online marketplaces collect advertising fees and do not display items purely organically (i.e., by consumer's utility, price, or past sales) as a default ranking mechanism.

Sorting by expected revenue (i.e., expected demand \times price), on the other hand, increases the platform's profits, as the increase in transaction commissions is greater than the decrease in advertising commissions. This result suggests that the ranking algorithm (and fee structure) currently in place is suboptimal, thus

motivating the next question: How can the online marketplace better balance the trade-off between commissions and ad fees by changing the ranking algorithm and fee structure in a manner that accounts for both consumers' and advertisers' responses? We address this question next.⁴²

6.3. Advertiser Model Simulations

To assess how changes in the ranking algorithm and fee structure affect consumers' and advertisers' behaviors and the marketplace's profits, we conduct 4 simulations: (i) changing the product-ranking algorithm in isolation; (ii) changing the fee structure for clicks and transactions in isolation; (iii) changing both the ranking algorithm and fee structure together via auctions on clicks (CPC) coupled with displaying products by (expected clicks \times bids); and (iv) conducting auctions on clicks (CPC) in only the top 5 positions (i.e., limiting advertising slots) and changing the ranking algorithm to sort by expected revenue (i.e., expected demand \times price) in slots 6+. Counterfactual (i) focuses on rankings, counterfactual (ii) focuses on pricing mechanism, and counterfactuals (iii and iv) consider both.

6.3.1. The Effect of Increased Advertising Weight in the Marketplace Ranking Algorithm. Increasing the weight of advertising in the product-ranking algorithm will provide a greater incentive to advertise. This yields greater advertising revenue. On the other hand, to the extent the advertised goods do not align with preferences, advertising is more likely to disrupt search, thus yielding lower revenue from transactions. To explore this trade-off, we consider the case where the position of an advertised product is improved by 10% over the current policy by adjusting the weight in the ranking algorithm (which converts to a median increase of about 200 slots).

Consistent with a ranking algorithm that makes advertising more effective by increasing the lift in rank for advertised products, the mean advertising probability increases by 3%. The increased incentive to advertise is offset to some degree by the competitive response of other sellers, who are also likely to increase their advertising, thus mitigating the rank increase from advertising in the absence of such competitive response. Further, as competition intensifies, seller welfare falls 7.3%, reinforcing the importance

Table 8. Effect of Ranking Strategy on Platform Profits

Platform profits (% change)	Ranking rule			
	Utility	Price	Past sales volume	Revenue
Commissions from transactions (f^T)	—	—	+	+
Commissions from advertising (f^A)	—	—	—	—
Overall platform profits	-24.0	-84.6	-9.8	148.5

of capturing competition in the advertiser model. Overall, the increase in advertiser spending generates more revenue for the marketplace.

On the consumer side, however, consumers' browsing lengths, clicks, and purchases decrease by 0.3%, 0.5%, and 5%, respectively, and their ex-post consumption utility lessens by 3.2%. This negative effect on consumption utility can be explained by the finding that organically weaker (less popular) products have higher marginal valuation for advertising, and sellers are more prone to advertise these goods. In this regard, heavier weight on advertising disrupts consumers' search processes, as the likelihood of finding goods they want within their browsing lengths decreases.

Contrasting the two effects, we find that the effect of increased advertising revenue offsets the loss in transaction revenue on the consumer side and that the platform's profit increases by 3.5% due to this increase in commissions from advertising. In contrast, sellers' overall welfare decreases by 7.3%, as they face higher advertising competition and pay more for advertising commissions.

6.3.2. The Effect of the Marketplace Fee Structure: Combining CPA and CPC. As various fee structures differentially affect each stage of the purchase funnel (impressions, clicks, and purchases), a question of general interest is which pricing mechanism should be used by the online marketplace platform. Hence, we first explore the implication of a fixed cost-per-click basis and a percentage of the sale basis (cost-per-action) as a next counterfactual analysis keeping the current ranking algorithm.⁴³

To find the (pareto) optimal fee structure for this online marketplace platform, we conduct a coarse grid search combined with a steepest descent method on the profit objective function as a function of CPC and CPA fees. Findings are presented in the second column of Table 9. The optimal fee structure turns

out to be setting zero cost-per-action ($f_T = 0, f'_A = 0$) coupled with a more substantial \$0.35 charge for the click (CPC).⁴⁴ Although sellers are less likely to advertise (−11.2%), reducing CPA and instead charging advertising fees based on CPC (and/or CPM) has the potential for pareto improvement, leading to positive outcomes for both sellers and the platform. Sellers gain in overall welfare, as they do not face negative valuation on demand when advertising ($1 - f_T - f'_A - \delta = 1 - 0.13 - 0 - 0.76 = 0.11 > 0$). Intuitively, this finding suggests that the marginal fees of advertising ($f_A = 0.17$) are set too high under the current pricing scheme relative to the marginal gains from advertising and that advertiser valuations are better monetized via clicks.

6.3.3. The Effect of the Marketplace Fee Structure and Ranking: Auction on Clicks. Although the platform in consideration charges a fixed CPA (f^A), a common advertising fee structure adopted in practice is the generalized second-price auction on clicks (Edelman et al. (2007)). To explore the impact of such mechanism, we consider the following setting: The advertisers bid for clicks (CPC), and the platform ranks the products optimally by the rank score (i.e., expected clicks \times bid).⁴⁵ Advertising payment per click is set to be equal to (next highest rank score \div expected clicks).⁴⁶

In the third column of Table 9, the platform's profits increase, which is consistent with the theory that auction mechanisms can yield higher profits than the fixed pricing, especially when there are many bidders competing (Krishna 2009). Sellers are worse off, as the platform extracts more of the sellers' surplus, whereas consumers are better off, as the platform integrates the expected clicks (reflecting consumers' preference) into the product ranking.

6.3.4. The Effect of the Marketplace Fee Structure and Ranking: Combining CPA and Auction on Clicks. CPC auctions leverage fees from advertising while foregoing

Table 9. Effect of Platform Strategies on Consumers and Advertisers

Policy	Ranking rule	Advertiser side policy			
		+10%	–	Auction Revenue	Hybrid
Manipulation	Fee	—	$f'_c = 0.35$	CPC Auction	5 Positions
Response (% Change)	Consumer				
	Browsing	−0.3	0.7	−0.9	−0.3
	Clicking	−0.5	3.6	−0.9	−0.26
	Purchase Probability	−5.0	5.3	17.7	6.8
	Choice Utility	−3.2	7.2	24.8	15.0
	Advertiser				
	Prob (Ad)	2.6	−11.2	N/A	N/A
	Seller Welfare	−7.3	2.8	−8.7	−399
Platform					
	Profits	3.5	156	177	181

the revenues from transactions. We conjecture that the platform outcome might be further improved if we combine an auction pricing mechanism with a different ranking policy. Thus, similar to Amazon's current practice, we explore an alternative that combines transaction commissions with a click auction.⁴⁷ Specifically, we simulate a generalized second-price auction on clicks for the top 5 slots, while retaining a transaction commission level of ($f_T = 0.13$). The platform is assumed to rank the first 5 products by (expected clicks \times bid) and by the expected revenue (expected demand \times price) for the remaining list from slots 6 and lower (similar to Section 6.2). Note that this simulation is designed to enhance both transaction revenue and advertising revenue. Transaction revenue is enhanced by ranking slots 6 and lower, whereas advertising revenue comes from the sellers with the highest valuations for advertising.

The fourth column of Table 9 indicates that the platform's profits are the highest under this counterfactual and that almost all of the sellers' surplus are extracted by the platform, perhaps explaining the ubiquity of this "top slots" advertising mechanism in practice for online marketplaces. From this, we conclude that combining CPA and auction on clicks best balances the trade-off the platform faces between revenues from transactions and advertising. This strategy yields the largest profit gains for the platform.

7. Conclusion

This paper considers the monetization of online marketplaces. To achieve this aim, we consider all three agents in the two-sided network: (i) the platform that sets the advertising fees (CPM, CPC, and CPA) and placement of items listed on the market; (ii) the sellers who jointly make advertising decisions conditioned on the platform's policies and expected consumer behavior; and (iii) consumers who search (browse and click) and make purchase decisions given their preferences, search (browsing/clicking) costs, and the list of products displayed.

This research offers a number of advances with regard to the prior literature on consumer search and advertising in online environments. On the consumer side, our approach integrates browsing, clicking, and purchase behaviors in an online marketplace. On the seller side, we map each type of consumer engagement to the advertiser valuation thereof (CPM, CPC, and CPA) and model the strategic interactions of advertisers in response to the platform's ranking algorithm and fee structure.

On the consumer side, we find that price, the number of pictures, and clicking and browsing costs affect the length of search, formation of consideration set, and ultimately the products purchased by the consumers.

The average marginal cost of browsing and clicking are \$0.89 and \$3.90, respectively, and there exists considerable heterogeneity across consumers.

On the seller side, we find that the combined marginal cost of goods and opportunity costs of selling elsewhere for the sellers on this platform is substantial (74% of the selling price). As a result, the valuation from unit demand is negative (−4% of the transaction amount) for the sellers who advertise. This negative valuation is due to the high CPA-based advertising fees that may incentivize sellers to redirect consumers to buy their product on other venues. The median seller valuation from a click is estimated to be \$0.13, and sellers rarely gain positive valuations for impressions. In other words, sellers appear to value the potential for clicks more than selling an item on the marketplace, under the current fee structure.

On the platform side, we consider two strategies: changing the ranking algorithm and changing the advertising pricing mechanism. A trade-off between ad revenue and sales revenue must be balanced in these strategies. For example, increased advertising can interrupt consumer search, thereby leading to lower sales. Alternatively, advertising decisions, to the extent that they reflect product quality, can improve the consumer search experience (e.g., Sahni and Zhang 2019). With regard to ranking strategy, ordering products by consumer utility or from low to high price increases items sold but decreases platform profits, as those items that are sold are lower-price items relative to the prices of goods sold under the current ranking algorithm. Although sorting by past sales increases transaction commissions, sorting also decreases the platform's profits due to a decrease in advertising fees. On the other hand, listing items by expected revenue enhances platform profits as the increase in transaction commissions is the greatest.

With regard to the platform's pricing strategies, reducing CPA while charging advertising fees based on CPC (and/or CPM) has the potential for pareto improvement, wherein both advertisers' welfare and the platform's profits increase. This strategy also lowers the likelihood that advertisers will list items to gain clicks (possibly in the hope of own-site future sales) while hoping not to sell them on the platform. The platform can further enhance its revenue in equilibrium by auctioning the top 5 positions (i.e., limiting the advertising slots) based on CPC pricing, then ordering by expected revenue from position 6 and lower. Limiting the advertising slots extracts the rents from the advertisers with the highest valuations, and ordering items by expected revenue for slots 6 and below generates greater returns from sales—thus helping revenue on both sides of the platform.

Although this paper investigates a broad range of interactions among buyers, sellers, and the platform

in an online marketplace platform, a number of additional extensions are possible. First, on the buyer side, one can extend our model to incorporate the consumer's site-visit incidence decision that can depend on which sellers advertise and how the platform ranks advertised versus organic products. Another possible extension is to consider cross-category and cross-store browsing, clicking, and purchase, which will yield some novel insights on platform strategies. In addition, consumers may consider nonclicked items, thereby forming latent consideration sets. Future research is also warranted regarding which information should be presented to consumers on the product-listing page versus the product-detail page.⁴⁸

Second, sellers' pricing behavior is taken as given in our policy simulation, and we do not consider competition between e-commerce platforms. We believe this is a reasonable assumption in our empirical application where price is not found to be correlated with advertising decision and the varying fee structure of other platforms. Nonetheless, marketing implications of multihoming in two-sided online marketplaces represent an important direction for future analysis. With multihoming consumers, cross-promotion and advertising can produce potential benefits.

Last, our focus is on online merchandising platforms. The search model can also be applied to blogs and social media websites where visitors search a list of article titles in a top-to-bottom sequence and decide which ones to click on and read further. The search model is also suitable to the growing mobile-commerce environment, where only one or two products are visible on a screen and consumers scroll down in top-to-bottom fashion while deciding which products for which to gather further information. Presumably, the advertiser model could be applied to these contexts as well. Given the relatively nascent state of empirical research on online transactional platforms, we hope that our work will serve as a useful step in this rapidly growing context.

Acknowledgments

The authors thank Peter Arcidiacono, Bryan Bollinger, Garrett Johnson, Chris Nosko, Emily Wang, Sha Yang; and seminar participants at the 2016 International Choice Symposium, the 2016 Economics of Advertising Conference, the 2017 Summer Institute in Competitive Strategy, the 2017 National Bureau of Economics Research, the University of Alberta, Boston University, the University of Chicago, the University College London, the University of Colorado Boulder, Columbia University, Duke University, Duke-UNC Brown Bag, Emory University, Erasmus University, Harvard University, the University of Michigan, the University of Minnesota, Northeastern University, the University of Pittsburgh, the University of Rochester, the University of Southern California, Stanford University, and Yale University for comments and suggestions.

Endnotes

¹ Other examples of online marketplaces include Etsy, Yahoo! Shopping, eBay, Overstock, JD.com, CafePress, Zazzle, Oodle, eCrater, Bonanzle, and Fancy.

² See <http://www.statista.com/statistics/259782/third-party-seller-share-of-amazon-platform/>.

³ See <https://www.digitalcommerce360.com/article/infographic-top-online-marketplaces/>.

⁴ See <https://www.outerboxdesign.com/web-design-articles/mobile-e-commerce-statistics>.

⁵ Amazon, for example, charges 15% of the transaction price on average plus \$0.99 per item (or pay a monthly subscription fee of \$39.99, and the \$0.99 per item fee is waived).

⁶ Amazon uses an auction-based pricing model for each keyword, similar to keyword search engines. Etsy asks sellers to list several keywords and set one weekly maximum budget. Both charge sellers on a cost-per-click basis. On the other hand, the website in our empirical application asks sellers about the willingness to pay an extra 17% of the transaction price, and the platform has full discretion on how the sponsored products are displayed.

⁷ In some regards, this practice is similar to keyword advertising, which limits the number of sponsored search positions and orders the search ads by expected revenue.

⁸ The format is similar to Facebook's news feed or the design of the online marketplace Fancy (<https://fancy.com/>).

⁹ This assumption would lead to an underestimation of browsing costs should consumers actually browse fewer than the 10 last loaded items. We conduct a sensitivity analysis using the "overlay" data. The website also observes an "overlay" request when the consumer places the mouse pointer on top of the product picture. Hence, instead of assuming that the consumer browses all 10 items loaded last, we can define the last browse as the last overlay within that set. This alternative approach yields an upper bound on the browsing costs, and our estimates are robust to this alternative approach for inferring the end of browsing.

¹⁰ Although it is feasible to consider shopping across all categories, the problem becomes substantially more complex with little attendant insight. In this regard, restricting our attention to one portion of the site is much like other research that focuses on a single category rather than a choice across a basket of goods. Further discussion is included in Online Appendix A.1.

¹¹ If ads are clearly delineated, adding a sponsored ad indicator to the utility function can assess whether marks have an effect over and above rank (Sahni and Nair 2019). To the extent that ad indication affects consumer utility, sellers' expectations on consumers' responses, competition, and profits will change and, ultimately, sellers' decisions of whether to advertise.

¹² In the extreme case when every seller advertises, the resulting position will be the same as the organic position where no one advertises.

¹³ We define a new visit (search session) if a user comes to the website for the first time, is inactive for 24 hours, changes the category, or continues to search after purchasing an item.

¹⁴ Including users both with and without purchase (i.e., 72,030 users with 85,632 visits) yields qualitatively similar managerial implications. The full-sample estimation results and detailed comments are available in Online Appendix D.

¹⁵ The average conversion rate for an e-commerce website in Quarter 1 2017 in the United States is around 2.46% and internationally 2.48% (<http://www.smartinsights.com/e-commerce/e-commerce-analytics/e-commerce-conversion-rates/>). The conversion rate (#total demand/#total visits) in our sample is higher and is about 4.2%.

¹⁶ Ursu (2018) also finds that conditional on a click, higher rank does not generate more purchases.

¹⁷ A considerable literature supports top-to-bottom search behavior (Ansari and Mela 2003, Granka et al. 2004), and top-down can be rationalized when consumers search optimally by inferring an advertiser's quality from the position (Athey and Ellison 2011, Chen and He 2011). As such, the top-down search behavior assumption is often invoked in the sponsored search context (Aggarwal et al. 2008, Kempe and Mahdian 2008, Chan and Park 2015). We similarly adopt this assumption in our online marketplace context.

¹⁸ The percentage of visits with deviations from top-down click behavior conditional on multiple clicks, 21% ($= 1.78/8.4$), is lower than reported in Jeziorski and Moorthy (2017) (28%) and much lower than Jeziorski and Segal (2015) (57%). In Jeziorski and Moorthy (2017), brand prominence largely influences consumers searching for cameras. We conjecture that handmade goods predominantly include sellers (=brands) with few listings and little brand recognition, which may in part explain why our data exhibit stronger evidence for top-to-bottom browsing/clicking behavior.

¹⁹ Two potential reasons regarding why advertising decisions rarely change are as follows. First, it is possible for an advertiser to consider future outcomes but only upon the initial listing decision based on the net present value of the advertising decision. Second, there are potentially substantial costs to monitoring the states of the market each day to change advertising over the duration of a listing. Should these costs be sufficiently high, it might suffice to make a decision once and then not deviate from this initial choice.

²⁰ In Online Appendix A.2.2, we document some of the important observables that suggest different advertising valuations across products, and in Online Appendix A.2.3, we briefly discuss advertisers' pricing decisions. A key insight from this analysis is that advertising strategy appears independent of price, suggesting the plausibility of an exogenous pricing assumption.

²¹ Like many scanner-data papers, we focus on what happens conditional on store visit and take the shopping trip decision as given. With this simplifying assumption, we take the market size (consumer visits) to be fixed for the counterfactual exercises. Regarding the assumption of independence across visits, the website currently does not retarget or customize results by person. Using linear models of current browsing length, number of clicks, and purchase decisions regressed on past behaviors, we find no evidence of state dependence, suggesting that decisions are independent across sessions.

²² The platform does not provide refinement options like sorting and filtering on the main landing page. As such, our consumer model abstracts away from sorting and filtering decisions and instead considers the effect of sorting by price and past sales as counterfactual exercises.

²³ We define "consideration set" to be the set of products over which the consumer actively seeks (via clicking) information on attributes that enter their consumption utility when evaluating the items for purchase.

²⁴ Our approach is inspired by the oblivious equilibrium (Weintraub et al. 2006) and the approximate aggregation in Krusell and Smith (1998) and Lee and Wolpin (2006), but we consider a static environment. Recently, this method has also been adopted in analyzing ad exchange auctions (Iyer et al. 2014, Balseiro et al. 2015, Lu and Yang 2016).

²⁵ In Online Appendix A.2.3, we show that seller pricing is not correlated with the advertising decision. Because products are usually sold via multiple sales channels, it is plausible that the advertising strategy on this web platform is independent of the pricing decision set for all sales channels. Hence, we treat price as exogenous (which also has the benefit of substantially simplifying the supply-side analysis).

²⁶ In Online Appendix Section A.2.2, we show that firms who include a link to their own websites tend to advertise more, a finding suggestive of greater valuations for those who can more readily redirect exposed customers to their own sites and avoid paying cost-per-action fees to the platform.

²⁷ In Figure 3, as #browses (and #clicks) decrease exponentially with position, firms with products ranked closer to the top will have a much higher increase in #impressions and #clicks from advertising (e.g., going from position 10 to 1 will yield a much higher increase in #impressions and #clicks than going from position 1,000 to 990). Were advertiser valuations linear in impressions and clicks, advertisers organically positioned higher would advertise more because they gain more #impressions and #clicks from advertising. However, we find that the opposite holds (Figure 5), suggesting marginally decreasing returns from clicks and impressions. We conjecture that the information value of advertising becomes marginally lower as consumers become more aware of the products (Blake et al. 2015) and/or that sellers might value the first few clicks and impressions highly to the extent the first few purchases cover the fixed costs of production. Advertisers gain incremental clicks and impressions from high-search-cost consumers, and their search and purchase likelihoods might be lower. These rationales suggest the potential for diminishing marginal returns in our advertiser valuation model.

²⁸ The inclusion of category–seller-level unobservables and exclusion of product-level unobservables is motivated by data limitation and not by the functional form restrictions required for identification. In estimation, we add a category dummy for accessories (e.g. necklace, ring, or bracelet) and a dummy for large sellers (brands with more than 150 product listings).

²⁹ In the absence of $\eta_{i,t}^c$, the clicking decision becomes a deterministic function of the states. As such, the error is necessary to rationalize the act of clicking an item later in search results when a similar item that appeared earlier in search results is not clicked.

³⁰ In our data, the last click rarely coincides with the last browse. The percentage of visits with consumers browsing >10 more items even after the last click is 96%.

³¹ See Seiler (2013, p. 183) for a similar discussion, where a separate set of T1EV error terms are introduced for each decision stage in order to obtain an analytic solution for the value functions.

³² To validate this assumption, we show that the actual ranking by the platform's algorithm and the approximate ranking based on Equation (16) yield similar predictions, even though the latter assumes smaller information demands on the part of the advertiser (Figure C.3 in Online Appendix C.2.1).

³³ As we aggregate data to the product level, the identification of the diminishing marginal returns is achieved by assuming a common parameter, θ 's and δ , across sellers or across products within a seller.

³⁴ The BIC of the five-segment model is 18,377.

³⁵ Each segment's marginal cost is calculated by using a dollar metric weighted by the user segment-type probability. For example, for the 4-segment model average marginal cost of clicking $= \sum_g \Pr(\text{type}_g) \times \frac{\exp(\gamma_g^c)}{\exp(0.29)}$, where 0.29 is the coefficient for log(price).

³⁶ Chen and Yao (2017) report a click cost of about 13% of the average hotel price ($= \$21.54/\169) and a marginal browsing cost (as inferred from the slot coefficient in their model) of about \$1.01 ($= \exp(0.01)$). In our case, the marginal click cost is about 20% of the average product price ($= \$3.90/\19.5), and the marginal browsing cost is \$0.89. These numbers are quite close, with the differences reflecting more browsing and less clicking observed in our data (i.e., average clicks are 0.8 in our data as compared with 2.3 in Chen and Yao 2017).

³⁷ An alternative model, wherein search is modeled myopically (i.e., the discount factor is set to 0 at the browsing decision step, implying aimless consumer search), deteriorates model fit markedly,

with a substantially lower log-likelihood (−14,443). The lower fit suggests that consumers are forward-looking, incurring search costs in return for future gains.

³⁸ Related, in keyword-sponsored search context, Yao and Mela (2011) estimate the mean value of a click to be \$0.25 for software products with a typical retail price of \$22, and our click valuation is consistent with their findings.

³⁹ When ordering products by utility level, the available (listed) products are sorted by the choice utility (consumption utility) in Equation (1) based on the consumer model estimates. As the consumer model preference parameters in our empirical context are estimated to be from one segment, this sorting leads to a single-product display ranking across consumers. Thus, we do not consider rankings customized to the individual consumer.

⁴⁰ In Chen and Yao (2017), the average utility of hotels booked increases by 17% with the refinement tool (sorting/filtering) as compared with without one. The larger percentage gains in choice utility (165%, 57%, 15%) in our context arise from the default ranking system, which does not emphasize consumer preferences in the scoring algorithm. The (baseline) default ranking is predominantly influenced by “days listed (i.e., sorting by newest to oldest; Figure 2), followed by advertising and popularity scores. As a result, consumer utility is relatively low to start, enabling large potential gains. In contrast, Chen and Yao (2017, p. 4360) mention that “the default ranking of hotels is based on booking frequencies, which to some extent already reflects the average utility levels of these hotels among population. Consequently, even without refinement tools, the baseline level of consumer welfare is fairly high if consumers make decisions according to the default ranking.”

⁴¹ Note that past sales are not only correlated with consumers’ utility but also with sellers’ advertising decisions and the platform’s ranking algorithm in the past. Therefore, the results for sorting by past sales can differ from sorting by utility.

⁴² In calculating the platform’s profits for the counterfactual setting, we set $f^A = 0$, as advertising has no effect on ranking.

⁴³ As sellers in our empirical context rarely gain valuations from impressions and thus CPM, we focus our attention on CPA (purchase) and CPC (click) while setting CPM f^I to be zero.

⁴⁴ Although in a different context, our result is consistent with the average CPC (\$0.35) for Facebook Advertising in Korea (<http://www.rudibedy.com/blog/facebook-advertising-cpc-cpm-per-country/>).

⁴⁵ Other fees are set to be zero ($f_T = 0, f_A = 0, f_I = 0$) in this exercise.

⁴⁶ We model advertisers having diminishing marginal returns on clicks, and for simplicity, we assume that the bid equals the mean valuation (i.e., total expected valuation/total expected clicks).

⁴⁷ Amazon charges 15% of the transaction price on average as transaction commissions and uses an auction-based CPC pricing model for the limited top slots.

⁴⁸ Changes in the set of attributes presented on the product-listing page versus the product-detail page may affect browsing and clicking costs. With further variation in data (e.g., exogenous variation in which content is present on the product listing page versus the product-detail page), the consumer model could be extended to incorporate these potential changes in costs.

References

- Aggarwal G, Feldman J, Muthukrishnan S, Pál M (2008) Sponsored search auctions with Markovian users. *Internat. Workshop Internet Network Econom.* (Springer, Berlin), 621–628.
- Ansari A, Mela CF (2003) E-customization. *J. Marketing Res.* 40(2): 131–145.
- Arbatskaya M (2007) Ordered search. *RAND J. Econom.* 38(1):119–126.
- Arcidiacono P, Miller RA (2011) Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity. *Econometrica* 79(6):1823–1867.
- Armstrong M, Zhou J (2011) Paying for prominence. *Econom. J. (London)* 121(556):F368–F395.
- Armstrong M, Vickers J, Zhou J (2009) Prominence and consumer search. *RAND J. Econom.* 40(2):209–233.
- Athey S, Ellison G (2011) Position auctions with consumer search. *Quart. J. Econom.* 126(3):1213–1270.
- Balseiro SR, Besbes O, Weintraub GY (2015) Repeated auctions with budgets in ad exchanges: Approximations and design. *Management Sci.* 61(4):864–884.
- Blake T, Nosko C, Tadelis S (2015) Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica* 83(1):155–174.
- Bodapati AV (2008) Recommendation systems with purchase data. *J. Marketing Res.* 45(1):77–93.
- Bronnenberg BJ, Kim BJ, Mela CF (2016) Zooming in on choice: How do consumers search for cameras online? *Marketing Sci.* 35(5): 693–712.
- Chan TY, Park YH (2015) Consumer search activities and the value of ad positions in sponsored search advertising. *Marketing Sci.* 34(4):606–623.
- Chen Y, He C (2011) Paid placement: Advertising and search on the Internet. *Econom. J. (London)* 121(556):F309–F328.
- Chen Y, Yao S (2017) Sequential search with refinement: Model and application with click-stream data. *Management Sci.* 63(12): 4345–4365.
- Damiano E, Li H (2007) Price discrimination and efficient matching. *Econom. Theory.* 30(2):243–263.
- De los Santos B, Koulayev S (2017) Optimizing click-through in online rankings with endogenous search refinement. *Marketing Sci.* 36(4):542–564.
- De los Santos B, Hortacsu A, Wildenbeest MR (2012) Testing models of consumer search using data on web browsing and purchasing behavior. *Amer. Econom. Rev.* 102(6):2955–2980.
- Dinerstein M, Einav L, Levin J, Sundaresan N (2018) Consumer price search and platform design in Internet commerce. *Amer. Econom. Rev.* 108(7):1820–1859.
- Edelman B, Ostrovsky M, Schwarz M (2007) Internet advertising and the generalized second-price auction: Selling billions of dollars worth of keywords. *Amer. Econom. Rev.* 97(1):242–259.
- Fradkin A (2019) A simulation approach to designing digital matching platforms. Preprint, submitted January 24, <https://ssrn.com/abstract=3320080>.
- Ghose A, Ipeirotis PG, Li B (2012) Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Sci.* 31(3):493–520.
- Granka LA, Joachims T, Gay G (2004) Eye-tracking analysis of user behavior in www search. *Proc. 27th Annual Internat. ACM SIGIR Conf. Res. Development Inform. Retrieval* (ACM, New York), 478–479.
- Guadagni PM, Little JD (1983) A logit model of brand choice calibrated on scanner data. *Marketing Sci.* 2(3):203–238.
- Gupta S (1988) Impact of sales promotions on when, what, and how much to buy. *J. Marketing Res.* 25(4):342–355.
- Hagiu A, Jullien B (2011) Why do intermediaries divert search? *RAND J. Econom.* 42(2):337–362.
- Hong H, Shum M (2006) Using price distributions to estimate search costs. *RAND J. Econom.* 37(2):257–275.
- Honka E (2014) Quantifying search and switching costs in the US auto insurance industry. *RAND J. Econom.* 45(4):847–884.

- Honka E, Chintagunta P (2016) Simultaneous or sequential? Search strategies in the US auto insurance industry. *Marketing Sci.* 36(1): 21–42.
- Honka E, Hortaçsu A, Vitorino MA (2017) Advertising, consumer awareness, and choice: Evidence from the US banking industry. *RAND J. Econom.* 48(3):611–646.
- Horton JJ (2017) The effects of algorithmic labor market recommendations: Evidence from a field experiment. *J. Labor Econom.* 35(2):345–385.
- Huang JT (2018) Visibility policy, seller incentives, and pricing dynamics in a digital goods marketplace. Unpublished doctoral thesis, Stanford University, Stanford, CA.
- Iyer K, Johari R, Sundararajan M (2014) Mean field equilibria of dynamic auctions with learning. *Management Sci.* 60(12):2949–2970.
- Jeziorski P, Moorthy S (2017) Advertiser prominence effects in search advertising. *Management Sci.* 64(3):1365–1383.
- Jeziorski P, Segal I (2015) What makes them click: Empirical analysis of consumer demand for search advertising. *Amer. Econom. J. Microeconom.* 7(3):24–53.
- Kamakura WA, Russell G (1989) A probabilistic choice model for market segmentation and elasticity structure. *J. Marketing Res.* 26(4):379–390.
- Kempe D, Mahdian M (2008) A cascade model for externalities in sponsored search. Papadimitriou C, Zhang S, eds. *International Workshop on Internet and Network Economics*, Lecture Notes in Computer Science, vol. 5385 (Springer, Berlin), 585–596.
- Kim JB, Albuquerque P, Bronnenberg BJ (2010) Online demand under limited consumer search. *Marketing Sci.* 29(6):1001–1023.
- Kim JB, Albuquerque P, Bronnenberg BJ (2016) The probit choice model under sequential search with an application to online retailing. *Management Sci.* 63(11):3911–3929.
- Koulayev S (2014) Search for differentiated products: identification and estimation. *RAND J. Econom.* 45(3):553–575.
- Krishna V (2009) *Auction Theory* (Academic Press, New York).
- Krusell P, Smith AA Jr (1998) Income and wealth heterogeneity in the macroeconomy. *J. Political Econom.* 106(5):867–896.
- Lee D, Wolpin KI (2006) Intersectoral labor mobility and the growth of the service sector. *Econometrica* 74(1):1–46.
- Long F, Jerath K, Sarvary M (2018) The informational role of sponsored advertising on online retail marketplaces. Working paper, Columbia University, New York.
- Lu S, Yang S (2016) A two-sided market analysis of behaviorally targeted display advertising. Working paper, University of Houston, Houston.
- Mehta N, Rajiv S, Srinivasan K (2003) Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Sci.* 22(1):58–84.
- Mojir N, Sudhir K (2019) Price search across time and across stores. Cowles Foundation Discussion Paper No. 1492R, Harvard Business School, Boston.
- Moraga-González JL, Sándor Z, Wildenbeest MR (2018) Consumer search and prices in the automobile market. Working paper, VU University Amsterdam, Amsterdam.
- Resnick P, Varian HR (1997) Recommender systems. *Commun. ACM* 40(3):56–59.
- Sahni NS, Nair H (2019) Sponsorship disclosure and consumer deception: Experimental evidence from native advertising in mobile search. *Marketing Sci.*, ePub ahead of print July 8, <https://doi.org/10.1287/mksc.2018.1125>.
- Sahni NS, Zhang C (2019) Search advertising and information discovery: Are consumers averse to sponsored messages? Stanford University Graduate School of Business Research Paper No. 3441786, Stanford University, Stanford, CA. Accessed <https://ssrn.com/abstract=3441786>.
- Schafer JB, Konstan J, Riedl J (1999) Recommender systems in e-commerce. *Proc. 1st ACM Conf. Electronic Commerce* (ACM, New York), 158–166.
- Seiler S (2013) The impact of search costs on consumer behavior: A dynamic approach. *Quant. Marketing Econom.* 11(2):155–203.
- Sharma S, Abhishek V (2017) Effect of sponsored listings on online marketplaces: The role of information asymmetry. Preprint, submitted August 4, <https://ssrn.com/abstract=3013468>.
- Simonov A, Nosko C, Rao JM (2018) Competition and crowd-out for brand keywords in sponsored search. *Marketing Sci.* 37(2):200–215.
- Smith B, Linden G (2017) Two decades of recommender systems at amazon.com. *IEEE Internet Comput.* 21(3):12–18.
- Stigler GJ (1961) The economics of information. *J. Political Econom.* 69(3):213–225.
- Ursu RM (2018) The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Sci.* 37(4):530–552.
- Varian HR (2010) Computer mediated transactions. *Amer. Econom. Rev.* 100(2):1–10.
- Weintraub GY, Benkard L, Van Roy B (2006) Oblivious equilibrium: A mean field approximation for large-scale dynamic games. Weiss Y, Schölkopf B, Platt JC, eds. *Advances in Neural Information Processing Systems*, vol. 18 (Curran Associates, Red Hook, NY), 1489–1496.
- Weitzman ML (1979) Optimal search for the best alternative. *Econometrica* 47(3):641–654.
- Wilson CM (2010) Ordered search and equilibrium obfuscation. *Internat. J. Indust. Organ.* 28(5):496–506.
- Yao S, Mela CF (2011) A dynamic model of sponsored search advertising. *Marketing Sci.* 30(3):447–468.
- Yoganarasimhan H (2019) Search personalization using machine learning. *Management Sci.*, ePub ahead of print August 29, <https://doi.org/10.1287/mnsc.2018.3255>.
- Zhang S, Yao L, Sun A, Tay Y (2019) Deep learning based recommender system: A survey and new perspectives. *ACM Comput. Survey* 52(1):5.
- Zhou J (2011) Ordered search in differentiated markets. *Internat. J. Indust. Organ.* 29(2):253–262.