Appendix. Monetizing Online Marketplaces - Online Appendix

A. Data

A.1. The Buyers: Data Sample

The transactional site we consider has several "categories", including: i) a main landing page product feed with all variety of goods, ii) more specialized categories such as jewelry and handbag, and iii) various designer stores (brand stores). Hence, the considered site bears similarities to a retailer (such as a grocer or a department store) with many categories. We focus our attention on the main landing page category with the reasons discussed in sub-section 2.1.1 but provide some summary statistics for the entire platform across categories in this sub-section. Focusing on the individuals with purchasing history, Table A.1 presents the shares by (sub) categories. The main landing page category constitutes the largest share of visits and impressions and the second largest share of clicks and purchases.

Table A.1 Shares of Visits, Browses, Clicks and Purchases

Top Categories	% Visits	% Impressions	% Clicks	% Purchases
Main Page	18.9	21.8	12.5	12.3
Jewelry	7.4	9.8	6.5	5.8
Bracelet	5.8	16.0	13.4	9.1
Brand1	3.7	5.3	8.0	14.9
Clothes/Acc.	3.0	2.0	1.3	0.3
Brand2	2.5	2.4	4.3	10.0
Necklace	2.5	4.2	4.1	2.9
Ring	2.3	3.1	4.4	1.6
Brand3	2.0	3.4	2.4	3.9

A.2. The Advertisers

A.2.1. Product Listing and Advertising Decisions. Figure A.1 portrays the concentration of goods across sellers, the x-axis is the seller ID, and the y-axis is the number of products per seller. Most merchants are casual sellers with few listings (median 4), and there are only a couple of sellers with more than 50 items. We find our results to be robust to the exclusion of these large sellers. Overall, the non-concentrated nature of sellers suggests that each is sufficiently atomistic as to have little, if any, marginal impact on the observed advertising equilibrium outcomes.

A.2.2. Advertising Decisions. Table A.2 documents some observable aspects that suggest different valuations across products via a logit regression analysis of the products' advertising status against price and other observables. The first column of the table reports the estimates when price and shipping fees are separately included as covariates, whereas the second column considers the effective price (price + shipping fees). Price is not statistically significant in either specification, suggesting that pricing strategy is not primarily driven in relation to the advertising decision. Further discussion on pricing strategy is included in the next sub-section A.2.3.

Past sales is operationalized as an indicator that assumes a value of one if there have been prior sales of the item. Consistent with our previous discussion, popular products who can organically appear early in the

Figure A.1 Product Listings

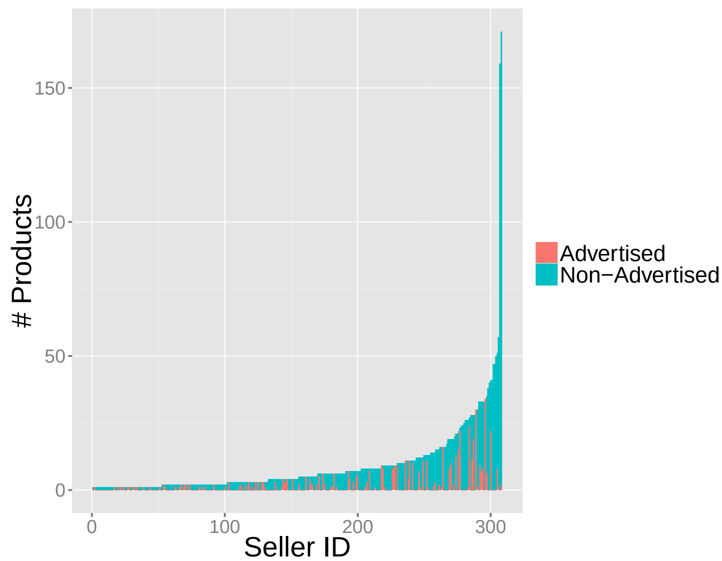


Table A.2 Other Observables and Advertising Decisions

DV: Advertising Decision (0/1)	Specification (1)	Specification (2)
Constant	0.16 (0.33)	**−0.60** (0.17)
Price	0.001 (0.002)	–
Shipping Fees	**−0.28** (0.11)	–
Effective Price	–	−0.0001 (0.002)
#Total Purchases >0	**−0.41** (0.17)	**−0.41** (0.17)
Include URL	**0.52** (0.11)	**0.52** (0.11)
Refundable	**0.48** (0.11)	**0.46** (0.10)
Quantities Available	0.0002 (0.0003)	0.0003 (0.0003)
Log(# Products per Brand)	**−0.44** (0.05)	**−0.44** (0.05)
Category Fixed Effects	Y	Y
Material Fixed Effects	Y	Y
BIC ($N=2853$)	2747.7	2748.0

search order advertise less (presumably because of the decreasing marginal returns to exposures). Analyzing text description in the product detail pages reveals that about 20% of the sellers include URL or sellers' website addresses to explicitly nudge consumers to redirect, and these are the sellers who seem to benefit more from advertising.

In addition, the products with non-refundable policy advertise less, and the estimates for category/material dummies suggest that bracelets, silver, and stone products are advertised more frequently. Last, the probability of advertising a given product is lower for sellers with many listings because the information value of advertising is likely lower for larger advertisers (Blake et al. (2015)). To control for the difference between casual sellers and big sellers, we add seller fixed effects in estimating the advertiser model.

A.2.3. Pricing Decisions.

Seller Pricing and Seller Advertising. Although seller pricing decisions are beyond the scope of this paper, our analysis presumes that the correlation between pricing and advertising decisions are modest.

Accordingly, we conduct an additional regression analysis of price on advertising status and include seller fixed effects (see Table A.3). We find no significant relationship, suggesting the plausibility of the exogenous pricing assumption we employ.

Table A.3 Pricing and Advertising Decisions

DV: Price	Estimates		
Constant	**19.5** (0.49)	**25.5** (2.52)	**25.5** (2.43)
Advertising (Opt In == 1, Opt Out == 0)	−0.85 (1.10)	−0.76 (1.25)	−0.89 (1.20)
Seller Fixed Effects	–	Y	Y
Material Fixed Effects	–	–	Y
Category Fixed Effects	–	–	Y
Adjusted $R^2 (N = 2853)$	0.000	0.61	0.64

Seller Pricing and Platform Fees. In our analysis we have presumed pricing strategy is exogenous to the decisions made by the marketplace platform. However, under the counterfactual scenario in which the fee structure is changed, it is possible that sellers significantly raise/lower prices in response to the changes in fee structure. To address this concern, we collect additional price information for 513 products we find listed in sellers' own websites or other selling channels (e.g. general e-commerce platforms, websites, or mobile apps). We note that the seller does not pay commissions if a product is sold on its own website but incur fees in various amounts if it is sold elsewhere. Accordingly, in Table A.4, we regress log price on own website dummy and product level fixed effects. The coefficient for own website dummy is not significant (with only 1.9% change in price), supporting our modeling assumption that a single price is exogenously set across all selling channels and is not adjusted in response to the different levels of fees imposed in different platforms.

Table A.4 Pricing across Selling Channels

DV: log(Price)	Estimates
Constant	**2.71** (0.11)
Own Website Fixed Effect	−0.019 (0.020)
Product Fixed Effects	Y

B. Model

B.1. State Transitions and Consumer Beliefs

We assume that consumers know the distribution of product characteristics available on the site and formulate rational beliefs based on product attribute transition. The states on product attribute transition include external attributes $\mathbf{Z}$ and internal attributes $\mathbf{X}$. Conditioned on the product attribute transition, the consumer's belief system can be characterized by the maximum utility of the items in the consideration set, u^*, and the information available on the product listing pages, $\mathbf{Z}$.

Attribute State Transitions. Let T be the total number of products available on the site and $h\left(\mathbf{Z}_{j(1)}, \mathbf{X}_{j(1)}, ..., \mathbf{Z}_{j(T)}, \mathbf{X}_{j(T)}\right)$ be the joint distribution of product attributes. In the context we consider, all consumers are presented with the same order of products, thus the distribution is not subscripted by i. To factor h, we assume a first order Markov process on $\{\mathbf{Z}_j, \mathbf{X}_j\}$ such that

$$\begin{aligned} h\left(\mathbf{Z}_{j(1)}, \mathbf{X}_{j(1)}, ..., \mathbf{Z}_{j(T)}, \mathbf{X}_{j(T)}\right) &= h(\mathbf{Z}_{j(1)}, \mathbf{X}_{j(1)}) \prod_{t=2}^{T} h(\mathbf{Z}_{j(t)}, \mathbf{X}_{j(t)} \mid \mathbf{Z}_{j(t-1)}, \mathbf{X}_{j(t-1)}) \\ &= h_1(\mathbf{X}_{j(1)} \mid \mathbf{Z}_{j(1)}) h_2(\mathbf{Z}_{j(1)}) \prod_{t=2}^{T} h_1(\mathbf{X}_{j(t)} \mid \mathbf{Z}_{j(t)}, \mathbf{Z}_{j(t-1)}, \mathbf{X}_{j(t-1)}) h_2(\mathbf{Z}_{j(t)} \mid \mathbf{Z}_{j(t-1)}, \mathbf{X}_{j(t-1)}) \end{aligned}$$

To simplify h_1, we assume that $\mathbf{Z}_{j(t)}$ is a sufficient statistic for $(\mathbf{Z}_{j(t-1)}, \mathbf{X}_{j(t-1)})$ in predicting $\mathbf{X}_{j(t)}$. That is, conditional on having the information on $\mathbf{Z}_{j(t)}$, $(\mathbf{Z}_{j(t-1)}, \mathbf{X}_{j(t-1)})$ is not informative of $\mathbf{X}_{j(t)}$. For example, this condition will be satisfied if the price information about $(t-1)$th product has no additional information in predicting the quality of t-th product when we have t-th product price information. Also, this condition will be satisfied if product attributes at position $(t-1)$ are independent of those at position t. Also for h_2, we consider ranking algorithm shown on product listing page such that h_2 only depends on $\mathbf{Z}_{j(t-1)}$ and is independent of $\mathbf{X}_{j(t-1)}$ (e.g., sort by price lowest to highest where price is shown on the product listing page as an external attribute). In sum, we simplify h and use

$$h\left(\mathbf{Z}_{j(1)}, \mathbf{X}_{j(1)}, ..., \mathbf{Z}_{j(T^r)}, \mathbf{X}_{j(T)}\right) = h_1(\mathbf{X}_{j(1)} \mid \mathbf{Z}_{j(1)}) h_2(\mathbf{Z}_{j(1)}) \prod_{t=2}^{T} h_1(\mathbf{X}_{j(t)} \mid \mathbf{Z}_{j(t)}) h_2(\mathbf{Z}_{j(t)} \mid \mathbf{Z}_{j(t-1)}) \quad \text{(B.1)}$$

Belief State Transitions. The belief state transitions can be expressed as

$$\begin{aligned} f_1^s\left(\mathbf{Z}_{j(t+1)} \mid \mathbf{Z}_{j(t)}\right) &= h_2(\mathbf{Z}_{j(t+1)} \mid \mathbf{Z}_{j(t)}) \\ f_0^c\left(u_{it}^* \mid u_{it-1}^*, \mathbf{Z}_{j(t)}\right) &= \mathbf{1}\left(u_{it}^* = u_{it-1}^*\right) \\ f_1^c\left(u_{it}^* \mid u_{it-1}^*, \mathbf{Z}_{j(t)}\right) &= f^u\left(u_{it}^* \mid u_{it-1}^*, \mathbf{Z}_{j(t)}\right) \end{aligned} \quad \text{(B.2)}$$

The first line indicates consumer's belief on transition of $\mathbf{Z}_{j(t+1)}$ when browsing continues, and the actual empirical distribution h_2 is used for this rational belief. The second line represents the belief state transition when the consumer does not click. In this case, the maximal utility in hand remains the same because the consumer simply moves on to browse the next item. The third line expresses the belief state transition when a consumer does click on product $j(t)$, in which case the maximal utility u_{it}^* is believed to transit with distribution $f^u\left(u_{it}^* \mid u_{it-1}^*, \mathbf{Z}_{j(t)}\right)$. That is, there is some likelihood of drawing an item better than those clicked before.

Using the iid $N(0, \sigma_\epsilon^2)$ assumption made on ϵ_{ij} and the additive separability of utility specification, $f^u\left(u_{it}^* \mid u_{it-1}^*, \mathbf{Z}_{j(t)}\right)$ can further be decomposed into

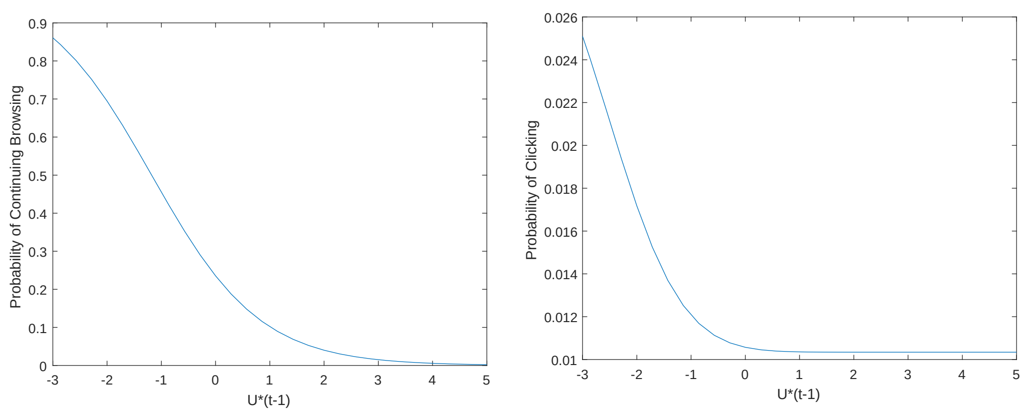
$$f^u\left(u_{it}^* \mid u_{it-1}^*, \mathbf{Z}_{j(t)}\right) = \begin{cases} \int_{\mathbf{X}_{j(t)}} \Phi\left(\frac{u_{it}^* - \mathbf{X}_{j(t)}\alpha - \mathbf{Z}_{j(t)}\beta}{\sigma_\epsilon}\right) h_1(\mathbf{X}_{j(t)} \mid \mathbf{Z}_{j(t)}) & \text{when } u_{it}^* = u_{it-1}^* \\ \int_{\mathbf{X}_{j(t)}} \frac{1}{\sigma_\epsilon}\phi\left(\frac{u_{it}^* - \mathbf{X}_{j(t)}\alpha - \mathbf{Z}_{j(t)}\beta}{\sigma_\epsilon}\right) h_1(\mathbf{X}_{j(t)} \mid \mathbf{Z}_{j(t)}) & \text{when } u_{it}^* > u_{it-1}^* \end{cases} \quad \text{(B.3)}$$

where ϕ and Φ are pdf and cdf of standard normal distribution, respectively. The first line indicates the probability that the clicked product yields a lower utility than u_{it}^*, the maximal utility in the consideration set formed prior to the click. The second line presents the probability of a click yielding a better product than previously discovered, u^*_{it-1}. The distribution for u^*_{it} (maximal utility in the consideration set at step t of search) is truncated from below by definition, with the truncation point given by u^*_{it-1}. As the u^*_{it-1} a consumer has in his hand weakly increases with the number of items previously clicked, the expected benefit of clicking also decreases in the number of items clicked, all others equal (e.g., the observed attributes).[1] Finally, $h_1(\mathbf{X}_{j(t)} \mid \mathbf{Z}_{j(t)})$ represents product attribute state transition as defined in Equation (1).

B.2. Existence and Uniqueness of the Consumer Model Solution

In our search model, a consumer is presented with an *exogenous* search sequence, and the optimal stopping problem closely resembles Rust's replacement model (Rust (1987), Seiler (2013)). As the maximum utility of the items in the consideration set, u^*_{t-1}, increases, the expected incremental increase in u^*_t from an additional browsing (or clicking) event decreases, which in turn decreases the probability of continuing browsing (or clicking an item) with respect to u^*_{t-1}. Eventually, this incremental increase in u^*_t becomes so small relative to a constant clicking costs that search stops. This guarantees the existence and the uniqueness of the solution.[2] In Figure 1, we plot probability of continuing browsing and clicking with respect to u^*_{t-1} for a given $\mathbf{Z}_{j(t)}$. Of note, though consumers are more likely to purchase the last item clicked in our model, free recall is also allowed, thereby capturing the pattern shown in data (Figure 4).

Figure B.2 Optimality of Search



C. Estimation

C.1. The Consumer Model

C.1.1. Derivation of Likelihood for Browsing, Clicking, and Purchase. In this section, we derive closed form expression for joint likelihood of browsing, clicking, and purchase.

[1] For example, if a consumer clicks and draws high ϵ_{ij}s in the beginning of the search process, this consumer will terminate search early, and the ϵ_{ij}s included in the consideration set will be truncated below. Similar discussion on this selection issue can be found in Chen and Yao (2016) and Honka (2014).

[2] Additionally, a single choice (purchase) assumption within a visit is required for the uniqueness of the solution. This assumption follows the definition of 'visit' we construct. In our data about 10% were multiple purchases (= 2 purchases) within the same visit (search session). In such cases, we assume that a new visit (search session) starts after purchasing the first item.

Clicking Decision Likelihood at Position t. The likelihood of observing click decision d^c_{it}, conditional on browsing and the (observed and unobserved) states can be defined as

$$\begin{aligned}&\mathcal{L}_t^{click|browse}(d^c_{it} \mid u^*_{it-1}, \mathbf{Z}_{j(t)}; \Theta_1)\\&= \left[p^c_0\left(u^*_{it-1}, \mathbf{Z}_{j(t)}; \Theta_1\right)\right]^{\mathbf{1}(d^c_{it}=0)} \times \left[1 - p^c_0\left(u^*_{it-1}, \mathbf{Z}_{j(t)}; \Theta_1\right)\right]^{\mathbf{1}(d^c_{it}=1)}\end{aligned} \tag{C.1}$$

where $p^c_0\left(u^*_{it-1}, \mathbf{Z}_{j(t)}; \Theta_1\right)$ is defined in Equation (4).

Browsing Decision Likelihood at Position t. The likelihood of observing browsing decision d^s_{it}, based on the (observed and unobserved) states can similarly be defined as

$$\begin{aligned}&\mathcal{L}_t^{browse}(d^s_{it} \mid u^*_{it}, \mathbf{Z}_{j(t)}; \Theta_1)\\&= \left[p^s_0\left(u^*_{it}, \mathbf{Z}_{j(t)}; \Theta_1\right)\right]^{\mathbf{1}(d^s_{it}=0)} \times \left[1 - p^s_0\left(u^*_{it}, \mathbf{Z}_{j(t)}; \Theta_1\right)\right]^{\mathbf{1}(d^s_{it}=1)}\end{aligned} \tag{C.2}$$

where $p^s_0\left(u^*_{it}, \mathbf{Z}_{j(t)}; \Theta_1\right)$ is given by Equation (8).

Consumer Purchase Decision Likelihood at Position t. Let T^s_i reference the position where individual i chooses to stop browsing such that $d^s_{iT^s_i} = 0$. Also denote T^p_i as the position in the browsing sequence where the purchased product is presented to the consumer, such that $d^p_{ij(T^p_i)} = 1$ (If the consumer chooses the outside option of not purchasing, then $d^p_{ij(T^p_i=0)} = 1$). The final consideration set $\Gamma_i = \Gamma_{iT^S_i}$ contains $K_{iT^S_i}$ number of products, and we index them as $\{1, ..., p^*, ...K_{iT^S_i}\}$ in the order encountered for consideration. Further we define $t(p)$ as the browsing sequence position of pth indexed product in the consideration set, such that $t(p^*) = T^p_i$.

This ordering suggests three partitionings for choice: first, those items that a consumer did not choose prior to finding the chosen alternative $\{1, ..., (p^* - 1)\}$; second, the chosen alternative $\{p^*\}$; and third, those items the consumer did not choose after finding the chosen alternative $\{(p^* + 1), ..., K_{iT^S_i}\}$. The cases of the clicked items not chosen prior to the chosen alternative differ from those clicked items encountered after the chosen alternative. More specifically, we know that all items clicked after the chosen item will not have higher utility than the highest so far (i.e., the chosen item). Thus, it is not possible for u^* to increase with click. However, for items not chosen prior to the chosen alternative, u^* can increase with each item clicked, even though u^* will not be higher than the chosen alternative. Therefore, when determining how choice affects the likelihood, we need to explicitly condition on the order in which the clicked item is encountered. In light of the foregoing discussion, we incorporate choice information into inference for the latent variable u^*_{it} transition as follows:

1. **Items clicked prior to the chosen item** : when $t(p) \leq T^p_i - 1$

   In this case, the reservation utility u^*_{it} weakly increases, and the transition probability of u^*_{it} can be characterized as[3]

$$f^u_1\left(u^*_{it} \mid u^*_{it-1}, \mathbf{Z}_{j(t)}, \mathbf{X}_{j(t)}\right) = \begin{cases} \Phi\left(\frac{u^*_{it} - \mathbf{X}_{j(t)}\alpha - \mathbf{Z}_{j(t)}\beta}{\sigma_\epsilon}\right) & \text{when } u^*_{it} = u^*_{it-1} \\ \frac{1}{\sigma_\epsilon}\phi\left(\frac{u^*_{it} - \mathbf{X}_{j(t)}\alpha - \mathbf{Z}_{j(t)}\beta}{\sigma_\epsilon}\right) & \text{when } u^*_{it} > u^*_{it-1} \end{cases}$$

[3] In the likelihood of unobserved state u^*_{it} transition, the product detail page information $\mathbf{X}_{j(t)}$ is included as a state space. This is different from the consumer's beliefs on u^*_{it}.

2. **The chosen item** : when $t(p) = T_i^p$

   If a product is bought at position $t(p)$, this product must yield the maximal utility among the ones clicked so far. If we consider a finely discretized space for u_{it}^* or a continuous case, u_{it}^* must be strictly greater than u_{it-1}^*.

$$f_2^u\left(u_{it}^* \mid u_{it-1}^*, \mathbf{Z}_{j(t)}, \mathbf{X}_{j(t)}\right) = \frac{1}{\sigma_\epsilon}\phi\left(\frac{u_{it}^* - \mathbf{X}_{j(t)}\alpha - \mathbf{Z}_{j(t)}\beta}{\sigma_\epsilon}\right) \text{ as } u_{it}^* > u_{it-1}^*$$

3. **Items clicked after the chosen item** : when $T_i^p < t(p) \leq T_i^s$

   If a product is clicked after T_i^p but has not been purchased, the associated utility found at position $t(p)$ should not be greater than $u_{iT_i^p}^*$.

$$f_3^u\left(u_{it}^* \mid u_{it-1}^*, \mathbf{Z}_{j(t)}, \mathbf{X}_{j(t)}\right) = \Phi\left(\frac{u_{it}^* - \mathbf{X}_{j(t)}\alpha - \mathbf{Z}_{j(t)}\beta}{\sigma_\epsilon}\right) \text{ as } u_{it}^* = u_{it-1}^*$$

Combining three cases, the likelihood from choice decision incorporated into the transition of unobserved u_{it}^*, can be written as

$$\begin{aligned}
&\mathcal{L}_t^{purchase|click,browse}\left(u_{it}^* \mid u_{it-1}^*, \mathbf{Z}_{j(t)}, \mathbf{X}_{j(t)}\right) \\
&= \left[\mathbf{1}\left(t \leq T_i^p - 1\right) f_1^u(\cdot) + \mathbf{1}\left(t = T_i^p\right) f_2^u(\cdot) + \mathbf{1}\left(T_i^p < t \leq T_i^s\right) f_3^u(\cdot)\right]^{1(d_{it}^c=1)} \\
&\times \left[\mathbf{1}\left(u_{it}^* = u_{it-1}^*\right)\right]^{1(d_{it}^c=0)} \qquad \text{(C.3)}
\end{aligned}$$

where the second line represents the case where t-th positioned product in the search sequence is not clicked, and hence $u_{it}^* = u_{it-1}^*$.

Combining Browsing, Clicking, and Choice. We define the total likelihood of observing the whole path of choices $\mathbf{d}_i = \{d_{i1}^c, ..., d_{iT_i^s}^c, d_{i1}^s, ..., d_{iT_i^s}^s, d_{i1}^p, ..., d_{iT_i^s}^p\}$ based on the (observed and unobserved) states as

$$\begin{aligned}
&\mathcal{L}(\mathbf{d}_i \mid u_{i0}^*, ..., u_{iT_i^s}^*, \mathbf{Z}, \mathbf{X}; \Theta_1) \\
&= \prod_{t=1}^{T_i^s} \mathcal{L}_t^{browse} \mathcal{L}_t^{click|browse} \mathcal{L}_t^{purchase|click,browse}
\end{aligned}$$

where $\mathcal{L}_t^{browse}$, $\mathcal{L}_t^{click|browse}$, and $\mathcal{L}_t^{purchase|click,browse}$ are defined in Equations (5), (4), and (6) respectively. This total likelihood is derived from multiplying over the likelihood of clicking and browsing decisions at $t = 1, ..., T_i^s$, and the transition of unobserved u^* is represented within $\mathcal{L}_t^{purchase|click,browse}$.

Integrating Out Unobservable States. Now we define the likelihood of observing $\mathbf{d}_i = \{d_{i1}^c, ..., d_{iT_i^s}^c, d_{i1}^s, ..., d_{iT_i^s}^s, d_{i1}^p, ..., d_{iT_i^s}^p\}$ based only on the observed states by integrating out over the unobservables $(u_{i1}^*, ..., u_{iT_i^s}^*)$.

$$L_i(\Theta_1^g) = \int_{u_{iT_i^s}^*} ... \int_{u_{i1}^*} \int_{u_{i0}^*} f^u\left(u_{i0}^*\right) \mathcal{L}(\mathbf{d}_i \mid u_{i0}^*, ..., u_{iT_i^s}^*, \mathbf{Z}, \mathbf{X}; \Theta_1^g)$$

The initial probability $f^u\left(u_{i0}^*\right)$ is the distribution of outside option value $f^u\left(\epsilon_{i0}\right) = \phi(\epsilon_{i0})$. Once we fix u_{i0}^*, the transition of $u_{it}^* | u_{it-1}^*$ is governed by $\mathcal{L}_t^{purchase|click,browse}$ as discussed above. This likelihood ensures that the purchased product has the highest utility among all clicked products. Further, the log-likelihood of the sample data is given by

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

where we integrate out latent class consumer heterogeneity.

C.1.2. Solving the Dynamic Problem. We specify the consumer decision to be an infinite horizon problem for three reasons. First, we find that the consumers in our data browse quite extensively, yet the browsing is never terminated at the last product available on the website. Thus, in our empirical setting, it is reasonable to assume that the consumer faces stationary value functions conditional on the states $(u_t^*, \mathbf{Z}_t)$. Second, we believe that the belief state transition can be represented as stationary conditional on the attributes Z. Third, although our estimation method can accommodate the finite horizon setting in which the future value terms are obtained via backward recursion for every search step t, the infinite horizon specification lowers the computational cost as the future value terms are computed using contraction mapping only once for a given set of parameters. Hence, we solve the dynamic search as an infinite horizon problem where stopping browsing is an absorbing state.

We estimate the consumer model using MLE 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 steps are as follows:[4]

1. Outer loop: Starting with the iteration step $iter = 0$, initialize the consumer model parameters $\Theta_1^{iter} \equiv (\alpha^{g,iter}, \beta^{g,iter}, \gamma_1^{g,iter}, \gamma_2^{g,iter}, \lambda^{g,iter})\ g = 1, ..., G$.
2. Inner loop: Starting with the iteration step $k = 0$, initialize the value functions, $Emax^{browse,k}$.
   (a) Given $Emax^{browse,k}$, compute the conditional value function for the click decision based on Equation (2). Then these conditional value functions are used to compute the conditional choice probability of no click, p_t^c, as defined in Equation (4) and also the expected future value of click, $Emax^{click,k}$, as defined in Equation (7).
   (b) Similarly given $Emax^{click,k}$ obtained in Step 2(a), compute the conditional value function for the browsing decision based on Equation (6). Then these conditional value functions are used to compute the conditional choice probability of ending browsing, p_t^s, as defined in Equation (8). Finally, the expected future value of browsing, $Emax^{browse,k+1}$, is updated for the next iteration step $(k+1)$ using the Equation (3).
3. Repeat Step 2(a) - Step 2(b) until convergence. This convergence will ensure that both the value functions and the conditional choice probabilities converge.
4. Compute the log-likelihood in Equation (15), based on the converged conditional choice probabilities. Optimize the log-likelihood to compute the new set of parameters Θ_1^{iter+1}
5. Repeat Step 2 - Step 4 until we find the global maximum.

[4] The value function states are discretized as follows. Price is discretized into 15 grid spaces based on their quantiles. The grid points for #likes include 0 and 1 as these are commonly observed states. In addition, the higher values for likes are discretized into 4 grid spaces based on their quantiles (hence, there is a total 6 grid spaces for the number of likes). We consider values of u^* that lie between $u^* \in [-3, 5]$ and discretize this interval into equidistant spaces of 30. The lower bound of the u^* range is based on the idea that the initial value is drawn from $u_{i0} = \epsilon_{i0} \sim N(0,1)$ and u^* can only increase as the search process progresses. The upper bound of the u^* is based on the maximum value of u^* over the potential range of the parameter spaces, i.e., $max\,(u_{ij} = \mathbf{X}_j\alpha + \mathbf{Z}_j\beta + \epsilon_{ij})$. At the parameter values estimated, $max\,(\mathbf{X}_j\alpha + \mathbf{Z}_j\beta) \approx 0.245$, so the upper limit of 5 for u^* does not generally bind. The discretization employed assumes that the states lie at the middle value of the respective grid space. We checked the robustness of the discretization by expanding the price, the likes, and u^* dimensions by 50, 15, and 50 grid spaces respectively. The end points of u^* range were also extended to $[-5, 10]$. In all cases, the estimates were stable.

C.1.3. Identification and Purchase Data. In this exercise, we consider homogeneous consumers and assume that there are 50 products on the platform, with a single dimension attribute for each Z and X. Z can be thought of as price displayed in the product listing page, and X can be thought of as the number of pictures available in the product detail page. One set of 50 products are randomly drawn from

$$(Z, X) \sim N\left(\begin{bmatrix} 5 \\ 2.5 \end{bmatrix}, \begin{bmatrix} 9 & 1 \\ 1 & 9 \end{bmatrix}\right)$$

A synthetic data set is generated with 100 simulations. The deep parameters used as a baseline and the estimated results are present in Table C.5. σ_ϵ is normalized to be one for identification purposes, and constant functional forms were used for clicking and browsing costs. The recovered parameters are all close to the true values with small standard errors.

Table C.5 Estimation Results from Simulated Data

	Parameter	True	Estimates (SE)	Estimates (SE) Without Purchase
α	Valuation on **X**	0.5	**0.4869** (0.0788)	**0.5696** (0.1825)
β	Valuation on **Z**	−1.2	**−1.2151** (0.1690)	**−1.3575** (0.7144)
γ_1	Clicking Cost	0.4	**0.3983** (0.0312)	**0.4029** (0.0342)
γ_2	Browsing Cost	0.3	**0.3082** (0.0093)	**0.3108** (0.0149)

The last column shows the results when purchase data are ignored and only browsing and clicking observations are used in estimation. Although clicking and browsing cost estimates are still close to the true values, the preference parameter estimates are much worse with at least twice the previous standard errors. This suggests that the identification of preference parameters are significantly enhanced when purchase data are consolidated. This is because purchase data provide additional information on how the unobserved maximal utility u_t^* transits as search progresses. For example, if a consumer clicks items in position $(1, 3, 5)$, we can infer that the maximal utility found so far increases weakly with $u_1^* \leq u_3^* \leq u_5^*$. However, if we also have purchase information that this consumer buys the item positioned at 3, we can further infer that the third product has the highest utility among the ones clicked, that is, $u_1^* < u_3^* = u_5^*$. This narrower bound on the transition of unobserved maximal utility significantly narrows down the bounds for preference parameters.

C.1.4. Model Fit. Table C.6 presents the in-sample and out-of-sample model fit of the consumer model. For the in-sample, 1000 set of 956 visits is simulated, and the key statistics are compared to those of the data. The heterogeneity on the cost parameters significantly improves the fit of the distribution (e.g., SD). For the out-of-sample, we hold out randomly selected 190 visits (about 20% of the sample) and then estimate the model using only 766 visits. Based on the new estimates, we simulate 1000 set of 956 visits to calculate the key statistics. Overall the model fits well.

C.2. The Advertiser Model

C.2.1. Beliefs on Product Placement. The platform's ranking algorithm displays products in the order of rank scores:

$$\begin{aligned} Rank_{j,t,d_j^a,\mathbf{d}_{-j}^a} &= Rank\left(\text{Own Score}_{jt}, \text{Others' Scores}_{-jt}\right) \\ &= Rank\left(\text{Popularity}_{jt}, \text{Slot Adjust}_{jt}, \text{Days Listed}_{jt}, \text{Advertising}_{jt}, \text{Others' Scores}_t\right) \end{aligned} \quad \text{(C.4)}$$

Table C.6 The Consumer Model Fit

# Per Visit		Median	Mean	SD
Browsing Length	Data	20	77.8	277.0
(Impressions)	1 Segment (In)	54.0 (2.49)	77.8 (2.43)	77.6 (3.58)
	4 Segments (In)	26.3 (1.28)	78.2 (7.13)	215.8 (28.8)
	4 Segments (Out)	27.1 (1.32)	90.7 (9.22)	283.9 (42.0)
Clicks	Data	0	0.8	3.0
	1 Segment (In)	0.00 (0.06)	0.83 (0.04)	1.23 (0.06)
	4 Segments (In)	0 (0)	0.83 (0.09)	2.61 (0.37)
	4 Segments (Out)	0 (0)	0.69 (0.07)	2.31 (0.33)
Purchase (Demand)	Data	0	0.04	0.2
	1 Segment (In)	0 (0)	0.05 (0.01)	0.21 (0.01)
	4 Segments (In)	0 (0)	0.04 (0.01)	0.20 (0.02)
	4 Segments (Out)	0 (0)	0.04 (0.01)	0.19 (0.01)
#Clicks	Data	0	1.14	2.9
/Browsing Length (%)	1 Segment (In)	0.00 (0.02)	1.07 (0.08)	2.34 (0.70)
	4 Segments (In)	0 (0)	1.15 (0.12)	3.54 (0.81)
	4 Segments (Out)	0 (0)	0.92 (0.11)	3.21 (0.84)
#Purchases	Data	0	7.4	22.2
/#Clicks (%)	1 Segment (In)	0 (0)	5.87 (0.98)	20.1 (1.94)
	4 Segments (In)	0 (0)	7.06 (1.37)	21.2 (2.59)
	4 Segments (Out)	0 (0)	7.13 (1.48)	21.6 (2.79)

where the second line reflects how the own rank score (Own Score) is a function of the popularity score, slot adjustment score, days listed, and the advertising score (= days listed × advertising status). Because advertisers do not observe all components or how they are combined, we need to generate a model of advertiser beliefs, denoted $\widehat{Rank}$. Of all the components that enter the rank function, advertisers know only their own advertising status (d_j^a) and days listed. They do not know their popularity score or slot adjustment score. Roughly speaking, the slot adjustment score depends on rank and days listed, whereas the popularity score is a function of rank, days listed, as well as other unobserved characteristics that drive more clicks and likes conditional on the product position. Substituting these into the rank function yields

$$\begin{aligned} Rank_{j,t,d_j^a,\mathbf{d}_{-j}^a} = Rank\,(&\text{Popularity}(\text{Rank}_{\boldsymbol{jt}}, \text{Days Listed}_{jt}, \text{Unobserved}_j), \\ &\text{Slot Adjust}\ (\text{Rank}_{\boldsymbol{jt}}, \text{Days Listed}_{jt}), \\ &\text{Days Listed}_{jt}, \text{Advertising}_{jt}, \text{Others' Scores}_t\big) \\ Rank_{j,t,d_j^a,\mathbf{d}_{-j}^a} = Rank\,&\big(\text{Unobserved}_j, \text{Days Listed}_{jt}, \text{Advertising}_{jt}, \text{Others' Scores}_t\big) \end{aligned}$$

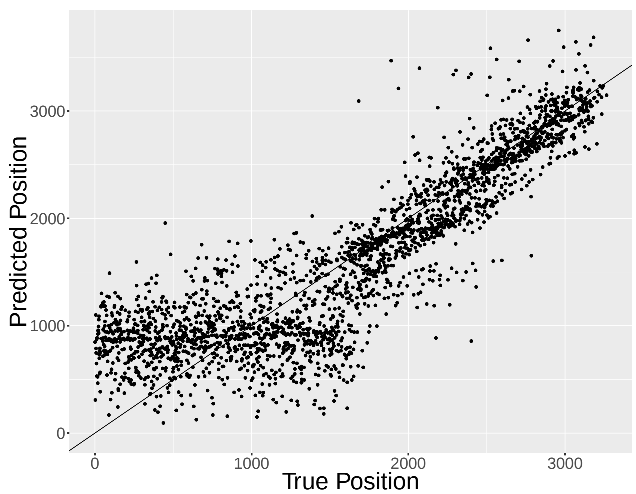
As the unobserved part is a component of popularity score that is not explained by the rank and days listed, a proxy measure called "organic strength" is constructed for each product j by regressing the popularity score at product-day level on days listed and observed product position using a generalized additive model, then taking the mean of the residuals. Although the seller may not know the underlying popularity score for each day, we presume that the seller knows its own product's inherent general popularity level (unobserved characteristics) with respect to others, which is captured by including the organic strength term into the seller's information set. Last, advertisers do not know other advertisers' scores, so we presume advertisers instead condition on the aggregate states of competing advertisers' decisions when forming own rank beliefs.

The rationale is that a greater number of competing advertisers leads to a lower rank. In sum, the advertiser's belief on the product placement for a given day t is assumed to be based on a generalized additive model

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

where d_j^a is own advertising strategy, $E_t(\mathbf{d}_{-j}^a)$ is the *aggregate states* of others' advertising strategies, and J_t is the total number of products available. Interaction terms of the input elements are also included in this model (e.g., interaction between organic strength and days listed). Figure C.3 plots product position based on the platform's algorithm on the x-axis and the sellers' beliefs about product placements on the y-axis on a given day. Although the sellers' beliefs are based on only aggregate and own individual states, sellers' approximated beliefs hew closely to rational expectations.

Figure C.3 Beliefs on Product Placement



C.2.2. Solving the Advertiser Problem. We estimate the advertiser model in three stages, and these stages are described next.

Stage 1 - Estimate Seller's Beliefs About Platform Ranking Algorithm. First, estimate the function governing sellers' beliefs on product placement as described in sub-section 4.2.1, that is, we estimate g function in Equation (16).

Stage 2 - Estimate Effect of Advertising on Product Placement and Consumer Responses.

1. Compute product placement for each advertising decision

   On a given day t, given seller's information set $\left(d_j^a,\, E_t(\mathbf{d}_{-j,}^a),\, J_t,\, \text{Days Listed}_{jt},\, \text{Organic Strength}_j\right)$, compute the belief about product $j's$ placement when advertising $\left(\widehat{Rank}_{j,t,d_j^a=1,\mathbf{d}_{-j}^a}\right)$ and not advertising $\left(\widehat{Rank}_{j,t,d_j^a=0,\mathbf{d}_{-j}^a}\right)$ using the function g estimated in Stage 1. For estimation, we compute $\left(E_t(\mathbf{d}_{-j}^a),\, J_t\right)$ under the observed advertising strategies and use these two statistics as the aggregate beliefs.

2. Compute consumer responses based on product placement beliefs $\left(\widehat{Rank_{j,t,d^a_j=0,\mathbf{d}^a_{-j}}}, \widehat{Rank_{j,t,d^a_j=1,\mathbf{d}^a_{-j}}}\right)$
Using the consumer demand model, simulate consumer demand, click, and impressions $\left(\widehat{D_{j,t,,d^a_j,\mathbf{d^a_{-j}}}}, \widehat{C_{j,t,d^a_j,\mathbf{d^a_{-j}}}}, \widehat{I_{j,t,d^a_j,\mathbf{d^a_{-j}}}}\right)$ by displaying product j at position $\left(\widehat{Rank_{j,t,d^a_j,\mathbf{d}^a_{-j}}}\right)$. When simulating consumers' behaviors for a given product, we further assume that consumers' belief state transitions (e.g., $f(Z_{t+1}|Z_t)$) under the new ranking are common knowledge. In other words, we assume that the distribution of observed products' attributes under the new ranking is known to sellers. This is done at the daily level, and these simulated responses are aggregated across time periods to form product j's lifetime demand, clicks, and impressions, which are entered into Equations (11) and (12).[5]

3. Accounting for uncertainty in $\left(\widehat{Rank_{j,t,d^a_j=0,\mathbf{d}^a_{-j}}}, \widehat{Rank_{j,t,d^a_j=1,\mathbf{d}^a_{-j}}}\right)$
The seller faces uncertainty regarding $\left(E_t(\mathbf{d}^a_{-j,}), J_t\right)$ and therefore ultimately $\left(\widehat{Rank_{j,t,d^a_j=0,\mathbf{d}^a_{-j}}}, \widehat{Rank_{j,t,d^a_j=1,\mathbf{d}^a_{-j}}}\right)$. This uncertainty arises because sellers do not know ξ_j, but instead only know its distribution. To account for the uncertainty in the sellers' beliefs regarding rank, we simulate 1000 sets of ξ_j, generating 1000 sets of $\left(E_t(\mathbf{d}^a_{-j,}), J_t\right)$, leading to 1000 sets of $\left(\widehat{Rank_{j,t,d^a_j=0,\mathbf{d}^a_{-j}}}, \widehat{Rank_{j,t,d^a_j=1,\mathbf{d}^a_{-j}}}\right)$ and then ultimately 1000 sets of $\left(\widehat{D_{j,t,,d^a_j,\mathbf{d^a_{-j}}}}, \widehat{C_{j,t,d^a_j,\mathbf{d^a_{-j}}}}, \widehat{I_{j,t,d^a_j,\mathbf{d^a_{-j}}}}\right)$. We compute the expected value of $\left(\widehat{D_{j,t,,d^a_j,\mathbf{d^a_{-j}}}}, \widehat{C_{j,t,d^a_j,\mathbf{d^a_{-j}}}}, \widehat{I_{j,t,d^a_j,\mathbf{d^a_{-j}}}}\right)$ to account for the uncertainty in sellers' beliefs.

Stage 3 - Estimate Seller Model Parameters.

1. Starting with the iteration step $iter = 0$, initialize the advertiser model parameters $\Theta_2^{iter} \equiv (\theta^{iter}, \theta^{D,iter}, \theta^{C,iter}, \theta^{I,iter}, \delta)$.

2. Using Equation (13), compute the advertising probability for product j based on the aggregated consumer responses obtained in Stage 2, when advertising $\left(\widehat{D_{j,d^a_j=1,\mathbf{d}^a_{-j}}}, \widehat{C_{j,d^a_j=1,\mathbf{d}^a_{-j}}}, \widehat{I_{j,d^a_j=1,\mathbf{d}^a_{-j}}}\right)$ and not advertising $\left(\widehat{D_{j,d^a_j=0,\mathbf{d}^a_{-j}}}, \widehat{C_{j,d^a_j=0,\mathbf{d}^a_{-j}}}, \widehat{I_{j,d^a_j=0,\mathbf{d}^a_{-j}}}\right)$ and the given set of parameters Θ_2^{iter}.

3. Compute the log-likelihood in Equation (17), based on the advertising probabilities computed. Optimize the log-likelihood to compute the new set of parameters Θ_2^{iter+1}.

4. Repeat Step 2 - Step 3 until we find the global maximum.

C.2.3. Computing Equilibrium Advertising Strategies for the Policy Simulations. As described in Stage 2 above, in estimation we use the actual advertising strategies to compute $\left(E(\mathbf{d}^a_{-j}), J\right)$. However, these strategies will change as the site changes its policies. Hence, in policy simulations, we need to iterate over the sellers' beliefs and the advertising decisions until convergence. This convergence will ensure that the aggregate beliefs are consistent with the underlying advertisers' decisions in equilibrium.[6] The steps follow:

1. Estimate sellers' beliefs about platform ranking algorithm
For the policy simulation where we do not change the ranking algorithm (i.e. where we only change

[5] We aggregate consumer responses up to the point the (belief on) product position reaches 2000. As consumers median browsing length is 20 (mean 79), this constraint does not impact aggregation.

[6] Although we do not provide proof for existence, we did not encounter convergence issue in our implementation. Related, in a dynamic auction setting Iyer et al. (2014) proves existence of mean field equilibrium under mild assumptions.

the fee structure), we use the same g function (Equation (16)) used in the estimation. For the policy simulation where we do vary the ranking algorithm, g function is updated. That is, the product position on the left-hand side of Equation (16) is simulated based on the score inputs and the platform's new ranking algorithm under the counterfactual scenario, then new sellers' beliefs are constructed by estimating this g function again.

2. Starting with the iteration step $k = 0$, initialize the advertising strategies $\mathbf{d}^{a,k}$. We start from the observed advertising strategies in the data.
3. For each product j, obtain the aggregate beliefs $\left(E(\mathbf{d}_{-j}^{a,k}), J\right)$ given $\mathbf{d}^{a,k}$. We also update consumers' belief transition in Equations (2) and (3) based on $\mathbf{d}^{a,k}$ and the platform's actual ranking algorithm.
4. Next step is to estimate the effect of advertising on product placement and consumer responses (impressions, clicks, and purchases). To compute this, we run Steps 1 - 3 in Stage 2 of sub-section C.2.2.
5. Compute the new advertising strategy for product j, $d_j^{a,k+1}$. This can be achieved by running Step 2 in Stage 3 of sub-section C.2.2, based on the estimated parameters $\Theta_2 = (\theta, \theta^C, \theta^C, \theta^I, \delta)$ from the advertiser model. Changing the fee structure can affect the sellers' listing behavior (for example, a seller would delist an item if the expected listing fees are higher than the expected gains from listing). To account for the change in the seller's listing behavior, we impose a participation constraint that each seller's utility is greater than the minimum of the seller utilities estimated in the actual fee structure setting.
6. Stack the updated advertising probabilities $d_j^{a,k+1}$ into $\mathbf{d}^{a,k+1}$.
7. Iterate Step 3 - Step 6 above until convergence. This ensures the individual decisions are consistent with the aggregate expectations.

D. Full Sample Results

In the analysis reported in the paper, we restrict our attention to the users with at least one purchase (within the estimation period, across all categories) in our analyses. Arguably, those that do not make purchases generate advertiser value via impressions and clicks. To obtain a better sense of the magnitude of potential bias arising from the sample selection, we re-estimate our demand side model with the 'full sample' of consumers (including both with and without purchase) and use these new estimates to infer advertiser valuations. In the full sample, we observe $72,030$ individuals meeting our criteria, with a total of $85,632$ visits. An individual makes 1.2 visits in average (median 1) during the sample period. These consumers browse $2,256,244$ times in total, among which $24,870$ are considered, and 40 are purchased within the main page product feed.

D.1. The Consumer Model

Except for the constant, the preference parameters for the full sample are within 2 standard deviations (do not appear to significantly differ) from the purchase sample estimates. The lower constant reflects the data pattern, where the mean purchase rate for the full sample is lower than that of the purchase sample. The average marginal costs of browsing and clicking are \$0.94 and \$3.92, respectively, which are higher than those estimated using the purchase sample. Higher browsing/clicking costs are also consistent with the data pattern; the consumers in the full sample are less likely to browse and click within a visit because they are less interested in purchasing the products.

Table D.7 The Consumer Model Estimates

		Purchase	Full
Preference			
Type1	# Pictures (X)	**0.18** (0.07)	**0.11** (0.06)
	Log (Price) (Z)	−**0.29** (0.10)	−**0.27** (0.08)
	# Likes (Z)	−0.00 (0.01)	−0.01 (0.00)
	Constant	−**2.67** (0.09)	−**4.39** (0.30)
Cost			
Type1	Clicking	**1.42** (0.02)	**1.66** (0.00)
	Browsing	**0.16** (0.00)	**0.20** (0.00)
Type2	Clicking	**1.77** (0.03)	**1.73** (0.01)
	Browsing	**0.15** (0.00)	**0.15** (0.00)
Type3	Clicking	**1.16** (0.04)	**1.29** (0.01)
	Browsing	**0.20** (0.00)	**0.17** (0.00)
Type4	Clicking	**1.76** (0.05)	**0.87** (0.01)
	Browsing	**0.17** (0.00)	**0.25** (0.00)
Type Probability			
Type1		**0.05** (0.01)	**0.87** (0.01)
Type2		**0.04** (0.01)	**0.06** (0.00)
Type3		**0.20** (0.04)	**0.03** (0.00)
N		74,400	2,256,244
LL		−9,077	−477,604
BIC		18,344	955,457
AIC		18,344	955,242

D.2. The Advertiser Model

The estimates for the advertiser model predicated upon the full sample results (i.e., full sample demand estimates and simulations of full sample consumer behaviors) are reported in the second column in Table D.8.

Table D.8 The Advertiser Model Estimates

	Parameter	Purchase	Full
θ^D	Demand	**0.62** (0.37)	0.01 (0.007)
θ^C	log(Clicks)	**0.017** (0.005)	0.000 (0.000)
θ_1^I	log(Impressions (in thousand))	0.000 (0.000)	0.000 (0.000)
δ	Marginal Cost	**0.74** (0.03)	**0.74** (0.03)
θ	Constant	−**1.01** (0.06)	−**0.96** (0.06)
	Include URL	**0.28** (0.06)	**0.22** (0.09)
	Silver	**0.44** (0.15)	**0.43** (0.16)
	Stone	**0.20** (0.09)	**0.30** (0.07)
	Bracelet	**0.29** (0.07)	**0.29** (0.06)
	Refundable	**0.26** (0.06)	**0.26** (0.06)
	Seller Fixed Effects	Y	Y
	LL ($N = 2853$)	−850.0	−874.3

All estimates are similar, except for the marginal valuation for purchase, click, and impression. This difference in estimates for marginal value arises because the total number of expected purchases, clicks, and impressions for a given product are larger for the full sample, leading to decreased mean values *per* each purchase, click, and impression. Although the per click (or impression) value is smaller, we find that the *total* advertiser valuations from purchases, clicks, and impressions using the full sample are indeed similar to the total valuations from the purchase sample.

As total advertiser valuation is largely unchanged, the advertisers' total willingness to pay to the platform (= smaller willingness to pay per click × larger total number of clicks in the full sample) will also be largely similar. Thus, the key counterfactual findings based on the advertiser model (e.g., percentage change in the total seller welfare and the platform's profits from the baseline) yield essentially the same insights regardless of whether one uses the purchase sample or the full sample.

The cautionary lesson from using the purchase sample to impute advertiser valuations lies in interpreting 'marginal valuation' and 'marginal fee'. The marginal valuations estimated using the purchase sample are the marginal valuations per purchase, click, and impression of those with at least one purchase. Accordingly, the changes in marginal fees (f^A, f^T, f^C, f^I) under the counterfactual exercises should be interpreted as the marginal fee charged to those with at least one purchase.

References

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

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

Honka E (2014) Quantifying search and switching costs in the us auto insurance industry. *The RAND Journal of Economics* 45(4):847–884.

Iyer K, Johari R, Sundararajan M (2014) Mean field equilibria of dynamic auctions with learning. *Management Science* 60(12):2949–2970.

Rust J (1987) Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society* 999–1033.

Seiler S (2013) The impact of search costs on consumer behavior: A dynamic approach. *Quantitative Marketing and Economics* 11(2):155–203.