

# A Dynamic Model of Sponsored Search Advertising

Song Yao  
Carl F. Mela<sup>1</sup>

March 19, 2009

<sup>1</sup>Song Yao (email: [song.yao@duke.edu](mailto:song.yao@duke.edu), phone: 919-660-2882) is a Ph.D. Candidate at Duke University, Durham, North Carolina, 27708. Carl F. Mela (email: [mela@duke.edu](mailto:mela@duke.edu), phone: 919-660-7767, fax: 919-681-6245) is a Professor of Marketing, The Fuqua School of Business, Duke University, Durham, North Carolina, 27708. The authors would like to thank seminar participants at Cornell University, Dartmouth College, Duke University, Emory University, Erasmus University, Georgia Institute of Technology, Georgia State University, Harvard Business School, INSEAD, London Business School, New York University, Northwestern University, Ohio State University, Stanford University, University of British Columbia, University of California at Berkeley, University of Chicago, University of Maryland, University of Rochester, University of Southern California, University of Texas, University of Tilburg, and the 2008 Marketing Science Conference as well as J.P. Dubé, Wes Hartmann, Günter Hitsch, Han Hong, Anja Lambrecht, Andrés Musalem, Harikesh Nair, Peter Rossi and Rick Staelin for their feedback. We gratefully acknowledge financial support from the NET Institute ([www.netinst.org](http://www.netinst.org)) and the Kauffman Foundation.

## Abstract

Sponsored search advertising is ascendant—Jupiter Research reports expenditures rose 28% in 2007 to \$8.9B and will continue to rise at a 26% CAGR, approaching 1/2 the level of television advertising and making it one of the major advertising trends to affect the marketing landscape. Yet little empirical research exists to explore how the interaction of various agents (*searchers*, *advertisers*, and the *search engine*) in keyword markets affects consumer welfare and firm profits. The dynamic structural model we propose serves as a foundation to explore these outcomes. We fit this model to a proprietary data set provided by an anonymous search engine. These data include consumer search and clicking behavior, advertiser bidding behavior, and search engine information such as keyword pricing and website design.

With respect to the *advertisers*, we find evidence of dynamic bidding behavior. Advertiser value for clicks on their sponsored links averages about 27 cents. Given the typical \$22 retail price of the software products advertised on the considered search engine, this implies a conversion rate (sales per click) of about 1.2%, well within common estimates of 1-2% (gamedaily.com). With respect to *consumers*, we find that frequent clickers place a greater emphasis on the position of the sponsored advertising link. We further find that 10% of consumers do 90% of the clicks.

We then conduct several policy simulations to illustrate the effects of changes in *search engine* policy. First, we find the search engine obtains revenue gains of nearly 1.4% by sharing individual level information with advertisers and enabling them to vary their bids by consumer segment. This also improves advertiser revenue by 11% and consumer welfare by 2.9%. Second, we find that a switch from a first to second price auction results in truth telling (advertiser bids rise to advertiser valuations), consistent with economic theory. However, the second price auction has little impact on search engine profits. Third, consumer search tools lead to a platform revenue increase of 3.7% and an increase of consumer welfare by 5.6%. However, these tools, by reducing advertising exposures, lower advertiser profits by 4.1%.

**Keywords:** Sponsored Search Advertising, Two-sided Market, Dynamic Game, Structural Models, Empirical IO, Customization, Auctions

# 1 Introduction

Sponsored search is one of the largest and fastest growing advertising channels. In January of 2009 alone, Internet users conducted 13.5B searches using the top 5 American search engines compared to 10.5B in the previous January, indicating a robust 29% year over year increase.<sup>1</sup> In the United States, annual expenditures on sponsored search advertising increased 28% to \$8.9B in 2007 and the number of firms using sponsored search advertising rose from 29% to 41%.<sup>2</sup> Moreover, advertising expenditures on sponsored search is forecast to grow to \$25B by 2012.<sup>3</sup> By contrast, overall 2007 television advertising spending in the United States is estimated to be \$62B, an increase of only 0.7% from the preceding year.<sup>4</sup> Hence, search engine marketing is becoming a central component of the promotional mix in many organizations.

The growth of this new medium can be ascribed to several factors. First, the increasing popularity of search engine sites relative to other media among consumers affords a greater advertising reach. In addition to searches on large general search engines such as `Google.com`, `MSN.com`, and `Yahoo.com`, search is also widespread on more focused ones (`dealttime.com` searches Internet stores, `kayak.com` searches travel products, `addall.com` searches books, etc.). The broad reach of search can be ascribed to the complexity of wading through an estimated 155 million sites to return relevant results in response to users' search queries.<sup>5</sup> By comparison, a top rated TV show such as "Desperate Housewives" only has about 25M viewers (IRI, 2007); and the growing popularity of DVR services offered by TiVo and cable companies have and will further decrease the audience base of traditional TV advertising. Second, more and more consumers use the Internet for their transactions (Ansari et al. (2008)), and Internet search is an especially efficient way to promote online channels. For example, Qiu et al. (2005) estimate that more than 13.6% of the web traffic is affected by search engines. Third, search advertising often targets consumers who are actively seeking information related to the advertisers' products. For example, a search of "sedan" and "automotive dealer" might signal an active purchase state. As a result of these various factors, Jupiter Research reports that 82% of advertisers were satisfied or extremely satisfied with search marketing ROI in 2006 and 65% planned to increase search spending in 2007.<sup>6</sup>

Given the increasing ubiquity of sponsored search advertising, the topic has seen substantially increased attention in marketing as of late (Ghose and Yang (2007), Rutz and Bucklin (2007), Rutz and Bucklin (2008), Goldfarb and Tucker (2008)). These recent advances focus upon the efficacy

---

<sup>1</sup> "January 2009 U.S. Search Engine Rankings," comScore, Inc. (<http://ir.comscore.com/releasedetail.cfm?ReleaseID=366442>). "January 2008 U.S. Search Engine Rankings," comScore, Inc. (<http://www.comscore.com/press/release.asp?press=2068>).

<sup>2</sup> "US Paid Search Forecast, 2007 to 2012," Jupiter Research, 2007.

<sup>3</sup> "US Interactive Marketing Forecast, 2007 to 2012," Forrester Research, 2007.

<sup>4</sup> "Insider's Report," 2007, McCann WorldGroup, Inc.; <http://www.tns-mi.com/news/01082007.htm>

<sup>5</sup> "January 2008 Web Server Survey," Netcraft Company ([http://news.netcraft.com/archives/2008/01/28/january\\_2008\\_web\\_server\\_survey.html](http://news.netcraft.com/archives/2008/01/28/january_2008_web_server_survey.html)).

<sup>6</sup> "US Paid Search Forecast, 2007 to 2012," Jupiter Research, 2007.

of an advertiser campaign. To date, empirical research on keyword search has been largely silent on the perspective of the search engine, the competition between advertisers, and the behavior of the searcher. Given that the search engine interacts with advertisers and searchers to determine the price and consumer welfare of the advertising medium (and hence its efficacy), our objective is to broaden this stream of research to incorporate the role of all three agents: the search engine, the advertisers, and the searchers. This exercise enables us to determine the role of search engine marketing strategy on the behavior of advertisers and consumers as well as the attendant implications for search engine revenues. Our key contributions are, accordingly, as follows:

1. From a theoretical perspective, we conceptualize and develop an integrated model of web searcher, advertiser and search engine behavior. To our knowledge, this is the first empirical paper focusing on the marketing strategy of the search engine. Much like Yao and Mela (2008), we construct a model of a two-sided network in an auction context. One side of the network includes the searchers who generate revenue for the advertiser. On the other side of the two-sided network are advertisers whose bidding behavior determines the revenue of the search engine. In the middle lies the search engine. The goal of the search engine is to price consumer information, set auction mechanisms, and design webpages to elucidate product information so as to maximize its profits. Some key insights from this model include:
  - Advertisers in our application have an average value per click of \$0.27. Given that the average price of software products advertised on the site in our data is about \$22, this implies these advertisers expect about 1.2% (i.e.,  $\$0.27/\$22$ ) of clicks will lead to a purchase. This is consistent with the industry average of 1-2% reported by **GameDaily.com**, suggesting good face validity for our model.
  - In addition, we find considerable heterogeneity in consumer response to sponsored search advertising. Frequent link clickers, who represent 10% of the population but 90% of the clicks, tend to be more sensitive to slot order – in part because slot position can signal product quality. These insights represent central inputs into our policy simulations alluded to below.
2. From a substantive point of view, we offer concrete marketing policy recommendations to the search engine. In particular, the two-sided network model of keyword search we consider allows us to address the effect of the following policy simulations (and would enable us to address many others) on auction house and advertiser profits as well as consumer welfare:
  - Search Tools. Many search engines, especially specialized ones such as **Shopping.com**, provide users options to sort/filter search results using certain criteria such as product prices. On one hand, the search tools may mitigate the desirability of bidding for advertisements because these tools can remove less relevant advertisements. This would tend

to lower search engine revenues. On the other hand, these tools can also attract more users to the site leading to a potential increase in advertising exposures and searchers. This would increase revenues. The trade-off leads to the question of how search tools impact consumer searching behavior, firms' advertising decisions, and search engine profits. Our analysis indicates that negative consumer effects on search engine profits ( $-6.4\%$ ) outweigh the corresponding positive advertiser effects on search engine profits ( $2.7\%$ ) and that overall the sort/filter options enhance platform profits by  $3.7\%$ . Consistent with this result, there is a corresponding loss in consumer welfare of  $5.6\%$  and an attendant increase in advertiser profits of  $4.1\%$ .

- **Segmentation and Targeting.** Most search engines auction keywords across all market segments. However, it is possible to auction keywords by segment. This targeting tends to reduce competition between advertisers within segments as markets are sliced more narrowly, leading to lower bids and hence lower potential revenues for the search engine. Yet targeting also enhances the efficiency of advertising, which tends to increase advertiser bids. Overall, we find that the latter effect dominates ( $2.2\%$ ) the former effect ( $-0.8\%$ ) and that search engine revenue increases  $1.4\%$  by purveying keywords by consumer market segments. Moreover, we find advertiser profits improve by  $11\%$  (from reduced competition in bidding and more efficient advertising) and consumer welfare (as measured by utility) increases  $2.9\%$ . Hence, this change leads to considerable welfare gains across all agents.
- **Market Intelligence.** Advertisers' knowledge about consumers changes if search engines sell consumer demographic and behavioral information to advertisers. This raises the question of whether and how information asymmetries between the engine and advertisers affect bidding behavior. We find that there is a negligible  $0.08\%$  increase in search engine profits accrued when this information asymmetry is erased. This finding suggests that advertisers are able to make reasonable inferences about the nature of heterogeneity in market response from the aggregate demand data, consistent with recent research regarding individual level inference from aggregate demand (Chen and Yang (2007), Musalem et al. (2008)). Moreover, the value of this information lies more in the advertiser's ability to exploit it by targeted bidding (as indicated in the preceding paragraph).
- **Mechanism Design.** The wide array of search pricing mechanisms raises the question of which auction mechanism is the best in the sense of incenting advertisers to bid more aggressively thereby yielding maximum returns for the search engine. We consider two common mechanisms: a first price auction (as used by the considered firm in our analysis) and a second price auction (wherein a firm pays the bid of the next lowest bidder). Virtually no revenue gains accrue to the platform from a second price auction ( $0.01\%$ ). However, advertiser bids under second price auction are close to bidders' true

values (bids average 98% of valuations), while bids under the first price auction are much lower (72%). This finding is consistent with theory that suggests first price auctions lead to bid shading and second price auctions lead to truth telling (Edelman et al. (2007)). Hence, we lend empirical validation to the theoretical literature on auction mechanisms in keyword search.

3. From a methodological view, we develop a dynamic structural model of keyword advertising. This dynamic is induced by the search engine’s use of past advertising performance when ranking current advertising bids. The dynamic aspect of the problem requires the use of some recent innovations pertaining to the estimation of dynamic games in economics (e.g., Bajari et al. (2007), Pesendorfer and Schmidt-Dengler (2008)). We extend this work to be Bayesian in implementation and apply it to a wholly new context. Overall, we find that there is a substantial improvement in model fit when the advertiser’s strategic bidding behavior is considered (the log marginal likelihood improves by 50.0), consistent with the view that their bidding behavior is dynamic. In addition, we find the posterior distributions of parameter estimates is non-normal; while classical methods assume asymptotic normality, our Bayesian approach does not.

Though we cast our model in the context of sponsored search, we note that the problem, and hence the conceptualization, is even more general. Any interactive, addressable media format (e.g., DVR, satellite digital radio) can be utilized to implement similar auctions for advertising. For example, with the convergence in media between computers and television in DVRs, simple channel or show queries can be accompanied by sponsored search, and this medium may help to offset advertising losses arising from ads skipping by DVR users. In such a notion, the research literature on sponsored search auctions generalizes to a much broader context, and our model serves as a basis for exploring search based advertising.

The remainder of this paper proceeds as follows. First we overview the relevant literature to differentiate our analysis from previous research. Given the relatively novel research context, we then describe the data to help make the problem more concrete. Next, we outline the details of our model, beginning with the clicking behavior of consumers and concluding with the advertiser bidding behavior. Subsequently, we turn to estimation and present our results. We then explore the role of information asymmetry, targeted bidding, advertising pricing, and webpage design by developing policy simulations that alter the search engine marketing strategies. We conclude with some future directions.

## 2 Recent Literature

Research on sponsored search, commensurate with the topic it seeks to address, is nascent and growing. Heretofore this literature can be characterized along two distinct dimensions: theoretical and empirical. The theoretical literature details how agents (e.g., advertisers) are likely to react to different pricing mechanisms. In contrast, the empirical literature measures the effect of advertising on consumer response in a given market but not the reaction of these agents to changes in the platform environment (e.g., advertising pricing, information state or the webpage design of the platform). By integrating the theoretical and empirical research streams, we develop a complete representation of the role of pricing and information in the context of keyword search. To elaborate on these points, we begin by surveying theoretical work on sponsored search and then proceed to discuss some recent empirical research.

Foundational theoretical analyses of sponsored search include Edelman et al. (2007) and Varian (2007) who examine the bidding behaviors of advertisers in this auction game. The authors assume the auction game as a complete information and simultaneous-move static game, in which exogenous advertising click-through rates increase with better placements. In equilibrium advertiser bidding behavior has the same payoff structure as a Vickrey-Clarke-Groves auction, where a winner's payment to the seller equals to those losing bidders' potential payoffs (opportunity costs) were the winner absent (Groves (1979)). Extending this work, Chen and He (2006) incorporate clicking behavior into their model and show that, under the Google bidding mechanism, consumers clicking behavior is affected by access to product information. In particular, they make inferences about product quality based on the ranking presented by the platform and search sequentially according to the ranking. As an equilibrium response, advertisers submit bids equal to their true values for the advertising. Athey and Ellison (2008) also consider a model that integrates both consumers and advertisers as in Chen and He (2006) but make several important extensions. They assume that advertisers only know the distribution of their competitors' valuations about sponsored slots, thus making the auction an incomplete information game. They also assume consumers engage in costly sequential searches that follow an optimal stopping rule (cf. Weitzman (1979)): consumers stop searching when the expected return of further searching is lower than the best choice in their consideration set. Thus, the better slot an advertiser gets, the higher the probability that her product will be purchased by consumers. Intuitively, this is because a better spot will give the product a higher probability to be included in the consumer's consideration set before she stops searching. So a better slot implies higher sales for the advertiser. The equilibrium bidding strategy in Athey and Ellison (2008) is different than those in previous literature. Although bids are still monotonically increasing in values (qualities), high value (quality) advertisers will bid more aggressively than low value (quality) advertisers. Katona and Sarvary (2008) further extend the analysis by relaxing several key assumptions such as the competition for traffic between sponsored

links and organic links, the heterogeneity of advertisers in term of their inherent attractiveness to consumers. The author shows multiple equilibria in this auction which do not have closed form solutions. Additional work by Iyengar and Kumar (2006), Feng (2008), and Garg et al. (2006) explicitly consider the effect of the various auction mechanisms on search engine profits. In particular, Iyengar and Kumar (2006) show that the Google pricing mechanism maximizes neither the search engine’s revenue nor the efficiency of the auction suggesting the potential to improve on this mechanism as we seek to do. Further, they show that the optimal mechanism is incumbent upon the characteristics of the market, thereby making it imperative to estimate market response as we intend to do in order to improve on pricing mechanisms. Summarizing the key insights from this stream of work, we note that i) there are three agents interacting in the sponsored search context, those who engage in keyword search, advertisers that bid for keywords, and the search platform, ii) searchers affect advertisers bidding behavior by reacting to the search engine’s web page design and hence advertiser payoffs, iii) bidders affect searcher behavior by the placement of their advertisements on the page, and iv) changes in advertiser and consumer behavior are incumbent upon the strategies of the platform.

In spite of these insights, several limits remain. First, because equilibrium outcomes are incumbent upon the parameters of the system, it is hard to characterize precisely how agents will behave. This implies it would be desirable to estimate a model of keyword search in order to measure these behaviors. Second, a static advertiser game over bidding periods is typically assumed, which is inconsistent with the pricing practices used by search engines. Search engines commonly use the preceding period’s click-throughs together with current bids to determine advertising placement, making this an inherently dynamic game. Third, this research typically assumes no asymmetry in information states between the advertiser and the search engine even though the search engine knows individual level clicking behaviors and the advertiser does not. We redress these issues in this paper.

Empirical research on sponsored search advertising is also proliferating. Notable among these papers, Rutz and Bucklin (2008) investigate the efficacy of different keyword choices by measuring the conversion rate from users’ clicks on ads to actual sales for the advertiser. In a related paper, Rutz and Bucklin (2007) considers how advertiser revenue is affected by click-throughs and exposures. This work is important because it demonstrates that advertiser valuations differ for various placements and keywords, and that the bids are likely to be related to placements. Ghose and Yang (2007) further investigate the relationships among different metrics such as click-through rate, conversion rate, bid price, and advertisement rank. Though extant empirical research on sponsored search establishes a firm link between advertising, slot position, and revenues – and indicates that these effects can differ across advertisers, some limitations of this stream of work remain. First, it emphasizes a single agent (one advertiser), making it difficult to predict how advertisers in an oligopolistic setting might react to a change in the auction mechanism, webpage design, or

information state regarding consumers. Competitive interactions are material to understanding the role of each agent in the context of sponsored search. For example, an advertiser’s value to the search engine pertains not only to its direct payment to the search engine but also to the indirect effect that advertiser has on the intensity of competition during bidding. The increased intensity of competition may serve to drive bids upward and hence increase search engine revenues. Second, the advertisers’ actions affect search engine users and vice-versa. For example, with alternative advertisers being placed at premium slots on a search result page, it is likely that users’ browsing behaviors will be different. As advertisers make decisions with the consideration of users’ reactions, any variations of users’ behaviors provide feedback on advertisers’ actions and thus will ultimately affect the search engine revenue.

Integrating these two research streams suggests it is desirable to both model and estimate the equilibrium behavior of all the agents in a network setting. In this regard, sponsored search advertising can be characterized as a two-sided market wherein searchers and advertisers interact on the platform of the search engine (Rochet and Tirole (2006)). This enables us to generalize a structural modeling approach advanced by Yao and Mela (2008) to study two-sided markets. These authors model bidder and seller behavior in the context of electronic auctions to explore the effect of auction house pricing on the equilibrium number of listings and closing prices. However, additional complexities exist in the keyword search setting including: i) the aforementioned information asymmetry between advertisers and the search engine and ii) the substantially more complex auction pricing mechanism used by search engines relative to the fixed fee auction house pricing considered in Yao and Mela (2008). Moreover, unlike the pricing problem addressed in Yao and Mela (2008), sponsored search bidding is inherently dynamic owing to the use of lagged advertising click rates to determine current period advertising placements. Hence we incorporate the growing literature of two-step dynamic game estimation (e.g., Hotz and Miller (1993); Bajari et al. (2007); Bajari et al. (2008)). Instead of explicitly solving for the equilibrium dynamic bidding strategies, the two-step estimation approach assumes that observed bids are generated by equilibrium play and then use the distribution of bids to infer underlying primitive variables of bidders (e.g., the advertiser’s expectation about the return from advertising). A similar method is also used in an auction context in Jofre-Bonet and Pesendorfer (2003). However, our approach is unique inasmuch as it is a Bayesian instantiation of these estimators, which leads to desirable small sample properties and enables considerable flexibility in modeling choices. Equipped with these advertiser primitives, we solve the dynamic game played by the advertiser to ascertain how changes in search engine policy affect equilibrium bidding behavior.

### 3 Empirical Context

The data underpinning our analysis is drawn from a major search engine for high technology consumer products. Within this broad search domain, we consider search for music management software because the category is relatively isolated in the sense that searches for this product do not compete with others on the site.<sup>7</sup> The category is a sizable one for this search engine as well. Along with the increasing popularity of MP3 players, the use of music management PC software is increasing exponentially, making this an important source of revenue. The goal of the search engine is to enable consumers to identify and then download trial versions of these software products before their final purchase.<sup>8</sup> It is important to note that the approach we develop can readily generalize to other contexts and that we consider this particular instantiation to be a particular illustration of a more general approach.

#### 3.1 Data Description

The data are comprised of three files, including:

- Bidding file. Bidding is logged into a file containing the bidding history of all active bidders from January 2005 to August 2007. It records the exact bids submitted, the time of each bid submission, and the resulting monthly allocation of slots. Hence, the unit of analysis is vendor-bid event. These data form the cornerstone of our bidding model.
- Product file. Product attributes are kept in a file that records, for each software firm in each month, the characteristics of the software they purvey. This file also indicates the download history of each product in each month.
- Consumer file. Consumer log files record each visit to the site and are used to infer whether downloads occur as well as browsing histories. A separate but related file includes registration information and detailed demographics for those site visitors that are registered. These data are central to the bidding model in the context of complete information.

We detail each of these files in turn.

---

<sup>7</sup>The search engine defines music management broadly enough that an array of different search terms (e.g., MP3, iTunes, iPod, lyric, etc.) yield the same search results for the software products in this category. Hence we consider the consumer decision of whether to search for music software on the site and whether to download given a search.

<sup>8</sup>A “click” and a “download” are essentially the same from the perspectives of the advertiser, consumer, and search engine. In the “click” case, a consumer makes several clicks to investigate and compare products offered by different vendors and then makes a final purchase. In the “download” case, a consumer downloads several products and makes the comparison before final purchase. Hence there is no difference for a “click” and a “download” in the current context. We use “click” and “download” interchangeably throughout the paper.

### 3.1.1 Bidding File

Most search engines yield “organic” search results that are often displayed as a list of links sorted by their relevance to the search query (Bradlow and Schmittlein (2000)). Sponsored search involves advertisements placed above or along side the organic search results (See Figure 1). Given that users are inclined to view the topmost slots in the page (Ansari and Mela (2003)), advertisers are willing to pay a premium for these more prominent slots (Goldfarb and Tucker (2008)).

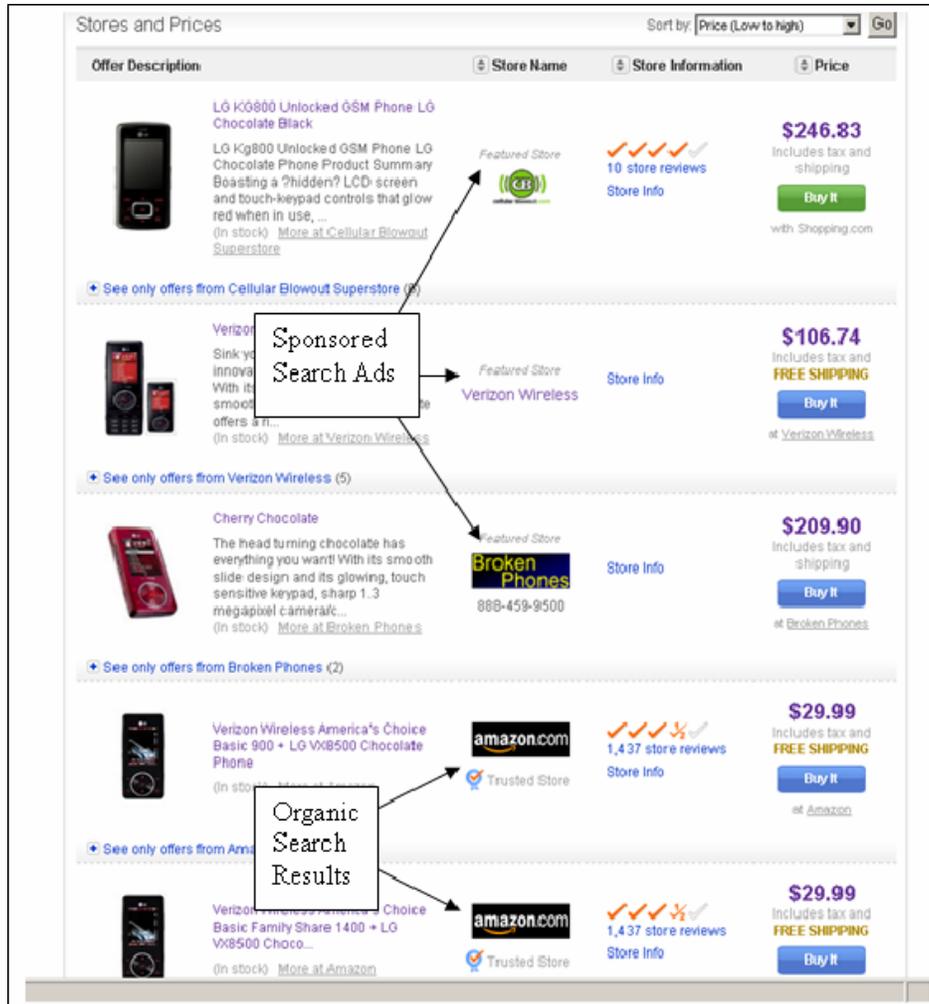


Figure 1: Searching “chocolate phone” Using A Specialized Search Engine

To capitalize on this premium, advertising slots are auctioned off by search engines. Advertisers specify bids on a per-click basis for a respective search term. Though most search engines use auctions to price advertisements, there is considerable variation in the nature of the auctions they use. For example Overture.com (who pioneered Internet search auctions and is now a part of

Yahoo!) adopted a first price auction wherein the advertiser bidding the highest amount per click received the most prominent placement at the cost of its own bid for each click.<sup>9</sup> First price auctions are still used by **Shopping.com** and a number of other Internet properties. **Google** has developed an algorithm which factors in not only the level of the bid, but the expected click-through rate of the advertiser. This enhances search engine revenue because these revenues depend not only on the per-click bid, but also the number of clicks a link receives. Another distinction of the **Google** practice is that advertisers pay the next bidder’s bid (adjusted for click-through rates) as opposed to their own bids.<sup>10</sup>

The mechanism used by the firm we consider is similar to that of **Google** except that the considered search engine uses a first price auction in place of a second price auction (we intend to compare the efficacy of this mechanism to that of **Google** in our policy experiments). Winning bids are denoted as *sponsored search results* and the site flags these as sponsored links. The site we consider affords up to five premium slots which is far less than the 400 or so products that would appear at the search engine. Losing bidders and non-bidders are listed beneath the top slots on the page and like previous literature we denote these listings as *organic search results*.

These advertiser bidding histories by month are all captured in a database. In particular, the search engine collects bidding and demographic data on all advertisers (products attributes, products download history, and bids from active bidders). We detail these data next. Table 1 reports summary statistics for the bidding files. At this search engine, bids were submitted on a monthly basis. Over the 32 months from January 2005 to August 2007, 322 bids (including zeros) were submitted by 21 software companies.<sup>11</sup> As indicated in Table 1, bidders on average submitted about 22 positive bids in this interval (slightly less than once per month). The average bid amount (conditioned on bidding) was \$0.20 with a large variance across bidders and time.

Table 1: Bids Summary Statistics

|                          | Mean  | Std. Dev. | Minimum | Maximum |
|--------------------------|-------|-----------|---------|---------|
| Non-zero Bids ( $\phi$ ) | 19.55 | 8.32      | 15      | 55      |
| Non-zero Bids/Bidder     | 21.78 | 10.46     | 1       | 30      |
| All Bids ( $\phi$ )      | 8.14  | 11.04     | 0       | 55      |
| Bids/Bidder              | 23.13 | 9.68      | 1       | 32      |

<sup>9</sup>In the economics literature, such an auction with multiple items (slots) where bidders pay what they bid is sometimes termed as discriminatory auction (Krishna (2002)).

<sup>10</sup>With a simplified setting, Edelman et al. (2007) show that the Google practice may result in an equilibrium with bidders’ payoffs equivalent to the Vickrey-Clarke-Groves (VCG) auction, whereas VCG auction has been proved to maximize total payoffs to bidders (Groves (1979)). Iyengar and Kumar (2006) further show that under some conditions the Google practice induces VCG auction’s dominant “truth-telling” bidding strategy, i.e., bidders will bid their own valuations.

<sup>11</sup>Since some products were launched after January 2005, they were not observed in all periods.

### 3.1.2 Product File

Searching for a keyword on this considered site results in a list of relevant software products and their respective attributes. Attribute information is stored in a product file along with the download history of all products that appeared in this category from January 2005 to August 2007. In total, these data cover 394 products over 32 months. The attributes include the price of the non-trial version of a product, backward compatibility with preceding operating systems (e.g., Windows 98 and Windows Server 2003), expert ratings provided by the site, and consumer ratings of the product.<sup>12</sup> Trial versions typically come with a 30-day license to use the product for free, after which consumers are expected to pay for its use. Expert ratings at the site are collected from several industrial experts of these products. The consumer rating is based on the average feedback score about the product from consumers. Tables 2 and 3 give summary statistics for all products as well as active bidders' products. Based on the compatibility information, we sum each product's compatibility dummies and define this summation as a measure for that product's compatibility with older operating systems. This variable is later used in our estimation.

Table 2: Product Compatibility

|                     | Percentage |
|---------------------|------------|
| All Products        |            |
| Windows NT 4.0      | 54         |
| Windows 98          | 64         |
| Windows Me          | 66         |
| Windows 2000        | 91         |
| Windows Server 2003 | 43         |
| Bidders' Products   |            |
| Windows NT 4.0      | 67         |
| Windows 98          | 67         |
| Windows Me          | 71         |
| Windows 2000        | 85         |
| Windows Server 2003 | 57         |

Overall, active bidders' products have higher prices, better ratings, and more frequent updates.

### 3.1.3 Consumer File

The consumer file contains the log files of consumers from May 2007 to August 2007. This file contains each consumer's browsing log when they visit the search engine both within the search site and across Internet properties owned by the search site. The consumer file also has the registration information for those that register.

---

<sup>12</sup>We further considered file size but found many missing values. Moreover, in light of increased Internet speed, file size has become somewhat inconsequential in the download decision and thus is omitted from our analysis.

Table 3: Product Attributes and Downloads

|                                     | Mean    | Std. Dev. | Minimum | Maximum |
|-------------------------------------|---------|-----------|---------|---------|
| All Products                        |         |           |         |         |
| Non-trial Version Price \$          | 16.65   | 20.43     | 0       | 150     |
| Expert Rating (if rated)            | 3.87    | 0.81      | 2       | 5       |
| Average Consumer Rating (if rated)  | 3.89    | 1.31      | 1       | 5       |
| Months Lapse Since Last Update      | 15.31   | 9.88      | 1       | 31      |
| Compatibility Index                 | 3.29    | 1.47      | 0       | 5       |
| Number of Downloads/(Product×Month) | 1367.29 | 9257.16   | 0       | 184442  |
| Bidders' Products                   |         |           |         |         |
| Non-trial Version Price \$          | 21.97   | 15.87     | 0       | 39.95   |
| Expert Rating (if rated)            | 4       | 0.50      | 3       | 5       |
| Average Consumer Rating (if rated)  | 4.06    | 0.91      | 2.5     | 5       |
| Months Lapse Since Last Update      | 2.38    | 0.66      | 1       | 3       |
| Compatibility Index                 | 3.51    | 1.51      | 0       | 5       |
| Number of Downloads/(Product×Month) | 1992.12 | 6557.43   | 0       | 103454  |

The browsing log of a consumer records the entry time, browsing path, and duration of the visit. It also indicates whether the consumer made downloads and, if yes, which products she downloaded. Upon a user viewing the search results of software products, the search engine allowed the consumer to sort the results based on some attributes such as the ratings; consumers can also filter products based on some criteria such as whether a product’s non-trial version is free. The browsing log records the sorting and filtering actions of each consumer. Prior to sorting and filtering, the top five search results are allocated to sponsored search slots and the remaining slots are ordered by how recently the software has been updated. There is a small, discrete label indicating whether a search result is sponsored, and sorting and filtering will often remove these links from the top five premium slots.

As the demographic information upon the registration is only optional, the dataset provides little if any reliable demographics of consumers. Hence we focus instead upon whether a consumer is a registered user of the search engine and on their past search behavior at the other website properties, in particular whether they visited any music related site (which should control for the consumers’ interests in music).

## 4 Model

The model incorporates behaviors of the agents interacting on the search engine platform: i) advertisers who bid to maximize their respective profits and ii) utility maximizing consumers who decide whether to click on the advertiser’s link. For any given policy applied by the search engine, this integrated model enables us to predict equilibrium revenues for the search engine (the consumer-

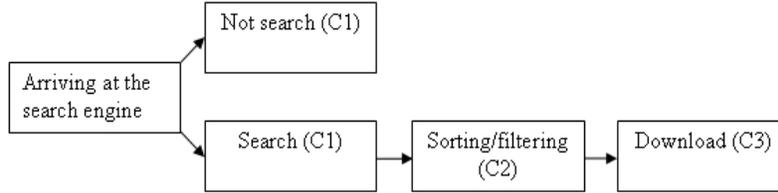


Figure 2: Consumer Decisions

advertiser interactions are analogous to a sub-game contingent on search engine behavior). The behavior of the bidder (advertiser) is dependent on the behavior of the consumer as consumer behavior affects advertiser expectations for downloads and, hence, their bids. The behavior of the consumer is dependent upon the advertiser because the rank of the advertisement affects the behavior of the consumer. Hence, the behaviors are interdependent. Because advertisers move prior to consumers' actions in this game, we first expost the consumer model and then solve the bidder problem in a backward deduction manner.

## 4.1 Consumer Model

Advertiser profit (and therefore bidding strategy) is incumbent upon their forecast of consumer downloads for their products  $d_j^t(k, X_j^t; \Omega_c)$ , where  $k$  denotes the position of the advertisement on the search engine results page,  $X_j^t$  indicate the attributes of the advertiser  $j$ 's product at time  $t$ , and  $\Omega_c$  are parameters to be estimated. Thus, we seek to develop a forecast for  $d_j^t(k, X_j^t; \Omega_c)$  and the attendant consequences for bidding. To be consistent with the advertisers information set, we begin by basing these forecasts of consumer behavior solely on statistics observed by the advertiser: the aggregate download data and the distribution of consumers characteristics. Later, in the policy section of the paper, we assess what happens to bidding behavior and platform revenues when disaggregate information is revealed to advertisers by the platform. We begin by describing the consumer's download decision process and how it affects the overall number of downloads.

### 4.1.1 The Consumer Decision Process

Figure 2 overviews the decisions made by consumers. In any given period  $t$ , the consumer's problem is whether and which software to select in order to maximize their utility. The resolution of this problem is addressed by a series of conditional decisions.

First, the consumer decides whether she should search on the category considered in this analysis (C1). We presume that the consumer will search on the site if it maximizes her expected utility.<sup>13</sup>

Conditioned upon engaging a search, the consumer next decides whether to sort and/or filter the results (C2). The two search options lead to the following 4 options for viewing the results:  $\kappa = \{0 \equiv \text{neither}, 1 \equiv \text{sorting but not filtering}, 2 \equiv \text{not sorting but filtering}, 3 \equiv \text{sorting and filtering}\}$ . For each option, the set of products returned by the search engine differs in terms of the number and the order of products. Consumers choose the sorting/filtering option that maximizes their expected utility.

Third, the consumer then chooses which, if any products to download (C3). We presume that consumers choose to download software if it maximizes their expected utility. We discuss the modeling details for this process in a backward induction manner (C3–C1).

**Download** We assume that consumers exhibit heterogeneous preferences for the products and that these consumers choose products to download to maximize their expected payoffs. Consumer  $i$  of preference segment  $g$  ( $g = 1, 2, \dots, G$ ) has some underlying latent utility  $u_{ijt}^{g\kappa}$  for downloading software  $j$  in period  $t$ , conditional on her sorting/filtering choice  $\kappa$ . A product will be downloaded if and only if  $u_{ijt}^{g\kappa} \geq 0$ . Let  $a$  index product attributes

$$u_{ijt}^{g\kappa} = \tilde{\alpha}_j^g + \sum_a x_{jat}^\kappa \tilde{\beta}_a^g + \tilde{\varepsilon}_{ijt}^{g\kappa} \quad (1)$$

where

- $x_{jat}^\kappa$  is the observed attribute  $a$  of product  $j$ ; product attributes also includes product  $j$ 's slot  $k$  on the search page that may vary conditional on sorting/filtering choice  $\kappa$  (hence the superscript  $\kappa$ );
- $\tilde{\beta}_a^g$  is consumer  $i$ 's "taste" regarding product attribute  $a$ , which is segment specific;
- $\tilde{\varepsilon}_{ijt}^{g\kappa}$ 's are individual idiosyncratic preference shocks, realized *after* the sorting/filtering decision. They are independently and identically distributed over individuals, products and periods as zero mean normal random variables.

To allow the variance of the download ( $\tilde{\varepsilon}_{ijt}^{g\kappa}$ ) and sorting/filtering errors ( $\xi_{it}^{g\kappa}$ , which will be detailed below) to differ, both must be properly scaled (cf., Train (2003), Chapter 2). Hence, we invoke the following assumption:

**Assumption 1:**  $\tilde{\varepsilon}_{ijt}^{g\kappa}$ 's are independently and identically distributed normal random variables with mean 0 and variance normalized to  $(\delta^g)^2$ .  $\xi_{it}^{g\kappa}$ 's are independently and identically distributed Type I extreme value random variables.

---

<sup>13</sup>Though we do not explicitly model the consumer's decision to search across different terms, product categories or competitors, our model incorporates an "outside option" that can be interpreted as a composite of these alternative behaviors.

Under assumption 1, we may re-define the utility in Equation 1 as

$$u_{ijt}^{g\kappa} = \delta^g \left( \underbrace{\alpha_j^g + \sum_a x_{jat}^\kappa \beta_a^g}_{\bar{u}_{ijt}^{g\kappa}} + \varepsilon_{ijt}^{g\kappa} \right) \quad (2)$$

where  $\{\alpha_j^g, \beta_a^g, \varepsilon_{ijt}^{g\kappa}\} = \{\tilde{\alpha}_j^g, \tilde{\beta}_a^g, \tilde{\varepsilon}_{ijt}^{g\kappa}\} / \delta^g$ ;  $\bar{u}_{ijt}^{g\kappa}$  is the scaled “mean” utility and  $\varepsilon_{ijt}^{g\kappa} \sim N(0, 1)$ . The resulting choice process is a multivariate probit choice model.<sup>14</sup> Let  $d_{ijt} = 1$  stand for downloading and  $d_{ijt} = 0$  stand for not downloading. We have

$$d_{ijt} = \begin{cases} 1 & \text{if } u_{ijt}^{g\kappa} \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

and the probability of downloading conditional on parameters  $\{\alpha_j^g, \beta_a^g\}$  is

$$\begin{aligned} \Pr(d_{ijt} = 1) &= \Pr(u_{ijt}^{g\kappa} \geq 0) \\ &= \Pr(\delta^g (\bar{u}_{ijt}^{g\kappa} + \varepsilon_{ijt}^{g\kappa}) \geq 0) \\ &= \Pr(-\varepsilon_{ijt}^{g\kappa} \leq \bar{u}_{ijt}^{g\kappa}) \\ &= \Phi(\bar{u}_{ijt}^{g\kappa}) \end{aligned} \quad (4)$$

where  $\Phi(\cdot)$  is the standard normal distribution CDF.

**Sorting and Filtering** Prior to making a download decision, consumers face several filtering and sorting options which we index as  $\kappa = 0, 1, 2, 3$ . We expect consumers to choose the option that maximizes their expected download utility. Although consumers know the distribution of the product utility error terms ( $\tilde{\varepsilon}_{ijt}^{g\kappa}$ ), these error terms do not realize before the sorting/filtering. Hence consumers can only form an expectation about the total utilities of all products under a given sorting/filtering option  $\kappa$  before choosing that option. Let  $U_{it}^{g\kappa}$  denote the total expected utility from products under option  $\kappa$ , which can be calculated based on Equation 1:

$$U_{it}^{g\kappa} = \sum_j E_\varepsilon(u_{ijt}^{g\kappa} | u_{ijt}^{g\kappa} \geq 0) \Pr(u_{ijt}^{g\kappa} \geq 0). \quad (5)$$

This definition reflects that a product’s utility is realized only when it is downloaded. Hence, the expected utility  $E_\varepsilon(u_{ijt}^{g\kappa} | u_{ijt}^{g\kappa} \geq 0)$  is weighted by the download likelihood,  $\Pr(u_{ijt}^{g\kappa} \geq 0)$ . The expectation,  $E_\varepsilon(\cdot)$ , is taken over the random preference shocks  $\varepsilon_{ijt}^{g\kappa}$ .

In addition to  $U_{it}^{g\kappa}$ , individuals may accrue additional benefits or costs for using sorting/filtering option  $\kappa$ . These benefits or costs may arise from individual differences of efficiency or experience

---

<sup>14</sup>It can be shown that, under very weak assumptions, download decisions across multiple products with the purpose of maximizing total expected utility can be represented by a multivariate binary choice probit model.

in terms of engaging the various options for ordering products. We denote such benefits or costs by random terms  $\xi_{it}^{g\kappa}$ 's. As indicated in assumption 1,  $\xi_{it}^{g\kappa}$ 's are i.i.d. Type I extreme value.  $\xi_{it}^{g\kappa}$  is not observed by researchers but known to individual  $i$ . Note that these sorting/filtering benefits or costs do not materialize during the consumption of the products. Therefore, they do not enter the latent utility in Equation (1). The total utility of search option  $\kappa$  is thus given by

$$z_{it}^{g\kappa} = U_{it}^{g\kappa} + \xi_{it}^{g\kappa}. \quad (6)$$

Consumers choose the option of sorting/filtering that leads to the highest total utility  $z_{it}^{g\kappa}$ .

With  $\xi_{it}^{g\kappa}$  following a Type I extreme value distribution, the choice of sorting/filtering becomes a logit model such that

$$\Pr(\kappa)_{it}^g = \frac{\exp(U_{it}^{g\kappa})}{\sum_{\kappa'=0}^3 \exp(U_{it}^{g\kappa'})} \quad (7)$$

To better appreciate the properties of this model, note that  $U_{it}^{g\kappa}$  in Equation 5 can be written in a closed form:<sup>15</sup>

$$\begin{aligned} U_{it}^{g\kappa} &= \sum_j E_\varepsilon(u_{ijt}^{g\kappa} | u_{ijt}^{g\kappa} \geq 0) \cdot \Pr(u_{ijt}^{g\kappa} \geq 0) \\ &= \delta^g \sum_j \left( \bar{u}_{ijt}^{g\kappa} + \frac{\phi(\bar{u}_{ijt}^{g\kappa})}{\Phi(\bar{u}_{ijt}^{g\kappa})} \right) \cdot \Phi(\bar{u}_{ijt}^{g\kappa}). \end{aligned} \quad (8)$$

With such a formulation, the factors driving the person's choice of filtering or sorting become more apparent:

- Filtering eliminates options with negative utility, such as highly priced products (because consumer price sensitivity is negative). As a result, the summation in Equation 8 for the filter option will increase as the negative  $\bar{u}_{ijt}^{g\kappa}$  are removed. This raises the value of the filter option suggesting that price sensitive people are more likely to filter on price.

<sup>15</sup>For a normal random variable  $x$  with mean  $\mu$ , standard deviation  $\sigma$  and left truncated at  $a$  (Greene (2003)),  $E(x|x \geq a) = \mu + \sigma \lambda(\frac{a-\mu}{\sigma})$ , where  $\lambda(\frac{a-\mu}{\sigma})$  is the hazard function such that  $\lambda(\frac{a-\mu}{\sigma}) = \frac{\phi(\frac{a-\mu}{\sigma})}{1 - \Phi(\frac{a-\mu}{\sigma})}$ .

Hence with  $u_{ijt}^{g\kappa} \sim N(\delta^g \bar{u}_{ijt}^{g\kappa}, (\delta^g)^2)$ , we have

$$\begin{aligned} &E(u_{ijt}^{g\kappa} | u_{ijt}^{g\kappa} \geq 0) \\ &= (\delta^g \cdot \bar{u}_{ijt}^{g\kappa} + \delta^g \cdot \frac{\phi(-\frac{\delta^g \cdot \bar{u}_{ijt}^{g\kappa}}{\delta^g})}{1 - \Phi(-\frac{\delta^g \cdot \bar{u}_{ijt}^{g\kappa}}{\delta^g})}) \\ &= \delta^g (\bar{u}_{ij}^{g\kappa} + \frac{\phi(\bar{u}_{ij}^{g\kappa})}{\Phi(\bar{u}_{ij}^{g\kappa})}) \end{aligned}$$

- Sorting re-orders products by their attribute levels. Products that appear low on a page will typically have lower utility regardless of their product content (because consumer slot rank sensitivity is negative). For example, suppose a consumer relies more on product ratings. By moving more desirable items that have high ratings up the list, sorting can increase the  $\bar{u}_{ijt}^{g\kappa}$  for these items, thereby increasing the resulting summation in Equation 8 and the value of this sorting option.<sup>16</sup>

**Keyword Search** The conditional probability of keyword search takes the form

$$\Pr(\text{search}_i^g) = \frac{\exp(\lambda_0^g + \lambda_1^g IV_{it}^g)}{1 + \exp(\lambda_0^g + \lambda_1^g IV_{it}^g)} \quad (9)$$

where  $IV_i^g$  is the inclusive value for searching conditional on the segment membership.  $IV_{it}^g$  is defined as

$$IV_{it}^g = \log\left[\sum_{\kappa} \exp(U_{it}^{g\kappa})\right]. \quad (10)$$

This specification can be interpreted as the consumer making a decision to use a keyword search based on the rational behavior of utility maximization (McFadden (1977); Ben-Akiva and Lerman (1985)).<sup>17</sup> A search term is more likely to be invoked if it yields higher expected utility.

**Segment Membership** Recognizing that consumers are heterogeneous in behaviors described above, we apply a latent class model in the spirit of Kamakura and Russell (1989) to capture heterogeneity in consumer preferences. Heterogeneity in preference can arise, for example, when some consumers prefer some features more than others. We assume  $G$  exogenously determined segments.<sup>18</sup> Note that our specification implies a dependency across decisions within the same segment that is not captured via the stage-specific decision errors, and therefore captures the effect of unobserved individual specific differences in search behavior.

The prior probability for user  $i$  being a member of segment  $g$  is defined as

$$pg_{it}^g = \exp(\gamma_0^g + Demo'_{it}\gamma^g) / \sum_{g'=1}^G \exp(\gamma_0^{g'} + Demo'_{it}\gamma^{g'}) \quad (11)$$

where  $Demo'_{it}$  is a vector of attributes of user  $i$  such as demographics and past browsing history; vector  $\{\gamma_0^g, \gamma^g\}_{\forall g}$  contains parameters to be estimated. For the purpose of identification, one segment's parameters are normalized to zero.

<sup>16</sup>In particular, in the data over 80% of consumers who used the sorting option chose ratings to re-order products. Thus, we suspect that consumers who rely on ratings are more likely to use the sorting option to see which items are the most popular ones.

<sup>17</sup>This specification is consistent with the consumer information structure such that  $\xi_i^{g\kappa}$  is not observed by researchers but known to consumer  $i$ .

<sup>18</sup>It is possible to allow for continuous mixtures of heterogeneity as well. In our application, many consumers enter only once, making it difficult to identify a consumer specific term for them.

### 4.1.2 Consumer Downloads

The search, sort/filter, and download models can be integrated to obtain an expectation of the number of downloads that an advertiser receives for a given position of its keyword advertisement. Advertisers must form this expectation predicated on observed *aggregate* download totals,  $d_j^t$  (in contrast to the search engine who observes  $y_{ijt}, \kappa_{it}$  and  $Demo_{it}$ ).<sup>19</sup>

To develop this aggregate download expectation, we begin by noting that the download utility  $u_{ijt}^{g\kappa}$  is a function of consumer specific characteristics and decisions  $\zeta_{ijt} = [\varepsilon_{ijt}^{g\kappa}, \xi_{it}^{g\kappa}, search_{it}^g, \text{segment } g \text{ membership}, Demo_{it}']$  and that an advertiser needs to develop an expectation of downloads over the distribution of these unobserved (to the advertiser) individual characteristics. Define

$$A_{ijt} = \{\zeta_{ijt} : u_{ijt}^{g\kappa} \geq 0\},$$

i.e.,  $A_{ijt}$  is the set of values of  $\zeta_{ijt}$  which will lead to the download of product  $j$  in period  $t$ .

Let  $\mathcal{D}(\zeta_{ijt})$  denote the distribution of  $\zeta_{ijt}$ . The likelihood of downloading product  $j$  in period  $t$  can be expressed as

$$P_j^t = \int_{\zeta_{ijt} \in A_{ijt}} \mathcal{D}(\zeta_{ijt}) \quad (12)$$

$$= \int_{Demo_{it}} \sum_g \sum_{\kappa} [\Phi(\bar{u}_{ijt}^{g\kappa}) \frac{\exp(U_{it}^{g\kappa})}{\sum_{\kappa'=0}^3 \exp(U_{it}^{g\kappa'})}] \Pr(search_{it}^g) pg_{it}^g d\mathcal{D}(Demo_{it}) \quad (13)$$

where the first term in the brackets captures the download likelihood, the second term captures the search strategy likelihood, and the first term outside the brackets captures the likelihood of search.  $pg_{it}^g$  is the probability of segment  $g$  membership and  $\mathcal{D}(Demo_{it})$  is the distribution of demographics.

Correspondingly, the advertiser with attributes  $X_j^t$  has an expected number of downloads for appearing in slot  $k$ ,  $d_j^t(k, X_j^t; \Omega_c)$ , which can be computed as follows

$$d_j^t(k, X_j^t; \Omega_c) = M_t P_j^t \quad (14)$$

where  $\Omega_c$  is the set of consumer preference parameters;  $M_t$  is the market size in period  $t$ ; and  $X_j^t$  is the vector of the  $a$  product characteristics,  $x_{jat}$ .

Product attributes are posted on the search engine and are therefore common knowledge to all advertisers and consumers. We assume these  $X_j^t$  are exogenous within the scope of our sponsored search analysis for several reasons. First, advertisers distribute and promote their products through multiple channels and they do so over longer periods of time than considered herein. Hence, product

<sup>19</sup>We discuss the corresponding advertiser download expectation under complete information in section 7.3.

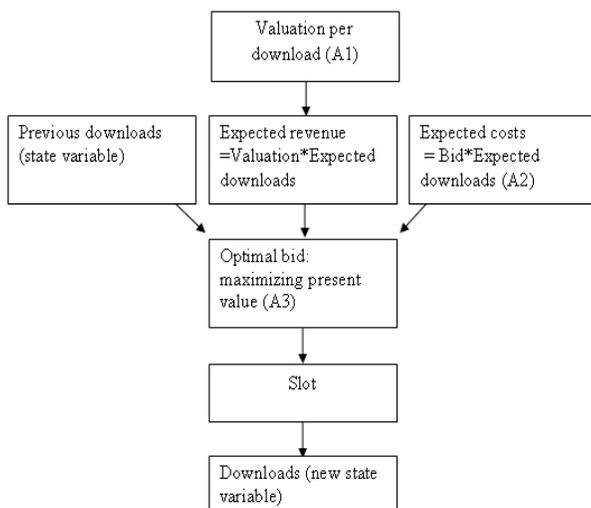


Figure 3: Advertiser Decisions

attributes are more likely to be determined via broader strategic considerations than the particular auction game and time frame we consider. Second, the attribute levels for each product are stable over the duration of our data and analysis. We would expect more variation in attribute levels if they were endogenous to the particular advertiser and search engine decisions we consider. Third, because there is little or no variation in product attributes over time, it is not feasible to estimate endogenous attribute decision making with our data.

## 4.2 Advertiser Model

Figure 3 overviews the dynamic game played by the advertiser. Advertiser  $j$ 's problem is to decide the optimal bid amount  $b_j^t$  with the objective of maximizing discounted present value of payoffs.<sup>20</sup> Higher bids lead to greater revenues because they yield more favorable positions on the search engine, thereby yielding more click-throughs for the advertiser. However, higher bids also increase costs (payments) leading to a trade-off between costs and revenues. The optimal decision of whether and how much to bid is incumbent upon the bidding mechanism, the characteristics of the advertiser, the information available at the time of bidding (including the state variables), and the nature of competitive interactions.

An advertiser's period profit for a download is the value it receives from the download less the costs (payments) of the download. Though we do not observe the value of a download, we

<sup>20</sup>Because the search engine used in our application has the dominant market share in the considered category, we do not address advertiser bidding on other sites. Also, it would be difficult to obtain download data from these more minor competitors. We note this is an important issue and call for future research.

infer this value by noting the observed bid can be rationalized only for a particular value accrued by the advertiser. We presume this value is drawn from a distribution known to all firms. The total period revenue for the advertiser is then the value per download times the expected number of downloads.<sup>21</sup> The total period payment upon winning is the number of downloads times the advertiser’s bid. Hence, the total expected period profit is the number of downloads times the profit per download (i.e., the value per downloads less the payment per download).

Of course, the bid levels and expected download rates are affected by rules of the auction. Though we elaborate in further details on the specific rules of bidding below, at this point we simply note that the rules of the auction favor advertisers whose products were downloaded more frequently in the past since such products are more likely to lead to higher revenues for the platform.<sup>22</sup> Current period downloads are, in turn, affected by the position of the advertisement on the search engine. Because past downloads affect current placement, and thus current downloads, the advertiser’s problem is inherently dynamic; and past downloads are treated as a state variable.

Finally, given the rules of the auction, we note that all advertisers move simultaneously. While we presume a firm knows its own value, we assume competing firms know only the distribution of this value.

The process is depicted in Figure 3. We describe the process with more details as follows: Section 4.2.1 details the rules of the auction that affect the seller costs (A2), section 4.2.2 details the advertisers’ value distribution (A1), and section 4.2.3 indicates how period values and costs translate to discounted profits and the resulting optimal bidding strategy (A3).

#### 4.2.1 Seller Costs and the Bidding Mechanism

We begin by discussing how slot positions are allocated with respect to bids and the effect of these slot positions on consumer downloads (and thus advertiser revenue).

Upon a consumer completing a query, the search engine returns  $k = 1, 2, \dots, K, \dots, \bar{N}$  slots covering the products of all firms. Only the top  $K = 5$  slots are considered as premium slots. Auctions for these  $K$  premium slots are held every period ( $t = 1, 2, \dots$ ). An advertiser seeks to appear in a more prominent slot because this may increase demand for the advertiser’s product. Slots  $K + 1$  to  $\bar{N}$  are non-premium slots which compose a section called organic search section.

There are  $N$  advertisers who are interested in the premium slots ( $N \leq \bar{N}$ ). In order to procure a more favorable placement, advertiser  $j$  submits bid  $b_j^t$  in period  $t$ . These bids, submitted simultaneously, are summarized by the vector  $\mathbf{b}^t = \{b_1^t, b_2^t, \dots, b_N^t\}$ .<sup>23</sup> Should an advertiser win slot  $k$ ,

<sup>21</sup>The expected number of downloads is inferred from the consumer model and we have derived this expression in section 4.1.2.

<sup>22</sup>This is because the payment made to the search engine by an advertiser is the advertiser’s bid times its total downloads.

<sup>23</sup>For the purpose of a clear exposition, we sometimes use boldface notations or pairs of braces to indicate vectors whose elements are variables across all bidders. For example,  $\mathbf{d}^t = \{d_j^t\}_{\forall j}$  is a vector whose elements are  $d_j^t, \forall j$ .

the realized number of downloads  $d_j^t$  is a random draw from the distribution with the expectation  $d_j^t(k, X_j^t; \Omega_c)$ . The placement of advertisers into the  $K$  premium slots is determined by the ranking of their  $\{b_j^t d_j^{t-1}\}_{\forall j}$ , i.e., the product of current bid and last period realized downloads; the topmost bidder gets the best premium slot; the second bidder gets the second best premium slot; and so on. A winner of one premium slot pays its own bid  $b_j^t$  for each download in the current period. Hence, the total payment for winning the auction is  $b_j^t d_j^t$ .

If an advertiser is not placed at one of the  $K$  premium slots, it will appear in the organic section; advertisers placed in the organic section *do not* pay for downloads from consumers. The ranking in the organic search section is determined by the product update recency at period  $t$ , which is a component of the attribute vector of each product,  $X_j^t$ . Other attributes include price, consumer ratings, and so on.

Given that the winners are determined in part by the previous period's downloads, the auction game is inherently dynamic. Before submitting a bid, the commonly observed state variables at time  $t$  are the realized past downloads of all bidders from period  $t - 1$ ,<sup>24</sup>

$$\mathbf{s}^t = \mathbf{d}^{t-1} = \{d_1^{t-1}, d_2^{t-1}, \dots, d_N^{t-1}\}. \quad (15)$$

#### 4.2.2 Seller Value

The advertiser's bid determines the cost of advertising and must be weighed against the potential return when deciding how much to bid. We denote advertiser  $j$ 's valuation regarding one download of its product in period  $t$  as  $v_j^t$ . We assume that this valuation is private information but drawn from a normal distribution that is commonly known to all advertisers. Specifically,

$$\begin{aligned} v_j^t &= v(X_j^t; \theta) + f_j + r_j^t \\ &= X_j^t \theta + f_j + r_j^t \end{aligned} \quad (16)$$

where  $\theta$  are parameters to be estimated and reflect the effect of product attributes on valuation. The  $f_j$  are firm-specific fixed effect terms assumed to be identically and independently distributed *across advertisers*. This fixed effect term captures heterogeneity in valuations that may arise from omitted firm-specific effects such as more efficient operations. The  $r_j^t \sim N(0, \psi^2)$  are private shocks to an advertiser's valuation in period  $t$ , assumed to be identically and independently distributed *across advertisers and periods*. The sources of this private shock may include: (1) temporary increases in the advertiser's valuation due to some events such as a promotion campaign; (2) unexpected shocks to the advertiser's budget for financing the payments of the auction; (3) temporary production capacity constraint for delivering the product to users; and so on. The random shock  $r_j^t$  is realized

---

<sup>24</sup>Though state variables can be categorized as endogeneous (past downloads) and exogenous (product attributes), our exposition characterizes only downloads as state variables because these are the only states whose evolution is subject to a dynamic constraint.

at the beginning of period  $t$ . Although  $r_j^t$  is private knowledge, we assume the distribution of  $r_j^t \sim N(0, \psi^2)$  is common knowledge among bidders. We further assume the fixed effect  $f_j$  of bidder  $j$  is known to all bidders but not to researchers. Given bidders may observe opponents' actions for many periods, the fixed effect can be inferred among bidders (Greene (2003)).

### 4.2.3 Seller Profits: A Markov Perfect Equilibrium (MPE)

Given  $v_j^t$  and state variable  $\mathbf{s}^t$ , predicted downloads and search engine's auction rules, bidder  $j$  decides the optimal bid amount  $b_j^t$  with the objective of maximizing discounted present value of payoffs. In light of this, every advertiser has an expected period payoff, which is a function of  $\mathbf{s}^t$ ,  $\mathbf{X}^t$ ,  $r_j^t$  and all advertisers' bids  $\mathbf{b}^t$

$$\begin{aligned}
& \mathbf{E} \pi_j (\mathbf{b}^t, \mathbf{s}^t, \mathbf{X}^t, r_j^t; \theta, f_j) & (17) \\
= & \mathbf{E} \sum_{k=1}^K \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot (v_j^t - b_j^t) \cdot d_j^t(k, X_j^t; \Omega_c) \\
& + \mathbf{E} \sum_{k=K+1}^{\bar{N}} \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot v_j^t \cdot d_j^t(k, X_j^t; \Omega_c) \\
= & \mathbf{E} \sum_{k=1}^K \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot (X_j^t \theta + f_j + r_j^t - b_j^t) \cdot d_j^t(k, X_j^t; \Omega_c) \\
& + \mathbf{E} \sum_{k=K+1}^{\bar{N}} \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot (X_j^t \theta + f_j + r_j^t) \cdot d_j^t(k, X_j^t; \Omega_c)
\end{aligned}$$

where the expectation for profits is taken over other advertisers' bids  $\mathbf{b}_{-j}^t$ .  $\Pr(k|\cdot)$  is the conditional probability of advertiser  $j$  getting slot  $k$ ,  $k = 1, 2, \dots, \bar{N}$ .  $\Pr(k|\cdot)$  depends not only on bids, but also on states  $\mathbf{s}^t$  (the previous period's downloads) and product attributes  $\mathbf{X}^t$ .<sup>25</sup> This is because: i) the premium slot allocation is determined by the ranking of  $\{b_j^t d_j^{t-1}\}_{\forall j}$ , where  $\mathbf{d}^{t-1}$  are the state variables and ii) the organic slot allocation is determined by product update recency, an element of  $\mathbf{X}^t$ .

In addition to the current period profit, an advertiser also takes its expected future payoffs into account when making decisions. In period  $t$ , given the state vector  $\mathbf{s}^t$ , advertiser  $j$ 's discounted expected future payoffs evaluated prior to the realization of the private shock  $r_j^t$  is given by

$$\mathbf{E} \left[ \sum_{\tau=t}^{\infty} \rho^{\tau-t} \pi_j (\mathbf{b}^\tau, \mathbf{s}^\tau, \mathbf{X}^\tau, r_j^\tau; \Omega_{aj}) | \mathbf{s}^t \right] \quad (18)$$

where  $\Omega_{aj} = \{\theta, \psi, f_j\}$ , with  $a$  denoting advertiser behavior (in contrast to the parameters  $\Omega_c$  in the consumer model). Further, we denote  $\Omega_a = \{\Omega_{aj}\}_{j=1,2,\dots,N} = \{\theta, \psi, f_j\}_{j=1,2,\dots,N}$ . The parameter  $\rho$  is a common discount factor. The expectation is taken over the random term  $r_j^t$ , bids in period  $t$

<sup>25</sup>Note that  $\mathbf{s}^t$  and  $\mathbf{X}^t$  are observed by all bidders before bidding.

as well as all future realization of  $\mathbf{X}^t$ , shocks, bids, and state variables. The state variables  $\mathbf{s}^{t+1}$  in period  $t + 1$  is drawn from a probability distribution  $P(\mathbf{s}^{t+1}|\mathbf{b}^t, \mathbf{s}^t, \mathbf{X}^t)$ .

We use the concept of a pure strategy Markov perfect equilibrium (MPE) to model the bidder's problem of whether and how much to bid in order to maximize the discounted expected future profits (Bajari et al. (2007); Ryan and Tucker (2008); Dubé et al. (2008); and others). The MPE implies that each bidder's bidding strategy only depends on the then-current profit-related information, including state,  $\mathbf{X}^t$  and its private shock  $r_j^t$ . Hence, we can describe the equilibrium bidding strategy of bidder  $j$  as a function  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t) = b_j^t$ .<sup>26</sup> Given a state vector  $\mathbf{s}$ , product attributes  $\mathbf{X}$  and prior to the realization of current  $r_j$  (with the time index  $t$  suppressed), bidder  $j$ 's expected payoff under the equilibrium strategy profile  $\boldsymbol{\sigma} = \{\sigma_1, \sigma_2, \dots, \sigma_N\}$  can be expressed recursively as:

$$V_j(\mathbf{s}, \mathbf{X}; \boldsymbol{\sigma}) = \mathbf{E} \left[ \pi_j(\boldsymbol{\sigma}, \mathbf{s}, \mathbf{X}, r_j; \Omega_a) + \rho \int_{\mathbf{s}'} V_j(\mathbf{s}', \mathbf{X}'; \boldsymbol{\sigma}) dP(\mathbf{s}'|\mathbf{b}, \mathbf{s}, \mathbf{X}) | \mathbf{s} \right] \quad (19)$$

where the expectation is taken over current and future realizations of random terms  $\mathbf{r}$  and  $\mathbf{X}$ . To test the alternative theory that advertiser's may be myopic in their bidding, we will also solve the advertiser problem under the assumption that period profits are maximized independently over time.

The advertiser model can then be used in conjunction with the consumer model to forecast advertiser behavior as we shall discuss in the policy simulation section. In a nutshell, we presume advertisers will choose bids to maximize their expected profits. A change in information states, bidding mechanisms, or webpage design will lead to an attendant change in bids conditioned on the advertisers value function, which we estimate as described next.

## 5 Estimation

### 5.1 An Overview

Though it is standard to estimate dynamic MPE models via a dynamic programming approach such as a nested fixed point estimator (Rust (1994)), this requires one to repetitively evaluate the value function (Equation 19) through dynamic programming for each instance in which the parameters of the value function are updated. Even when feasible, it is computationally demanding to implement this approach. Instead, we consider the class of two-step estimators. The two-step estimators are predicated upon the notion that the dynamic program can be estimated in two steps that dramatically simplify the estimation process by facilitating the computation of the value function.

---

<sup>26</sup>The bidding strategies are individual specific due to the fixed effect  $f_j$  (hence the subscript  $j$ ). For the purpose of clear exposition, we use  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$  instead of  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t; f_j)$  throughout the paper. Multiple observations for each advertiser allows the identification of  $\sigma_j, j = 1, 2, \dots, N$ .

Specifically, in this application we implement the two-step estimator proposed by Bajari et al. (2007) (BBL henceforth).

As can be seen in equation 19, the value function is parameterized by the primitives of the value distribution  $\Omega_a$ . Under the assumption that advertisers are behaving rationally, these advertiser private values for clicks should be consistent with observed bidding strategies. Therefore, in the second step estimation, values of  $\Omega_a$  are chosen so as to make the observed bidding strategies congruent with rational behavior. We detail this step in Section 5.3 below.

However, as can be observed in equations 19 and 17, computation of the value function is also incumbent upon i) the bidding policy function that maps bids to the states (downloads), product attributes, and private shocks  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t) = b_j^t$ ; ii) the expected downloads  $d_j^t(k, X_j^t; \Omega_c)$ ; and iii) a function that maps the likelihood of future states as a function of current states and actions  $P(\mathbf{s}^{t+1} | \mathbf{b}^t, \mathbf{s}^t, \mathbf{X}^t)$ . These are estimated in the first step as detailed in Section 5.2 below and then substituted into the value function used in the second step estimation.

## 5.2 First Step Estimation

In the first step of the estimation we seek to obtain:

1. A “partial” policy function  $\tilde{\sigma}_j(\mathbf{s}, \mathbf{X})$  describing the equilibrium bidding strategies as a function of the observed state variables and product attributes,  $\mathbf{X}$ . We estimate the policy function by noting that players adopt equilibrium strategies (or decision rules) and that behaviors generated from these decision rules lead to correlations between i) the observed states (i.e., past period downloads) and product characteristics and ii) advertiser decisions (i.e., bids). The partial policy function captures this correlation. In our case, we use a fixed effects Tobit model to link bids to states and product characteristics as described in Section A.1.1 of the Appendix. Subsequently, the full policy function  $\sigma_j(\mathbf{s}, \mathbf{X}, r_j^t)$  can be inferred based on  $\tilde{\sigma}_j(\mathbf{s}, \mathbf{X})$  and the distribution of private random shocks  $r_j^t$ . The partial policy function can be thought of as the marginal distribution of the full policy function. Inferences regarding the parameters of the full policy function can be made by finding the distribution of  $r_j^t$  that, when “integrated out,” leads to the best rationalization for the observed bids. We discuss our approach to infer the full policy functions from the partial policy function in Appendix A.1.1.
2. The expected downloads for a given firm at a given slot,  $d_j^t(k, X_j; \Omega_c)$ . The  $d_j^t(k, X_j; \Omega_c)$  follows directly from the consumer model. Hence, the first step estimation involves i) estimating the parameters of the consumer model and then ii) using these estimates to compute the expected number of downloads. The expected total number of downloads as a function of slot position and product attributes is obtained by using the results of the consumer model to forecast the likelihood of each person downloading the software and then integrating

these probabilities across persons.<sup>27</sup> We discuss our approach for determining the expected downloads in Section A.1.2 of the Appendix.

3. The state transition probability  $P(\mathbf{s}'|\mathbf{b}, \mathbf{s}, \mathbf{X})$  which describes the distribution of future states (current period downloads) given observations of the current state (past downloads), product attributes, and actions (current period bids). These state transitions can be derived by i) using the policy function to predict bids as a function of past downloads, ii) determining the slot ranking as a function of these bids, past downloads and product attributes, and then iii) using the consumer model to predict the number of current downloads as a function of slot position. Details regarding our approach to determining the state transition probabilities is outlined in Section A.1.3 of the Appendix.

With the first step estimates of  $\sigma_j(\mathbf{s}, \mathbf{X}, r_j^t)$ ,  $d_j^t(k, X_j; \Omega_c)$ , and  $P(\mathbf{s}'|\mathbf{b}, \mathbf{s}, \mathbf{X})$ , we can compute the value function in Equation 19 as a function with only  $\Omega_a$  unknown. In the second step, we estimate these parameters.

### 5.3 Second Step Estimation

The goal of the second step estimation is to recover the primitives of the bidder value function,  $\Omega_a$ . The intuition behind how the second-stage estimation works is that true parameters should rationalize the observed data. For bidders' data to be generated by rational plays, we need

$$V_j(\mathbf{s}, \mathbf{X}; \sigma_j, \boldsymbol{\sigma}_{-j}; \Omega_a) \geq V_j(\mathbf{s}, \mathbf{X}; \sigma'_j, \boldsymbol{\sigma}_{-j}; \Omega_a), \forall \sigma'_j \neq \sigma_j \quad (20)$$

where  $\sigma_j$  is the equilibrium policy function and  $\sigma'_j$  is some deviation from  $\sigma_j$ . This equation means that any deviations from the observed equilibrium bidding strategy will not result in more profits. Otherwise, the strategy would not be optimal. Hence, we first simulate the value functions under the equilibrium policy  $\sigma_j$  and the deviated policy  $\sigma'_j$  (i.e., the left hand side and the right hand side of equation 20). Then we choose  $\Omega_a$  to maximize the likelihood that Equation 20 holds. We describe the details of this second step estimation in Appendix A.2.

### 5.4 Sampling Chain

With the posterior distributions for the advertiser and consumer models established, we estimate the models using MCMC approach as detailed in Appendix B. This is a notable deviation from prior research that uses a gradient based technique. The advantage of using a Bayesian approach,

---

<sup>27</sup>As an aside, we note that advertisers have limited information from which to form expectations about total downloads because they observe the aggregate information of downloads but not the individual specific download decisions. Hence, advertisers must infer the distribution of consumer preferences from these aggregate statistics. In a subsequent policy simulation we allow the search engine to provide individual level information to advertisers in order to assess how it affects advertiser behavior and, therefore, search engine revenues.

as long as suitable parametric assumptions can be invoked, is that it facilitates model convergence, has desirable small sample properties, increases statistical efficiency, and enables the estimation of a wide array of functional forms (Rossi et al. (2005)). Indeed, posterior sampling distributions for many of the parameters are highly skewed and/or have thin tails, and Kolmogorov-Smirnov tests indicate these posterior distributions are often not normally distributed. Hence, we seek to make a methodological contribution to the burgeoning literature on two-step estimators for dynamic games.

## 6 Results

### 6.1 First Step Estimation Results

Recall, the goal of the first step estimation is to determine the policy function,  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$ , the expected downloads  $d_j^t(k, X_j^t; \Omega_c)$ , and the state transition probabilities  $P(\mathbf{s}^{t+1} | \mathbf{b}^t, \mathbf{s}^t, \mathbf{X}^t)$ . To determine  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$ , we first estimate the partial policy function  $\tilde{\sigma}_j(\mathbf{s}^t, \mathbf{X}^t)$  and then compute the full policy function. To determine  $d_j^t(k, X_j^t; \Omega_c)$ , we first estimate the consumer model and then compute the expected downloads. Last  $P(\mathbf{s}^{t+1} | \mathbf{b}^t, \mathbf{s}^t, \mathbf{X}^t)$  is derived from the consumer model and the partial policy function. Thus, in the first stage we need only to estimate the partial policy function and the consumer model. With these estimates in hand, we compute  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$ ,  $d_j^t(k, X_j^t; \Omega_c)$ , and  $P(\mathbf{s}^{t+1} | \mathbf{b}^t, \mathbf{s}^t, \mathbf{X}^t)$  for use in the second step. Thus, below, we report the estimates for the partial policy function and the consumer model on which these functions are all based.

#### 6.1.1 Partial Policy Function $\tilde{\sigma}_j(\mathbf{s}, \mathbf{X})$

The vector of independent variables  $(\mathbf{s}, \mathbf{X})$  for the partial policy function (i.e., the Tobit model of advertiser behavior that captures their bidding policy as outlined in Appendix section A.1.1) contains the following variables:

- Product  $j$ 's state variable, last period download  $d_j^{t-1}$ . We reason that high past downloads increase the likelihood of a favorable placement and, therefore, affect bids.
- Two market level variables: the sum of last period downloads from all bidders and the number of bidders in last period. Since we only have 322 observations of bids, it is infeasible to estimate a parameter to reflect the effect of each opponent's state (i.e., competition) on the optimal bid. Moreover, it is unlikely a bidder can monitor every opponent's state in each period before bidding because such a strategy carries high cognitive and time costs. Hence, summary measures provide a reasonable approximation of competing states in a limited information context. Others in the literature who have invoked a similar approach include Jofre-Bonet and Pesendorfer (2003) and Ryan (2006). Like them, we find this provides a fair model fit. Another measure of competitive intensity is the number of opponents. Given that bidders

cannot directly observe the number of competitors in the current period, we used a lagged measure of the number of bidders.

- Product  $j$ 's attributes in period  $t$  ( $X_j^t$ ), including its non-trial version price, expert rating, consumer rating, update recency, and compatibility with an older operating system. We expect that a higher quality product will yield greater downloads thereby affecting the bidding strategy.
- An advertiser specific constant term to capture the impact of the fixed effect  $f_j$  on bidding strategy.
- To control the possible effect of the growth of ownership of MP3 players, we also collect the average lagged price of all new MP3 players in the market from a major online retailing platform ([www.pricegrabber.com](http://www.pricegrabber.com)).

Table 4 reports the estimation results for the Tobit model.<sup>28</sup> As a measure of fit of the model, we simulated 10,000 bids from the estimated distribution. The probability of observing a positive simulated bid is 42.0%; the probability of observing a positive bid in the real data is 41.6%. Conditional on observing a positive simulated bid, these bids have a mean of \$0.21 with a standard deviation of \$0.07. In the data, the mean of observed positive bids is \$0.20 and the standard deviation is \$0.08. We also estimate the same model only using 70% (227/322) of the observations and use the remaining 30% as a holdout sample. The change in estimates is negligible. We then use the holdout to simulate 10,000 bids. The probability of observing a positive bid is 40.2%, while there are 42.4% positive bids in the holdout sample. Among the positive simulated bids, the mean is \$0.23 and the standard deviation is \$0.06. The corresponding statistics in the holdout are \$0.21 and \$0.07. Overall, the fit is good.

The estimates yield several insights into the observed bidding strategy. First, the bidder's state variable ( $d_j^{t-1}$ ) is negatively correlated with its bid amount  $b_j^t$  because the ranking of the auction is determined by the product of  $b_j^t$  and  $d_j^{t-1}$ . All else being equal, a higher number of lagged downloads means a bidder can bid less to obtain the same slot. Second, the total number of lagged downloads in the previous period ( $\sum_{j'} d_{j'}^{t-1}$ ) and the lagged number of bidders both have a positive impact on a bidder's bid. We take this to mean increased competition leads to higher bids. Third, bids are increasing in the product price. One possible explanation is that a high priced product yields more value to the firm for each download, and hence the firm competes more aggressively for a top slot. Similarly and fourth, a high price for MP3 players reflects greater value for the downloads also leading to a positive effect on bids. Fifth, "Lapse Since Last Update" has a negative effect on bids. Older products are more likely obsolete, thereby generating lower value for consumers. If

---

<sup>28</sup>To conserve space, we do not report the estimates of fixed effects. These are available upon request from the authors.

Table 4: Bidding Function Estimates

|                                  | Median | 95% Interval   |
|----------------------------------|--------|----------------|
| $\varphi$                        |        |                |
| Lagged Downloads $_{jt}/10^3$    | -0.12* | (-0.16, -0.09) |
| Total Lagged Downloads $_t/10^3$ | 0.04*  | (0.01, 0.08)   |
| Lagged Number of Bidders $_t$    | 0.04*  | (0.01, 0.06)   |
| Lapse Since Last Update $_{jt}$  | -0.31* | (-0.70, -0.02) |
| Non-trial Version Price $_{jt}$  | 0.36*  | (0.33, 0.39)   |
| Expert Ratings $_{jt}$           | 0.47   | (-0.26, 1.23)  |
| Consumer Ratings $_{jt}$         | 0.81*  | (0.10, 1.50)   |
| Compatibility Index $_{jt}$      | -0.91* | (-1.74, -0.20) |
| Lagged MP3 Player Price $_t$     | 0.03*  | (0.02, 0.03)   |
| $\tau$                           | 8.14*  | (6.78, 11.37)  |
| Log Marginal Likelihood          |        | -1123.59       |

this is the case, firms can reasonably expect fewer final purchases after downloads and, therefore, bid less for these products. Likewise and sixth, higher compatibility with prior software versions reflects product age leading to a negative estimate for this variable. Finally, ratings from consumers and experts (albeit not significant for experts) have a positive correlation with bid amounts – these again imply greater consumer value for the goods, making it more profitable to advertise them.

### 6.1.2 Consumer Model

The consumer model is estimated using MCMC approach based on the posterior distribution described in Appendix A.1.2. We consider the download decisions for each of the 21 products who entered auctions, plus the top 3 products who did not. Together these firms constitute over 80% of all downloads. The remaining number of downloads are scattered across 370 other firms, each of whom has a negligible share. Hence, we exclude them from our analysis.

Table 5: Alternative Numbers of Latent Segments

|             | Log Marginal Likelihood |
|-------------|-------------------------|
| 1 Segment   | -12769.3                |
| 2 Segments* | -12511.9                |
| 3 Segments  | -12571.1                |
| 4 Segments  | -12551.4                |

Note: \* indicates the model with the best fit.

We calibrate the model by estimating an increasing number of latent segments until there is no significant improvement in model fit. We use log marginal likelihood as the measurement for model fit. In Table 5 we report the comparison of the log marginal likelihoods for models with up

to four segments. The model with two segments gives the best result in terms of maximizing the log marginal likelihood.

Table 6: Consumer Model Estimates

|  | Segment 1 (89.5%)<br>(Infrequent searcher) | Segment 2 (10.5%)<br>(Frequent searcher and slot sensitive) |
|--|--|---|
|  | Median<br>(95%Interval)                    | Median<br>(95%Interval)                                     |
| $\beta^g$ (utility parameters)         |  |   |
| Constant                               | -0.09<br>(-0.11,0.001)                     | 0.35<br>(0.31,0.38)   |
| Slot Rank                              | -0.08<br>(-0.06,-0.09)                     | -0.51<br>(-0.52,-0.50)                                      |
| Non-trial Version Price                | 0.03<br>(0.03,0.04)                        | -0.04<br>(-0.04,-0.03)                                      |
| Expert Ratings                         | 0.16<br>(0.15,0.17)                        | 0.06<br>(0.06,0.07)   |
| Consumer Ratings                       | 0.11<br>(0.11,0.12)                        | 0.03<br>(0.03,0.05)   |
| Compatibility Index                    | -0.08<br>(-0.09,-0.07)                     | 0.16<br>(0.16,0.17)   |
| Total Download Percentage              | 0.01<br>(-0.02,0.05)                       | 0.09<br>(0.08,0.10)   |
| $\delta^g$ (sorting/filtering scaling) | 1.52<br>(1.48,1.55)                        | 1.87<br>(1.78,1.99)   |
| $\lambda^g$ (search probability)       |  |   |
| $\lambda_0^g$ (base)                   | -10.22<br>(-10.75,-9.60)                   | -0.78<br>(-1.21,-0.54)                                      |
| $\lambda_1^g$ (1-correlation)          | 0.02<br>(0.01,0.02)                        | 0.03<br>(0.01,0.04)   |
| $\gamma^g$ (segment parameters)        |  |   |
| Constant                               | -  | -4.01<br>(-4.74,-2.87)                                      |
| Music Site Visited                     | -  | 7.66<br>(5.77,10.18)  |
| Registration Status                    | -  | -0.24<br>(-1.91,0.86)                                       |
| Product Downloaded<br>in Last Month    | -  | -0.40<br>(-1.57,-0.02)                                      |

Table 6 presents the estimates of the model with two segments. Conditional on the estimated segment parameters and demographic distribution, we calculate the segment sizes as 89.5% and 10.5%, respectively. Based on the parameter estimates in Table 6, Segment 1 is less likely to initiate a search (low  $\lambda_0^g$ ). The primary basis of segmentation is whether a customer has visited a music website at other properties owned by the download website; these customers are far more likely to be in the frequent download segment. Moreover, upon engaging a search, this segment appears to be less sensitive to slot ranking but more sensitive to consumer and expert ratings than segment 2. Segment 2, composed of those who search more frequently, relies more heavily on the slot order when downloading. Overall, we speculate that segment 1 are the occasional downloaders who base their download decisions on others’ ratings and tend not to exclude goods of high price. In contrast, segment 2 contains the “experts” or frequent downloaders who tend to rely on their

own assessments when downloading. Of interest is the finding that those in segment 2 rely more on advertising slot rank. This is consistent with a perspective that frequent downloaders might be more strategic; knowing that higher quality firms tend to bid more and obtain higher ranks, those who download often place greater emphasis on this characteristic.

More insights on this difference in download behavior across segments can be gleaned by determining the predicted probabilities of searching and sorting/filtering by computing  $\Pr(search_i^g) = \frac{\exp(\lambda_0^g + \lambda_1^g IV_{it}^g)}{1 + \exp(\lambda_0^g + \lambda_1^g IV_{it}^g)}$  and  $\Pr(\kappa)_{it}^g = \frac{\exp(U_{it}^{g\kappa})}{\sum_{\kappa'=0}^3 \exp(U_{it}^{g\kappa'})}$  in Equations 9 and 7, respectively. Table 7 reports these probabilities for both segments.

Table 7: Searching Behavior of Consumers

|                          | Segment 1 | Segment 2 |
|--------------------------|-----------|-----------|
| Searching                | 0.4%      | 60.8%     |
| No sorting or filtering  | 78.7%     | 86.1%     |
| Sorting but no filtering | 21.3%     | 8.2%      |
| No sorting but filtering | → 0       | 5.5%      |
| Sorting and filtering    | → 0       | 0.3%      |

Table 7 confirms the tendency of those in segment 2 to be more likely to initiate a search in the focal category. Though comprising only 10.5% of all consumers, they represent 90% of all searches. The increased searching frequency suggests that members of segment 2 are ideal customers to target because more searches lead to more downloads.

Moreover, segment 2 (heavy downloaders) is more likely to be influenced by sponsored advertising. To see this, note that segment 1 consumers put more weight on the ratings of products (e.g., expert and consumer ratings) than do segment 2 consumers. As a consequence segment 1 consumers engage in far more sorting. Sorting eliminates the advantage conferred by sponsored advertising because winners of the sponsored search auction may be sorted out of desirable slots on the page.

However, Table 7 indicates consumers in segment 1 (occasional downloaders) seldom filter. Filtering occurs when consumers seek to exclude negative utility options from the choice set (e.g., omitting a product not compatible with a certain operating system). Given the high sensitivity to rank order, segment 2 consumers are more prone to eliminate advertised options by filtering. We suspect this segment, by virtue of being a frequent visitor, searches for very specific products that conform to a particular need. Overall, however, segment 1 is more likely to sort and/or filter than segment 2 (21.3% vs. 13.9%) suggesting that segment 2 is more valuable to advertisers. We will explore this conjecture in more detail in our policy analysis.

## 6.2 Second Step Estimation Results

### 6.2.1 Alternative Models

In addition to our proposed model, we consider two alternative models of seller behavior: i) myopic bidding and ii) heterogeneous valuations across consumer segments.<sup>29</sup> Table 8 reports the fit of each model. In the first alternative model, advertisers maximize period profits independently as opposed to solving the dynamic bidding problem given in Equation 19. This model yields a considerably poorer fit. Hence, we conclude that advertisers are bidding strategically. This strategic behavior might result from dynamics in the bidding process coupled with non-linearities in advertising response. Similar dynamic behavior has been evidenced in the face of non-linear advertising demand systems with dynamics in advertising carryover (Bronnenberg (1998)).

The second model considers the case wherein advertiser valuations for clicks differ across segments. In this model, we augment Equation 16 by allowing it to vary by segment and then integrate this heterogeneity into the seller profit function given by Equation 17. This model leads to only a negligible increase in fit. Closer inspection of the results indicates little difference in valuations across segments, implying advertisers perceive that the conversion rates of each segment are essentially the same. Hence, we adopt the more parsimonious single valuation model. It is further worth noting that all of our subsequent results and policy simulations evidence essentially no change across these two models.

Table 8: Alternative Models

| Model   | Log Marginal Likelihood |
|---|-------------------------|
| Base Model  | -1651.3                 |
| Base Model Without Advertiser Dynamics            | -1701.3                 |
| Base Model With Heterogeneous Customer Valuations | -1645.1                 |

### 6.2.2 Valuation Model Results

Table 9 shows the results of second step estimation for the favored model.<sup>30</sup> With respect to the advertiser value function, we find that newer, more expensive and better rated products yield greater values to the advertiser. This is consistent with our conjecture in Section 6.1.1 that firms bid more aggressively when having higher values for downloads. We find that, after controlling for observed product characteristics, 95% of the variation in valuations across firms is on the order of \$0.02. We attribute this variation in part due to differences in the operating efficiency of the firms.

<sup>29</sup>We do not estimate the discount factor  $\rho$ . As shown in Rust (1994), the discount factor is usually unidentified. We fix  $\rho = 0.95$  for our estimation. We also experiment at  $\rho = 0.90$  and see minimal difference in the results.

<sup>30</sup>Advertiser specific constant terms  $f_j$  are not reported to save space.

Table 9: Value per Click Parameter Estimates

|                                 | Median | 95% Interval   |
|---------------------------------|--------|----------------|
| $\theta$                        |        |                |
| Lapse Since Last Update $_{jt}$ | -0.98* | (-1.02, -0.95) |
| Non-trial Version Price $_{jt}$ | 0.24*  | (0.23, 0.25)   |
| Expert Ratings $_{jt}$          | 0.46*  | (0.38, 0.65)   |
| Consumer Ratings $_{jt}$        | 0.97*  | (0.94, 0.98)   |
| Compatibility Index $_{jt}$     | -0.27* | (-0.29, -0.26) |
| Lagged MP3 Player Price $_t$    | 0.01*  | (0.01, 0.02)   |
| $\psi$ , random shock std. dev. | 1.45*  | (1.43, 1.51)   |
| Log Marginal Likelihood         |        | -1651.33       |

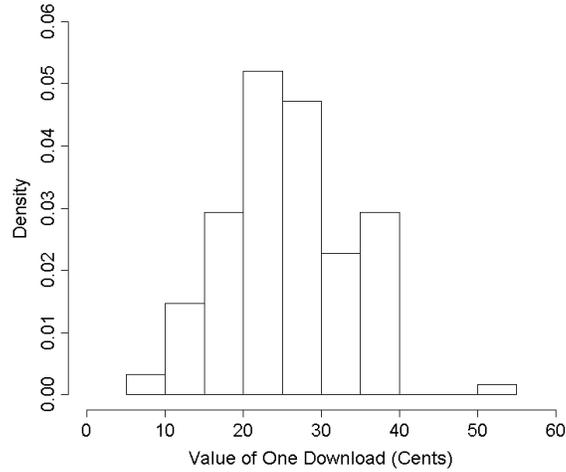


Figure 4: Distribution of Values per Download

Given the second step results, we can further estimate the value of a download to a firm. We estimate the advertiser’s value for a download in each period. The distribution of these estimates across time and advertisers is depicted in Figure 4. As indicated in the figure there is substantial variation in the valuation of downloads. Table 9 explains some of this variation as a function of the characteristics of the software and firm specific effects. Overall, the mean value of a download to these advertisers is \$0.27. This compares to an average bid of \$0.20 as indicated in Table 1. Hence, on average, each click implies an expected return to a firm of about \$0.07. To our knowledge, this is the first paper to impute the advertiser’s return from a click in a keyword search context. One way to interpret these results is to consider the firm’s expected sales per download to rationalize the bid. The firm’s profit per click is roughly  $CR_j^t \cdot P_j^t - b_j^t$ , where  $CR_j^t$  indicates the download-sale conversion rate (or sales per download) and  $P_j^t$  is the non-trial version price. Ignoring dynamic effects and setting this profit per click equal to  $v_j^t - b_j^t$  yields a rough approximation of the conversion rate as  $CR_j^t = v_j^t / P_j^t$ . Viewed in this light, the effect of higher quality software, which raises  $v_j^t$ , leads to a higher implied conversion rate. Noting that the average price of the software is \$22, this average per-click valuation implies that 1.2% of all clicks lead to a purchase (that is, the conversion rate is  $0.27/22 = 1.2\%$ ). This estimate lies within the industry average conversion rate of 1 – 2% reported by Gamedaily.com, suggesting our findings have high face validity.<sup>31</sup>

Another interesting observation about this result is that the mean value is very close to the mean bid (\$0.27 vs. \$0.20), which brings the net profits for advertising almost to a break-even point. To better understand this observation, note that an advertiser’s profitability of advertising depends on the comparison between its marginal return and marginal cost. The marginal return for advertising is the advertiser’s value per click, which is common across all clicks. The marginal cost is the bid because the marginal cost of posting software for download is negligible. If an advertiser’s marginal return is higher than its marginal cost, it will increase its bid level to procure a better slot – depending on the bid levels of its competitors. In equilibrium, the bid level (marginal cost) will approach the value per click (marginal return), consistent with our estimate. This difference will decrease as competition increases.

## 7 Policy Simulations

Given the behavior of consumers and advertisers, we can predict how changes in search engine policy affect overall bidding, downloads, consumer welfare, and revenues. The advertiser-consumer behaviors are analogous to a subgame conditioned on search engine policy. To assess the effect of changes in policy, we recompute the equilibrium behavior of consumers and advertisers conditioned on the new policy. One might ask whether these deviations in policy are valid as the initial strategies

---

<sup>31</sup> “Casual Free to Pay Conversion Rate Too Low.” Gamedaily.com (<http://www.gamedaily.com/articles/features/magid-casual-free-to-pay-conversion-rate-too-low/70943/?biz=1>).

might reflect optimal behavior on the part of the search engine. However, extensive interactions with the search site makes it clear that they have neither considered using these alternative policies nor have they tried them in the past in order to obtain a sense of the strategies' impacts. Hence, it is hard to argue that they are behaving strategically and thus we think these are reasonable policy simulations to consider. Alternatively, estimating a model incorporating the engine's behavior invokes rather strong assumptions of rationality due to the complexity and novelty of the problem. Also, we observe no variation in the considered behaviors of the search engine, meaning there is no means to identify the primitives driving such behaviors.

We describe four policy simulations: i) the effect of alternative webpage designs on search engine revenues, ii) the value of targeting (i.e., not only allowing advertisers to bid on keywords but also on market segments.), iii) the value of disaggregate consumer data, and iv) the effect of alternative pricing mechanisms on search engine revenue. As we can no longer assume the optimal policy estimated in stage one of our two-step estimator remains valid in the face of a change in search engine policy, the following policy simulations involve explicitly solving the dynamic programming problem. Though an involving task, recent advances pertaining to approximate dynamic programming makes solving high-dimension DP problems become possible (Powell (2007)). More details regarding the implementation of the policy simulations are presented in Appendix C. Hence, we limit our discussion to the objectives and insights from these simulations.

## 7.1 Policy Simulation I: Alternative Webpage Design

The goal of the search engine's sorting/filtering options is to provide consumers with easier access to price and rating information across different products. As shown in section 4.1 and evidenced by our results, sorting and filtering play a crucial role in consumer decision process. In light of this outcome, it is possible to consider an alternative webpage design of the search engine – eliminating the option of sorting and filtering for consumers – and assessing the resulting impact on consumer search, advertiser bidding, and the search engine's revenues. Because this change can have contrasting effects on consumer behavior (consumer should be less likely to search on the site because of the decrease in utility arising from fewer search options) and advertiser behavior (advertisers should bid more because of the decreased likelihood their advertisements will be sorted or filtered off of the search results), the overall effect is unclear. Using our model, it can be tested which effect dominates. We do this by setting the probability of consumer choosing no sorting/filtering option in equation 7 as 1. This manipulation mimics the scenario in which the sorting/filtering option is disabled. Under this new policy, we find that the search engine's revenue decreases by 3.7%, suggesting the consumer effect is larger.

Next, to more precisely measure these contrasting effects, we apportion the revenue change across consumers and advertisers. Let  $D_{j0}^t$  ( $D_{j1}^t$ ) denote the number of downloads for product  $j$  in period  $t$  before (after) the change of the webpage. Let  $B_{j0}^t$  ( $B_{j1}^t$ ) denote the bid from advertiser  $j$

in period  $t$  before (after) the new policy. Accordingly we can calculate (i) the revenue effect arising solely from changes in consumer behavior by holding advertiser behavior fixed,  $(\sum_{j,t} B_{j1}^t D_{j1}^t - \sum_{j,t} B_{j1}^t D_{j0}^t)$  and (ii) the effect arising from changing advertiser behavior by holding consumer behavior fixed,  $(\sum_{j,t} B_{j1}^t D_{j0}^t - \sum_{j,t} B_{j0}^t D_{j0}^t)$ . Using this decomposition, we find the effect arising from consumers  $(\sum_{j,t} B_{j1}^t D_{j1}^t - \sum_{j,t} B_{j1}^t D_{j0}^t) / \sum_{j,t} B_{j0}^t D_{j0}^t$  is  $-6.4\%$  while the effect from advertisers  $(\sum_{j,t} B_{j1}^t D_{j0}^t - \sum_{j,t} B_{j0}^t D_{j0}^t) / \sum_{j,t} B_{j0}^t D_{j0}^t$  is  $2.7\%$ . Consistent with this result, consumer welfare as measured by their overall utility, *declines*  $5.6\%$  when the search tools are removed while advertiser profits *increase*  $4.1\%$ . Thus, for the search engine, the disadvantage of this new policy to consumers outweighs the advantages resulting from more aggressive advertiser bidding.

## 7.2 Policy Simulation II: Segmentation and Targeting

It can be profitable for advertisers to target specific consumers. In this instance, instead of a single bid on a keyword, an advertiser can vary its bids across market segments. For example, consider two segments,  $A$  and  $B$ , wherein segment  $B$  is more sensitive to product price and segment  $A$  is more sensitive to product quality. Consider further, two firms,  $X$  and  $Y$ , where firm  $X$  purveys a lower price, but lower quality, product. Intuitively, firm  $X$  should bid more aggressively for segment  $B$  because quality sensitive segment  $A$  will not likely buy the low quality good  $X$ . This should lead to higher revenues for the search engine. On the other hand, there is less bidding competition for firm  $X$  within segment  $B$  because  $Y$  finds this segment unattractive – this dearth of competition can drive the bid of  $X$  down for segment  $B$ . This would place a downward pressure on search engine profits. Hence, the optimal revenue outcome for the search engine is likely to be incumbent upon the distribution of consumer preferences and the characteristics of the goods being advertised. Our approach can assess these effects of segmentation and targeting strategy on the search engine’s revenue.

To implement this policy simulation, we enable the search engine to serve a different advertisement to each market segment and allow advertisers to bid differentially each period for these keyword slots across the two consumer segments (see Appendix C.2 for details). We find the search engine’s resulting revenue increases by  $1.4\%$ . Using a similar decomposition mentioned in section 7.1, we find the revenue effect arising from the consumer side of the market is  $2.2\%$ . We attribute this effect mainly to the enhanced efficiency of advertisements under targeting. In other words, targeting leads to more desirable advertisements for consumers thereby yielding increased downloads. In contrast, the effect arising from advertisers is  $-0.8\%$  as a result of diminished competitive intensity. Overall, the consumer effect of targeting is dominant, and a net gain in profitability is indicated.

This policy also benefits advertisers in two ways: by increasing the efficiency of their advertising and reducing the competitive intensity of bidding within their respective segments. Overall, we project an  $11\%$  increase in advertiser revenue under the targeting policy. Consistent with this view

of consumer gains, consumer welfare increases by 2.9%. In sum, every agent finds this new policy to be an improvement.

### 7.3 Policy Simulation III: Incorporating Disaggregate Level Data

Advertisers and search engines are endowed with different levels of information. The search engine knows all of the clicks made by visitors to its site. The advertiser knows only the total number of downloads all the advertisers received. Hence, there is an information asymmetry arising from the different level of market intelligence accruing to each respective agent. Given this disaggregate consumer information is owned by the search engine but not observed by advertisers, it is relevant to ask how the information revelation from the platform to advertisers will affect advertiser behavior and, hence, platform revenues. In practice, this means that the platform is interested in whether to sell or give this information to advertisers and how it should be priced. More generally, this counterfactual exemplified the value of market intelligence and how it can be computed in the context of a structural model.

Accordingly, we implement a counterfactual scenario under which advertisers have access to the click histories of consumers similar to that described in Section 7.2, except that advertisers are not able to bid by segment. We then assess i) how bidding behavior and returns to advertisers change under this counterfactual information structure and ii) how the resulting revenues change at the search engine. We find little change in bidding behavior, and that search engine profits increase only 0.08%. The finding suggests that advertisers are able to form consistent inferences about consumer behavior from the aggregate download behavior. The finding is interesting because it suggests there is little value to selling this information to advertisers and that aggregate data can be useful for making inferences about consumer heterogeneity (Chen and Yang (2007), Musalem et al. (2008)). Further, the result suggests the value of this disaggregate information lies not in the more efficient estimates, but rather in the ability to exploit this heterogeneity by targeting bids by segment as outlined in Section 7.2.

### 7.4 Policy Simulation IV: Alternative Auction Mechanisms

Auction mechanism design has been an active domain of research since the seminal work of Vickrey (1961). Optimal mechanism design involves several aspects including the rules of the auction, the efficiency of the auction in terms of allocation surplus across players, and so forth. We focus on the payment rules in this investigation. In particular, while the focal search engine currently charges winning advertisers their own bids, many major search engines such as `Google.com` and `Yahoo.com` are applying a “generalized second-price auction” as termed in Edelman et al. (2007). Under the generalized second-price auction rules, winners are still determined by the ranking of  $\{b_{j'}^t, d_{j'}^{t-1}\}_{\forall j'}$ . However, instead of paying its own bid amount, the winner of a slot pays the highest losing bidder’s

bid adjusted by their last period downloads.<sup>32</sup> For example, suppose bidder  $j$  wins a slot with the bid of  $b_j^t$  and last period download  $d_j^{t-1}$ , its payment for each download will be  $b_j^t d_{j'}^{t-1} / d_j^{t-1}$ , where  $j'$  is the highest losing bidders for the slot bidder  $j$  wins.

Though “generalized second-price auction” is widely adopted by major search engines, the optimality of such a mechanism has not been substantiated (Iyengar and Kumar (2006), Katona and Sarvary (2008)). By implementing a policy simulation that contrasts the search engine and advertiser revenues under the two different mechanisms, we find little difference in revenues for the advertiser or search engine (for example, search engine revenues increase 0.01%). However, the advertisers’ bids for clicks approach their values for clicks. Under second price auction, the median ratio of bid/value is 0.97 compared to 0.78 under first price auction. This is consistent with the theory that in equilibrium bidders bid their true values under “generalized second-price auction” (Edelman et al. (2007)). This offers empirical support for the contention that generalized second price auctions yield truth telling – though we find little practical consequence in terms of auction house revenue.

## 8 Conclusion

Given the \$9B firms annually spend on keyword advertising and its rapid growth, we contend that the topic is of central concern to advertisers and platforms that host advertising alike. In light of this growth, it is surprising that there is little extant empirical research pertaining to modeling the demand and pricing for keyword advertising in an integrated fashion across advertisers, searchers, and search engines. As a result, we develop a dynamic structural model of advertiser bidding behavior coupled with an attendant model of search behavior. The interplay of these two agents has a number of implications for the search engine platform that hosts them. Because we need to infer advertiser and consumer valuations and use these estimates to infer the effects of a change in search engine strategy, we develop a structural model of keyword search as a two-sided network. In particular, we consider i) how the platform or search engine should price its advertising via alternative auction mechanisms, ii) whether the platform should accommodate targeted bidding wherein advertisers bid not only on keywords, but also behavioral segments (e.g., those that purchase more often), iii) whether and how the search engine should sell information on individual clicking histories, and iv) how an alternative webpage design of the search engine with less product information would affect bidding behavior and the engine’s revenues.

Our *model of advertiser bidding* behavior is predicated on the advertiser choosing its bids to maximize the net present value of its discounted profits. The period profits contain two components: i) the advertiser’s value for a given click times the number of clicks on the advertisement and ii) the payment in form of the bid per click times the number of clicks on the advertisement. Whereas

---

<sup>32</sup>In the paper by Edelman et al. (2007), the adjustment using last period downloads is not considered.

the advertiser’s costs are determined by their bids, we infer the advertiser’s valuation for clicks. Specifically, we estimate valuations by choosing them such that, for an observed set of bids, the valuations rationalize the bidding strategy, that is, making profits as high as possible. In this sense, our structural model “backs out” the advertiser’s expectation for the profit per click. Given an estimate of these valuations, it becomes possible to ascertain how advertiser profits are affected by a change in the rules of the auction, a change in the webpage design, or a change in the information state of the advertiser.

The advertiser model presents some computational challenges in estimation because it is a dynamic problem. In particular, the advertiser’s past downloads affects its subsequent advertising rank and, therefore, its subsequent downloads. Accordingly, we use recent advances in economics wherein a two-step estimator is applied to the problem (BBL). The first step is used to infer the bidding policy and consumer clicking behavior. The second step is used to infer the advertiser valuations conditioned on the bidding strategy and the consumer clicking decisions. Our approach departs from previous work on two-step estimators via our Bayesian instantiation. Like all MCMC approaches, this innovation enables one to estimate a broader set of models and does not rely on asymptotic for inference (Rossi et al. (2005)).

We find that the advertiser model leads to improvements in fit over a model that treats advertiser decisions myopically. Further, we find that advertiser valuations are positively correlated with product attributes that may enhance product quality. One possible explanation is that a higher quality leads to a greater likelihood of a product purchase after a trial download, thus enhancing the advertiser’s valuation for a download; hence, the more aggressive bidding. Also, estimated valuations for downloads/clicks are consistent with a download to sales ratio of 1.2%, well within industry estimates of 1% to 2%.

As noted above, a central component to the calculation of advertiser profits is the expectation of the number of clicks on its advertisement received from consumers. This expectation of clicks is imputed from our *consumer search and clicking model*. This model, which involves three steps (the choice of whether to search, whether to use search tools, and whether to download), follows from the standard random utility theory (McFadden (1977)), and is computed using traditional MCMC methods adapted to our context. Key findings include the presence of two segments: one which downloads frequently and places greater weight on keyword search order, and another that downloads less often and puts more emphasis on consumer and expert ratings. The latter group constitutes 90% of the market but only 10% of all downloads.

Using the consumer and advertiser model, we conduct *policy simulations pertaining to search engine policy*. Relating to the consumer side, we explore the effect of changing the search engine’s website design in order to reduce usability but increase advertising exposures. Such a change would i) benefit advertisers and lead to higher bids – suggesting an increase in search engine revenues but ii) reduce usability leading to lower traffic – suggesting lower search engine revenues.

Hence, the net impact on search engine profits is unclear. We manipulate usability by removing the sorting and filtering feature on the search engine site. By removing these features, advertiser bids per download increase along with a corollary increase in their profits of 4.1%. In contrast, there is a resulting diminution in search site traffic from loss of consumer usability that engenders a decrease in consumer welfare ( $-5.6\%$ ). Combined, these contrasting impacts lead to an overall reduction of 3.7% in search engine revenue, suggesting it would not be prudent to change the site. Second, we consider the possibility of allowing advertisers to bid by segment and allowing advertising slot rankings to differ by segment. Though this reduces competition within segments (suggesting lower bidding intensity and lesser search engine profits), targeting also enhances the expected number of downloads by increasing the relevance of the advertisements (suggesting larger search engine profits). Overall, the latter effect dominates, leading to an increase in search engine revenues of 1.4%. More efficient targeting also leads to i) increases in advertiser revenues on the order of 11% (because of more efficient consumer response and less intense bidding competition for advertisements) and ii) gains in consumer welfare of 2.9%. Hence, this policy change strictly increases the welfare of all agents. Third, we explore alternative auction designs. We find that a generalized second price auction leads to advertiser bids that are consistent with the advertiser’s valuations for those clicks (i.e., advertisers bid their valuations) while first price auctions lead to lower bids (advertisers shade their bids). This result is consistent with previous theoretical research. However, there is little material difference of this truth telling from the perspective of the search engine; revenues are essentially unchanged.

Several extensions are possible. First, we use a two-step estimator to model the dynamic bidding behavior of advertisers without explicitly solving for the equilibrium bidding strategy. Solving explicitly for this strategy could provide more insights into bidder behavior in this new marketing phenomenon. For example, following the extant literature we assume that a bidder’s return of the advertising only comes from consumers’ clicks. It is possible that advertisers also accrue some values from the exposures at the premium slots. A clear characterization of bidding strategy can better facilitate our understanding about how advertisers value sponsored advertising in term of clicks and exposures and, hence, present a better guideline for search engines to design their pricing schedule. Second, our analysis focuses upon a single category. Bidding across multiple keywords is an important direction for future research. In particular, the existence of multiple keywords auctions may present opportunities for collusions among bidders. For example, advertisers may collusively diverge their bids to different keywords. By doing so, they can find a more profitable trade-off between payments to the search engine and clicks across keywords. In a theoretical paper by Stryzowska (2005) the author shows an equilibrium where bidders implicitly collude across multiple auctions in the context of online auctions such as `eBay.com`. One managerial implication is how to detect and discourage collusion and reduce its negative impact on search engine revenues. Third, competition between search engines over advertisers is not modeled. Though our

data provider has a dominant role in this specific category, inter-engine competition is unattended in the literature. To some extent, sponsored search advertising can be understood as advertisers purchasing products (media) from search engines through auctions. An advertiser makes discrete choice about search engines before entering auctions. Little research has been done on the advertiser's choice problem, even though there is abundant discrete choice research that can be applied (Palma et al. (1992)). Accordingly, the inter-engine competition deserves future attention. Finally, our analysis is predicated on a relatively short duration of bidding behavior. Over the longer-term, there may be additional dynamics in bidding and download behavior that might arise from consumer learning or the penetration of search marketing into the market place, the so called "durable goods problem," (Horsky and Simon (1983)). Overall, we hope this study will inspire further work to enrich our knowledge of this new marketplace.

## References

- Ansari, A. and C. F. Mela (2003). E-customization. *Journal of Marketing Research* 40(2), 131–145.
- Ansari, A., C. F. Mela, and S. A. Neslin (2008). Customer channel migration. *Journal of Marketing Research* 45(1), 60–76.
- Athey, S. and G. Ellison (2008). Position auctions with consumer search. *Working Paper*.
- Bajari, P., C. L. Benkard, and J. Levin (2007). Estimating dynamic models of imperfect competition. *Econometrica* 75(5), 1331–1370.
- Bajari, P., V. Chernozhukov, H. Hong, and D. Nekipelov (2008). Nonparametric and semiparametric analysis of a dynamic game model. *Working Paper*.
- Ben-Akiva, M. and S. Lerman (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.
- Bradlow, E. T. and D. C. Schmittlein (2000). The little engines that could: Modeling the performance of world wide web search engines. *Marketing Science* 19(1), 43–62.
- Bronnenberg, B. J. (1998). Advertising frequency decisions in a discrete markov process under a budget constraint. *Journal of Marketing Research* 35(3), 399–406.
- Chen, Y. and C. He (2006). Paid placement: Advertising and search on the internet. *Working Paper*.
- Chen, Y. and S. Yang (2007). Estimating disaggregate models using aggregate data through augmentation of individual choice. *Journal of Marketing Research* 44(4), 613–621.
- Dubé, J.-P., G. J. Hitsch, and P. Chintagunta (2008). Tipping and concentration in markets with indirect network effects. *Working Paper*.
- Edelman, B., M. Ostrovsky, and M. Schwarz (2007). Internet advertising and the generalized second price auction: Selling billions of dollars worth of keywords. *The American Economic Review* 97, 242–259.
- Feng, J. (2008). Optimal mechanism for selling a set of commonly-ranked objects. *Marketing Science* 27(3), 501–512.
- Garg, D., Y. Narahari, and S. S. Reddy (2006). An optimal mechanism for sponsored search auctions and comparison with other mechanisms. *Working Paper*.
- Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis* (2 ed.). Chapman & Hall/CRC.

- Ghose, A. and S. Yang (2007). An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Working Paper*.
- Goldfarb, A. and C. Tucker (2008). Search engine advertising: Pricing ads to context. *Working Paper*.
- Greene, W. H. (2003). *Econometric analysis*. Prentice Hall.
- Groves, T. (1979). Efficient collective choice when compensation is possible. *The Review of Economic Studies* 46(2), 227–241.
- Horsky, D. and L. S. Simon (1983). Advertising and the diffusion of new products. *Marketing Science* 2(1), 1–17.
- Hotz, V. J. and R. A. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies* 60(3), 497–529.
- Iyengar, G. and A. Kumar (2006). Characterizing optimal adword auctions. *Working Paper*.
- Jofre-Bonet, M. and M. Pesendorfer (2003). Estimation of a dynamic auction game. *Econometrica* 71(5), 1443–1489.
- Judd, K. L. (1998). *Numerical methods in economics*. MIT Press.
- Kamakura, W. A. and G. J. Russell (1989, November). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* 26(4), 379–390.
- Katona, Z. and M. Sarvary (2008). The race for sponsored links: A model of competition for search advertising. *Working Paper*.
- Krishna, V. (2002). *Auction Theory*. Academic Press.
- McFadden, D. L. (1977). Modeling the choice of residential location. *Cowles Foundation Discussion Paper No. 477*.
- Musalem, A., E. T. Bradlow, and J. S. Raju (2008). Who’s got the coupon? estimating consumer preferences and coupon usage from aggregate information. *Journal of Marketing Research* 45(6), 715–730.
- Palma, A. D., S. P. Anderson, and J. F. Thisse (1992). *Discrete Choice Theory of Product Differentiation*. Massachusetts Institute of Technology Press.
- Pesendorfer, M. and P. Schmidt-Dengler (2008). Asymptotic least squares estimators for dynamic games. *Review of Economic Studies* 75, 901–928.

- Powell, W. B. (2007). *Approximate Dynamic Programming: Solving the Curses of Dimensionality*. Wiley, Inc. Publication.
- Qiu, F., Z. Liu, and J. Cho (2005). Analysis of user web traffic with a focus on search activities. In *Proceedings of the International Workshop on the Web and Databases (WebDB)*, pp. 103–108.
- Rochet, J.-C. and J. Tirole (2006). Two-sided markets: A progress report. *Rand Journal of Economics* 37(3), 645–667.
- Rossi, P. E., G. M. Allenby, and R. E. McCulloch (2005). *Bayesian statistics and marketing*. John Wiley and Sons.
- Rust, J. (1994). Structural estimation of markov decision processes. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume IV. Amsterdam: Elsevier Science.
- Rutz, O. J. and R. E. Bucklin (2007). A model of individual keyword performance in paid search advertising. *Working Paper*.
- Rutz, O. J. and R. E. Bucklin (2008). From generic to branded: A model of spillover dynamics in paid search advertising. *Working Paper*.
- Ryan, S. (2006). The costs of environmental regulation in a concentrated industry. *Working Paper, Massachusetts Institute of Technology*.
- Ryan, S. and C. Tucker (2008). Heterogeneity and the dynamics of technology adoption. *Working Paper, Massachusetts Institute of Technology*.
- Spiegelhalter, D. J., N. G. Best, W. R. Gilks, and H. Inskip (1996). Hepatitis b: a case study in mcmc methods. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, Chapter 2, pp. 21–43. Chapman & Hall/CRC.
- Stryszowska, M. (2005). Simultaneous vs. overlapping internet auctions. *Working Paper*.
- Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Varian, H. R. (2007). Position auction. *International Journal of Industrial Organization* 25, 1163–1178.
- Vickrey, W. (1961). Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance* 16(1), 8–37.
- Weitzman, M. L. (1979, May). Optimal search for the best alternative. *Econometrica* 47(3), 641–654.
- Yao, S. and C. F. Mela (2008). Online auction demand. *Marketing Science* 27(5), 861–885.

# Appendix

## A Two Step Estimator

### A.1 First Step Estimation

#### A.1.1 Estimating the Advertiser’s Policy Function

**The Partial Policy Function** The partial policy function links states ( $\mathbf{s}$ ) and characteristics ( $\mathbf{X}$ ) to decisions ( $\mathbf{b}$ ). Ideally this relation can be captured by a relatively flexible parametric form and estimated via methods such as maximum likelihood or MCMC to obtain the partial policy function parameter estimates. The exact functional form is typically determined by model fit comparison among multiple specifications (e.g., Jofre-Bonet and Pesendorfer (2003)). We considered several different specifications for the distribution of bids and found the truncated normal distribution gives the best fit in terms of marginal likelihoods.<sup>33</sup> Specifically, we allow

$$\begin{aligned} b_j^t &= \begin{cases} y_j^{t*} & \text{if } y_j^{t*} \geq \chi \\ 0 & \text{otherwise} \end{cases} & (\text{A1}) \\ y_j^{t*} &\sim N([\mathbf{s}^t, \mathbf{X}_j^t] \cdot \boldsymbol{\varphi} + \varphi_j, \tau^2) \end{aligned}$$

where  $[\mathbf{s}^t, \mathbf{X}_j^t]$  is the vector of independent variables;  $\tau$  is the standard deviation of  $y_{jt}^*$ ;  $\varphi_j$  is a bidder specific constant term due to the fixed effect  $f_j$  in valuations (Equation 16);  $\chi$  is the truncation point, which is set at 15 to be consistent with the 15¢ minimum bid requirement of the search engine.

One possible concern when estimating the partial policy function  $\tilde{\boldsymbol{\sigma}}(\mathbf{s}, \mathbf{X})$  (and the full policy function  $\boldsymbol{\sigma}(\mathbf{s}, \mathbf{X}, r_j^t)$  next) is that there may be multiple equilibrium strategies; and the observed data are generated by multiple equilibria. If this were the case, the policy function would not lead to a unique decision and would be of limited use in predicting advertiser behavior. It is therefore necessary to invoke the following assumption (BBL).

**Assumption 2 (Equilibrium Selection):** The data are generated by a single Markov perfect equilibrium profile  $\boldsymbol{\sigma}$ .

Assumption 2 is relatively unrestrictive since our data is generated by auctions of *one* keyword and from *one* search engine. Given data are from a single market, the likelihood is diminished that different equilibria from different markets are confounded. We note that this assumption is often employed in such contexts (e.g., Dubé et al. (2008)).

---

<sup>33</sup>We experimented with alternative specifications including a Beta distribution and a Weibull distribution whose scale, shape, and location parameters are functions of  $(\mathbf{s}, \mathbf{X})$ . The current specification gives the best fit in terms of marginal likelihoods.

This partial policy function is then used to impute the full policy function  $b_j = \sigma_j(\mathbf{s}, \mathbf{X}, r_j^t)$  as detailed below based on  $r_j^t$ 's distribution parameter  $\psi$ .

**Full Policy Functions**  $\sigma_j^t(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$ . To evaluate the value function of this dynamic game, we need to calculate bids as a function of not only  $(\mathbf{s}^t, \mathbf{X}^t)$  but also the unobserved shocks  $r_j^t$  (see section 4.2.3). To infer this full policy function  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$  from the estimated partial policy function,  $\tilde{\sigma}_j(\mathbf{s}^t, \mathbf{X}^t)$ , we introduce one additional assumption.

**Assumption 3 (Monotone Choice):** For each bidder  $j$ , its equilibrium strategy  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$  is increasing in  $r_j^t$  (BBL).

Assumption 3 implies that bidders who draw higher private valuation shocks  $r_j^t$  will bid more aggressively.

To explore these two assumptions, note that the partial policy function  $\tilde{\sigma}(\mathbf{s}^t, \mathbf{X}^t)$  presents distributions for bid  $b_j^t$  and the latent  $y_j^{t*}$ , whose CDF's we denote as  $F_b(b_j^t | \mathbf{s}^t, \mathbf{X}^t)$  and  $F(y_j^{t*} | \mathbf{s}^t, \mathbf{X}^t)$ , respectively.<sup>34</sup> According to the model in equation A1, the population mean of  $y_j^{t*}$  across bidders and periods is  $[\mathbf{s}^t, \mathbf{X}_j^t] \cdot \varphi + \varphi_j$ . Around this mean, the variation across *bidders and periods* can be captured by the variance term  $\tau^2$ . With assumption 3, we can attribute  $\tau^2$  to the random shocks  $r_j^t$ .

Given the normal distribution assumption of the random shock  $r_j^t \sim N(0, \psi^2)$ , we may impute the  $y_j^{t*}$  (and hence  $b_j^t$ ) for each combination of  $(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$ , i.e., the full policy function. To see this, note that since  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$  is increasing in  $r_j^t$ ,<sup>35</sup>

$$F(y_j^{t*} | \mathbf{s}^t, \mathbf{X}^t) = \Pr(\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t) \leq y_j^{t*} | \mathbf{s}^t, \mathbf{X}^t) = \Phi(\sigma_j^{-1}(y_j^{t*}, \mathbf{s}^t, \mathbf{X}^t) / \psi)$$

where  $\sigma_j^{-1}(y_j^{t*}, \mathbf{s}^t, \mathbf{X}^t)$  is the inverse function of  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$  with respect to  $r_j^t$  and  $\Phi(\cdot)$  is the CDF of standard normal distribution. In equilibrium, we have  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t) = y_j^{t*}$ . By substitution

<sup>34</sup>To be more specific, we estimate a continuous distribution  $F(y_j^{t*} | \mathbf{s}^t, \mathbf{X}^t)$  for  $y_j^{t*}$  from equation A1; then conditional on the truncation point  $\chi$ , we can back out the (discontinuous) distribution  $F_b(b_j^t | \mathbf{s}^t, \mathbf{X}^t)$  for  $b_j^t$ .

<sup>35</sup>In this Appendix, we are abusing the notation of  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t)$ . For the purpose of a clear exposition, we define  $\sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t) = b_j^t$  in the paper. To match the bidding function estimated in equation A1, the more accurate definition should be

$$\begin{aligned} b_j^t &= \begin{cases} y_j^{t*} & \text{if } y_j^{t*} \geq \chi \\ 0 & \text{otherwise} \end{cases} \\ y_j^{t*} &= \sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t). \end{aligned}$$

and rearrangement we get

$$\begin{aligned}
y_j^{t*} &= \sigma_j(\mathbf{s}^t, \mathbf{X}^t, r_j^t) \\
&= F^{-1}\left(\Phi\left(\sigma_j^{-1}(y_j^{t*}, \mathbf{s}^t, \mathbf{X}^t)/\psi\right) \mid \mathbf{s}^t, \mathbf{X}^t\right) \\
&= F^{-1}\left(\Phi\left(r_j^t/\psi\right) \mid \mathbf{s}^t, \mathbf{X}^t\right)
\end{aligned} \tag{A2}$$

where  $\sigma_j^{-1}(y_j^{t*}, \mathbf{s}^t, \mathbf{X}^t) = r_j^t$ ;  $r_j^t/\psi$  has a standard normal distribution.

Therefore there is a unique mapping between the likelihood of observing a given valuation shock  $r_j^t$  and the  $y_j^{t*}$ . Each  $r_j^t$  drawn by a firm implies a corresponding quantile on the  $r_j^t$ 's distribution; this quantile in turn implies a  $y_j^{t*}$  from the distribution represented by that firm's partial bidding function  $\tilde{\sigma}_j(s^t, X^t)$ . However, because we do not know  $\psi$  and, thus, the distribution of  $r_j^t$ , we have to make draws from an alternative distribution  $r_j^t/\psi$  that has a one-one quantile mapping to  $r_j^t$ . To do this, we first draw a random shock  $r_j^t/\psi$  from  $N(0, 1)$  for each advertiser  $i$  in period  $t$ . Next, we determine  $F(y_j^{t*} \mid \mathbf{s}^t, \mathbf{X}^t)$  using results estimated in Equation A1 and looking at the distribution of its residuals to determine  $F$ . That is, for each value of  $y_j^{t*}$ , we should be able to compute its probability for a given  $\mathbf{s}^t$  and  $\mathbf{X}^t$  using  $F$ . Accordingly,  $F^{-1}$  links probabilities to  $y_j^{t*}$  (therefore  $b_j^t$ ) for a given  $\mathbf{s}^t$  and  $\mathbf{X}^t$ . We then use  $F^{-1}$  to link the probability  $\Phi(r_j^t/\psi)$  to  $b_j^t$  for a particular  $\mathbf{s}^t$  and  $\mathbf{X}^t$ . In this manner we ensure the bids and valuations in Equation A11 comport. In Appendix A.2.1, when evaluating the value function for a set of given parameter values of  $\psi$  in Equation A11 or evaluating base functions defined in Equation A12, we integrate out over the unobserved shocks  $r_j^t$  by drawing many  $r_j^t/\psi$  from  $N(0, 1)$ .

### A.1.2 Consumer Model Estimation

We derive the consumer model conditioned on the information state of the advertiser as described in section 4.1. Given that advertisers do not observe what each person downloaded or the characteristics of these persons, they must infer consumer behavior from aggregate instead of individual level data.

Advertisers do observe the aggregate data in the form of download counts  $d_j^t = \{d_1^t, d_2^t, \dots, d_N^t\}$  in period  $t$ . A single  $d_j^t$  follow a binomial distribution. Given the download probabilities  $P_j^t$  in Equation 12, a single  $d_j^t$ 's probability mass function is  $\binom{M_t}{d_j^t} [P_j^t]^{d_j^t} [1 - P_j^t]^{M_t - d_j^t}$ , where  $M_t$  is the consumer population size in period  $t$ . Hence the likelihood of observing  $d^t$  is

$$L(\mathbf{d}^t \mid \Omega_c) = \prod_j \binom{M_t}{d_j^t} [P_j^t]^{d_j^t} [1 - P_j^t]^{M_t - d_j^t}$$

where  $\Omega_c \equiv \{\alpha^g, \beta^g, \delta^g, \gamma^g, \lambda_0^g, \lambda_1^g\}_g$  are parameters to be estimated.

Naturally, the full posterior distribution of the model will be the product of  $L(d^t|\Omega_c)$  across periods and  $p(\Omega_c)$ , the prior distributions of parameters, i.e.,

$$p(\Omega_c|data) \propto \prod_t L(\mathbf{d}^t|\Omega_c) \cdot p(\Omega_c). \quad (\text{A3})$$

An advertiser's predicted downloads  $d_j^t(k, X_j^t; \Omega_c)$  can readily be constructed using the parameter estimates as shown in equation 14

$$d_j^t(k, X_j^t; \widehat{\Omega}_c) = M_t \widehat{F}_j^t. \quad (\text{A4})$$

This prediction is then used to forecast expectations of future downloads and slot positions in the firm's value function in the second step estimation.

### A.1.3 State Transition Function $P(\mathbf{s}'|b_j, \mathbf{b}_{-j}, \mathbf{s}, X)$

To compute the state transition, note that the marginal number of expected downloads is given by the expected downloads given a slot position multiplied by the probability of appearing in that slot position and then summed across all positions:

$$P(\mathbf{s}'|b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}) = \sum_k d(k, X; \Omega_c) \Pr(k|b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}). \quad (\text{A5})$$

The expected downloads given a slot position in A5 is defined in 14. We can decompose the likelihood of appearing in slot  $k$  as follows

$$\begin{aligned} & \Pr(k|b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}) \\ = & \Pr_{\{k \leq K\}}(k|b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}) I\{k \leq K\} + \Pr_{\{k > K\}}(k|b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}) I\{k > K\} \end{aligned} \quad (\text{A6})$$

where  $\Pr_{\{k \leq K\}}(k|b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X})$  is the probability of appearing in slot  $k$  of the sponsored search section (i.e.,  $k \leq K$ ), and  $\Pr_{\{k > K\}}(k|b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X})$  is the likelihood of appearing in slot  $k$  of the organic search section (i.e.,  $k > K$ ). We discuss these two probabilities next.

**Likelihood of Premium Slot  $k \leq K$ .** Let us first consider the likelihood of winning one of the premium slots  $k$  ( $k \leq K$ ),  $\Pr_{\{k \leq K\}}(k|b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X})$  as an order statistic reflecting the relative quality of the advertiser's bid, which is defined as  $b_j d_j^{(-1)}$ . Higher quality bids are more likely to be assigned to better slots. Denote  $\Psi_{bd}(b_{j'}, d_{j'}^{(-1)} | \mathbf{s}, \mathbf{X})$  as the distribution CDF of  $b_{j'} d_{j'}^{(-1)}, \forall j'$ , where  $d_{j'}^{(-1)}$  is from the state vector and  $b_{j'}$  has a distribution depending on the strategy profile  $\sigma(\cdot)$ .<sup>36</sup> For bidder  $j$  to win a premium slot  $k$  by bidding  $b_j$ , it implies that (1) among all of the other  $N - 1$

<sup>36</sup>It is difficult to write a closed form solution for  $\Psi_{bd}$ , but we may use the sample population distribution to approximate  $\Psi_{bd}$ .

competing bidders, there are  $k - 1$  bidders who have a higher ranking than  $j$  in terms of  $b_{j'}d_{j'}^{(-1)}$  and (2) the other ones have a lower ranking than  $j$ . The probability of having a higher ranking than  $j$  is  $[1 - \Psi_{bd}(b_j d_j^{(-1)} | s, \mathbf{X})]$ . Thus the probability of bidder  $j$  winning slot  $k$  by bidding  $b_j$  is simply an order statistics as shown below; note that the combination  $\binom{N-1}{k-1}$  in the equation is because any  $(k - 1)$  out of the  $(N - 1)$  competing bidders can have a higher ranking than  $j$ .<sup>37</sup>

$$\begin{aligned}
& \Pr_{\{k \leq K\}}(k | b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}) & (A7) \\
&= \binom{N-1}{k-1} [1 - \Psi_{bd}(b_j d_j^{(-1)} | \mathbf{s}, \mathbf{X})]^{k-1} [\Psi_{bd}(b_j d_j^{(-1)} | \mathbf{s}, \mathbf{X})]^{(N-1)-(k-1)} \\
&= \binom{N-1}{k-1} [1 - \Psi_{bd}(b_j d_j^{(-1)} | \mathbf{s}, \mathbf{X})]^{k-1} [\Psi_{bd}(b_j d_j^{(-1)} | \mathbf{s}, \mathbf{X})]^{N-k}
\end{aligned}$$

**Likelihood of Organic Slot  $k > K$ .** Next we consider what happens when an advertiser does not win this auction and is placed in the organic search section. In this case, by the rules of the auction, the bidder's slot is determined by its update recency compared to all products in the organic search section. For bidder  $j$  to be placed in organic slot  $k > K$  it implies that (1) there are  $K$  bidders who have a higher ranking of  $b_{j'}d_{j'}^{(-1)}$  than bidder  $j$  (i.e.,  $j$  loses the auction) and (2) among the other  $\bar{N} - K - 1$  products (i.e., all products at the search engine less those who win premium slots and  $j$  itself), there are  $k - K - 1$  products that have a higher update recency than  $j$  and (3) the other ones have a lower ranking than  $j$ . Hence,

$$\begin{aligned}
& \Pr_{\{k > K\}}(k | b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}) \\
&= \Pr(k > K | b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}) \cdot \Pr(k | b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}, k > K) & (A8)
\end{aligned}$$

where the first term is the probability of losing the auction (condition 1) and the second term denotes the likelihood of appearing in position  $k > K$  (condition 2 and 3). Note that the main reason for the difference between A7 and A8 is the change of ranking mechanisms. The ranking is based on  $b_{j'}d_{j'}^{(-1)}$  for  $k \leq K$  and update recency when  $k > K$ . The first term in A8 does not appear as an order statistics (as shown below) since when  $k > K$  the order of  $b_{j'}d_{j'}^{(-1)}$  becomes meaningless. Instead, the update recency is affecting the ranking. The two terms in A8 can be expressed as follows.

---

<sup>37</sup>An alternative interpretation of equation A7 is the probability mass function (PMF) of a binomial distribution. Among  $N - 1$  competing bidders, there are  $k - 1$  higher than bidder  $j$  and  $(N - 1) - (k - 1)$  lower than  $j$ , and the probability of higher than  $j$  is  $[1 - \Psi_{bd}(b_j d_j^{(-1)} | \mathbf{s}, \mathbf{X})]$ . Hence, we may consider the expression in A7 as the PMF of a binomial distribution.

Losing the auction implies that among  $j$ 's  $N - 1$  opponents, there are  $K$  bidders who have a higher ranking than  $j$  in terms of  $b_j, d_j^{(-1)}$ . Hence,

$$\Pr(k > K | b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}) = \binom{N-1}{K} [1 - \Psi_{bd}(b_j, d_j^{(-1)} | \mathbf{s}, \mathbf{X})]^K. \quad (\text{A9})$$

The conditional probability of being placed in an organic slot  $k > K$  (condition 2 and 3) is, again, an order statistics.<sup>38</sup> This distribution is incumbent upon the update recency of all  $\bar{N}$  products exclusive of the  $K$  winners in the sponsored search section. Denoting the distribution of update recency of all products as  $\Psi_{up}$ , which can be approximated from the sample population distribution observed in the data, we obtain the following:

$$\begin{aligned} & \Pr(k | b_j, \mathbf{b}_{-j}, \mathbf{s}, \mathbf{X}, k > K) & (\text{A10}) \\ = & \binom{\bar{N} - K - 1}{k - K - 1} [1 - \Psi_{up}]^{k-K-1} [\Psi_{up}]^{(\bar{N}-K-1)-(k-K-1)} \\ = & \binom{\bar{N} - K - 1}{k - K - 1} [1 - \Psi_{up}]^{k-K-1} [\Psi_{up}]^{\bar{N}-k} \end{aligned}$$

Combining Equations A10 and A9 into A8, and then A8 and A7 into A6, yields the state transition equation.

Given that we have detailed the estimation of the first step functions  $(\sigma_j(\mathbf{s}, \mathbf{X}, r_j^t), d_j^t(k, X_j^t; \Omega_c), P(\mathbf{s}' | \mathbf{b}, \mathbf{s}, \mathbf{X}))$ , we now turn to the second step estimator, which is incumbent upon these first step functions.

## A.2 Second Step Estimation of Bidder Model

In this Appendix we detail how to estimate the parameters in the value function. This is done in two phases: first, we simulate the value function conditioned on  $\Omega_a$ , and second, we construct the likelihood using the simulated value function conditioned on  $\Omega_a$ .

### A.2.1 Phase 1: Simulation of Value Functions Given $\Omega_a$

To construct the value function we first simplify its computation by linearization, and second using this simplification, we simulate the expected value function conditioned on  $\Omega_a$  by integrating out over draws for  $\mathbf{s}^t, \mathbf{X}^t$ , and  $\tilde{r}_j^t$ .

---

<sup>38</sup>This order statistics can again be interpreted as the PMF of a binomial distribution similar to A7.

**Linearize the Value Function** We simplify the estimation procedure by relying on the fact that Equation 17 is linear in the parameters  $\Omega_a$ . We can rewrite Equation 17 by factoring out  $\Omega_a$ .

$$\begin{aligned}
& \mathbf{E}\pi_j(\mathbf{b}^t, \mathbf{s}^t, \mathbf{X}^t, r_j^t; \Omega_a) & (A11) \\
&= \sum_{k=1}^K \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot (v(X_j^t; \theta) + f_j + r_j^t - b_j^t) \cdot d_j^t(k, X_j^t; \Omega_c) \\
&\quad + \sum_{k=K+1}^{\bar{N}} \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot (v(X_j^t; \theta) + f_j + r_j^t) \cdot d_j^t(k, X_j^t; \Omega_c) \\
&= \left[ \sum_{k=1}^{\bar{N}} \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot d_j^t(k, X_j^t; \Omega_c) \cdot X_j^t \right] \cdot \theta \\
&\quad + \left[ \sum_{k=1}^{\bar{N}} \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot d_j^t(k, X_j^t; \Omega_c) \right] \cdot f_j \\
&\quad + \left[ \sum_{k=1}^{\bar{N}} \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot d_j^t(k, X_j^t; \Omega_c) \cdot \tilde{r}_j^t \right] \cdot \psi \\
&\quad - b_j^t \sum_{k=1}^K \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) d_j^t(k, X_j^t; \Omega_c) \\
&= Base_{j1}^t [\theta', f_j]' + Base_{j2}^t \psi - Base_{j3}^t
\end{aligned}$$

where

$$\begin{aligned}
Base_{j1}^t &\equiv \left[ \begin{array}{c} \sum_{k=1}^{\bar{N}} \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot d_j^t(k, X_j^t; \Omega_c) \cdot X_j^t \\ \sum_{k=1}^{\bar{N}} \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot d_j^t(k, X_j^t; \Omega_c) \end{array} \right]' & (A12) \\
Base_{j2}^t &\equiv \left[ \sum_{k=1}^{\bar{N}} \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) \cdot d_j^t(k, X_j^t; \Omega_c) \cdot \tilde{r}_j^t \right] \\
Base_{j3}^t &\equiv b_j^t \sum_{k=1}^K \Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t) d_j^t(k, X_j^t; \Omega_c) \\
\tilde{r}_j^t &= r_j^t / \psi \sim N(0, 1).
\end{aligned}$$

Note that the values of  $\{Base_{j1}^t, Base_{j2}^t, Base_{j3}^t\}_{\forall t}$  are conditionally independent of  $\theta, f_j$  and  $\psi$ . This enables us to first evaluate  $\{Base_{j1}^t, Base_{j2}^t, Base_{j3}^t\}_{\forall t}$  and keep them constant when drawing  $\theta, f_j$  and  $\psi$  from their posterior distributions. By doing so, we reduce the computational burden of estimation as described next.

**Simulate the Value Functions Given  $\Omega_a$ .** After the linearization, given a set of advertiser parameters  $\Omega_a = \{\theta, f_{j=1,2,\dots,N}, \psi\}$  and Equation A11, the value function depicted in Equation 19 can also be written as the following with period index  $t$  invoked:

$$\begin{aligned}
V_j(\mathbf{s}^0, \mathbf{X}^0; \boldsymbol{\sigma}; \Omega_a) &= \mathbf{E}_{\mathbf{s}, \mathbf{X}, r} \left[ \sum_{t=0}^{\infty} \rho^t \pi_j(\boldsymbol{\sigma}, \mathbf{s}^t, \mathbf{X}^t, r_j^t; \Omega_a) \right] \\
&= \mathbf{E} \left[ \sum_{t=0}^{\infty} (\rho^t Base_{j1}^t \begin{bmatrix} \theta \\ f_j \end{bmatrix} + Base_{j2}^t \psi - Base_{j3}^t) \right] \\
&= [\mathbf{E} \sum_{t=0}^{\infty} \rho^t Base_{j1}^t] \begin{bmatrix} \theta \\ f_j \end{bmatrix} + [\mathbf{E} \sum_{t=0}^{\infty} \rho^t Base_{j2}^t] \psi - [\mathbf{E} \sum_{t=0}^{\infty} \rho^t Base_{j3}^t]
\end{aligned} \tag{A13}$$

where the expectation is taken over current and future private shocks, future states  $\mathbf{s}^t$ , future  $\mathbf{X}^t$  and  $R^t$ .

An estimated value function  $\widehat{V}_j(\mathbf{s}^0, \mathbf{X}^0; \boldsymbol{\sigma}; \Omega_a)$  can then be obtained by the following steps:

1. Draw private shocks  $\tilde{r}_j^t$  from  $N(0, 1)$  for all bidders  $j$  in period 0; draw initial choice of  $\mathbf{s}^0$  from the distribution of state variables derived from the observed data; draw  $\mathbf{X}^0$  from the observed distribution of product attributes.
2. Starting with the initial state  $\mathbf{s}^0, \mathbf{X}^0$  and the  $\tilde{r}_j^t$  step 1, calculate  $\widehat{b}_j^0$  for all bidders using the inversion (equation A2) described in Appendix A.1.1.
3. Use  $\mathbf{s}^0, \mathbf{X}^0$  and  $\widehat{\mathbf{b}}^0$  to determine the slot ranking, whose distribution is  $\Pr(k|b_j^t, \mathbf{b}_{-j}^t, \mathbf{s}^t, \mathbf{X}^t)$  in Equation A6 in Appendix A.1.3; using  $d(k, X_j^0; \Omega_c)$  in Equation 14, obtain a new state vector  $\mathbf{s}^1$ , whose distribution is  $P(\mathbf{s}^1|\widehat{\mathbf{b}}^0, \mathbf{s}^0, \mathbf{X}^0)$  in Equation A5 in Appendix A.1.3; draw  $\mathbf{X}^1$  from the observed distribution of product attributes.
4. Repeat step 1-3 for  $T$  periods for all bidders to compute all  $\mathbf{s}^t, \mathbf{X}^t, \tilde{r}_j^t$ , and  $\mathbf{b}^t$  for all periods;  $T$  is large enough so that the discount factor  $\rho^T$  approaches 0.
5. Using  $\mathbf{s}^t, \mathbf{X}^t, \tilde{r}_j^t, d_j^t(k, X_j^t; \Omega_c)$ , and  $\mathbf{b}^t$ , evaluate  $\{Base_{j1}^t, Base_{j2}^t, Base_{j3}^t\}_{t=0, \dots, T}$  and  $\left\{ \left[ \sum_{t=0}^T \rho^t Base_{j1}^t \right], \left[ \sum_{t=0}^T \rho^t Base_{j2}^t \right], \left[ \sum_{t=0}^T \rho^t Base_{j3}^t \right] \right\}$ .
6. The resulting values of

$$\{Base_{j1}^t, Base_{j2}^t, Base_{j3}^t\}_{t=0, \dots, T}$$

and

$$\left\{ \left[ \sum_{t=0}^T \rho^t Base_{j1}^t \right], \left[ \sum_{t=0}^T \rho^t Base_{j2}^t \right], \left[ \sum_{t=0}^T \rho^t Base_{j3}^t \right] \right\}$$

depend on the random draws of  $\mathbf{s}^t, \mathbf{X}^t, r^t$ . To compute

$$\left\{ \left[ \mathbf{E} \sum_{t=0}^{\infty} \rho^t Base_{j1}^t \right], \left[ \mathbf{E} \sum_{t=0}^{\infty} \rho^t Base_{j2}^t \right], \left[ \mathbf{E} \sum_{t=0}^{\infty} \rho^t Base_{j3}^t \right] \right\},$$

repeat step 1-6 for  $NR$  times so as to integrate out over the draws. Note that when  $T$  is large enough,  $[E \sum_{t=0}^T \rho^t Base_{j.}^t]$  is a good approximation of  $[E \sum_{t=0}^{\infty} \rho^t Base_{j.}^t]$  since  $\rho^T$  approaches 0.

7. Conditional on a set of parameters  $\Omega_a$  and

$$\left\{ \left[ \mathbf{E} \sum_{t=0}^{\infty} \rho^t Base_{j1}^t \right], \left[ \mathbf{E} \sum_{t=0}^{\infty} \rho^t Base_{j2}^t \right], \left[ \mathbf{E} \sum_{t=0}^{\infty} \rho^t Base_{j3}^t \right] \right\},$$

we may evaluate  $\widehat{V}_j(\mathbf{s}^0, \mathbf{X}^0; \boldsymbol{\sigma}; \Omega_a)$  from Equation A13.

An estimated deviation value function  $\widehat{V}_j(\mathbf{s}^0, \mathbf{X}^0; \boldsymbol{\sigma}'_j, \boldsymbol{\sigma}_{-j}; \Omega_a)$  with an alternative strategy  $\boldsymbol{\sigma}'_j$  other than  $\boldsymbol{\sigma}_j$  can be constructed by following the same procedure. We draw a deviated strategy  $\boldsymbol{\sigma}'_j$  by adding disturbance to the estimated policy function from Step 1. In particular, we add a normally distributed random variable (*mean* = 0; *s.d.* = 0.3) to each parameter.

We implement this process by first drawing  $NS = 10$  initial states for each bidder and  $\{X^t\}_{t=0,1,\dots,T}$  of all  $T = 200$  periods. Then for each combination of bidder and initial state, we use this process to compute the base value functions and  $ND = 100$  perturbed base functions. In Step 6, we use  $NR = 100$ . The discount factor  $\rho$  is fixed as 0.95.

The computational burden is reduced tremendously since we have linearized the value functions and factored out the parameters  $\Omega_a$ . We do not need to re-evaluate the value functions for each set of parameters  $\Omega_a$ . Instead, we only evaluate the base functions in Equation A12 once using step 1-6 and keep them fixed. Then for each draw of  $\Omega_a$  from the posterior distribution we may evaluate the value functions (step 7) so as to recover  $\Omega_a$  as described below.

### A.2.2 Phase 2: Recover $\Omega_a$

Recall our goal is to assess the likelihood that 20 holds. Define  $P_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \boldsymbol{\sigma}, \boldsymbol{\sigma}'_j; \Omega_a)$  as the probability of the event

$$\left\{ \widehat{V}_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \boldsymbol{\sigma}_j, \boldsymbol{\sigma}_{-j}; \Omega_a) \geq \widehat{V}_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \boldsymbol{\sigma}'_j, \boldsymbol{\sigma}_{-j}; \Omega_a) \right\}, \quad (\text{A14})$$

where  $s_{(ns)}^0$  stands for the  $ns$ -th initial state of bidder  $j$ . This event means that the estimated value function for the given initial state  $s_{(ns)}^0$  with observed strategy  $\boldsymbol{\sigma}_j$  is greater than the estimated value function with a deviation  $\boldsymbol{\sigma}'_j$ . For observed data to be rational, we should have

$P_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \boldsymbol{\sigma}, \sigma'_j; \Omega_a)$  converging to 1 under the true parameters, in the sense that all  $ND$  draws should result in the event of equation A14.

Note that  $P_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \boldsymbol{\sigma}, \sigma'_j; \Omega_a)$  is not observed, but it can be approximated with the sample analog from the simulated  $ND$  draws of  $\widehat{V}_j(s_{(ns)}^0, X^0; \sigma'_j, \sigma_{-j}; \Omega_a)$  as the follows:

$$\begin{aligned} & \widehat{P}_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \boldsymbol{\sigma}, \sigma'_j; \Omega_a) \\ &= \frac{1}{ND} \sum_{nd=1}^{ND} I \left\{ \widehat{V}_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \sigma_j, \boldsymbol{\sigma}_{-j}; \Omega_a) \geq \widehat{V}_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \sigma'_j, \boldsymbol{\sigma}_{-j}; \Omega_a)_{(nd)} \right\} \end{aligned} \quad (\text{A15})$$

where the subscript  $(nd)$  indices the  $nd$ -th simulated  $\widehat{V}_j(s_{(ns)}^0, X^0; \sigma'_j, \sigma_{-j}; \Omega_a)$ .

By pooling together all  $\widehat{P}_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \boldsymbol{\sigma}, \sigma'_j; \Omega_a)$ 's across bidders and  $(ns)$ , we are able to construct a simulated likelihood function as the following

$$L = \prod_{j,(ns)} \widehat{P}_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \boldsymbol{\sigma}, \sigma'_j; \Omega_a). \quad (\text{A16})$$

Denoting the prior of  $\Omega_a$  as  $p(\Omega_a)$ , the posterior can be written as

$$p(\Omega_a | data) \propto \prod_{j,(ns)} \widehat{P}_j(\mathbf{s}_{(ns)}^0, \mathbf{X}^0; \boldsymbol{\sigma}, \sigma'_j; \Omega_a) p(\Omega_a) \quad (\text{A17})$$

There may be some efficiency loss by using Bayesian method in estimation since the differences between  $V_j(s_{(ns)}^0, X^0; \sigma_j, \sigma_{-j}; \Omega_a)$  and  $\widehat{V}_j(s_{(ns)}^0, X^0; \sigma'_j, \sigma_{-j}; \Omega_a)$  also provide some information about the estimates and are not utilized (BBL). However, Bayesian method provides desirable small sample properties, increases statistical efficiency, and enables the estimation of a wide array of functional forms (Rossi et al. (2005)). We consider the benefits of Bayesian method to outweigh the loss of efficiency.

## B Sampling Chain

### B.1 Advertiser Model

#### B.1.1 Priors

To facilitate explication, denote the vector  $[\mathbf{s}^{t'}, \mathbf{X}_j^{t'}] \equiv Z_j^t$ , the matrix  $[Z_j^t]_{\forall j,t} \equiv Z$ , and the vector  $[y_j^{t*}]_{\forall j,t} \equiv y^*$ . We also denote the number of bidders as  $N$  and the number of observations for bidder  $j$  as  $N_j$ . So the total number of observations is  $\sum_j N_j$ ; the dimension of  $Z_j^t$  is 1 by  $d$  (the dimension of  $[\mathbf{s}^{t'}, X_j^{t'}]$ ); the dimension of  $Z$  is  $\sum_j N_j$  by  $d$ ; and the dimension of  $y^*$  is  $\sum_j N_j$ . The vector of bidder specific fixed effects,  $\varphi^{fe}$ , is a column vector with a length of the number of bidders.  $\Lambda$  is a matrix whose dimension is  $\sum_j N_j$  by  $N$ . Suppose rows  $a$  to  $b$  of matrix  $Z$  are the observations of

bidder  $j$ , then rows  $a$  to  $b$  of the  $j$ -th column of  $\Lambda$  are 1. Other elements of  $\Lambda$  are 0. Also denote  $\Gamma \equiv [\Lambda, Z]$ , whose dimension is  $\sum_j N_j$  by  $(N + d)$ .

The advertiser model is specified as

$$\begin{aligned} b_j^t &= \begin{cases} y_j^{t*} & \text{if } y_j^{t*} \geq \chi \\ 0 & \text{otherwise} \end{cases} \\ y_j^{t*} &\sim N([\mathbf{s}^t, \mathbf{X}_j^t] \cdot \boldsymbol{\varphi} + \varphi_j, \tau^2). \end{aligned} \tag{A18}$$

We iterate the sampling chain for 20,000 and use the second half of the chain to make inference. The priors use a diffused variance of 100; examinations of the posteriors shows that the choice of the variance is the order of magnitude greater than posterior distributions, which assures a proper but diffused prior (Spiegelhalter et al. (1996), Gelman et al. (2004)).

| Priors  |   | Selected Value  |
|---|---|---|
| $\boldsymbol{\varphi} = [\varphi', \varphi_1, \dots, \varphi_N]'$ | $\boldsymbol{\varphi} \sim N(\boldsymbol{\varphi}_0, I_\varphi \sigma_\varphi^2)$ | $\boldsymbol{\varphi}_0$ : estimates of a classical Tobit model of bids on $\Gamma = [\Lambda, Z]$ with the truncation at 15<br>$\sigma_\varphi^2 = 100$<br>$I_\varphi$ is an identity matrix with the dimension of $N + d$ |
| $\tau$  | $\tau \sim TN_{(0,+\infty)}(\mu_\tau, \sigma_\tau^2)$                             | $\mu_\tau = 5, \sigma_\tau^2 = 100$   |

### B.1.2 Conditional Posteriors

- $y_j^{t*}$

$y_j^{t*}$  is determined by the following

$$\begin{aligned} y_j^{t*} &= b_j^t, \text{ if } b_j^t > 0 \\ y_j^{t*} &\sim TN_{(-\infty, 15]}(Z_j^t \boldsymbol{\varphi} + \varphi_j, \tau^2), \text{ if } b_j^t = 0 \end{aligned}$$

$y_j^{t*}$  is right truncated at 15 when  $b_j^t = 0$ ; this is consistent with the 15¢ minimum bid requirement of the search engine.

- $\varphi$

$$\begin{aligned}
\text{Prior } \varphi &\sim N(\varphi_0, I_\varphi \sigma_\varphi^2) & (A19) \\
\text{Likelihood } L &\propto \prod_{j,t} \exp\left(-\frac{[y_j^{t*} - (Z_j^t \varphi + \varphi_j)]^2}{2\tau^2}\right) \\
\text{Posterior } (\varphi|\cdot) &\sim N(\mu_\varphi, \Sigma_\varphi) \\
\Sigma_\varphi &= [\Gamma' \Gamma \tau^{-2} + \sigma_\varphi^{-2}]^{-1} \\
\mu_\varphi &= \Sigma_\varphi \cdot \{\Gamma' \cdot y^* \tau^{-2} + \varphi_0 \sigma_\varphi^{-2}\}
\end{aligned}$$

- $\tau$

$$\text{Prior } \tau \sim TN_{(0,+\infty)}(\mu_\tau, \sigma_\tau^2)$$

A random walk proposal density is used in the  $(r)$ -th iteration,  $\tau^{(r)} \sim TN_{(0,+\infty)}(\tau^{(r-1)}, \sigma_p^2)$ , where  $\tau^{(r-1)}$  is the value from the  $(r-1)$ -th iteration;  $\sigma_p^2$  is the tuning variance which is chosen so that the acceptance rate is between 15% – 50%.

The acceptance probability  $pr^* = \min(1, pr)$  and

$$pr = \frac{L(\tau^{(r)}|\cdot)p(\tau^{(r)}|\mu_\tau, \sigma_\tau^2)\eta(\tau^{(r-1)}|\tau^{(r)}, \sigma_p^2)}{L(\tau^{(r-1)}|\cdot)p(\tau^{(r-1)}|\mu_\tau, \sigma_\tau^2)\eta(\tau^{(r)}|\tau^{(r-1)}, \sigma_p^2)}$$

where  $p(\tau^{(\cdot)}|\mu_\tau, \sigma_\tau^2)$  is the density of  $\tau^{(\cdot)}$  evaluated using the prior.  $L(\tau^{(\cdot)}|\cdot)$  is the likelihood evaluated at  $\tau^{(\cdot)}$ . To be specific

$$L(\tau^{(\cdot)}|\cdot) \propto \prod_{j,t} \pi(y_j^{t*}; Z_j^t \varphi + \varphi_j, \tau^{(\cdot)})$$

where  $\pi(y_j^{t*}; Z_j^t \varphi + \varphi_j, \tau^{(\cdot)})$  is the normal density of  $y_j^{t*}$  evaluated with mean  $Z_j^t \varphi + \varphi_j$  and standard deviation  $\tau^{(\cdot)}$ .

## B.2 Consumer Model

### B.2.1 Priors

|  | Priors   | Selected Value   |
|--|--|--|
| $B^g \equiv \{\{\alpha_j^g\}'_{\forall j}, \{\beta_a^g\}'_{\forall a}\}$ | $B^g \sim N(0, I_B \cdot \sigma_B^2)$                | $I_B$ is an identity matrix of dimension product attributes and intercepts; $\sigma_B^2 = 100$ |
| $\delta^g$   | $\delta^g \sim TN_{(0,+\infty)}(1, \sigma_\delta^2)$ | $\sigma_\delta^2 = 100$  |
| $\lambda_0^g$  | $\lambda_0^g \sim N(0, \sigma_\lambda^2)$            | $\sigma_\lambda^2 = 100$   |
| $\lambda_1^g$  | $\lambda_1^g \sim TN_{(0,1)}(0.5, \sigma_\lambda^2)$ | $\sigma_\lambda^2 = 100$   |
| $\gamma_0^g$   | $\gamma_0^g \sim N(0, \sigma_\gamma^2)$              | $\sigma_\gamma^2 = 100$  |
| $\gamma^g$   | $\gamma^g \sim N(0, I_\gamma \cdot \sigma_\gamma^2)$ | $I_\gamma$ is an identity matrix of dimension of 3; $\sigma_\gamma^2 = 100$                    |

### B.2.2 Conditional Posteriors

- $B^g = \{\{\alpha_j^g\}'_{\forall j}, \{\beta_a^g\}'_{\forall a}\}$

$$\text{Prior } B^g \sim N(0, I_B \cdot \sigma_B^2)$$

A random walk proposal density is used in the  $(r)$ -th iteration,  $B^{g(r)} \sim N(B^{g(r-1)}, \sigma_{pB}^2)$ , where  $B^{g(r-1)}$  is the value from the  $(r-1)$ -th iteration;  $\sigma_{pB}^2$  is the tuning variance which is chosen so that the acceptance rate is between 15% – 50%.

The acceptance probability  $pr^* = \min(1, pr)$  and

$$pr = \frac{L(B^{g(r)}|\cdot)p(B^{g(r)}|0, I_B \cdot \sigma_B^2)}{L(B^{g(r-1)}|\cdot)p(B^{g(r-1)}|0, I_B \cdot \sigma_B^2)}$$

where  $p(B^{g(\cdot)}|0, I_B \cdot \sigma_B^2)$  is the density of  $B^{g(\cdot)}$  evaluated using the prior.  $L(B^{g(\cdot)}|\cdot)$  is the likelihood evaluated at  $B^{g(\cdot)}$ . To be specific

$$L(B^{g(\cdot)}|\cdot) \propto \prod_{j,t} \binom{M_t}{d_j^t} [P_j^t]^{d_j^t} [1 - P_j^t]^{M_t - d_j^t}$$

where  $P_j^t$  is defined in equation 12 and evaluated at  $B^{g(\cdot)}$ .  $d_j^t$  is the actual downloads observed in the data and  $M_t$  is the market size. We set this to the mean total monthly number of visitors to the search engine and all other Internet properties owned by its parent (hence  $M_t = M$ ).

- $\delta^g$

Prior  $\delta^g \sim TN_{(0,+\infty)}(1, \sigma_\delta^2)$

A random walk proposal density is used in the  $(r)$ -th iteration,  $\delta^{g(r)} \sim TN_{(0,+\infty)}(\delta^{g(r-1)}, \sigma_{p\delta}^2)$ .

The acceptance probability  $pr^* = \min(1, pr)$  and

$$pr = \frac{L(\delta^{g(r)}|\cdot)p(\delta^{g(r)}|1, \sigma_\delta^2)\eta(\delta^{g(r-1)}|\delta^{g(r)}, \sigma_{p\delta}^2)}{L(\delta^{g(r-1)}|\cdot)p(\delta^{g(r-1)}|1, \sigma_\delta^2)\eta(\delta^{g(r)}|\delta^{g(r-1)}, \sigma_{p\delta}^2)}$$

where  $p(\delta^{g(\cdot)}|1, \sigma_\delta^2)$  is the density of  $\delta^{g(\cdot)}$  evaluated using the prior.  $L(\delta^{g(\cdot)}|\cdot)$  is the likelihood evaluated at  $\delta^{g(\cdot)}$ .  $\eta(\delta^{g(r-1)}|\delta^{g(r)}, \sigma_{p\delta}^2)/\eta(\delta^{g(r)}|\delta^{g(r-1)}, \sigma_{p\delta}^2)$  is the correction ratio since the proposal density is asymmetric.

- $\lambda_0^g$

Prior  $\lambda_0^g \sim N(0, \sigma_\lambda^2)$

A random walk proposal density is used in the  $(r)$ -th iteration,  $\lambda_0^{g(r)} \sim N(\lambda_0^{g(r-1)}, \sigma_{p\lambda}^2)$ .

The acceptance probability  $pr^* = \min(1, pr)$  and

$$pr = \frac{L(\lambda_0^{g(r)}|\cdot)p(\lambda_0^{g(r)}|1, \sigma_\lambda^2)}{L(\lambda_0^{g(r-1)}|\cdot)p(\lambda_0^{g(r-1)}|1, \sigma_\lambda^2)}$$

where  $p(\lambda_0^{g(\cdot)}|1, \sigma_\lambda^2)$  is the density of  $\lambda_0^{g(\cdot)}$  evaluated using the prior.  $L(\lambda_0^{g(\cdot)}|\cdot)$  is the likelihood evaluated at  $\lambda_0^{g(\cdot)}$ .

- $\lambda_1^g$

Prior  $\lambda_1^g \sim TN_{(0,1)}(0.5, \sigma_\lambda^2)$

A random walk proposal density is used in the  $(r)$ -th iteration,  $\lambda_1^{g(r)} \sim TN_{(0,1)}(\lambda_1^{g(r-1)}, \sigma_{p\lambda}^2)$ .

The acceptance probability  $pr^* = \min(1, pr)$  and

$$pr = \frac{L(\lambda_1^{g(r)}|\cdot)p(\lambda_1^{g(r)}|0.5, \sigma_\lambda^2)\eta(\lambda_1^{g(r-1)}|\lambda_1^{g(r)}, \sigma_{p\lambda}^2)}{L(\lambda_1^{g(r-1)}|\cdot)p(\lambda_1^{g(r-1)}|0.5, \sigma_\lambda^2)\eta(\lambda_1^{g(r)}|\lambda_1^{g(r-1)}, \sigma_{p\lambda}^2)}$$

where  $p(\lambda_1^{g(\cdot)}|1, \sigma_\lambda^2)$  is the density of  $\lambda_1^{g(\cdot)}$  evaluated using the prior.<sup>39</sup>  $L(\lambda_1^{g(\cdot)}|\cdot)$  is the likelihood evaluated at  $\lambda_1^{g(\cdot)}$ .  $\eta(\lambda_1^{g(r-1)}|\lambda_1^{g(r)}, \sigma_{p\lambda}^2)/\eta(\lambda_1^{g(r)}|\lambda_1^{g(r-1)}, \sigma_{p\lambda}^2)$  is the correction ratio since the proposal density is asymmetric.

- $\gamma_0^g$

$$\text{Prior } \gamma_0^g \sim N(0, \sigma_\gamma^2)$$

A random walk proposal density is used in the  $(r)$ -th iteration,  $\gamma_0^{g(r)} \sim N(\gamma_0^{g(r-1)}, \sigma_{p\gamma}^2)$ .

The acceptance probability  $pr^* = \min(1, pr)$  and

$$pr = \frac{L(\gamma_0^{g(r)}|\cdot)p(\gamma_0^{g(r)}|0, \sigma_\gamma^2)}{L(\gamma_0^{g(r-1)}|\cdot)p(\gamma_0^{g(r-1)}|0, \sigma_\gamma^2)}$$

where  $p(\gamma_0^{g(\cdot)}|0, \sigma_\gamma^2)$  is the density of  $\gamma_0^{g(\cdot)}$  evaluated using the prior.  $L(\gamma_0^{g(\cdot)}|\cdot)$  is the likelihood evaluated at  $\gamma_0^{g(\cdot)}$ .

- $\gamma^g$

$$\text{Prior } \gamma^g \sim N(0, I_\gamma \cdot \sigma_\gamma^2)$$

A random walk proposal density is used in the  $(r)$ -th iteration,  $\gamma^{g(r)} \sim N(\gamma^{g(r-1)}, I_\gamma \cdot \sigma_{p\gamma}^2)$ .

The acceptance probability  $pr^* = \min(1, pr)$  and

$$pr = \frac{L(\gamma^{g(r)}|\cdot)p(\gamma^{g(r)}|0, I_\gamma \cdot \sigma_\gamma^2)}{L(\gamma^{g(r-1)}|\cdot)p(\gamma^{g(r-1)}|0, I_\gamma \cdot \sigma_\gamma^2)}$$

where  $p(\gamma^{g(\cdot)}|0, I_\gamma \cdot \sigma_\gamma^2)$  is the density of  $\gamma^{g(\cdot)}$  evaluated using the prior.  $L(\gamma^{g(\cdot)}|\cdot)$  is the likelihood evaluated at  $\gamma^{g(\cdot)}$ .

## B.3 Second Step Estimation

### B.3.1 Priors

To facilitate explication denote the vector  $[\theta', f_1, f_2, \dots, f_N]'$  as  $\Delta$ .

---

<sup>39</sup>Note that  $\lambda_1^g$  is truncated to  $(0, 1)$  and has a large variance (100) for its prior. Hence, the prior is essentially an uninformative uniform distribution.

|          | Priors  | Selected Value  |
|----------|---|---|
|          |   | $\Delta_0$ : a vector of zeros with the length of the number of product attributes plus $N$ , the number of bidders         |
| $\Delta$ | $\Delta \sim N(\Delta_0, I_\Delta \sigma_\Delta^2)$   | $\sigma_\Delta^2 = 100$<br>$I_\Delta$ is an identity matrix with the dimension of the number of product attributes plus $N$ |
| $\psi$   | $\psi \sim TN_{(0,+\infty)}(\mu_\psi, \sigma_\psi^2)$ | $\mu'_\psi = 1, \sigma_\psi^2 = 100$  |

### B.3.2 Conditional Posteriors

- $\Delta$

$$\text{Prior } \Delta \sim N(\Delta_0, I_\Delta \sigma_\Delta^2)$$

A random walk proposal density is used in the  $(r)$ -th iteration,  $\Delta^{(r)} \sim N(\Delta^{(r-1)}, I_\Delta \sigma_{p\Delta}^2)$ , where  $\Delta^{(r-1)}$  is the value from the  $(r-1)$ -th iteration;  $\sigma_{p\Delta}^2$  is a scalar and functions as the tuning variance.

The acceptance probability  $pr^* = \min(1, pr)$  and

$$pr = \frac{L(\Delta^{(r)}|\cdot)p(\Delta^{(r)}|\Delta_0, I_\Delta \sigma_\Delta^2)}{L(\Delta^{(r-1)}|\cdot)p(\Delta^{(r-1)}|\Delta_0, I_\Delta \sigma_\Delta^2)}$$

where  $p(\Delta^{(\cdot)}|\Delta_0, I_\Delta \sigma_\Delta^2)$  is the density of  $\Delta^{(\cdot)}$  evaluated using the prior.  $L(\Delta^{(\cdot)}|\cdot)$  is the likelihood evaluated at  $\Delta^{(\cdot)}$ . The likelihood is defined in equation A16.

- $\psi$

$$\psi \sim TN_{(0,+\infty)}(\mu_\psi, I_2 \sigma_\psi^2)$$

A random walk proposal density is used in the  $(r)$ -th iteration,  $\psi^{(r)} \sim TN_{(0,+\infty)}(\psi^{(r-1)}, I_2 \sigma_{p\psi}^2)$ , where  $\psi^{(r-1)}$  is the value from the  $(r-1)$ -th iteration;  $\sigma_{p\psi}^2$  is a scalar and functions as the tuning variance.

The acceptance probability  $pr^* = \min(1, pr)$  and

$$pr = \frac{L(\psi^{(r)}|\cdot)p(\psi^{(r)}|\mu_\psi, I_2 \sigma_\psi^2)\eta(\psi^{(r-1)}|\psi^{(r)}, \sigma_{p\psi}^2)}{L(\psi^{(r-1)}|\cdot)p(\psi^{(r-1)}|\mu_\psi, I_2 \sigma_\psi^2)\eta(\psi^{(r)}|\psi^{(r-1)}, \sigma_{p\psi}^2)}$$

where  $p(\psi^{(\cdot)}|\mu_\psi, I_2 \sigma_\psi^2)$  is the density of  $\psi^{(\cdot)}$  evaluated using the prior.  $L(\psi^{(\cdot)}|\cdot)$  is the likelihood evaluated at  $\psi^{(\cdot)}$ . The likelihood is defined in equation A16. Since the proposal density is asymmetric, the ratio  $\frac{\eta(\psi^{(r-1)}|\psi^{(r)}, \sigma_{p\psi}^2)}{\eta(\psi^{(r)}|\psi^{(r-1)}, \sigma_{p\psi}^2)}$  is used to adjust the acceptance probability.

## C Policy Simulations

It is reasonable to expect that advertisers will change their bidding strategy in response to changes in a search engine’s new policy. Thus, the advertiser bidding rules estimated in the first stage of our analysis are not likely to reflect advertiser behavior under the new policy. Hence, we need to solve the new optimal bidding strategy for advertisers conditional on the primitives estimated off the data ( $\Omega$ ). This requires explicitly solving the dynamic programming problem (DP) for advertisers. Because of the dimension of the state space and the interaction across advertisers, solving the DP imposes a tremendous computational burden. In this appendix, we first outline our general approach to solving this dynamic advertiser bidding problem and then detail the manipulations underpinning each specific policy simulation.

### C.1 Computational Considerations

We use Approximate DP approach (cf. Judd (1998), Powell (2007)) to solve the advertiser bidding problem. To be specific, we impose a parametric decision rule to reflect the optimal bidding policy for advertisers and then find the parameters of the bidding rule that maximize the advertisers profits. Specifically, we proceed as follows:

1. Specify a parametric form for the bidding function of advertisers. We use the same functional form as the one estimated in the advertiser model (Equation A1).
2. Suppose  $r_j^t, t = 1, \dots, T, j = 1, \dots, N$  is a sequence of random shocks drawn from the distribution  $N(0, \psi)$ . Also, randomly seed an initial state  $s^0$  and  $\mathbf{X}^0$ . For an initial guess of the bidding rule parameters  $\varphi^{new} = [\varphi^{new}, \varphi_1^{new}, \dots, \varphi_N^{new}]'$ , we can approximate the advertiser value functions (as the initial guess, we choose the estimates of  $\varphi$  from the first stage advertiser model)

$$\begin{aligned} & \widehat{V}_j(s^0, \mathbf{X}^0, r_j^t; \sigma^{new}(\varphi^{new}); \Omega_a) \\ = & \pi_j(\sigma^{new}(\varphi^{new}), s^0, \mathbf{X}^0, r_j^0; \Omega_a) \\ & + \sum_{t=1}^T \rho^t \pi_j(\sigma^{new}(\varphi^{new}), s^t, \mathbf{X}^t, r_j^t; \Omega_a) P(s^t | \widehat{\mathbf{b}}^{t-1}, s^{t-1}, \mathbf{X}^{t-1}). \end{aligned}$$

3. Repeat step 2 for  $NR = 100$  draws of  $r_j^t$  sequence, and take the average; then we have a good approximation of value functions for advertisers

$$\widehat{V}_j(s^0, \mathbf{X}^0; \sigma^{new}; \Omega_a) = \frac{1}{NR} \sum_{nr=1}^{NR} \widehat{V}_j(s^0, \mathbf{X}^0, r_j^{t(nr)}; \sigma^{new}; \Omega_a).$$

4. Optimize  $\varphi^{new}$  to get the new bidding policy. The optimization rule we use is choosing  $\varphi^{new}$  to maximize the mean of the value functions across all advertisers. As a robustness check, we also consider maximizing the value function of (i) the most frequent bidder in the data, (ii) the least frequent bidder in the data, and (iii) a randomly picked bidder. For each policy simulation, we check the difference across these optimization rules and the results are all similar.
5. To ascertain that the estimated  $\varphi^{new}$  are not sensitive to the choice of  $s^0$  and  $X^0$ , as another robustness check we consider 3 alternative draws for  $s^0$  and  $X^0$ . Changes to the results are minimal.

## C.2 Policy Simulation: Segmentation and Targeting

Neither the search engine nor advertisers actually observes the segment memberships of consumers to help with targeting. However, it is possible for the advertiser to infer the posterior probability of consumer  $i$ 's segment membership conditional on its choices. These estimates can then be used to improve the accuracy and effectiveness of targeting.

More specifically, suppose the search engine observes consumer  $i$  in several periods. Let us consider consumer  $i$ 's binary choices over downloading, sorting/filtering, and searching in those periods. Denote these observations as  $H_i(\{y_{ijt}\}_{j,t}, \{\kappa_{it}\}_t, \{search_{it}\}_t)$ . The likelihood of observing  $H_i(\{y_{ijt}\}_{j,t}, \{\kappa_{it}\}_t, \{search_{it}\}_t)$  is

$$\begin{aligned} & L(H_i(\{y_{ijt}\}_{j,t}, \{\kappa_{it}\}_t, \{search_{it}\}_t)) \\ &= \sum_g \prod_t L(H_i(\{y_{ijt}\}_j, \{\kappa_{it}\}_t, \{search_{it}\}_t) | g_{it}) \cdot pg_{it}^g \end{aligned} \quad (\text{A20})$$

where

$$\begin{aligned} & L(H_i(\{y_{ijt}\}_{j,t}, \{\kappa_{it}\}_t, \{search_{it}\}_t) | g_{it}) \\ &= \prod_j \int_{u_{ijt}^{g\kappa}} \int_{\mathbf{z}_{it}^g} \pi(y_{ijt} | u_{ijt}^{g\kappa}, \kappa_{it}^g, g_{it}) \pi(\kappa_{it}^g | \mathbf{z}_{it}^g, g_{it}) du_{ijt}^{g\kappa} d\mathbf{z}_{it}^g \Pr(search_{it}^g) \end{aligned} \quad (\text{A21})$$

Hence, the posterior probability of segment membership for consumer  $i$  can be updated in a Bayesian fashion,

$$\begin{aligned} \Pr(i \in g | H_i(\{y_{ijt}\}_{j,t}, \{\kappa_{it}\}_t, \{search_{it}\}_t)) \\ &= \frac{\prod_t L(H_i(\{y_{ijt}\}_j, \{\kappa_{it}\}_t, \{search_{it}\}_t) | g_{it}) \cdot pg_{it}^g}{\sum_{g'} \prod_t L(H_i(\{y_{ijt}\}_j, \{\kappa_{it}\}_t, \{search_{it}\}_t) | g'_{it}) \cdot pg'_{it}} \end{aligned} \quad (\text{A22})$$

As a consequence, the engine will have a more accurate evaluation about the segment membership of that consumer.

On the other hand, suppose some consumers only visit the engine once. Before they make the product choices, the search engine cannot obtain a posterior distribution outlined in Equation A22 since their choices of products are still unavailable. Still, it is possible to establish a more informative prediction about their memberships based on their  $\kappa_{it}$ 's before their product choices. Similar to Equation A22, the posterior in this case is

$$\Pr(i \in g | H_i(\kappa_{it})) = \frac{L(H_i(\kappa_{it}) | g_{it}) \cdot pg_{it}^g}{\sum_{g'} L(H_i(\kappa_{it}) | g'_{it}) \cdot pg_{it}^{g'}} \quad (\text{A23})$$

where

$$L(H_i(\kappa_{it}) | g_{it}) = \int_{z_{it}^{g\kappa}} \pi(\kappa_{it}^g | z_{it}^g, g_{it}) dz_{it}^{g\kappa}$$

We can construct an analysis to consider the benefits of targeting as follows. First, we compute the return to advertisers when advertisers can only bid on keywords for all segments. Second, we compute the return accruing to advertisers when they can bid for keywords at the segment level using the approach detailed in section C.1. The difference between the two returns can be considered as a measure for the benefits of targeting. At the same time, we may calculate the advertisers' returns under the two scenarios and the difference may be a measure for the value of market intelligence.